Data Privacy Risk

Daxuan Cheng [a] Ding Ding [b] Yin Liao [a] Zheyao Pan [a*]

[This Version: October 2025]

Abstract

Firms' growing reliance on personal data has exposed them to a critical yet difficult-to-measure dimension of intangible risk: data privacy risk – the risk of financial, operational, and reputational harm arising from firms' data practices involving the collection, storage, and use of personal data. We develop a firm-level measure of data privacy risk (*DPRisk*) by applying textual analysis to earnings call discussions. We rigorously validate *DPRisk* through multiple tests, demonstrating its conceptual distinctiveness from existing risk measures. The measure rises meaningfully around major regulatory events and high-profile privacy incidents, is systematically higher in data-intensive industries and among GDPR-exposed firms, and correlates with prior privacy violations. Finally, we show that *DPRisk* is economically meaningful: firms with elevated *DPRisk* earn lower abnormal earnings announcement returns and face higher spreads on bank loans. These results suggest that data privacy risk is a priced and persistent dimension of corporate risk in the modern digital economy.

Keywords: Data privacy, Privacy regulation, Risk management, Intangible risk, Earnings calls

JEL Classification: D22, G14, G21, G28, G32, K24, L51

[*] We are grateful for constructive comments and suggestions from Andrew Ainsworth, Mingze Gao, Hieu Nguyen, Xiaofei Pan, Qian Wang, Michael Weber, conference participants from 2025 FMA Annual Meeting, and seminar participants at Macquarie University and the University of Wollongong. We also thank Zenan Sun for generously advising on computational linguistic matters. All errors remain our own.

[a] Department of Applied Finance, Macquarie Business School, Macquarie University. Emails: daxuan.cheng@mq.edu.au; yin.liao@mq.eu.au; terry.pan@mq.edu.au.

[b] Research School of Finance, Actuarial Studies & Statistics, Australian National University, Email: ding.ding@anu.edu.au.

Data Privacy Risk

Abstract

Firms' growing reliance on personal data has exposed them to a critical yet difficult-to-measure dimension of intangible risk: data privacy risk – the risk of financial, operational, and reputational harm arising from firms' data practices involving the collection, storage, and use of personal data. We develop a firm-level measure of data privacy risk (DPRisk) by applying textual analysis to earnings call discussions. We rigorously validate DPRisk through multiple tests, demonstrating its conceptual distinctiveness from existing risk measures. The measure rises meaningfully around major regulatory events and high-profile privacy incidents, is systematically higher in data-intensive industries and among GDPR-exposed firms, and correlates with prior privacy violations. Finally, we show that DPRisk is economically meaningful: firms with elevated DPRisk earn lower abnormal earnings announcement returns and face higher spreads on bank loans. These results suggest that data privacy risk is a priced and persistent dimension of corporate risk in the modern digital economy.

Keywords: Data privacy, Privacy regulation, Risk management, Intangible risk, Earnings calls

JEL Classification: D22, G14, G21, G28, G32, K24, L51

1. Introduction

Modern firms increasingly rely on digitized personal data to create value. From technology giants to traditional retailers, firms collect, analyze, and monetize consumer data to derive competitive advantages (Demirer et al., 2024; Goldberg et al., 2024; Martin et al., 2017). As data becomes a core economic asset, it also exposes firms to a new class of risks centered on privacy (Acemoglu et al., 2022). Unlike traditional cybersecurity threats which often arise from external attacks(Florackis et al., 2023; Crosignani et al., 2023), data privacy risk frequently originates from a firm's internal practices related to the collection, use, sharing, and governance of personal information. A firm may be fully insulated from external breaches and still face significant legal, reputational, and operational consequences from how it handles data. Data privacy risk has thus become a distinct and pervasive form of intangible risk, shaped by evolving technology, regulations, consumer scrutiny, and strategic business decisions.

The economic stakes of data privacy risk exposure are broad and growing. As the digital transformation of our social and economic activities accelerates and public awareness increases, data governance has moved from a back-office concern to a major source of a firm's reputational capital – or liability (Jones & Tonetti, 2020; Li et al., 2024). Regulators have responded with sweeping privacy rules[1]. The European Union's General Data Protection Regulation (GDPR; Regulation (EU) 2016/679) grants regulators unprecedented extraterritorial authority[2] over the commercial use of personal data. In the U.S., more than twenty states – including California, Colorado, Connecticut, Utah, and Virginia – have enacted similar laws, such as the California Consumer Protection Act (CCPA). Firms now operate in an environment where perceived privacy failures, even in the absence of an actual breach, can lead to fines, class-action lawsuits, and loss of market value (Jaworski & Schmeltzer, 2022). Data privacy is no longer merely a legal compliance issue; it is a fundamental, evolving economic risk.

Despite this transformation, we lack a systematic understanding of which firms are most vulnerable to this risk, how stakeholders perceive it, and the extent to which capital markets price it (Acquisti et al.,

[1] According to the United Nations Conference on Trade and Development, almost all developed markets have legislation in privacy and data protection, and well more than half of the remaining markets do so. See: United Nations Conference on Trade and Development (no date) *Data protection and privacy legislation worldwide*. Available at: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide (Accessed: 16 June 2025).

[2] For example, one of the noticeable features of GDPR is its extraterritoriality principle, allowing it to supervise organizations from all over the world as long as they operate in Europe and pulls EU consumer data in any form or step of business.

2016). This gap reflects the inherent challenge of quantifying firm-level data privacy risk. Data privacy risk imposes both direct costs and indirect, harder-to-measure consequences that complicate corporate decision-making. Direct costs include substantial compliance expenses from dedicating corporate resources to meet evolving regulatory requirements; financial penalties for violations; litigation costs; and potential operational disruptions tied to governance failures. Indirect costs arise through reputational damage that can erode customer trust and drive customer loss, increasing acquisition costs and reducing revenues (Frey and Presidente, 2024; Goldberg et al., 2024). Firms may also face reduced operational efficiency and constrained data-driven strategies as concerned customers limit data sharing (Demirer et al., 2024; Ferracuti et al., 2024). The multifaceted and evolving nature of these costs makes data privacy risk challenging to measure with traditional structured data or existing broad regulatory proxies, which tend to be sparse, backward-looking, and ill-suited for capturing forward-looking firm-level exposure.This paper addresses this gap by developing a text-based, forward-looking measure of firm-level data privacy risk. We define data privacy risk (*DPRisk*) as the threat of financial, operational, and reputational harm to firms from their collection, storage, and use of personal data, shaped by evolving regulations, customer expectations, and potential leaks. Using a corpus of over 150,000 quarterly earnings call transcripts spanning U.S. listed firms from 2010 to 2023, we adopt the framework of Hassan et al. (2019) to identify discussions of risk and uncertainty associated with data privacy issues. To construct our privacy dictionary, we combine a seed library drawn from the GDPR and CCPA legal texts with semantically similar bigram expansions identified through word embedding techniques trained on earnings call discussions (Li et al., 2021). This hybrid approach enables us to track how privacy-related risks are actually discussed by sophisticated market participants in forward-looking business meetings while being informed by regulatory discourse.

Our *DPRisk* measure demonstrates unique economic relevance and robustness across multiple dimensions. First, it captures unique informational content. *DPRisk* exhibits minimal correlation with existing measures of other intangible risks using similar methodologies that include political risk, climate change risk, supply chain risk (Ersahin et al., 2024; Hassan et al., 2019; Sautner et al., 2023), as well as cybersecurity risk (Jiang et al., 2024; Florackis et al., 2023), while showing strong co-movement with contemporaneous stock return volatility. This orthogonality suggests that *DPRisk* captures a distinct dimension of intangible corporate risk. Second, *DPRisk* shows economically meaningful time-series evolution, rising notably around major data breach events and the adoption of landmark privacy regulations such as the GDPR and the CCPA. Third, DPRisk exhibits strong persistence at the industry level while displaying significant cross-sectional variation both across industries and geographies. Firms in industries that are inherently more reliant on consumer data or handle more sensitive personal information consistently show higher average DPRisk values, reflecting greater exposure to regulatory scrutiny such as the General Data Protection Regulation (GDPR) (Frey and Presidente, 2024). Geographically, firms with substantial operations in the European Union exhibit significantly higher DPRisk scores following the official announcement of the GDPR's adoption in 2016, underscoring the role of jurisdictional regulatory regimes in shaping privacy risk. Additionally, firms with a history of privacy violation incidents also tend to have higher DPRisk scores on average, indicating the importance of firm-level compliance records in determining privacy risk exposure.

We then test whether DPRisk is systematically priced in financial markets, beginning with an examination of how capital markets respond to cross-sectional differences in firms' privacy risk exposures. We find that firms exhibiting elevated *DPRisk* experience economically significant declines in equity valuation, showing lower short-term cumulative abnormal returns that range from 14.2 to 15.6 basis points for a one-standard-deviation increase in *DPRisk* disclosed in their quarterly earnings calls. These return effects persist after controlling for standard firm characteristics, earnings surprise, and general sentiment in privacy-related discussions in the call. We also find no evidence of stock return reversal in the two months after the earnings call, suggesting that the price impact associated with high data privacy risk disclosure is not transitory. The evidence suggests that equity market investors respond quickly to data privacy risk and treat it as a credible signal for unfavorable exposure.

In the credit market, we find that firms with higher *DPRisk* face higher borrowing costs. A one-standard-deviation increase in annual *DPRisk* is associated with an increase of 6.11 to 8.50 basis points[3] in spreads on newly originated bank loans in the following year, which is equivalent to an incremental loan cost of approximately 2.7%-3.7%. The results remain robust after controlling for firm and loan characteristics, as well as fixed effects in different model specifications. The evidence suggests that

[3] The economic size of this effect is meaningful and comparable with the recent findings in the bank loan literature. For instance, Gad et al. (2024) find a 8.31 basis point increase in loan spread given a one-standard-deviation increase in earnings-call-based political risk. Smilarly, the effect of customer concentration is about 10 basis points as documented in Campello and Gao (2017), and the effect ot lender trust on loan spreas is about 3 basis points (Hagendorff et al., 2023).

sophisticated creditors explicitly price in the financial and operational implications of perceived increases in exposure to data privacy risks, as consistent with previous studies that focused on bank responses to data privacy laws and data breaches (Gupta et al., 2024; Huang & Wang, 2021).

This paper contributes to three intersecting literature streams on corporate risk in the digital economy, intangible asset management, and the pricing of non-traditional risks. First, we extend work on technology-related corporate risks by focusing on regulatory data privacy risk, a distinct channel rooted in internal data governance and evolving legal obligations, which remains underexplored relative to cybersecurity breaches that reflect exogenous attacks on corporate systems (Cong et al., 2025; Florackis et al., 2023; Jamilov et al., 2021; Jiang et al., 2024). Second, we contribute to the literature on intangible asset management by highlighting how privacy risk represents a growing, hard-to-measure liability that can erode firms' reputational capital and strategic flexibility—complementing research on brand capital, customer trust, and organizational capital as value-relevant but intangible assets (Belo et al., 2017; Eisfeldt and Papanikolaou, 2013 & 2022). Third, we advance the study of how capital markets price non-traditional risks by providing novel evidence that investors systematically incorporate firm-level exposures to regulatory privacy risk, filling a gap in work examining the valuation effects of operational disruptions, supply-chain shocks, and technology-related vulnerabilities (Barrot and Sauvagnat, 2016; Kamiya et al., 2021).

The remainder of this paper is organized as follows: Section 2 introduces the institutional background and reviews the related literature. Section 3 describes the construction of *DPRisk* and other data. Section 4 validates the measure and presents the basic properties of *DPRisk*. Section 5 examines the effect of *DPRisk* in both the equity and credit markets. Section 6 concludes.

2. Institutional Background,Theoretical Concept, and Related Literature

2.1 Data Privacy Regulations

In the digital economy era, the rapid growth of data-driven business models has prompted a global wave of strengthened data privacy regulation, aimed at safeguarding personal information amid expanding collection, use, and sharing practices. The European Union's General Data Protection Regulation (GDPR), implemented in 2018, set a global standard with its broad territorial scope, requiring firms to ensure lawful, transparent, and limited data collection, secure processing, and clear data subject rights, including consent management and breach notification. It requires firms to obtain clear and informed consent for data collection and use, conduct Data Protection Impact Assessments (DPIAs) for high-risk processing activities, and ensure data subject rights—including access, rectification, erasure ("right to be forgotten"), portability, and objection. Companies must also implement robust technical and organizational measures to secure personal data and notify authorities and affected individuals of breaches within strict timeframes. Notably, the GDPR's extraterritorial scope extends its requirements to any firm, regardless of location, that processes personal data of individuals in the EU, thereby setting a global benchmark for privacy compliance and introducing substantial financial penalties for violations.

Following the GDPR's lead, many jurisdictions worldwide have developed or strengthened their own data privacy laws to address growing public and regulatory expectations in the digital economy. In the United States, the California Consumer Privacy Act (CCPA, effective 2020) and its amendment, the California Privacy Rights Act (CPRA), established some of the strongest state-level requirements, granting consumers the rights to know, access, delete, and opt out of the sale of personal information, while imposing disclosure and governance obligations on firms. Brazil's Lei Geral de Proteção de Dados (LGPD) closely mirrors the GDPR with requirements for lawful basis of processing, data subject rights, and breach notification. Across Asia-Pacific, Japan amended its Act on the Protection of Personal Information (APPI) to enhance cross-border data transfer requirements and strengthen individual rights, while South Korea's Personal Information Protection Act (PIPA) is considered among the world's most rigorous, with high standards for consent and use limitation. Australia, Singapore, China, and other countries have also introduced or updated data privacy laws to impose clearer requirements on firms regarding the collection, use, sharing, and secure storage of personal data. Collectively, this global expansion of regulation has transformed data privacy compliance into a critical area of corporate governance and risk management for firms operating in the digital economy.

2.2 Defining Data Privacy Risk

The pervasive digital transformation of the global economy has elevated data to the status of a core, strategic production input, like labor and capital (Begenau et al., 2018; Farboodi & Veldkamp, 2020). In particular, user-generated data that ranges from personal information to behavioral insights has become central to firm value creation through targeted marketing, product optimization, and strategic decision-making. However, the very activities employed by firms to make data a strategic asset – its large-scale collection, storage, and monetization – also expose firms to growing risks (Bian et al., 2021). These risks stem not only from technological vulnerabilities but also from rapidly evolving regulatory pressure and increasing public concern over the misuse and unauthorized leakage of personal information (Acquisti et al., 2016; Goldfarb & Que, 2023). As a result, data privacy risk has become a salient form of intangible corporate risk, distinct from the more traditional notions of cybersecurity and information governance (Kamiya et al., 2021; Gomes et al., 2024).

We identify three primary channels through which data privacy risk materializes at the firm level:

Compliance Costs – Data privacy regulations set clear limits on how firms collect, store, and use sensitive user data that impose nontrivial compliance burdens. These include but are not limited to obligations related to user consent, data minimization and portability, storage limitations, and the right to know and erasure (Barrett, 2018; Jones & Kaminski, 2020). As firms adapt, they incur costs across several dimensions: (i) operational expenses from redesigning or modifying existing products and services to meet new privacy standards (Gupta et al., 2024); (ii) investments in upgrading technological infrastructure for data anonymization, encryption, and secure data processing (Demirer et al., 2024; Zhang et al., 2021); and (iii) human resource expenditures for establishing dedicated privacy teams, recruiting specialized roles[4], and training employees on data privacy norms (Kim et al., 2024).

Violation Costs - Non-compliance with privacy mandates can trigger aggressive investigations, substantial fines, costly litigations, and in extreme cases, business suspensions. Recent enforcement cases under the GDPR against major technology firms like Amazon, Google, Meta, and LinkedIn highlight the escalating financial repercussions from regulatory violations and class-action settlements. For example, Meta has faced extensive legal scrutiny and considerable financial penalties globally for data privacy violations, accumulating over $2.5 billion in fines to date[5]. On the other side, the

[4] For example, the GDPR requires organizations to set a Data Protection Officer (DPO) position if they have significant data processing and data subject monitoring needs. The DPO is designed to be the primary person responsible for data governance and compliance in the organization, and its role is to ensure that the business takes appropriate measures to properly protect personal data.

[5] Businesses that fail to meet the requirements of the GDPR face serious legal sanctions, such as a hefty fine of 20 million euros or 4% of global turnover, whichever is higher. Two most recent cases supported by GDPR are that the Irish Data Protection Commission fined Meta for 91 million euros in September 2024, after the company failed to notify the commission of a personal data breach and improper password storage; the Commission also fined LinkedIn (a Microsoft-owned career platform) for 310 million euros in October 2024 regarding its misconduct on using users' personal data for behavioral analysis and targeted advertising. (See more details in https://www.wsj.com/tech/meta-handed-100-million-fine-in-ireland-over-password-storage-e569ccc6?msockid=182f026ed2546cc236881752d3be6ddf and https://www.wsj.com/tech/linkedin-fined-more-than-300-million-in-ireland-over-personal-data-processing-ec0fb24e?st=GL2wPU&reflink=desktopwebshare_permalink.)

commercial value of consumer data has made it a frequent target of privacy violation and causes immediate ex-post costs (Agarwal et al., 2024; Wei & Zhu, 2024), including extensive crisis management efforts, damage containment through public relations campaigns, direct compensation to affected users, and mandatory regulatory reporting. These cases reflect a clear shift: regulators and the public are increasingly willing to swiftly impose severe penalties. For firms operating in the digital economy, the reputational and legal fallout from data privacy risks poses material consequences not only to operations but also to shareholders.

Consumer Trust and User Attrition – Privacy concerns have become a critical determinant of consumer sentiment and behavior, introducing a market-driven channel of risk (Fainmesser et al., 2023; Goldfarb & Que, 2023). High-profile data privacy incidents – most notably the Cambridge Analytica scandal involving Facebook's unauthorized sharing of user data[6] – have intensified public concern over data misuse and manipulative practices (Acemoglu et al., 2025). The resulting erosion of consumer trust has been shown to diminish customer lifetime value, prompting users to disengage from affected firms and transfer their patronage to competitors (Acquisti et al., 2016; Li et al., 2024; Martin et al., 2017). Users often withhold personal information from firms that are perceived as negligent, undermining the effectiveness of data-driven analytics and even corporate strategic decisions (Acemoglu et al., 2022; Ferracuti et al., 2024; Goldfarb & Tucker, 2019). These dynamics can lead to immediate declines in market share, sales, and profitability[7] (Aridor et al., 2020; Bian et al., 2021). Over time, data privacy failures inflict lasting brand reputation damage and sustained performance setbacks that extend well beyond immediate financial penalties.

2.3 Relating Data Privacy Risk to Other Risks

While data privacy risk is interconnected with other emerging forms of information risk, data privacy risk is a conceptually and operationally distinct risk in comparison to cybersecurity, data quality, and

[6] In March 2018, Facebook admitted to the unauthorized sharing of 87 million users' personal information with Cambridge Analytica, an English private consulting company, that used the data to influence U.S. election and Brexit vote in 2016. In response, a small Californian privacy group demanded a swift statewide ballot initiative which gained majority support leading to enactment of the California Consumer Privacy Act (CCPA) in state law (Gupta et al., 2024).

[7] For example, Bian et al. (2021) document significant negative stock market reactions when the mobile applications of firms are tagged by Apple AppStore with "privacy labels" (indicating that they collects device-tracking and user-privacy information), while the effect is stronger when those applications collect more amount of data or are more privacy-invasive.

data breach. Understanding these distinctions is fundamental to appreciating the unique contribution of data privacy risk and the *DPRisk* measure.

The occurrence of cybersecurity incidents is related to data privacy risk, as an attack, successful or not, may fuel public concerns about system vulnerabilities. However, despite this connection, cybersecurity risk (Jiang et al., 2024; Florackis et al., 2023; Jamilov et al., 2021) typically centers on a firm's entire digital infrastructure and the potential for operational disruption, threats that often arise from external attacks (Crosignani et al., 2023; Kamiya et al., 2021). In contrast, data privacy risk is fundamentally concerned with a firm's handling and governance of user information, including issues of consent, disclosure, and policy compliance, which are areas where internal failures or strategic misuse can be just as damaging as external threats.

Similarly, while data quality and breach risk emphasize integrity (Gomes et al., 2023) or protection of data assets, data privacy risk highlights the intricate dynamics between corporate behavior, regulatory oversight, and public awareness and scrutiny over the boundaries of acceptable data use. Data privacy risk thus arises from the tensions between data-driven value creation and societal expectations for responsible data stewardship – an interaction that defines and elevates data privacy risk as a distinctive and salient form of firm-level vulnerability that requires precise measurement and analysis.

2.4 Related Literature

The proliferation of digitized personal data has spurred a rapidly expanding field of inquiry into the economics of data privacy. Research in this literature explores the trade-offs between data utility (for firms, consumers, and society) and inherent privacy concerns that arise from data practices (data collection, storage, processing, and sharing). Privacy can be conceptualized as a scarce resource that influences how economic actors make decisions (Acquisti et al., 2016). This work sets the foundation for information asymmetry in privacy decisions, where firms may possess superior knowledge about their data practices compared to external actors such as consumers and investors, leading to market failures or suboptimal outcomes.

Following this, a body of recent empirical research quantifies the economic implications of data privacy and privacy violations, many studies using event study designs to exploit exogenous variations from privacy shocks. Studies have documented that data breaches can impact a wide range of firm outcomes, including declines in stock prices (Campbell et al., 2015) and increased executive turnover (Banker & Feng, 2019). Following landmark legislations like the GDPR and the CCPA, imposing stringent compliance requirements and significant penalties for violations, research has also investigated the economic consequences of these regulations on firms. Findings include higher production costs (Choe et al., 2024; Demirer et al., 2024; Gupta et al., 2024), lower profitability (Frey & Presidente, 2024; Goldberg et al., 2024), declines in investment and operational efficiency (Ferracuti et al., 2024; Jia et al., 2025; Maex, 2022), constrained firm growth (Boroomand et al., 2022; Farboodi et al., 2019), and higher cash flow risk that leads to greater bank loan costs (Agarwal et al., 2024; Huang & Wang, 2021). There exist considerable variations in these effects, as firms vary in their exposures to privacy laws and public scrutiny in terms of firm size, technological intensity, governance quality, and geography (Aridor et al., 2020; Frey & Presidente, 2024; Li et al., 2024). For example, following the GDPR's introduction, Demirer et al. (2024) document a 20% increase in data costs for European firms, relative to their U.S. counterparts, while Campbell et al. (2015) show that smaller firms tend to bear a disproportionate share of the compliance burden, due to larger firms having access to better infrastructure and more resources that enable them to absorb regulatory costs and mitigate legal risks (Johnson et al., 2023).

We extend this research in data governance and privacy regulation by moving beyond discrete "breach events" or specific regulatory impacts to develop a continuous measure of perceived data privacy risk. Given the real-time, forward-looking nature of earnings calls (Li et al., 2021), our measure captures the perceived data privacy risk of managers and sophisticated market participants as they discuss strategies and responses before, during, and after a breach event or widespread regulatory enforcement. At a quarterly frequency that is available across a broad panel of publicly listed firms, our *DPRisk* measure overcomes the limitations of event-based identification strategies and enables us to analyze variations in privacy risk at the firm level across time and sectors. This contributes a scalable and generalizable tool for understanding privacy-related risk exposure and how it influences valuation and finance as it evolves with the changing regulatory landscape and public scrutiny.

More broadly, our work contributes to prior work on intangible capital and corporate social responsibility risks. This body of work recognizes that intangible assets – such as intellectual property, brand reputation, human capital, and increasingly, data – are growing generators of firm value. A natural extension to this line of inquiry is the literature on intangible risks, which explores the financial market implications of various intangible risks. These include reputational risk (Dyck et al., 2010; Makridis, 2021), social responsibility and environmental risk (Asante-Appiah & Lambert, 2023; El Ghoul et al., 2011), and corporate governance risk (Gompers et al., 2003). Cybersecurity risk, related to data privacy risk, has also received significant attention, with studies demonstrating its impact on firm valuation and cost of capital following reported breaches or vulnerabilities (e.g., Florackis et al., 2023). Our paper distinguishes data privacy risk as a unique and important risk dimension that fundamentally pertains to the *handling of personal information* and its associated regulatory and societal implications, whereas cybersecurity broadly encompasses the protection of all digital assets from malicious external attacks. By developing a distinct measure and validating its statistical and economic properties, we show that data privacy risk conveys unique information to financial markets, complementing existing understandings of intangible risks.

Intangible assets are difficult to measure and value using traditional financial accounting information. A growing and expansive literature highlights the role of non-financial disclosures in enabling our understanding of intangible assets and risks that affect investor behavior, cost of capital, and firm valuation. In particular, textual analysis, utilizing computational linguistics and machine learning techniques, has become a powerful methodology for extracting nuanced insights from qualitative information that would otherwise be unobservable. This burgeoning field has enabled researchers to construct quantitative measures for a wide array of financial constructs, including managerial tone and sentiment (e.g., Loughran & McDonald, 2011; Tetlock et al., 2008), economic uncertainty (e.g., Baker et al., 2016), risk factor exposure (e.g., Hoberg & Phillips, 2016), and corporate culture (e.g., Li et al., 2021). Other work extracts meaningful signals from corporate disclosures (e.g., Jiang et al., 2019; Noh & Zhou, 2022; Wu, 2024). Together, the stream of literature extends the border of how investors and market stakeholders could acquire and process sophisticated information from textual data sources (Blankespoor, 2022; Bybee et al., 2023).

Earnings call transcripts, in particular, have emerged as a rich source of unstructured textual data that offers frequent and direct records of managerial communication and market interaction (Brown et al., 2004; Matsumoto et al., 2011). Studies have leveraged textual analysis of earnings calls to predict future earnings, explain stock price crashes, and more recently to understand managerial behaviors such as cultural signaling (Li et al., 2021), strategic communication (Gow et al., 2021; Hollander et al., 2010), investment expectation (Jha et al., 2025), and CEO optimism (Davis et al., 2015). Building on methodologies developed by Hassan et al. (2019, 2023; 2024a; 2024b) and Sautner et al. (2023), we use natural language processing (NLP) to extract latent signals from earnings calls to capture intangible risks that are not easily captured in accounting data. While previous work has quantified firm exposure to political, regulatory, environmental, and pandemic risks, none to our knowledge has focused specifically on data privacy. Our methodology contributes to this vibrant literature with a hybrid approach that develops a novel, domain-specific dictionary – instead of relying on a broad lexicon or purely statistical topic models – from combining legal texts (GDPR, CCPA) to achieve regulatory precision with corpus-driven expansions through Word2Vec, and filters for proximity to risk-oriented language. In contrast to Boroomand et al. (2022), who apply Latent Dirichlet Allocation (LDA) topic modeling to 10-K filings to measure attention to privacy, our approach captures privacy-related risk – not just attention – that leverages the timeliness and spontaneity of earnings calls. Our paper thus adds to this vibrant methodological literature with a new measure for a previously underexplored risk dimension, which ensures the *DPRisk* measure is both highly targeted and reflective of the evolving corporate discourse surrounding privacy. This approach yields a more precise and robust indicator of this specific intangible risk.

Our empirical analyses to evaluate the impact of *DPRisk* on equity valuation and the cost of debt are founded on the theories of information asymmetry and corporate disclosure. Market efficiency is predicated on the rapid and unbiased incorporation of new information into asset prices (Fama, 1970). Corporate disclosures such as earnings calls serve as an important mechanism to reveal new and important signals to the market, mitigating any information asymmetries between corporate insiders and external market participants (e.g., Grossman & Stiglitz, 1980; Healy & Palepu, 2001). By providing an empirical evaluation of how qualitative discussions of corporate privacy risks influence corporate valuation and financing, our paper offers a framework for understanding and measuring data privacy risk, highlighting its position as a distinct and economically significant determinant of corporate outcomes.

3. Measuring Data Privacy Risk

3.1 Data Overview

Our primary data is a comprehensive dataset of earnings call transcripts for U.S. publicly listed firms from 2010 to 2023, sourced from S&P Capital IQ. We begin our sample in 2010 for two reasons. First, coverage of earnings call transcripts becomes substantially more comprehensive after 2010. Second, this period coincides with increasing regulatory and commercial emphasis on data privacy, exemplified by the establishment of major privacy regulations and intensified incidences of class-action lawsuits on data privacy issues (Demirer et al., 2024; Goldfarb & Que, 2023; Martin et al., 2017).

We have a raw sample of 159,066 transcripts from 5,737 unique firms. Following established practice in the literature (e.g., Hassan et al., 2019; Li et al., 2021; Sautner et al., 2023), we analyze the full content of each transcript[8], including both the management presentation and the Q&A session with analysts.

We choose to conduct textual analysis on earnings call transcripts instead of other textual data, such as 10-K report files, due to the following reasons. First, the earnings call of a public firm is usually held immediately after the quarterly earnings announcement, thus it is updated more frequently and reflects up-to-date information on firm operations (Campbell et al., 2025; Hassan et al., 2019), which is important considering the rapid change of data privacy regulation environment and intensive occurrence of privacy violation events in the recent decade. Second, compared with regulatory report files, the oral discussion feature of earnings calls also makes it less constrained thus reducing the concern of information manipulation (Gow et al., 2021; Li et al., 2021). For example, content in the 10-K file and the other mandatory reports is usually highly formatted and decided internally by the firm management, which could be involved with self-selection issues in the disclosure process. Last but not least, the Q&A session of earnings calls is a unique disclosure channel in response to the information demand of public investors, because it documents various questions raised by analysts (Brown et al., 2004; Matsumoto et

[8] Following the literature, we also measure the alternative exposure data based on Q&A and presentation sessions separately in the unreported robustness tests. But unless indicated otherwise, our main tests are all based on full-transcript measurement considering the non-random nature of zero-question occasions in the conference calls (Chen et al., 2016; Gow et al., 2021), which may cause bias for the exposure calculation.

al., 2011), thus the diverse coverage of earnings calls makes it exceptionally suitable to capture the stakeholders' concerns about emerging topics like data privacy.

We supplement the earnings call data with several standard financial datasets. Stock return data of U.S. listed firms are obtained from the Center for Research in Security Prices (CRSP), and firm-level financial and accounting data come from Compustat. Bank loan information is sourced from DealScan and matched to the other firm-level data using the Chava and Roberts (2008) link table. Operation segments of U.S. listed firms are identified through Compustat Customer Segments. Firm-level credit ratings are from S&P Credit Ratings. Institutional holding data is obtained from the Thomson Reuters 13F database. We obtain macroeconomic indicators from the Board of Governors of the Federal Reserve System. We also use the OECD Inter-Country Input-Output (ICIO) Tables[9] from the United Nations Statistics Division to construct country-industry level GDPR exposure following Frey and Presidente (2024). Finally, incident-level data of privacy violation records is obtained from the RepRisk database.

3.2 The Data Privacy Bigram Library

As discussed in Section 2.1, data privacy risk arises from multiple sources that include ex-ante costs in compliance requirements, ex-post costs from the consequences of privacy violation behaviors, as well as long-term reputational damage and sustained performance setbacks triggered by public concern on data privacy. The complexity of these component risks is difficult to quantify using traditional methods or financial & accounting data alone. To address this, we design a novel textual analysis methodology to quantify firm-level data privacy risk by leveraging real-time narratives in earnings call transcripts to capture forward-looking discussions and risk perceptions by firm managers and analysts.

Our approach builds on a growing literature[10] that utilizes NLP and machine learning to extract discussion topics in earnings call transcripts and quantify risk exposures for specific topics. To accomplish this, we construct a training library that allows us to identify data privacy topics based on a

[9] NAICS2017US - ISIC Rev.4 Correspondence Table (https://unstats.un.org/unsd/classifications/Econ/ISIC), accessed and downloaded in October 4th, 2024.

[10] Recent representative literature that adopts similar approach to establish firm-level topic indices includes political uncertainty risk of Hassan et al. (2019), pandemic disease risk of Hassan et al. (2023), supply chain risk of Ersahin et al. (2024) as well as Wu (2024), corporate culture of Li et al. (2021), and climate change risk of Sautner et al. (2023), among others.

multi-source approach. We then apply the training library to our sample of transcripts to quantify data privacy risk exposure.

Identifying precise thematic information from unstructured text presents a methodological challenge (Sautner et al., 2023; Wu, 2024). Take data privacy content as an example, earnings call discussions usually contextualize data privacy considerations within topics such as regulation and litigation, technology development and innovation, as well as future opportunity and business performance, generating ambiguity for precise identification of data privacy risk. We adopt a multi-source strategy to build a domain-specific library of bigrams (two-word combinations) that are indicative of data privacy discussions. This strategy follows the methodology in recent work such as Hassan et al. (2019, 2023), Li et al. (2021), and Wu (2024).

3.2.1 Domain-Specific Regulatory Bigrams

To improve the precision and relevance of our data privacy bigram library, we begin by constructing a domain-specific bigram seed library from the full statutory texts of the General Data Protection Regulation (GDPR; Regulation (EU) 2016/679) and the California Consumer Privacy Act of 2018 (CCPA). These two statutes are the most comprehensive data privacy regulations in the EU and the U.S., respectively, and also serve as normative benchmarks for global data governance. The GDPR establishes a regulatory architecture for proactive corporate accountability and comprehensive individual data rights; the CCPA focuses on consumer control through transparency, opt-out rights, and reasonable security practices (Barrett, 2018; Goldberg et al., 2024; Li et al., 2024). Bigrams extracted from these primary legislative texts are directly tied to the regulatory vocabulary governing privacy principles, compliance standards, enforcement mechanisms, and jurisdictional reach. This approach ensures that our seed bigrams are not merely linguistically frequent but legally and conceptually anchored in data privacy concerns, providing a well-founded foundation for identifying meaningful and precise references to privacy concerns in corporate discussions.

We tokenize the full statutory texts into bigrams after standard pre-processing procedures, including the removal of stop words, punctuations, and numeric tokens. We then form two separate corpora of adjacent bigram pairs – one for each legislation – that preserve the legal and semantic features of privacy-related discourse.

3.2.2 Earnings Call Contextual Embeddings

Business discussions of data privacy concerns are often implicit and embedded within broader discussions of operational or technological strategy. To capture such references as they naturally arise in managerial discussions, we extend our training library by incorporating linguistic patterns drawn directly from the earnings call transcripts. Following the embedding-based approach of Li et al. (2021), we compile a corpus of 40,000 randomly selected earnings call transcripts from our full sample. This corpus is pre-processed using standard natural language processing techniques: tokenize the texts into sentences and words, lemmatize words to their base forms, and replace named entities (e.g., company names, specific patents, and dates) with pre-defined tags. We then train a Word2Vec model[11] (Mikolov et al., 2013) on this pre-processed corpus, allowing us to identify bigrams that are semantically similar to a set of seed words. Our seed set includes “data”, “privacy”, “privacy regulation”, and “privacy law”, which are characteristics of the definitions of data privacy risk and reflect the risk’s relevance to information structure. For each seed word, we extract the top 500 bigrams with the highest cosine similarity scores and review these manually to remove incorrect or irrelevant matches. This approach allows us to detect more nuanced and contextualized references to data privacy in corporate communications that may not be as explicit as in statutory texts but are nevertheless relevant to the oral discussion of underlying data privacy issues we aim to capture in the earnings calls.

[Insert Figure 1 About Here: Word Cloud of Data Privacy Bigrams]

3.2.3 The Combined Data Privacy Dictionary

We begin to build a comprehensive training dictionary drawn from the three sources discussed: 1,979 from the CCPA corpus, 1,836 from the GDPR corpus, and 951 from the earnings-call-based Word2Vec corpus. Figure 1 presents word clouds illustrating the most prominent bigrams from each source.

To ensure that our library captures bigrams specific to data privacy, we follow Ersahin et al. (2024) and Hassan et al. (2019) by using *Financial Accounting* (Libby et al., 2014), a widely adopted finance and

[11] The word2vec model adopts word-embedding algorithm and uses a neural network to learn dense, low-dimensional vectors that represent the meanings of words based on their context location. Following the specification of Li et al. (2021), we set the dimension of word vectors to 300, define two words as neighbors if they are no farther apart than five words in a sentence, and omit words that appear fewer than five times in the corpus. As a result, the model compute the cosine similarity between each word in the corpus for each n-grams (including unigram, bigram, and trigram) and therefore could effectively identifies synonyms based on the surrounding words.

accounting textbook, to identify bigrams commonly found in standard business and financial discourse. Bigrams appearing in both the textbook and our training corpus are removed, retaining only those terms most likely to reflect data privacy concerns.

The final, filtered training library contains 4,647 unique bigrams. Conditional on the training library, we observe that both the average number and proportion of data privacy bigrams appearing in earnings calls increase steadily over time, and reach a peak in 2020 when CCPA was implemented. The trend points to the rising importance of data privacy issues for managers and market participants, highlighting the need for a targeted, firm-level measure of data privacy risk.

3.3 Measuring Data Privacy Risk Exposure and Sentiment

We proceed to measure firm-level data privacy risk exposure by identifying the co-occurrence of privacy-related bigrams and risk-related language in each earnings call transcript. To capture data privacy discussions in contexts where managers and analysts express concerns over potential threats or uncertainties, we tokenize and clean each transcript (removing stop words, punctuation, and numeric characters), and count the number of bigrams from our data privacy dictionary that appear within a ten-word window surrounding any synonym[12] for “risk” or “uncertainty”. Formally, for transcript *j* of firm *i* in quarter *t*, we define the following:

$$DPRisk_{i,j,t} = \frac{1}{B_{i,j,t}} \sum_{b}^{B_{i,j,t}} I[b \in S \backslash N] \times I(|b - r| < 10) \qquad (1)$$

where b denotes a bigram in the earnings call transcript, r is the position of the nearest synonym for “risk” or “uncertainty”, S is the set of bigrams in the data privacy library, and N is the set of excluded, generic bigrams, $I[\cdot]$ is an indicator function, and $B_{i,j,t}$ is the total number of bigrams in an earnings call transcript j for firm i in quarter t.

Therefore, the *DPRisk* measures the intensity of data privacy risk discussions[13] normalized by transcript length, providing a measure that can be compared across calls. Based on the measure, we find that more

[12] The list of synonyms of risk and uncertainty words is derived from the Oxford Dictionary, following Hassan et al. (2019).

[13] Since we use a training bigram library constructed from multiple sources, we are not able to generate inverse document frequency as the term weight of each bigrams (Hassan et al., 2019; Li et al., 2021) to adjust for the importance/association strength of different bigrams. Thus we choose to use straight sum instead of weighed sum method. However, in an unreported robustness test, we find the main results do not change much if we applies the weight of bigrams as their normalized frequency from the original textual source. See also the discussion of Hassan et al. (2019) in Section III.A for this issue. Also note that there are some other studies (Jamilov et al., 2021; Wu, 2024) adopt this straight sum approach to build textual-based indicators.

than 65% of call transcripts in our sample discuss data privacy risk at least once, and about 24% of Q&A sessions raise at least one question related to data privacy risk topics. For ease of interpretation in subsequent analyses, we multiply *DPRisk* values by 1,000. At the firm-year level, we aggregate call-level *DPRisk* across all available transcripts for firm *i* in a given year.

Next, following the same procedure as in the construction of *DPRisk*, we utilize the training bigram library to measure the general sentiment of data privacy discussions, *DPSentiment*. Specifically, rather than identifying synonyms of risk or uncertainty, we construct the *DPSentiment* conditioning on proximity to positive and negative words (Loughran & McDonald, 2011) as equation (2):

$$DPSentiment_{i,j,t} = \frac{\sum_{b}^{B_{i,j,t}} I[b \in S \backslash N] \times \sum_{c=b-10}^{b+10} S(c)}{B_{i,j,t}} \tag{2}$$

where $S(c)$ is a function that assigns a value of +1 if bigram c is associated with positive sentiment, a value of -1 if bigram c is associated with negative sentiment, and zero otherwise. In other words, $\sum_{c=b-10}^{b+10} S(c)$ calculates the net sentiment among the ten words surrounding bigram b from the data privacy training library.

Taken at face value, our *DPRisk* metric distinguishes and quantifies the intensity of risk-related privacy discussions by managers and analysts, while the *DPSentiment* metric captures the directional tone that they use when discussing data privacy issues in general. Although *DPRisk* and *DPSentiment* appear to move in opposite directions over time, their correlation is close to zero (-0.002) across our sample, suggesting that they capture distinct and independent sources of variation in firm-level data privacy discussions. Both measures exhibit substantial cross-sectional dispersion across firms and industries, as expected given the variations in data handling practices, regulatory exposure, and strategic positioning in our sample firms. For example, *DPSentiment* may be positive when firms view regulatory changes such as the implementation of the GDPR or CCPA as opportunities to differentiate through improved compliance, enhanced data governance, or the launch of privacy-enhancement services. Conversely, DPSentiment turns negative when firms highlight rising compliance costs, operational complexity, or legal risks associated with evolving regulations. In this sense, the sentiment measure effectively reflects whether managers and analysts frame data privacy as a source of strategic opportunity or a regulatory burden, a relevant distinction for interpreting investor reactions.

[Insert Table 1 About Here: Summary Statistics]

Table 1 reports summary statistics for *DPRisk*, *DPSentiment*, and other firm financials used in the subsequent regression analysis. The mean (standard deviation) value of *DPRisk* is 2.172 (2.844), and the mean (standard deviation) value of *DPSentiment* is 1.043 (4.486) in the unfiltered transcript sample. We also measure *DPRisk* for different segments of earnings calls separately, as *DPRisk* in the presentation (Q&A) segment has a mean value of 2.592 (1.512) and a standard deviation of 3.905 (4.109) in the unfiltered transcript sample. We note that while we primarily use *DPSentiment* to support the interpretation of *DPRisk* in this paper, this measure may be of independent interest for a variety of future applications, such as assessing managerial tone, investor expectations, or firm responses to regulatory changes or privacy incidents. To maintain focus, the remainder of this paper centers on validating and applying *DPRisk*, and refers to *DPSentiment* only where directly relevant.

4. Validation Tests and Properties of *DPRisk*

An effective measure of data privacy risk must capture a distinct dimension of firm-level risk that is both forward-looking and economically meaningful. In this section, we assess the statistical and economic properties of *DPRisk*. We compare *DPRisk* to related, existing measures and conduct a series of empirical tests, showing that *DPRisk* is a new and salient measure that captures non-redundant information reflecting firm-level exposure to data privacy risks.

4.1 Comparison with Related Measures

4.1.1 Training Library Comparison

An informative measure of data privacy risk captures a distinct economic concept, one that extends the literature beyond existing themes such as cybersecurity, data risk, or digitalization. To establish the conceptual uniqueness of the *DPRisk* measure, we begin with a comparative analysis of our data privacy library against those used in recent textual studies that construct firm-level measures to address other aspects of firms' use of information technology infrastructure. We show that our library is designed to reflect the legal, regulatory, and operational features that characterize data privacy concerns and risk exposures.

Appendix II presents a side-by-side comparison of our data privacy library (or term list) with those developed to measure cyber risk (Jamilov et al., 2021), cybersecurity and data breach risk (Florackis et al., 2023; Gomes et al., 2023), firm digitalization (Chen & Srinivasan, 2024), and attention to data privacy (Boroomand et al., 2022). Our library is distinct in three important aspects. First, it is explicitly designed to capture the regulatory and operational language that defines corporate exposure to data privacy oversight. While cyber and data breach risk dictionaries typically emphasize external threats and technical vulnerabilities, our dictionary includes terms such as "*consumer privacy*", "*privacy policy*", "*personal data*", and "*deidentified information*" that directly reference a firm's compliance obligations. Second, our library draws from three complementary sources: earnings call-based Word2Vec, the GDPR, and the CCPA. This hybrid approach ensures both domain relevance and regulatory precision. For example, bigrams such as "*clinical datum*", "*sensitive personal*", and "*client priority*" are rarely used in cybersecurity or digitalization discussions, yet they are central to how firms manage data privacy and reflect evolving regulatory and public expectations for personal data use. Third, compared to terms extracted from standard 10-K disclosures about data privacy in Boroomand et al. (2022), our library is more focused on the regulatory and public oversight environment. Moreover, by analyzing discussions in earnings calls, we obtain insights into how firms communicate data privacy issues and risks under real-time scrutiny from investors and analysts. For instance, bigrams such as "*supervisory authority*", "*user identity*", and "*private regulation*" highlight this focus on regulation, enforcement, compliance, and accountability.

Collectively, these features position our library to capture the institutional and operational perspectives specific to data privacy. By focusing explicitly on regulatory statutes and discussions of data privacy in unscripted, investor-facing communications, *DPRisk* is by design a distinct and policy-relevant measure of firm-level risk exposure, enabling us to pinpoint how data privacy concerns translate into economically significant firm-level risk. This measure opens new avenues for understanding how regulatory exposure, compliance behavior, and public perception interact in modern data governance and firm disclosure.

4.1.2 Correlation with Related Existing Indicators

Having identified the conceptual uniqueness of our hybrid training library used to generate *DPRisk* measure, we now show that *DPRisk* is statistically different from other textual-based firm-level risk measurements constructed from similar methodologies. Table 2 presents a correlation matrix between *DPRisk* and numerous established indicators, including political risk (*PRisk*) and general all-content risk and sentiment measures from earnings calls (*Risk* and *Sentiment*) by Hassan et al. (2019); climate change risk (*CCRisk*) by Sautner et al. (2023); and supply chain risk (*SCRisk*) by Ersahin et al. (2024). Aggregated to the firm-year level, we find the correlations between *DPRisk* and these existing indicators are consistently below 0.16, suggesting that *DPRisk* is statistically distinct and captures a unique dimension of firm-level risks in spite of sharing methodological similarities from being built on the framework of Hassan et al. (2019). In addition, we note a modest negative correlation (-0.089) between *DPRisk* and the general all-content sentiment (*Sentiment*), as consistent with risk discussions being inherently associated with a negative tone.

[Insert Table 2 About Here: Correlation Matrix of *DPRisk* and Other Textual-based Risks]

Table 2 also reports the correlation between *DPRisk* and the cybersecurity risk measure (*CyberRisk*) of Florackis et al. (2023), given their conceptual similarity. The negligible correlations between *DPRisk* and *CyberRisk* (-0.039) are consistent with conceptual distinctions outlined in Section 2.1 and Section 4.1. Although both measures relate to firms' use of technology and data, they reflect fundamentally different sources of risk[14]. Focusing on the vulnerability of information technology infrastructure, cybersecurity risk typically captures exposure to external threats and does not necessarily involve data privacy violations. Data privacy risk, in contrast, often arises from internal misuse or improper data practices, thus representing a novel and economically meaningful dimension of firm-level risk.

4.2 Perturbation Tests on Bigram Selection Robustness

To evaluate the sensitivity of the *DPRisk* measure to individual bigrams in the training library, we perform a perturbation test following the methodology of Sautner et al. (2023). We begin by identifying the top 10 most frequently occurring bigrams from each of our three sources (GDPR, CCPA, and earnings call-based Word2Vec model). We then successively exclude each of these bigrams from the overall training library at a time, recomputing *DPRisk* based on the modified training library. This process yields 30 alternative versions of the *DPRisk* measure. After aggregating the measure to the firm-

[14] Alternatively, we also obtain machine-learning-based *Cyber Risk* indicator developed by Jiang et al. (2024) and re-run the test by replacing the cybersecurity risk measure. The results stay the same with a low correlation.

year level, we calculate the correlation of each of these measures with the original *DPRisk*. The average correlation exceeds 90%, which suggests that *DPRisk* is robust to variations in the training library and not driven by sensitivity to any individual high-frequency bigram. We also re-estimate all main empirical tests presented in Section 5 using these perturbed measures, confirming that all results remain robust with similar significance levels.

4.3 Data Privacy Risk Discussed In Call Excerpts

We further validate the *DPRisk* measure by examining its relationship with firms' discussion of data privacy topics during earnings calls, which represent a distinct and more interactive disclosure channel compared to formal written filings. Intuitively, if *DPRisk* meaningfully captures a firm's underlying exposure to data privacy risk, companies with higher *DPRisk* scores should devote greater attention to these issues in their spoken communications with analysts and investors, whether through prepared remarks or responses during the Q&A segment. To test this, we analyze earnings call transcripts across different industries to assess the frequency and depth of data privacy discussions. This approach provides an out-of-sample check on the measure's validity, ensuring it reflects substantive managerial awareness rather than boilerplate text. To further illustrate how data privacy risk is communicated in practice, we present a selection of representative excerpts from firms in the top 1% of *DPRisk* scores in Appendix III.

Our analysis of these excerpts confirms that data privacy discussions are often a significant component of earnings call discussions, reinforcing the validity of our risk measure. Firms across various industries, particularly those in highly regulated sectors such as finance, healthcare, and technology, consistently emphasize data privacy as a critical business concern. These discussions broadly span three cost channels: compliance challenges and operational burdens, violation penalties and extra expenditures, as well as the impact of consumer aversion towards potential privacy loss. In addition, strategic opportunities are also frequently mentioned as a recurring dimension in the discussion.

First, a common theme across calls is the discussion of resource-intensive efforts associated with data privacy risk management. Many firms cite significant direct compliance costs, including investments in data security infrastructure, legal expertise, and cross-departmental compliance enforcement. For example, Trend Micro Incorporated described infrastructure changes driven by consumer privacy protection as “the biggest pain” for their customers. Similarly, in May 2017, Varonis Systems, Inc. explained in response to a GDPR-related question that employees across its business, IT, and security teams were working together to jointly evaluate the implications of the GDPR in order to strengthen internal data protection. Furthermore, many call participants express concerns over the growing complexities of privacy regulations, especially the need to comply with evolving regulations, including region-specific mandates such as GDPR and CCPA. Managers frequently highlight the legal uncertainty surrounding cross-border data transfers and the inherent challenges of aligning compliance efforts across jurisdictions. This uncertainty increases both compliance costs and operational challenges associated with the secure management of customer data. For example, in July 2014, Digital Realty Trust, Inc. noted that privacy concerns and data sovereignty requirements in Europe, the Middle East, and Africa introduced substantial risks for its geographically dispersed data center operations. Also, an analyst expressed his doubt about the potential impacts of CCPA on the market position and strategy decisions in a conference call with Change Healthcare Inc. (February 2020).

Second, firms also widely acknowledge concerns over substantial indirect costs, including potential fines, reputational damage, and business disruptions arising from data breaches or regulatory non-compliance. This observation suggests that firms perceive data privacy risks not merely as a legal obligation that impacts financial performance, but also need cautious responses and thus alter their risk management. In one example, Bank of Montreal (May 2018) placed emphasis on their privacy commitment and referred to the recent customer data leakage events as an emerging risk that has become a priority of its daily operation. In August 2017, the management of Qualys Inc. referred to the rigorous penalty of GDPR if firms fail to ensure the security and privacy of customer data when they are introducing the extensive effort of firms on the importance of privacy protection. Likewise, Absolute Software Corporation (February 2013) stated in its presentation session that it had dedicated substantial effort to reconfigure its organizational and IT infrastructure to meet the dual changes in the rising number of related violation events and growing risk management compliance standards in privacy protection.

Third, we find that firms have exhibited their awareness of the consumers’ privacy-related behaviour, and their motivational changes on corporate operations and reputation. For instance, Roku, Inc. stated in May 2021 that to build a good relationship with consumers, it is important to seek their opt-in approval before tracking their data or targeting them under the increased regulatory environment. The Kroger Co. pointed out in November 2023 that they are committed to using patient data appropriately under privacy laws, so that they could maintain customer loyalty and their trust. We also find evidence that the privacy preference of different customer groups affects the infrastructure setup and operational strategy. In October 2015, Check Point Software Technologies Ltd. disclosed its flexible deployment options (on either cloud service or in-house appliance) as a responsive solution to satisfy users from the U.S. and Europe. In addition, the privacy regulation also plays a vital role now for firms to obtain a good reputation and an external positive impression. Multiple firms, including FinVolution Group (August 2022) and Mimecast Services Limited (August 2019), all mentioned their effort in privacy protection practice and how it benefits them in social impact establishment and industry positioning.

Beyond risk mitigation, we also observe that firms strategically frame data privacy as a source of competitive advantage and discuss it in a positive tone. A number of managers leverage call discussions to announce privacy-focused solutions or products, aiming to differentiate their business models to build customer trust and respond to growing market demand for privacy protection. In some cases, privacy regulations are even viewed as a catalyst for innovation and growth as businesses recognize opportunities in privacy-enhancing technologies and compliance services. We note various such cases within our transcript sample: Cisco Systems, Inc. (August 2017), Deutsche Telekom AG (November 2021), and Mastercard Incorporated (July 2022) all emphasized that stricter privacy regulations had stimulated the development of new products and strategic partnerships focused on privacy-enhancing technologies and services.

Taken together, these excerpts demonstrate that data privacy considerations are deeply embedded in corporate discussions, influencing both risk assessments and strategic decisions. The prevalence and depth of these discussions in high-*DPRisk* transcripts support the economic validity of our methodology, confirming that firms facing higher data privacy risks are indeed more likely to communicate these issues actively and publicly in their earnings calls.

4.4 Association with GDPR Adoption

Given the critical role of the GDPR in shaping global data privacy standards, we validate the economic relevance of *DPRisk* by examining firms' exposure to the GDPR and the intensity of data privacy risk discussions in earnings calls.

Our first approach is to examine the relationship between call-level *DPRisk* with the country-industry-level GDPR exposure developed by Frey and Presidente (2024), which is based on trade flow data from the OECD Inter-Country Input-Output (ICIO) Tables[15]. For each country-industry pair, the GDPR exposure measure captures the degree to which firms are exposed to GDPR-induced compliance obligations through international trade links. Firms with higher GDPR exposure, particularly those in the information and communication technology (ICT) industry, experience greater compliance costs, increased operational expenses, and declined profits (Demirer et al., 2024; Frey & Presidente, 2024). If *DPRisk* effectively reflects stakeholders' concerns about data privacy, we expect such firms to discuss more related issues during their earnings calls.

[Insert Table 3 About Here: Data Privacy Risk and GDPR Exposure]

To this end, we estimate contemporaneous regression of call-level *DPRisk* on GDPR exposure in Table 3. Column (1) shows a statistically and economically significant positive association between *DPRisk* and GDPR exposure, consistent with our hypothesis. This relationship is notably concentrated in the Q&A segment: Column (3) shows larger and more statistically significant coefficients on the GDPR exposure compared to Column (2). The results suggest that data privacy discussions are more likely to be initiated through analyst questions rather than proactively discussed in prepared remarks in the presentation segment. This asymmetry is consistent with two potential explanations: information frictions, wherein managers may underestimate the salience of GDPR exposures to firm fundamentals; or strategic obfuscation, whereby managers may deliberately downplay or withhold potential risks in their scripted disclosures.

[15] The tables follow similar format of standard input-output tables, but also provide a breakdown of of products and services sold directly to final consumers, which fits right in with the nature of the GDPR's regulations on individual consumer data. For each country-industry pair, the GDPR exposure calculates the share of products and services sold to EU consumers over total sales to consumers. In an unreported test, we further adjust the industry-level GDPR exposure into firm-level using the weight of firm-market sales ratio, and the results stay similar under this adjustment.

Our second validation exercise is to examine the responsiveness of *DPRisk* to shifts in exposure to the GDPR. Based on Maex (2022), we conduct the following event study model:

$$DPRisk_{i,j,t} = \beta_1 GDPR\_Segment_{i,t-1} * Post + \beta_2 GDPRSegment_{i,t-1} + \beta_3 Controls + Fixed\ Effects + \varepsilon_{i,t} \qquad (3)$$

where $DPRisk_{i,j,t}$ is the estimated data privacy risk measure of earnings call j for firm i at year t. $GDPR\_Segment_{i,t-1}$ is a firm-level indicator constructed from Compustat Customer Segment data, which is equal to one if firm *i* reports at least one segment operating in the European market subject to the GDPR[16]. $Post$ is a time dummy set to equal one for calls occurring on and after April 14th, 2016, the date the European Parliament formally adopted the GDPR proposal. We include lagged firm-level controls consisting of firm size, profitability (ROA), cash holding ratio, research and development expenditure, selling general and administration fees, growth opportunities (Tobin's Q), and sales ratio. Detailed variable definitions of all variables are provided in Appendix I We also include firm fixed effects and industry-year fixed effects to account for unobservable firm-specific influences and time-varying industry-specific shocks. Heteroskedasticity-robust standard errors are clustered by firm.

Our coefficient of interest is β_1, which captures the effect on *DPRisk* from exposure to GDPR-regulated markets following its adoption. A positive and significant β_1 would indicate that firms that are more affected by GDPR's adoption, through their EU-based business operations, are expected to have more discussions related to data privacy risks.

[Insert Table 4 About Here: Data Privacy Risk and the GDPR Adoption]

Table 4 reports the estimation results. Column (1) shows that data privacy risk discussions in earnings calls increased significantly after the GDPR adoption for firms with operational segments in the European market. Consistent with earlier results in Table 3, the increase is concentrated in the Q&A segment of the calls as Column (3) shows stronger responses compared to the presentation segment results in Column (2). The results consistently show that earnings call discussions shift systematically

[16] The GDPR is directly applicable for all 27 European Union member states and European Economic Area (EEA) countries. Beyond the EU member states, Iceland, Liechtenstein, and Norway began applying GDPR on July 2018 as party to the EEA agreement. The United Kingdom enacted the Data Protection Act 2018 on 23 May 2018 (before the GDPR become effective on 25 May 2018) to incorporate the GDPR into domestic law as "the UK version of GDPR". Considering the timeline of the GDPR adoption, we construct the *GDPR_Segment* based on the 27 countries, but our reported results in Tabe 2 remain robust to the inclusion of the extra four countries.

in response to the GDPR's adoption, particularly among firms with greater regulatory exposure. Call participants, especially analysts, become more focused on regulatory shocks, prompting more frequent and targeted discussions about data privacy risk. These dynamics highlight the value of earnings call transcripts, especially the real-time and unscripted discussions in the Q&A, to be a timely and informative source for capturing firm-level exposure to emerging regulatory risks, as measured by *DPRisk*.

4.5 Association with Privacy Violation Incidents

To further validate the economic and informational content of the *DPRisk* measure, our next test examines its relationship with realized data privacy violation incidents. We focus on all available privacy violation incident records in the RepRisk database, and estimate an OLS regression in which the dependent variable is our call-level *DPRisk* measure and the key explanatory variable is *Previous_violation_dummy*, an indicator of whether a firm has any privacy violation records before the earnings conference call. Following the methods of Florackis et al. (2023), we include a series of lagged firm-level controls consisting of firm size, growth opportunities (Tobin's Q), profitability (ROA), research and development expenditure, tangibility, and cash flow volatility at the industry level. We also control for *DPSentiment*, and corporate governance variables including institutional holding ratio, number of institutional holders. We include year fixed effects and industry fixed effects in the estimation specification, and heteroskedasticity-robust standard errors are clustered by firm.

[Insert Table 5 About Here: Data Privacy Risk and Privacy Violation Incidents]

Table 5 reports the estimation results. Column (1) points out that firms with past privacy violations are significantly more likely to discuss data privacy risk in subsequent earnings calls, while the comparison of columns (2) and (3) illustrates that the pattern is especially salient for the discussions during the Q&A phase instead of the presentation phase. Taken together, the results of Table 5 support the view that our *DPRisk* measure is responsive to realized privacy incidents as a proxy for actual reputational exposure.

4.6 Association with Realized Risk

If *DPRisk* captures a real, material dimension of firm-specific risk arising from regulatory uncertainty and operational complexity, we expect that it correlates positively with realized firm-level stock risk.

We examine whether *DPRisk* exhibits this characteristic, using stock return volatility as a standard proxy for firm-level realized risk (Florackis et al., 2023; Hassan et al., 2019; Wu, 2024). We estimate the following regression:

$$VOL_{i,t} = \beta_1 DPRisk_{i,t} + \beta_2 Controls_{i,t} + \varepsilon_{i,t} \tag{4}$$

where $VOL_{i,t}$ is firm i's realized stock return volatility in quarter t. $Controls_{i,t}$ include firm size and profitability (Hassan et al., 2019; Wu, 2024), the CBOE VIX index, and the economic policy uncertainty (EPU) index (Baker et al., 2016). Industry- and year-fixed effects account for unobserved heterogeneity across industries and over time. Heteroskedasticity-robust standard errors are clustered by firm.

[Insert Table 6 About Here: Realized Stock Return Volatility and Data Privacy Risk]

Table 6 reports a consistently positive and statistically significant relationship between *DPRisk* and realized volatility. Column (1) presents the baseline specification without fixed effects. A one-standard-deviation increase in *DPRisk* is associated with a 0.013 standard deviation[17] increase in realized volatility, which is economically meaningful and comparable to estimates of other risk measures from Wu (2024), Baker et al. (2016), and Hassan et al. (2019). Column (2) adds year fixed effects. The *DPRisk* coefficients remain robust economically and statistically (*t*-stat = 2.47, or 3.13), suggesting that the association is not merely driven by aggregate economic cycles. Column (3) includes both year and industry fixed effects. The coefficients on *DPRisk* again maintain similar magnitude and statistical significance, suggesting that *DPRisk* captures risk exposure beyond temporal and sectoral heterogeneity. Columns (4) to (6) introduce additional controls for aggregate risk indicators, VIX and EPU. Again, the *DPRisk* coefficients remain robust, indicating that the relationship is not an artifact of aggregate economic uncertainty.

[17] The firm-quarter level *DPRisk* has a sample standard deviation of 2.741 and an estimated coefficient of 0.010 in our regression model, while the realized volatility has a standard devation of 2.069 in our sample. Thus, the increase in units of standard deviation of realized volatility is 0.010×(2.741/2.069) ≈ 0.013.

4.7 Time-series and Industry Distribution

We further validate *DPRisk* by examining its time-series behavior and variation across industries. Plot A of Figure 2 presents the quarterly average of *DPRisk* for the full transcript, as well as for the presentation and Q&A sessions separately. Several patterns emerge.

First, *DPRisk* exhibits considerable variation over time, marked by multiple prominent spikes. Increases in *DPRisk* match the timing of major data privacy events such as new developments in regulations and high-profile corporate incidents. The largest spike, in Q2 2020, follows the enforcement of the CCPA, which imposed more stringent privacy obligations on firms[18] (Gupta et al., 2024). Earlier spikes coincide with milestones in the EU's GDPR implementation timeline, including a parliamentary committee orientation vote on the initial proposal in Q3 2013 and enforcement in Q2 2018. The most recent peak, in Q2 2022, corresponds to the EU's adoption of the Digital Services Act and Digital Markets Act, both of which expanded regulatory oversight over digital privacy. In addition to regulatory uncertainty, *DPRisk* also responds to well-publicized firm developments, such as Apple's privacy policy change in Q4 2022 and the Capital One cyber breach in Q3 2019[19]. These patterns suggest that *DPRisk* captures economically meaningful, time-varying firm-level exposure to regulatory and data privacy shocks.

[Insert Figure 2 About Here: Quarterly Average *DPRisk* Exposure Over Time]

Cross-industry differences in *DPRisk* are evident in Plot B of Figure 2, which tracks the time-series evolution of *DPRisk* for select Fama-French 12 industries, and in Figure 3, which presents average *DPRisk* values over the sample period for the full set of industry portfolios. *DPRisk* exhibits both persistence and substantial cross-industry variation. It is consistently high in sectors such as Healthcare, Medical Equipment & Drugs, Money Finance, Business Equipment, and Chemicals & Allied Products. These industries depend heavily on consumer-facing operations and routinely process large volumes of sensitive user data, making them inherently more exposed to privacy-related reputational, legal, and compliance risks. In comparison, *DPRisk* tends to be lower in sectors whose firm fundamentals are tied

[18] We find supportive evidence for this argument again as the market-average data privacy sentiment (*DPSentiment)* and also experiences the largest shock right after the enactment of CCPA in 2020, see figure in Appendix IV.

[19] On December 7th 2022, Apple announced its "Advanced Data Protection" setting providing its highest-ever data protection service sine its "Differential Privacy" policy in June 2016; Capital One data breach announced on July 29, 2019 was one of the largest ever financial information leak in the history that enabled exfiltration of sensitive private credit card application data of 106 million individuals in the United States and Canada (see discussion of Khan et al. (2022) for a sysmetic review).

the most to production capacity, margins, or macro trends, such as Consumer Non-durables, Wholesale, Retail & Services[20], Manufacturing, as well as Energy and Oil & Gas.

Plot B also reveals strong industry-specific persistence over time. We observe that the Healthcare, Medical Equipment & Drugs sector maintains consistently higher average *DPRisk* values than Utilities or Consumer Non-Durables throughout the sample. Similar patterns hold when we use standard deviation instead of mean values (see figure in Appendix V).

[Insert Figure 3 About Here: Average *DPRisk* Exposure Across Industries]

The evidence suggests that *DPRisk* captures persistent and economically intuitive variations in firm-specific exposure to data privacy risk across industries, reflecting both operational reliance on user data and the perceived salience of data privacy concerns.

4.8 Data Privacy Risk Exposure and Firm Characteristics

In a final validation exercise, we investigate the relationship between *DPRisk* and firm-level financial characteristics. This analysis serves two objectives by confirming that *DPRisk* reliably reflects the concerns of stakeholders and offering novel economic insights into how data privacy concerns relate to corporate strategy and performance.

We merge firm-level *DPRisk* with quarterly financial data from Compustat. Our sample spans Q1 2010 to Q4 2023 and includes 62,843 firm-quarter observations across 2,973 unique firms, which covers over 60% of publicly traded U.S. firms in the Compustat universe and accounts for more than 70% of total market capitalization. All continuous variables are winsorized at the 1st and 99th percentiles by year to mitigate the influence of outliers.

[Inset Table 7 About Here: Data Privacy Risk and Firm Characteristics]

Table 7 reports results from a contemporaneous regressions of *DPRisk* on a range of firm characteristics. Based on observed industry trends from Section 4.7, we follow Wu (2024) and use a binary indicator, "*High_DPRisk_Industry*", to identify firms from the data-intensive Healthcare, Medical Equipment &

[20] We note that although the target of Wholesale, Retail & Service industry is highly involved with consumer activities, their service modes do not necessarily include private data collection in the traditional way. However, it should be emphasized that this pattern is rapidly transforming and combing with intensive data collection and analysis in the recent year, therefore we expect a forthcoming change of data privacy risk in the industry level in response.

Drugs, Money Finance, or Business Equipment sectors (according to Fama-French 12 Industry Classification). We control for year and quarter fixed effects in all specifications, with heteroskedasticity-robust standard errors clustered at the firm level.

The regression estimates are consistent with the visual patterns in Figure 3. As expected, firms belonging to the "*High_DPRisk_Industry*" have significantly higher *DPRisk* values, as expected. For the firm fundamentals, *DPRisk* is significantly positively associated with cash holdings, yet negatively associated with SG&A[21] and COGS expenses. This reflects the fact that managing data privacy risk often necessitates substantial operational expenditures and additional human capital investment, requiring firms facing greater data privacy risk to allocate additional financial buffers in anticipation of compliance costs while concurrently cutting or reallocating discretionary spending. *DPRisk* is also associated with lower ROA and Tobin's Q, suggesting that data privacy risk may constrain operating performance and growth opportunities. The relationship between *DPRisk* and size is less consistent across specifications, indicating that data privacy risks may affect firms of all sizes – not just the smaller ones – depending on their perceived operational exposure to privacy issues. The coefficient on R&D is significantly negative, which fits in with the intuition that it may require massive research and development investment to mitigate the risk arising from data privacy practice.

5. Economic Applications and Implications

Having established the statistical and economic validity of the *DPRisk* measure, we apply it to capital market outcomes. In the first application, we investigate whether data privacy risk triggers stock market reactions. Next, we show the effect extends to the debt market.

5.1 Stock Market Reaction

Under the efficient market hypothesis, stock prices reflect all available information in the capital market. Changes in firm valuation following firm disclosures therefore offer a window into how investors perceive emerging risk factors in their expectations about future firm operations. If data privacy risk exposure indeed represents a salient corporate risk, we expect a greater emphasis on data privacy concerns or higher *DPRisk* should trigger a negative market reaction around the earnings call date,

[21] Following the discussion of Ramalingegoowda et al. (2021), we deduct the amount of SG&A expenses with the amount of R&D expenses as the latter one is usually included in the reported SG&A item.

reflecting investor reassessment of firm value in light of potential legal, reputational, or operational costs.

We measure market reaction using cumulative abnormal returns (CARs) over two windows around the earnings call date. Following Price et al. (2012), we calculate CARs for each earnings call j:

$$CAR[-1,1]_{i,j} = \sum_{t=-1}^{1} AR_{i,j,t} \tag{5}$$

$$CAR[2,60]_{i,j} = \sum_{t=2}^{60} AR_{i,j,t} \tag{6}$$

where $AR_{i,j,t}$ is the abnormal return of firm i on day t around an earnings call j that occurs on day 0. Daily abnormal returns are calculated using the risk-adjusted alpha from a six-factor model (Fama-French 5-factor plus momentum) using a 60-day estimation window. The short-term CAR[-1,1] reflects the immediate 3-day market repricing in response to data privacy discourse from the earnings call. CAR[2,60] allows us to examine the persistence of any short-term reaction and test for post-announcement drift or reversal (Florackis et al., 2023; Price et al., 2012).

Figure 4 plots average CARs by *DPRisk* quintiles. We observe a clear and monotonic decline in CAR[-1,1] across quintiles as firms earn an average cumulative abnormal return of 0.2% in the lowest *DPRisk* quintile to -0.15% for those in the highest. While investors react negatively to greater data privacy exposure, the longer-term effect is less clear. CAR[2, 60] does not exhibit a clear, monotonic relationship, which suggests that market response is concentrated in the days immediately surrounding the call and does not systematically reverse in the following two months.

[Insert Figure 4 About Here: Stock Market Return and Data Privacy Risk]

To test this relationship more formally, we estimate the following regression at the earnings call level:

$$CAR_{i,j} = \beta_0 + \beta_1 DPRisk_{i,j} + \beta_2 Controls + \varepsilon_{i,j} \tag{7}$$

where $DPRisk_{i,j}$ is the call-level data privacy risk of firm i disclosed in earnings call j. Consistent with existing literature, we include a comprehensive set of lagged $(t-1)$ firm characteristics as controls: earnings surprise (SUE), firm size (Size), book-to-market ratio, ROA, leverage, trading volume, stock return volatility, and dividend announcement dummy (Declaration). To mitigate the impact of linguistic sentiment characteristics on measurement results, we further control the data privacy sentiment (*DPSentiment*). Variable definitions are reported in Appendix I. Standard errors are robust to heteroscedasticity following Newey and West (1987) and two-way clustered by firm and quarter following Petersen (2009).

[Insert Table 8 About Here: Stock Market Reaction and Data Privacy Risk]

Table 8 shows that across specifications in Columns (1) to (3), *DPRisk* is negatively and significantly associated with immediate market reactions as measured by CAR[-1,1], indicating that investors react negatively when firms engage in more data privacy discussions during their earnings calls. A one-standard-deviation increase in *DPRisk* is associated with a 14.2 to 15.6 basis point[22] reduction in the short-term CAR.

Columns (4) to (6) present the long-term response CAR[2,60]. While the coefficients on *DPRisk* are positive, their sizes are smaller and statistically insignificant. This suggests that the initial negative stock market reaction associated with greater emphasis on privacy discussions does not reverse in the two-month window following the call. We observe similar patterns when we use excess stock return (stock return minus the market return) as an alternative proxy for abnormal return in equations (5) and (6).

We then conduct a series of robustness checks to establish the result. First, we further consider a series of alternative proxies of the short-term CARs within different periods, including CAR[0, 1], CAR[0, 3], and CAR[0, 5] as in Jha et al. (2025) and re-examine the estimation model in the equation (8). Again, we observe significantly negative relationships between the CARs and the *DPRisk*. The related results are reported in Table 9. Second, patterns revealed in Table 9 remain barely changed if we alternatively use the simple stock excess return to proxy the abnormal return in equations (5) and (6).

[Insert Table 9 About Here: Robustness of Stock Market Reaction and Data Privacy Risk]

To sum up, the results in this section indicate that discussions of data privacy risk in earnings calls trigger an immediate and negative response in the equity market. Following the decline in firm value,

[22] The call-level *DPRisk* has a sample standard deviation of 3.46 and an estimated coefficient of -0.041 to -0.045 in our regression model. Thus the reduction in CAR[-1, 1] is ranging from 3.46×0.041 ≈ 0.142 percentage to 3.46×0.045 ≈ 0.156 percentage.

the absence of a reversal in the subsequent two months after the call suggests that the market reaction reflects a reassessment of firm fundamentals that is persistent, rather than a short-lived overreaction. Together, they support the interpretation of *DPRisk* as a priced risk factor with a significant risk premium.

5.2 Cost of Debt Capital

The equity market reactions suggest that stakeholders in the equity market respond quickly to the disclosure of data privacy risks. Naturally, lenders are also stakeholders with capital at risk, and they must also assess firm-specific vulnerabilities when pricing credit. While previous studies document that realized data breach incidents and regulatory changes increase borrowing costs (Agarwal et al., 2024; Gupta et al., 2024; Huang & Wang, 2021), we ask whether anticipated data privacy risk, regardless of a realized breach event, also commands a risk premium in the debt market?

Our hypothesis is motivated by the idea that data privacy risk, even without any actual compliance breach, reflects a latent exposure with potential financial and operational costs. If lenders believe that data privacy concerns are predictive of future legal liability, reputational damage, or operational disruptions, we expect them to price this risk in loan contracting just as they would for more traditional forms of credit risk. In this sense, *DPRisk* offers a forward-looking signal that could shape lender behavior even in the absence of a breach event. Prior work suggests that borrower-side costs may be transmitted to lenders through elevated credit risk, compliance costs, or reputational spillovers (Cheng et al., 2024; Gad et al., 2024).

We test this hypothesis by investigating whether data privacy risk disclosures influence the terms of bank loan contracting. Following the literature on bank loan contracting (Bharath et al., 2011; Campello & Gao, 2017; Gad et al., 2024), we focus on individual loan tranches (facilities) and use all-in-spread-drawn (*AISD*) to measure loan interest spread, which is the additional basis points over LIBOR that borrowers agree to pay. To match the Compustat-Dealscan linking table from Chava and Roberts (2008), our sample coverage starts from January 2010 and ends in June 2020. After merging, the final sample covers 4,525 loan tranches issued to 1,442 unique U.S. listed firms as borrowers. We estimate the following baseline model at the loan-facility level:

$$\ln\left(Loan\_Spread_{l,i,t}\right) = \beta_1 DPRisk_{i,t-1} + \beta_2 Controls + Fixed\ Effects + \varepsilon_{l,i,t} \quad (8)$$

where $Loan\_Spread_{l,i,t}$ is the AISD spread of loan l for the borrower firm i in year t. $DPRisk_{i,t-1}$ is the annual *DPRisk* measure for firm i in the year before the loan origination to ensure the information has been disclosed by the contracting year. $Controls$ represents a set of firm-level and loan-level characteristics. Lagged firm-level controls[23] include firm size, profitability, market-to-book ratio, asset tangibility, leverage, Altman's Z-score, cash holdings, and a credit rating dummy. Loan-level controls include facility size and maturity. We also include multiple levels of fixed effects. Industry fixed effects account for substantial cross-industry variations as documented in Section 4.7. Lender (bank) fixed effects control for differences in pricing strategy, credit screening, and borrower-lender relationships[24]. Year fixed effects, or alternatively, macroeconomic controls (credit spread, term spread, and GDP growth) account for unobservable time-varying factors such as changes in credit conditions over time. Following Campello and Gao (2017), we do not include firm fixed effects given that our unit of analysis is the individual loan tranche. Heteroskedasticity-robust standard errors are clustered by borrower firm and year.

[Insert Table 10 About Here: Bank Loan Spread and Data Privacy Risk]

Table 10 presents the estimation results. Consistent with our hypothesis, we find that borrower firms with larger *DPRisk* in the year preceding loan origination are charged significantly higher interest spreads. Columns (4) and (5) report our fully specified model, including the full set of controls and fixed effects, and the coefficient on *DPRisk* is approximately 0.016 and statistically significant at the 1% level. A one-standard-deviation increase in *DPRisk* is associated with an increase in the load spread by 6.11 to 8.50 basis points[25], representing a 2.7%-3.7% increase in borrowing cost given the sample average loan spread is 224.38 basis points. The magnitude is comparable to the pricing effects of other

[23] To ensure that lenders use the most current accounting information to evaluate borrowers, we follow the modified matching procedure in Bharath et al. (2011) to merge firm-level control variables constructed from Compustat data in an alternative way. Particularly, if the loan is activated at least six months after the fiscal year ending months in the calendar year t, we use Compustat data from the fiscal year t. Otherwise, we keep using the data from the fiscal year t-1. However, we note that the significance and magnitude of results in Table 7 remain if we alternatively use simple matching procedure with one year lag.

[24] We report the empirical results using direct lender as lender fixed effect, although our baseline results stay robust with similar significance levels when we replace it with parent lender (the ultimate parent company of the bank) fixed effect.

[25] The firm-year level *DPRisk* has a sample standard deviation of 2.13 and an estimated coefficient of 0.015 to 0.021 in our regression model. Since the sample mean value of the natural logarithm of loan spread is 5.27, the reduction in loan spread is ranges from e^5.27 – e^(5.27−2.13×0.015) ≈ 6.11 basis points to e^5.27 – e^(5.27−2.13×0.021) ≈ 8.50 basis points.

established determinants of bank loan pricing documented in recent banking literature (Cheng et al., 2024; Gad et al., 2024).

As in the equity market, the results provide evidence that data privacy risk has material consequences for firms' cost of debt capital. Stakeholders in the credit market evidently treat data privacy risk disclosures as credible signals for higher borrower risk and adjust loan pricing accordingly. The findings complement our earlier stock market analysis, lending further support to *DPRisk* as a novel and priced firm-level risk factor that merits further investigation.

6. Conclusion

In this paper, we establish the foundation for incorporating data privacy considerations into corporate finance studies. As data becomes increasingly central to corporate value creation, data privacy risk will likely emerge as a critical determinant of capital allocation, investment decisions, and firm organization. We develop a novel, firm-level measure of data privacy risk by applying textual analysis to earnings call transcripts. The measure is built using a hybrid training library that draws from important data privacy regulations and reflects the language that market participants use in discussions about data privacy risks.

We validate the measure in several ways, demonstrating that it has intuitive time-series and cross-sectional properties. It rises around major regulatory changes and highly publicized privacy incidents, and is higher for industries that rely more on sensitive consumer data, and firms that operate in regions subject to stricter regulations or have prior privacy violation records. We also show that our measure is distinct from other firm-level risk measures established in recent literature with either similarity in concept or methodology, including cybersecurity risk, data breach risk, and digitalization.

We then study how markets respond to this risk. In the equity market, we find that firms with higher data privacy risk experience lower abnormal returns around earnings calls. In the credit market, those same firms are charged higher spreads on syndicated bank loans. Both sets of results show that shareholders and lenders respond to the variation of data privacy risk, and that this information has been priced into capital market outcomes.

Our data privacy risk measure is transparent, easy to implement, and applicable to any firm that holds regular earnings calls or engages in regular interactions with external stakeholders. It provides a useful application for studying how data privacy concerns affect firms and markets so that both researchers and practitioners can understand and manage this critical dimension of modern, intangible corporate risk. We hope future research will build on this framework to study firm responses to data privacy risk, differences across regulatory environments, and how privacy interacts with other forms of intangible risk in shaping corporate outcomes.

References

[1] Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–256. https://doi.org/10.1257/mic.20200200

[2] Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2025). When big data enables behavioral manipulation. *American Economic Review: Insights*, *7*(1), 19–38. https://doi.org/10.1257/aeri.20230589

[3] Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442–492. https://doi.org/10.1257/jel.54.2.442

[4] Agarwal, N., Agarwal, S., & Chatterjee, C. (2024). Data breach notification laws and the cost of private debt. *The British Accounting Review*, 101518. https://doi.org/10.1016/j.bar.2024.101518

[5] Aridor, G., Che, Y.-K., Nelson, W., & Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. *NBER Working Paper*. https://doi.org/10.2139/ssrn.3522845

[6] Asante-Appiah, B., & Lambert, T. A. (2023). The role of the external auditor in managing environmental, social, and governance (ESG) reputation risk. *Review of Accounting Studies*, *28*(4), 2589–2641. https://doi.org/10.1007/s11142-022-09706-z

[7] Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

[8] Banker, R. D., & Feng, C. (2019). The impact of information security breach incidents on CIO turnover. *Journal of Information Systems*, *33*(3), 309–329. https://doi.org/10.2308/isys-52532

[9] Barrett, L. (2018). Confiding in con men: U.S. privacy law, the GDPR, and information fiduciaries. *Seattle University Law Review*, *42*(3), 1057–1114.

[10] Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, 131(3), 1543–1592. https://doi.org/10.1093/qje/qjw018

[11] Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97, 71–87. https://doi.org/10.1016/j.jmoneco.2018.05.013

[12] Belo, F., Lin, X., & Vitorino, M. A. (2014). Brand capital and firm value. *Review of Economic Dynamics*, 17(1), 150–169. https://doi.org/10.1016/j.red.2013.05.001

[13] Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *The Review of Financial Studies*, *24*(4), 1141–1203. https://doi.org/10.1093/rfs/hhp064

[14] Bian, B., Ma, X., & Tang, H. (2021). The supply and demand for data privacy: Evidence from mobile apps. *Unpublished Working Paper. Social Science Research Network.* https://doi.org/10.2139/ssrn.3987541

[15] Blankespoor, E. (2022). Understanding investor interaction with firm information: A discussion of Lee and Zhong (2022). *Journal of Accounting and Economics*, *74*(2–3), 101523. https://doi.org/10.1016/j.jacceco.2022.101523

[16] Boroomand, F., Leiponen, A., & Vasudeva, G. (2022). Does the market value attention to data privacy? Evidence from U.S.-listed firms under the GDPR. *Working Paper*.

[17] Brown, S., Hillegeist, S. A., & Lo, K. (2004). Conference calls and information asymmetry. *Journal of Accounting and Economics*, *37*(3), 343–366. https://doi.org/10.1016/j.jacceco.2004.02.001

[18] Bybee, L., Kelly, B., & Su, Y. (2023). Narrative asset pricing: Interpretable systematic risk factors from news text. *The Review of Financial Studies*, *36*(12), 4759–4787. https://doi.org/10.1093/rfs/hhad042

[19] Campbell, J., Goldfarb, A., & Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*, *24*(1), 47–73. https://doi.org/10.1111/jems.12079

[20] Campbell, J. L., Zheng, X., & Zhou, D. (2025). Number of numbers: Does a greater proportion of quantitative textual disclosure reduce information risk? *Journal of Corporate Finance*, *94*, 102813. https://doi.org/10.1016/j.jcorpfin.2025.102813

[21] Campello, M., & Gao, J. (2017). Customer concentration and loan contract terms. *Journal of Financial Economics*, *123*(1), 108–136. https://doi.org/10.1016/j.jfineco.2016.03.010

[22] Chava, S., & Roberts, M. R. (2008). How does financing impact investment? The role of debt covenants. *The Journal of Finance*, *63*(5), 2085–2121. https://doi.org/10.1111/j.1540-6261.2008.01391.x

[23] Chen, S., Hollander, S., & Law, K. (2016). In search of interaction. *Unpublished Working Paper. Social Science Research Network.* https://doi.org/10.2139/ssrn.2449341

[24] Chen, W., & Srinivasan, S. (2024). Going digital: Implications for firm value and performance. *Review of Accounting Studies*, *29*(2), 1619–1665. https://doi.org/10.1007/s11142-023-09753-0

[25] Cheng, D., Gao, L., Liao, Y., & Pan, Z. (2024). Mind the cost of disturbance: Firm-level supply chain risk and the bank loan cost. *Working Paper*.

[26] Choe, C., Matsushima, N., & Shekhar, S. (2024). The bright side of the GDPR: Welfare-improving privacy management. *ISER Discussion Paper, No. 1246*. https://www.econstor.eu/handle/10419/302245

[27] Cong, W., Harvey, C., Rabetti, D., & Wu, Z.-Y. (2025). An anatomy of crypto-enabled cybercrimes. *Management Science*, 71(4), 3622–3633. https://doi.org/10.1287/mnsc.2023.03691

[28] Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, *20*(2), 639–673. https://doi.org/10.1007/s11142-014-9309-4

[29] Demirer, M., Hernández, D. J. J., Li, D., & Peng, S. (2024). Data, privacy laws and firm production: Evidence from the GDPR. *NBER Working Paper*.

[30] Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, 65(6), 2213–2253. https://doi.org/10.1111/j.1540-6261.2010.01614.x

[31] Eisfeldt, A. L., & Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4), 1365–1406. https://doi.org/10.1111/jofi.12034

[32] Eisfeldt, A. L., Kim, E., & Papanikolaou, D. (2022). Intangible value. *Critical Finance Review*, 11(2), 299–332. https://doi.org/10.1561/104.00000113

[33] El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance*, *35*(9), 2388–2406. https://doi.org/10.1016/j.jbankfin.2011.02.007

[34] Ersahin, N., Giannetti, M., & Huang, R. (2024). Supply chain risk: Changes in supplier composition and vertical integration. *Journal of International Economics*, *147*, 103854. https://doi.org/10.1016/j.jinteco.2023.103854

[35] Fainmesser, I. P., Galeotti, A., & Momot, R. (2023). Digital privacy. *Management Science*, *69*(6), 3157–3173. https://doi.org/10.1287/mnsc.2022.4513

[36] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

[37] Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big data and firm dynamics. *AEA Papers and Proceedings*, *109*, 38–42. https://doi.org/10.1257/pandp.20191001

[38] Farboodi, M., & Veldkamp, L. (2020). Long-run growth of financial data technology. *American Economic Review*, *110*(8), 2485–2523. https://doi.org/10.1257/aer.20171349

[39] Ferracuti, E., Koo, M., Lee, M., & Stubben, S. (2024). Acquisition of customer information and corporate decision making. *Unpublished Working Paper. Social Science Research Network.* https://doi.org/10.2139/ssrn.4682350

[40] Florackis, C., Louca, C., Michaely, R., & Weber, M. (2023). Cybersecurity risk. *The Review of Financial Studies*, *36*(1), 351–407. https://doi.org/10.1093/rfs/hhac024

[41] Frey, C. B., & Presidente, G. (2024). Privacy regulation and firm performance: Estimating the GDPR effect globally. *Economic Inquiry*, *62*(3), 1074–1089. https://doi.org/10.1111/ecin.13213

[42] Gad, M., Nikolaev, V., Tahoun, A., & Van Lent, L. (2024). Firm-level political risk and credit markets. *Journal of Accounting and Economics*, *77*(2–3), 101642. https://doi.org/10.1016/j.jacceco.2023.101642

[43] Goldberg, S. G., Johnson, G. A., & Shriver, S. K. (2024). Regulating privacy online: An economic evaluation of the GDPR. *American Economic Journal: Economic Policy*, *16*(1), 325–358. https://doi.org/10.1257/pol.20210309

[44] Goldfarb, A., & Que, V. F. (2023). The economics of digital privacy. *Annual Review of Economics*, *15*(Volume 15, 2023), 267–286. https://doi.org/10.1146/annurev-economics-082322-014346

[45] Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. https://doi.org/10.1257/jel.20171452

[46] Gomes, O., Mihet, R., & Rishabh, K. (2023). Data risk, firm growth, and innovation. *Unpublished Working Paper. Swiss Finance Institute Research Paper Series*. https://doi.org/10.2139/ssrn.4559921

[47] Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics*, *118*(1), 107–156. https://doi.org/10.1162/00335530360535162

[48] Gow, I. D., Larcker, D. F., & Zakolyukina, A. A. (2021). Non-answers during conference calls. *Journal of Accounting Research*, *59*(4), 1349–1384. https://doi.org/10.1111/1475-679X.12371

[49] Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408.

[50] Gupta, M., McGowan, D., & Ongena, S. (2024). The cost of data privacy law: Evidence from the California mortgage market. *Swiss Finance Institute Research Paper No. 23-25*. https://doi.org/10.2139/ssrn.4404636

[51] Hagendorff, J., Lim, S., & Nguyen, D. D. (2023). Lender trust and bank loan contracts. *Management Science*, *69*(3), 1758–1779. https://doi.org/10.1287/mnsc.2022.4371

[52] Hassan, T. A., Hollander, S., Lent, L. V., & Tahoun, A. (2024). The global impact of Brexit uncertainty. *The Journal of Finance*, *79*(1), 413–458. https://doi.org/10.1111/jofi.13293

[53] Hassan, T. A., Hollander, S., van Lent, L., Schwedeler, M., & Tahoun, A. (2023). Firm-level exposure to epidemic diseases: COVID-19, SARS, and H1N1. *The Review of Financial Studies*, *36*(12), 4919–4964. https://doi.org/10.1093/rfs/hhad044

[54] Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135–2202. https://doi.org/10.1093/qje/qjz021

[55] Hassan, T. A., Schreger, J., Schwedeler, M., & Tahoun, A. (2024). Sources and transmission of country risk. *Review of Economic Studies*, *91*(4), 2307–2346. https://doi.org/10.1093/restud/rdad080

[56] Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*(1), 405–440. https://doi.org/10.1016/S0165-4101(01)00018-0

[57] Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*. https://doi.org/10.1086/688176

[58] Hollander, S., Pronk, M., & Roelofsen, E. (2010). Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, *48*(3), 531–563. https://doi.org/10.1111/j.1475-679X.2010.00365.x

[59] Huang, H. H., & Wang, C. (2021). Do banks price firms' data breaches? *The Accounting Review*, *96*(3), 261–286. https://doi.org/10.2308/TAR-2018-0643

[60] Jamilov, R., Rey, H., & Tahoun, A. (2021). The anatomy of cyber risk. *NBER Working Paper*. https://doi.org/10.3386/w28906

[61] Jaworski, M. V., & Schmeltzer, P. F. (2022, October 25). *California, Colorado, Connecticut, Utah, Virginia, oh my: An enterprise wide data privacy solution to the state privacy law problem*. https://www.clarkhill.com/news-events/news/an-enterprise-wide-data-privacy-solution-to-the-state-privacy-law-problem/

[62] Jha, M., Qian, J., Weber, M., & Yang, B. (2025). ChatGPT and corporate policies. *NBER Working Paper*. https://doi.org/10.3386/w32161

[63] Jia, J., Jin, G. Z., Leccese, M., & Wagman, L. (2025). How does privacy regulation affect transatlantic venture investment? Evidence from GDPR. *NBER Working Paper*.

[64] Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, *132*(1), 126–149. https://doi.org/10.1016/j.jfineco.2018.10.001

[65] Jiang, H., Khanna, N., Yang, Q., & Zhou, J. (2024). The cyber risk premium. *Management Science*, 70(12), 8791–8817. https://doi.org/10.1287/mnsc.2022.02056

[66] Johnson, G. A., Shriver, S. K., & Goldberg, S. G. (2023). Privacy and market concentration: Intended and unintended consequences of the GDPR. *Management Science*, *69*(10), 5695–5721. https://doi.org/10.1287/mnsc.2023.4709

[67] Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*(9), 2819–2858. https://doi.org/10.1257/aer.20191330

[68] Jones, M. L., & Kaminski, M. E. (2020). An American's guide to the GDPR. *Denver Law Review*, *98*, 93.

[69] Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. Journal of Financial Economics, 139(3), 719–749. https://doi.org/10.1016/j.jfineco.2019.05.019

[70] Khan, S., Kabanov, I., Hua, Y., & Madnick, S. (2022). A systematic analysis of the Capital One data breach: Critical lessons learned. *ACM Trans. Priv. Secur.*, *26*(1), 3:1-3:29. https://doi.org/10.1145/3546068

[71] Kim, J.-B., Wang, C., & Wu, F. (Harry). (2024). Privacy breaches and the effect of customer notification. *MIS Quarterly*, *48*(4), 1483–1502. https://doi.org/10.25300/MISQ/2024/17540

[72] Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265–3315. https://doi.org/10.1093/rfs/hhaa079

[73] Li, Z., Lee, G., Raghu, T. S., & Shi, Z. (Michael). (2024). Impact of the general data protection regulation on the global mobile app market: Digital trade implications of data protection and privacy regulations. *Information Systems Research*. https://doi.org/10.1287/isre.2022.0421

[74] Libby, R., Libby, P. A., & Short, D. G. (2014). *Financial accounting* (Eighth edition). McGraw-Hill Education.

[75] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

[76] Maex, S. A. (2022). *Modern privacy regulation, internal information quality, and operating efficiency: Evidence from the general data protection regulation* [Ph.D., Temple University]. https://www.proquest.com/docview/2702195722/abstract/CEA15C8F74894C29PQ/1

[77] Makridis, C. A. (2021). Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018. *Journal of Cybersecurity*, *7*(1), Article 1. https://doi.org/10.1093/cybsec/tyab021

[78] Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, *81*(1), 36–58. https://doi.org/10.1509/jm.15.0497

[79] Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, *86*(4), 1383–1414. https://doi.org/10.2308/accr-10034

[80] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

[81] Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

[82] Noh, J., & Zhou, D. (2022). Executives' blaming external factors and market reactions: Evidence from earnings conference calls. *Journal of Banking & Finance*, *134*, 106358. https://doi.org/10.1016/j.jbankfin.2021.106358

[83] Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, *22*(1), 435–480. https://doi.org/10.1093/rfs/hhn053

[84] Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, *36*(4), 992–1011. https://doi.org/10.1016/j.jbankfin.2011.10.013

[85] Ramalingegowda, S., Utke, S., & Yu, Y. (2021). Common institutional ownership and earnings management. *Contemporary Accounting Research*, *38*(1), 208–241. https://doi.org/10.1111/1911-3846.12628

[86] Sautner, Z., Van Lent, L., Vilkov, G., & Zhang, R. (2023). Firm-level climate change exposure. *The Journal of Finance*, *78*(3), 1449–1498. https://doi.org/10.1111/jofi.13219

[87] Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437–1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

[88] Wei, Z., & Zhu, Y. (2024). Data breach notification laws and corporate payout policy. *Unpublished Working Paper. Social Science Research Network.* https://doi.org/10.2139/ssrn.4715249

[89] Wu, D. (Andrew). (2024). Text-based measure of supply chain risk exposure. *Management Science*, *70*(7). https://doi.org/10.1287/mnsc.2023.4927

[90] Zhang, Y., Zhang, C., & Xu, Y. (2021). Effect of data privacy and security investment on the value of big data firms. *Decision Support Systems*, *146*, 113543. https://doi.org/10.1016/j.dss.2021.113543

Appendix I: Definition of Variables

Variable	Definition	Source(s)
Altman Z-score	(1.2 × working capital + 1.4 × retained earnings + 3.3 × pretax-income + 0.999 × total sales) / total assets.	Compustat
Book-to-Market	Book equity / market value of equity.	Compustat, CRSP
Cash Holding Ratio	Cash and marketable securities / total assets.	Compustat
Cash Flow Volatility (Industry Level)	Industry average of the standard deviation of cash flow from operations to total assets [at]. The standard deviation is estimated for each firm on a rolling basis using information available in the past 5 years. The industry is defined at the two-digit SIC level	Compustat
Credit Rating	A dummy indicator that equals one if the firm has a public credit rating, zero otherwise.	Compustat
Credit Spread	Yield spread between average AAA and BBB rated corporate bonds in the U.S. market.	FRED
Declaration	A dummy indicator variable that equals one if the firm declares dividends within the 3-day window around the conference call (1, 1), and zero otherwise.	CRSP
DPRisk	Firm's exposure to data privacy risk, calculated as the appearance frequency of bigrams related to data privacy conditional on proximity to the synonyms of risk and uncertainty (Hassan et al., 2019), divided by the number of bigrams in the transcript.	Capital IQ
DPSentiment	Firm's net sentiment to data privacy, calculated as the appearance frequency of bigrams related to data privacy, conditional on proximity to words in the positive/negative sentiment dictionary of Loughran and McDonald (2011), divided by the number of bigrams in the transcript.	Capital IQ
EPU	Economic uncertainty index proxies for movements in policy-related economic uncertainty.	Baker et al. (2016)
GDP Growth	Quarterly average of the GDP growth rate of the year.	FRED
GDPR Exousure	The extent to which companies target EU consumers, replicated following the method of Frey and Presidente (2024).	OECD
GDPR Segment	A dummy indicator variable that equals one if the firm reports at least one individual segment operating in the European area that adopts the GDPR, and zero otherwise.	Compustat Customer Segment
Institutional Holding Ratio	Number of shares held by institutional shareholders that own more than 5% of a firm's equity to the total number of shares outstanding.	Thomson-Reuters 13F
Institutional Holder Number	Number of institutional shareholders that own more than 5% of a firm's equity to the total number of shares outstanding.	Thomson-Reuters 13F
Leverage	Total debt / total assets.	Compustat
Loan Maturity	Total number of months to maturity of a loan tranche.	DealScan
Loan Size	Total loan amount in USD million of a loan tranche, deflated as in 2005 dollars.	DealScan
Loan Spread	All-in-spread-drawn (AISD), the additional basis points required in loan contracts over LIBOR.	DealScan
Loan Type	A dummy indicator that equals one if the loan is a term loan contract, otherwise zero if it is a revolver loan contract.	DealScan
Long-term CAR	CAR[2, 60], the 2-month cumulative abnormal return, in percent, where day 0 is the conference call date, where the abnormal returns are estimated using risk-adjusted alpha calculated by Fama-French six factors plus momentum factor.	CRSP
Market-to-book	(Stock price × shares outstanding + total assets – book equity) / total assets.	Compustat
Previous Violation Dummy	A dummy indicator that equals one if a firm has privacy violation records before.	RepRisk
Profitability	EBITDA / total assets.	Compustat

R&D Intensity	Research and development expenditure / total assets.	Compustat
ROA	Net income / total assets.	Compustat
Secured	A dummy indicator that equals one if the loan tranche is secured, otherwise zero.	DealScan
SG&A Intensity	(Selling, general, and administrative expenses – research and development expenditure) / total assets.	Compustat
Short-term CAR	CAR[-1, 1], the 3-day cumulative abnormal return, in percent, where day 0 is the conference call date, where the abnormal returns are estimated using risk-adjusted alpha calculated by Fama-French six factors plus momentum factor.	CRSP
Size	The natural logarithm of total assets.	Compustat
Tangibility	Property, plant, and equipment / total assets.	Compustat
Term Spread	Yield spread between 10-year Treasury bonds and 3-month Treasury bills.	FRED
Tobin's Q	(Total assets – common/ordinary equity + market value of equity) / total assets.	Compustat
VIX	Volatility Index obtained from the Chicago Board Options Exchange.	CBOE
Volatility	The standard deviation of daily returns for the 90 trading day period ending 10 days before the conference call, scaled by one hundred.	CRSP
Volume	The log of the total share trading volume on the day of the conference call (trading volume for NASDAQ firms is divided by two to avoid double counting).	CRSP

Appendix II: Comparison of Training Library across Studies

This table compares the textual analysis training library (term list) of this paper with those from prior literature on related but different concepts. These include cybersecurity risk (Florackis et al., 2023), cyber risk (Jamilov et al., 2021), data risk (Gomes et al., 2023), digitalization (Chen & Srinivasan, 2024), and attention to data privacy (Boroomand et al., 2022).

Study	Theme	Textual Source	Training Library / Term List	Key Characteristics
Florackis et al. (2023)	Cybersecurity Risk	10-K Filings Item 1A Risk Factors	“security”, “system”, “information”, “result”, “business”, “breach”, “data”, “operation”, “customer”, “service”, “failure”, “loss”, “financial”, “damage”, “computer”, “include”, “technology”, “disruption”, “reputation”, “unauthorized” **(Top-20 most common words of the training set, Section 2.4)**	Focuses on **generic operational and technical risk terms** (e.g., “security,” “system,” “failure,” “disruption”) with broad applicability beyond privacy.
Jamilov et al. (2021)	Cyber Risk	Earnings Calls	“cybersecurity”, “network security”, “cyberattack”, “cybercrime”, “cyber threat”, “cyber incident”, “cyber event”, "data loss", "data integrity", "data security", "information theft", "data breach", "data theft", "data leak", "data compromise", "data fraud", "worm", "spyware", "phishing", "trojan", "malware", "ddos attack", "ransomware", "hacker", "hack", "hacked", "card fraud", "card breach", "system outage", "email compromise" **(All terms listed in Appendix B, Table 1)**	Emphasizes **external attack** and **technical threats** (e.g., “cyberattack,” “ransomware,” “phishing”), with less focus on legal obligations or user-related concerns.
Gomes et al. (2024)	Data (Breach) Risk	10-K Filings Item 1A Risk Factors	“computer attacks”, “computer intrusion”, “computer malware”, “cyber threats”, “data stealing”, “datacenter attack”, “digital breach”, “digital leak”, “digital loss”, “information system attacks”, “infrastructure attack”, “network attack”, “network integrity”, “network security”, “network threats”, “programs breach”, “services threat”, “software breach”, “system attack”, “system threat”, “systems attack”, “systems threat”, “third party breach” **(Part of terms listed in Appendix A.2)**	Focuses on data security issues such as **infrastructure vulnerability** and **system integrity** (e.g., “network threats,” “software breach”), and does not address regulatory or consumer-facing considerations.
Chen and Srinivasan (2024)	Digitalization (Level)	10-K Filings Business Description	“analytics”, “proprietary algorithm”, “virtual reality”, “automation”, “autonomous technology”, “artificial intelligence”, “intelligence”, “neural network”, “virtual assistant”, “cognitive computing”, “big data”, “data science”, “data mining”, “data lake”, “devops”, “digital twin”, “edge computing”, “cloud platforms”, “digitalization”, “digital strategy”, “digital marketing”, “business intelligence”, “biometric”, “deep learning”, “machine learning”, “NLP”, “image recognition”, “speech recognition” **(All terms listed in Appendix 1, Table 11)**	Targets **emerging tech and AI adoption** (e.g., “neural network,” “digital twin,” “big data”) as a measure of digital sophistication instead of data privacy issues.
Boroomand et al. (2022)	Attention to Data Privacy	10-K Filings Item 1A Risk Factors	“data”, “security”, “information”, “customer”, “breach”, “data security”, “system”, “security breach”, “access”, “employee”, “result”, “loss”, “business”, “operation”, “failure”, “damage”, “attack”, “center”, “data center”, “service”, “ability”, “facility”, “infrastructure”, “market”, “company”, “investment”, “certain”, “product service”, “insurance”, “law”, “regulation”, “privacy”, “protection”, “state”, “data protection”, “requirement”, “data base”, “application”, “software”, “technology”, “based”, “management”, “platform”, “user”, “party”, “third”, “third party”, “provider”, “service provider”, “processing”, “reply” **(Top terms of seven related topics as listed in Table 1)**	Captures broad **data-, IT-, and regulation-related general disclosure** (e.g., “data center,” “insurance,” “law,” “platform”) typically found in standardized 10-K filings.
This Paper	Data Privacy Risk	GDPR legal text CCPA legal text Earnings calls	“privacy law”, “privacy policy”, “cloud computing”, “private regulation”, “clinical datum”, “private cloud”, “user identity”, “safety datum”, “cloud service”, “personal data”, “client priority”, “datum center”, “datum analysis”, “supervisory authority”, “data subjects”, “data breach”, “processing activities”, “european data”, “consumer personal”, “sensitive personal”, “service provider”, “verifiable consumer”, “consumer privacy”, “information collected”, “medical information”, “privacy protection”, “deidentified information”, “ownership information”, “online privacy” **(Top bigrams from three sources: Earnings call Word2Vec, GDPR, CCPA)**	Tailored to capture material risk factors across **legal, regulatory, and operational dimensions of data privacy risk**. Combines policy-based corpora and unscripted business discussions, including terms like “privacy law,” “supervisory authority,” “consumer privacy,” “client priority”, “privacy protection” to reflect regulatory scrutiny, consumer concerns, and operational mandates.

Appendix III: Excerpts of Data Privacy Risk Discussions

This table presents a set of illustrative excerpts on data privacy issues from earnings call transcripts with data privacy risk (*DPRisk*) values in the top 1% of the distribution. The excerpts span both presentation and Q&A segments and are drawn from different industries. The transcripts are sourced from S&P Capital IQ.

Firm Name	Call Date	Phase	Transcript Text
Maximus, Inc.	2010/11/11	Presentation	The federal government is seeking to further safeguard consumers protected health information in areas that weren't contemplated on the HIPAA such as websites, web portals and social media sites. MAXIMUS was small but strategic contract with the Office of the National Coordinator or what's referred to as ONC to help assess privacy and security practices of entities that are not covered by HIPAA. Under the HITECH Act, ONC must prepare a report to Congress regarding personal health records and develop recommendations on the appropriate privacy and security requirements for the vendors who handle these records.
Absolute Software Corporation	2013/2/12	Presentation	Data privacy regulations are intensifying. While mobility is accelerating and data breaches are on the rise, the regulatory backdrop is intensifying. Organizations and their IT infrastructure must support these new mobile devices' use cases to ensure business success. However, they must all also ensure that governments' risk management compliance standards are fully met.
Kratos Defense & Security Solutions, Inc.	2014/3/11	Presentation	Kratos currently works with the health care organizations around the country to improve patient data security and privacy and to share compliance with HIPAA and HITECH, and our initial focus with Norse will be the delivery of live threat intelligence to help health care providers secure electronic protected health information or ePHI.
Informatica LLC	2014/7/24	Presentation	New data privacy regulations are fueling demand for Informatica Data Masking. To comply with South Africa's Protection of Personal Information Act, a leading cable TV provider selected Informatica to protect sensitive data.
Digital Realty Trust, Inc.	2014/7/29	Q&A	It is important, though, to understand that the privacy concerns of the European Union and the non-Union countries that are inside Europe does have a bit of an impact on operations of data centers through Continental Europe, as well as the islands. That actually ends up being a demand driver for some of our key customers, including some very large cloud providers who are taking a deployment strategy to eliminate the data sovereignty risk that other providers have. So we're happy to see folks adjust the data sovereignty requirements within EMEA.
Check Point Software Technologies Ltd.	2015/10/26	Q&A	I think in terms of the deployment option, we have more deployment option. It can be on our cloud infrastructure. It can be on the -- on a dedicated appliance service customer, ones that -- by the way, that we see a trend. In Europe, people prefer the in-house appliances, I think, for concern of privacy and exporting their data in the U.S. Most people prefer the cloud service, which is very economical and very easy to use as a utility.
Trend Micro Incorporated	2016/2/18	Presentation	Implementing security a lot of time costs a lot of resource and a lot of time for our customers, so we always need to evaluate the security breach pain is larger than the pain of implementing the security. So if we look at today's environment, the biggest customers' pain is always happen when there is new infrastructure change. And today, that infrastructure change in the cloud and data center, there's new environment that is happening. Or in the network because of the next generation data center, hybrid car, the network has become much more complicated. And for the consumer side or on the end-user side, because of all type of new devices, including all the IoTs, those are the new area that customer would have pain. And Trend Micro has been creating the solution for this 3 area for the cloud and data center, we create the solution for hybrid cloud security. For the consumer part or the user -- end-user part, we have complete user protection.

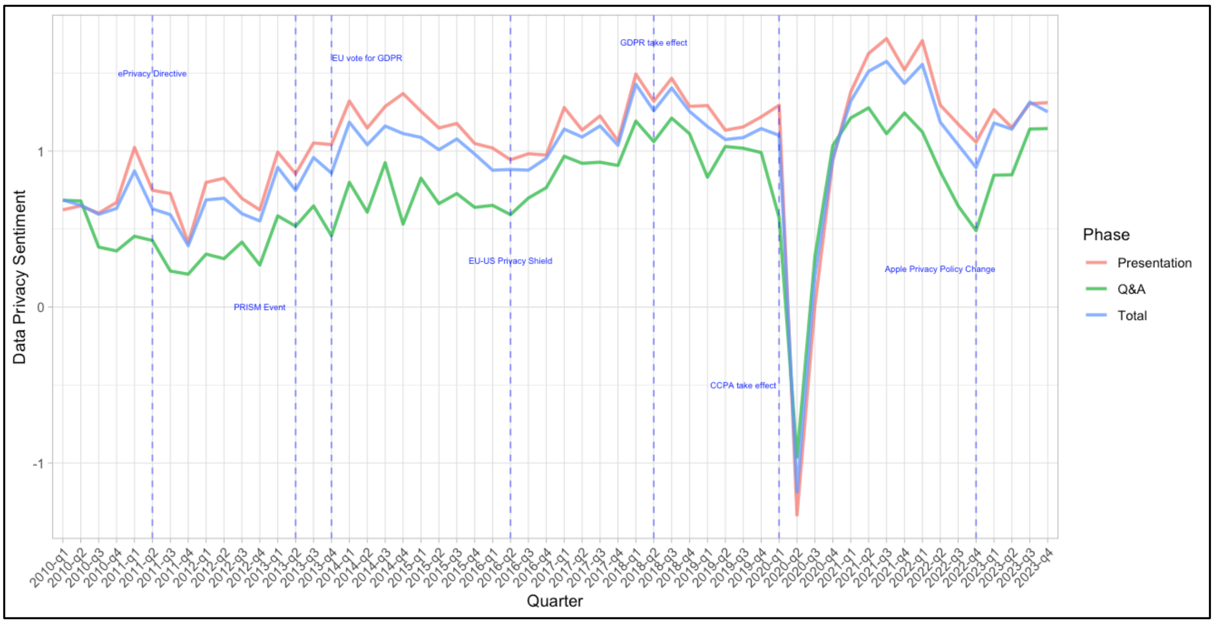
Guidance Software, Inc.	2016/5/5	Presentation	EnForce Risk Manager helps organizations identify, classify and remediate sensitive data based on their own information security policy. Performing these steps reduces an organization's overall information risk profile and makes it more difficult for an adversary to compromise or acquire their critical information. We think of this as reducing the information attack surface area. This approach also aligns with the most modern data protection mandates and global regulations. This product is definitely fulfilling a real customer need. In a survey we conducted with 580 organizations, 60% felt that they did not have well-established solutions to address data privacy and risk requirements, and approximately 85% felt this issue was a priority, and 46% felt it was a top 3 initiative this year.
Proofpoint, Inc.	2016/10/20	Q&A	So with the Privacy products, we had a really extraordinary Q2. Q3 was another great quarter. It didn't quite match Q2's performance, but we do continue to see strong demand for the Privacy product. And it's not just in regulated spaces like financial services or health care where there are actually financial penalties for inadvertently or otherwise sending out protected data on an encrypted basis in e-mail. We see an increasing interest from just broad rank-and-file companies who, among other things, see it as one more elements to help deal with blocking ex filtration threats.
Mandiant, Inc.	2016/11/3	Presentation	The primary market for the MVX Smart Grid are customers in industries that choose to not move their private data to the public cloud due to privacy concerns or regulations. These are industries such as the financial services, government agencies, health care and other regulated industries. Now for these customers, the separation of MVX can dramatically increase throughput, simplify our security architecture and reduce operational overhead. And it also leverages existing training and expertise on FireEye technology, and it protects the customers' investments in our appliances.
VEON Ltd.	2017/2/27	Presentation	Yes, you can do voice and data over these WiFi zones. Now what do you lose? You lose the mobility, and you also lose the security and the privacy that a carrier-grade mobile operator gives to you. Do you want your identity to be stolen? Do you want to take the risk of someone interfering or intercepting what you are just sending either to your bank or to your family or to someone else? No, you don't. We are dealing with a natural resource called spectrum. It is a limited natural resource. There are a few players, 3 to 4 usually in every market. People having a bundle is -- are paying for the privilege of having mobility at any time wherever you are. That is the basis of this industry's pricing.
Varonis Systems, Inc.	2017/5/4	Q&A	It's -- having an impact, it's a bit too early, but what's going on with GDPR, the way the regulation is really is structured, it's addressing data and the scope of the actual solution and the implication if you are not going to solve it. So this is not something that organizations can negotiate. And the other thing that we see is that what we did, it bought a lot of insider threat and data-centric security discussion from the business to IT to the security people.
SecureWorks Corp.	2017/6/6	Q&A	So we are seeing increased opportunities and increased demand across a whole host of verticals. And I think GDPR playing into that, particularly outside of North America is clearly a concern coming up in all of our contracts and all of our negotiations and discussions. We have a very focused effort to ensure that we implement all of the requirements and our compliance in the next 12 months. And we've had activities within the company underway across each of the different work streams to make this a success. I think -- by the way, I think it's an increased opportunity for us as we're addressing the market.
Qualys, Inc.	2017/8/2	Q&A	It is now accelerating the digital transformation of many of the large European companies, because of the risk they now incur to up to 4% of their revenues, as you know, if they cannot essentially demonstrate, in case of a breach, to the regulators that they've done everything possible to ensure the security and the privacy of the data of their customers. So these companies, and I did mention Societe Generale specifically last time, essentially, what they realize is that continuing securing their current infrastructure with these Enterprise point solutions that are difficult to deploy and to integrate and very costly, is in essence, a little bit futile.

Intersections Inc.	2017/8/10	Presentation	The future market opportunity for identity theft prevention and monitoring services is driven by 3 major trends. The first trend that will drive consumer and corporate demand for new and innovative identity theft protection services is the explosion of data breaches in the U.S.A. The identity theft resource [causes] 1,093 data breaches in 2016, a 40% increase over 2015. The culmination of these data breaches have now exposed so much PII of U.S. consumers that I believe consumers are faced with decades of potential misuse of their PII for a variety of nefarious purposes.
Cisco Systems, Inc.	2017/8/16	Presentation	From a security standpoint, the new network enables our customers to detect threats in encrypted traffic with unprecedented accuracy using Cisco's Encrypted Traffic Analytics and intelligence from Cisco's Talos cyber intelligence unit. We have created the only network that is designed for security while maintaining privacy, solving a previously unsolvable problem.
Bank of Montreal	2018/5/30	Presentation	From a risk perspective, we are now in an increasingly digital age that brings benefits to our customers but also new challenges for industries like us. Within this changing landscape, information and cyber security has been an ongoing priority for some time and shall remain so. We will continue to enhance our layered defenses, and learnings from incidents like this only strengthen us and our industry. Our commitment to customer privacy and security is unflinching.
Jianpu Technology Inc.	2018/11/19	Presentation	Our collaboration with financial service providers are not limited to facilitating loans and issuing credit cards. We increasingly focus on leveraging our big data and risk management technology to enable financial service providers to further enhance their operating efficiency and improve their decisioning capability in the last quarter. In July, we launched a joint modeling laboratory, where we can closely work together with financial institutions to develop statistical models that integrate our big data and AI technology into financial service providers application environment.
Mimecast Services Limited	2019/8/5	Presentation	Additionally, we believe the regulatory environment has raised the stakes for organizations regarding the cost of breach data as recent fines levied under GDPR are orders of magnitude larger than previously observed. Our archiving services help organizations comply with a wide range of regulatory requirements, including GDPR, HIPAA and FINRA. We were recently recognized as the best e-mail security solution by SC Awards Europe. And this builds on our reputation as a leader and positions the company well for customers across the region.
Change Healthcare Inc.	2020/2/13	Q&A	So I wanted to ask about the regulatory landscape around data privacy, particularly CCPA out here in California, the California Consumer Protection Act. How do you see this affecting your positioning in the market with respect to data and your go-to-market strategy? And I guess, net-net, is it a good thing or a bad thing for you?
Roku, Inc.	2021/5/6	Q&A	I'll just give you a simple example of how having a first-party relationship with the consumer helps in the world of increased privacy regulation. So often the way the regulation manifests itself is requiring that you get opt-in permission from consumers before you track your data or target them. And so for example, we have a feature -- we have ACR in our TVs, automatic content recognition. And that's an example of a feature where we felt like, okay, we should really get -- the regulation is a little unclear, but we felt like we should have consumers opt-in for that feature before we start collecting ACR data.
Deutsche Telekom AG	2021/11/12	Q&A	We need -- as well from a legislation perspective, the cloud act and the GDPR, the European -- is not asynchronized at the point in time, so there's a lot of legal uncertainty here which runs to decisions. So we need a kind of solution for the European environment. I think it is not about only infrastructure. It's as well on platform as a service.
Teladoc Health, Inc.	2022/2/22	Q&A	So I think we were really impressed to hear that BetterHelp customer acquisition cost is down. Can you maybe help us understand how some of the consumer privacy initiatives from Apple and Google, specifically Apple's app tracking transparency policy might be impacting your DTC advertising cost and productivity? And kind of what are you guys doing to mitigate the impact of those policies?

Wolverine World Wide, Inc.	2022/5/11	Presentation	As an example, Sweaty Betty will be launching a new SMS trial in the U.S. market, leveraging the work already done around this in our other brands. This functionality will support diversification in channel mix and offset privacy challenges in our current e-mail-based CRM programs. In addition, to build on the success of the insiders loyalty program, we are launching a new perk of the month to support trade and returning customer frequency.
Baozun Inc.	2022/5/26	Presentation	Another typical trend is brand's efforts in setting up China for China IT systems. In one of our recent China for China project with a leading international sportswear brand, we also launched a one-team methodology by fully integrating our e-commerce partner team, our IT team and the brand partner's team. We work together and we -- as one team. We focus not only on commercial and merchandising, but also on consumer privacy protection and lifetime value creation.
Mastercard Incorporated	2022/7/28	Q&A	The second piece is what do you do with all of that data, retail and commerce, travel lodging? I give you one example. Many other customers are trying to understand how to make -- run their business better using the payments data that is thrown off and we helped them with that. Our Dynamic Yield acquisition is one of them, where we help customers, retail and commerce customers, engage their end customers in a more targeted fashion. I imagine the landing page has now has personalized offers, in our case, always with strong consent from consumers and a focus on data privacy.
FinVolution Group	2022/8/23	Presentation	Furthermore, as part of our ongoing efforts in enhancing our data privacy and information security framework, we have obtained ISO 27001 certification of information security management system issued by DNV, a well-known international standard certification organization. We will continue to reinforce our ESG engagement on multiple levels, while leveraging FinVolution's innovative technology and differentiators in ways that improve and sustain communities for generations to come.
TechTarget, Inc.	2023/2/9	Q&A	When you look at our customers, their sales and marketing departments have to be modernized, and they really want to focus on first-party data, having access to real first-party data, not only at the account level but the individual prospect level, is really critical to our customers. And privacy and compliance concerns that continue to go -- address this market really put us -- those long-term elements and those long-term, I'll call it, tailwinds don't really go away even during a downturn in the macro.
Mastercard Incorporated	2023/4/27	Q&A	Fundamentally, though, I think we all have to be aware that the application of AI needs to be done in a principled way. We approach data privacy in a principal way. We approach crypto space in a principled way and the same thing applies here. So trustworthy the AI is clearly the focus. We've encouraged our employees to experiment with the technology but we set very clear guardrails and don't do it in production. But it's something that we cannot afford to ignore, we will not. We will lean in, but make sure that we are a trusted party when it comes to scaling it up.
The Kroger Co.	2023/11/30	Presentation	While we pride ourselves on data and insights, the patient data in our pharmacy operations is separate from our customer loyalty data and is protected by privacy laws. We are committed to using our data in an appropriate manner while not jeopardizing customers' trust.

Appendix IV: Time Series of Data Privacy Sentiment

This figure plots the quarterly time series of average data privacy sentiment (*DPSentiment*) from 2010 to 2023. Key developments in data privacy regulations and commercial data practices are marked by vertical dashed lines. *DPSentiment* is aggregated across all earnings call transcripts and is shown separately for the full transcript, the presentation segment, and the Q&A segment.



Appendix V: Dispersion of Data Privacy Risk across Industries

This figure shows the standard deviation of *DPRisk* across 12 industries classified based on the Fama-French 12 industry portfolios.

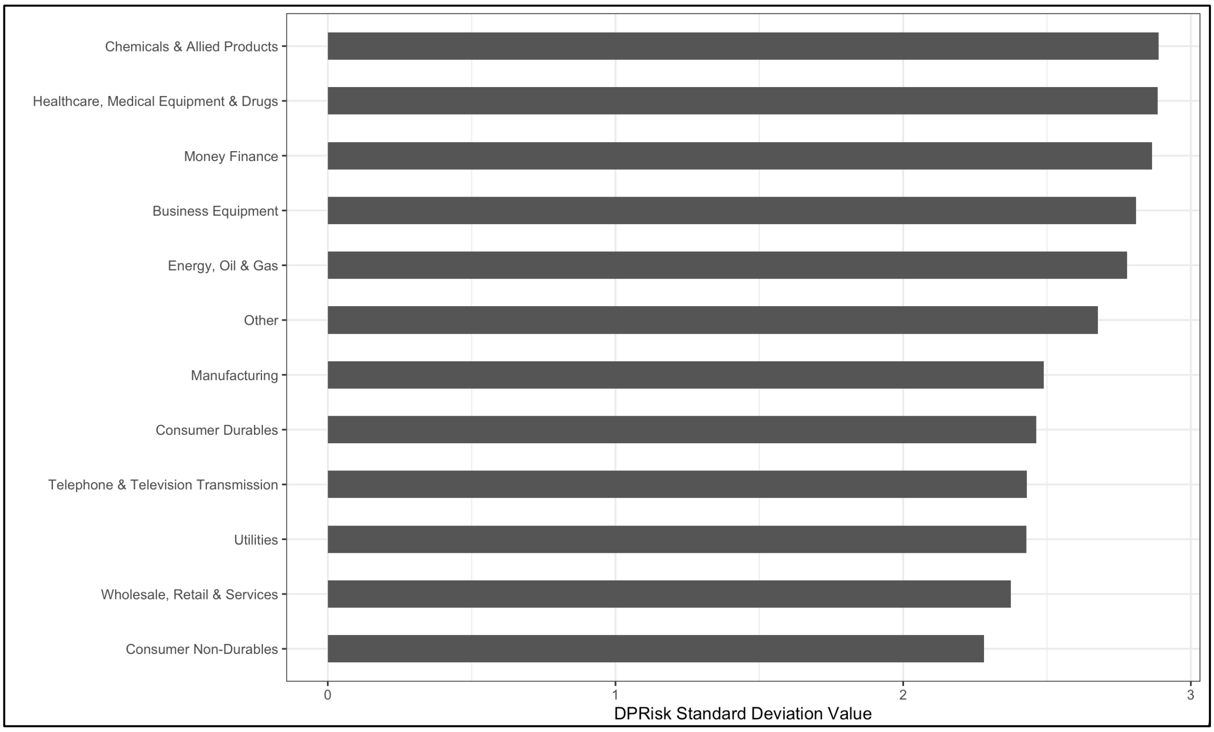
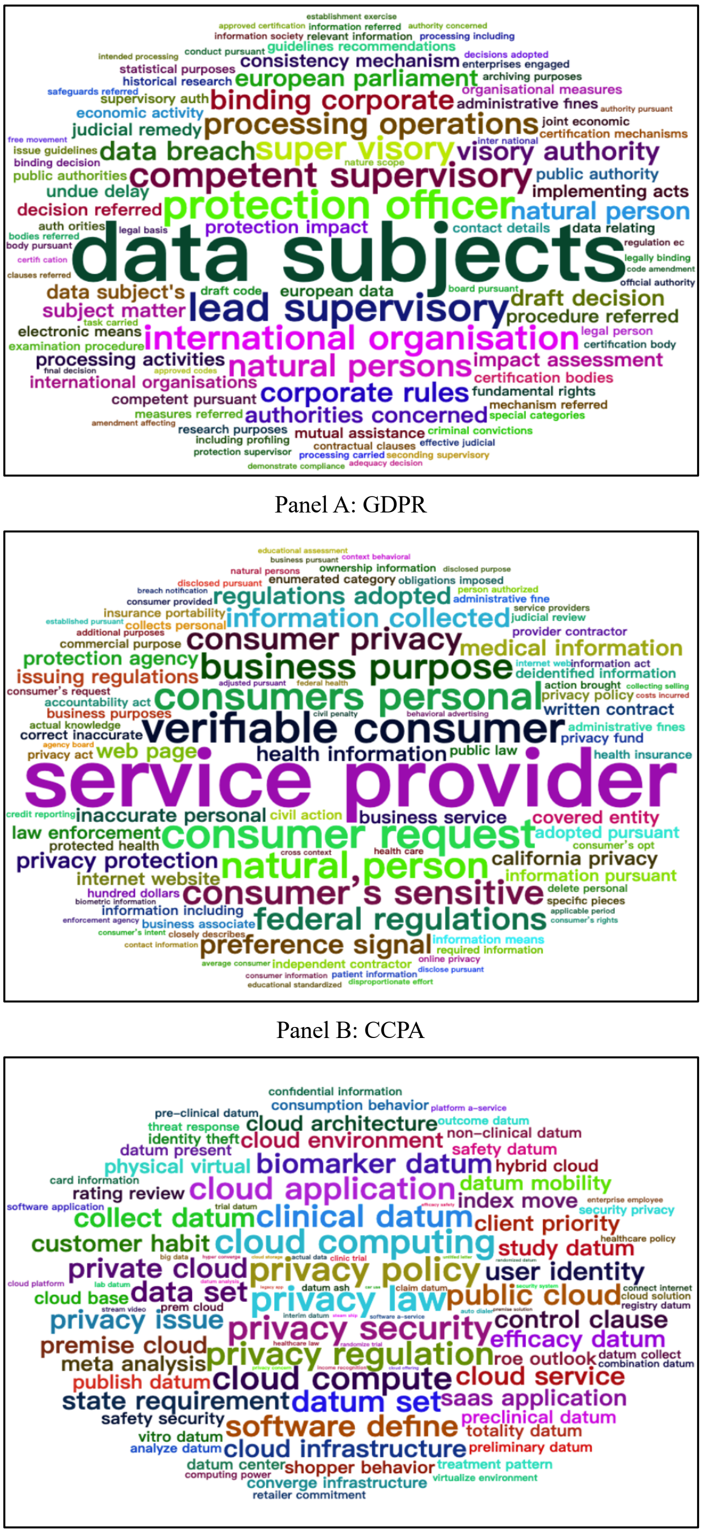


Figure 1: Word Clouds of Data Privacy Bigrams

This figure displays word clouds of the most frequent data privacy-related bigrams constructed from three different textual sources. Panel A is based on the GDPR legal text; Panel B on the CCPA legal text; and Panel C on a Word2Vec model trained on earnings call transcripts. Each word cloud highlights the most salient privacy-related bigrams specific to its respective source corpus.



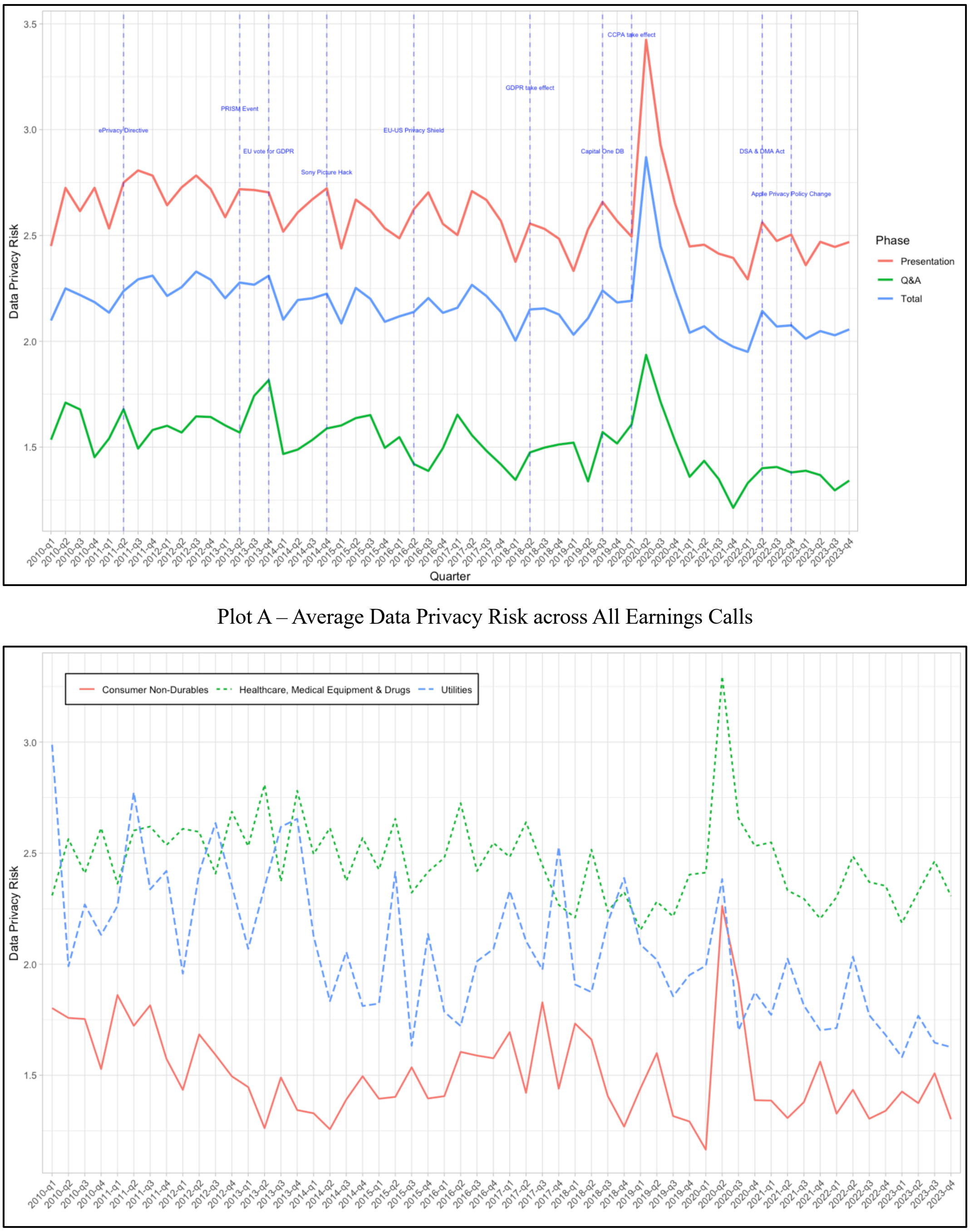
Panel A: GDPR

Panel B: CCPA

Panel C: Word2Vec

Figure 2: Time Series of Data Privacy Risk Exposure

This figure plots the quarterly time series of average data privacy risk (*DPRisk*) exposure from 2010 to 2023. Key developments in data privacy regulations and commercial data practices are marked by vertical dashed lines. Plot A presents *DPRisk* aggregated across all earnings call transcripts, shown separately for the full transcript, the presentation segment, and the Q&A segment. Plot B shows aggregate *DPRisk* for three selected industries: Consumer Non-Durables, Healthcare, Medical Equipment & Drugs, and Utilities. Industries are classified based on the Fama-French 12 industry portfolios. *DPRisk* is scaled by a factor of 1,000 for readability.



Plot A – Average Data Privacy Risk across All Earnings Calls

Plot B – Average Data Privacy Risk of Select Industries

Figure 3: Average Data Privacy Risk Exposure across Industries

This figure displays the average level of data privacy risk (*DPRisk*) exposure by industry. Firms are classified into 12 industries based on the Fama-French 12 industry portfolios. *DPRisk* is scaled by a factor of 1,000 for readability.

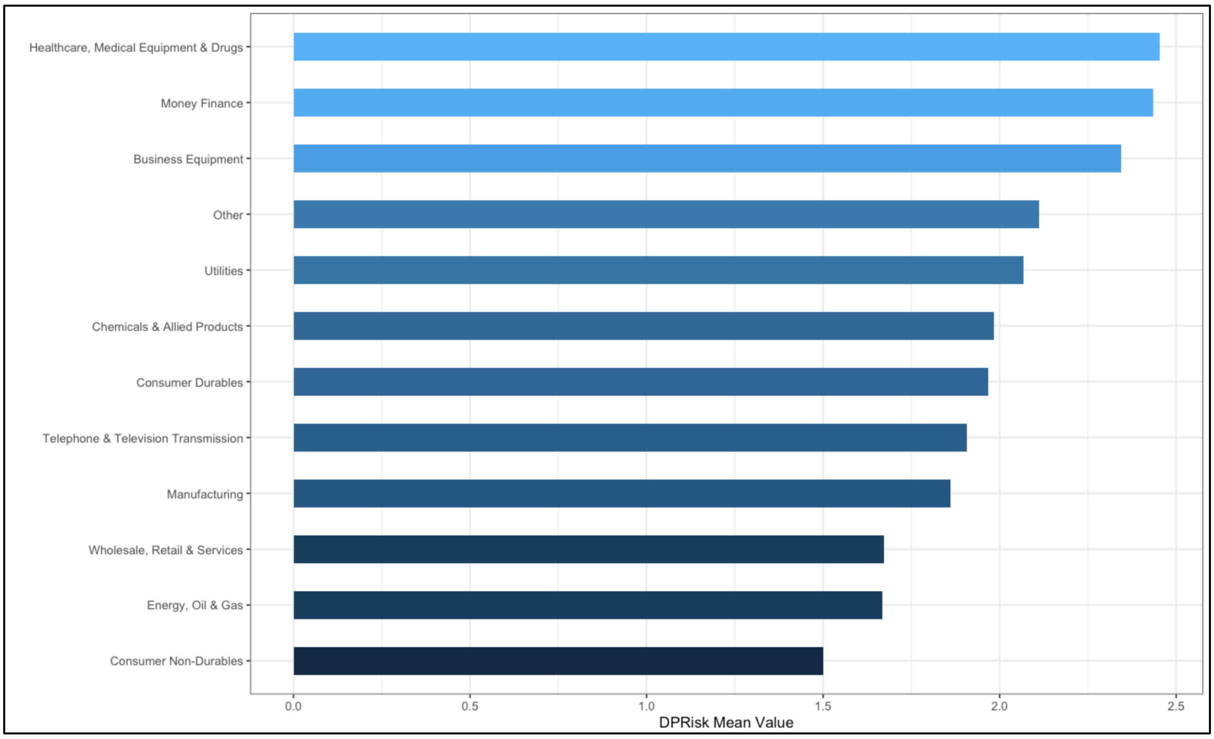


Figure 4: Stock Market Reaction and Data Privacy Risk

This figure displays the average cumulative abnormal return (CAR) following earnings calls. Sample observations are sorted into quintiles based on call-level data privacy risk (*DPRisk*). Panel A presents short-term stock market reactions, proxied by CAR[-1, 1] in percentage terms. Panel B presents long-term stock market reactions, proxied by CAR[2, 60]. CARs are estimated using the Fama-French five-factor model augmented with a momentum factor.

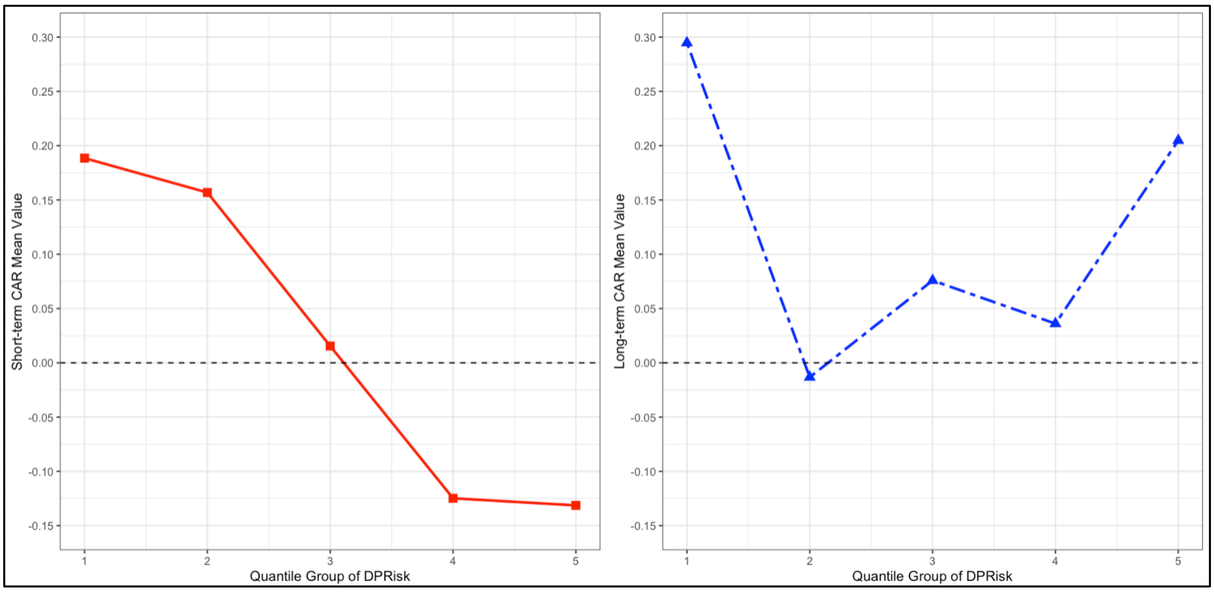


Table 1: Summary Statistics

This table reports summary statistics of variables, including mean, standard deviation, 25% percentile (P25), median, and 75% percentile (P75). Panel A reports firm-call level variables, Panel B reports firm-quarter level variables, and Panel C reports firm-year level variables. Detailed variable definitions of all variables are provided in Appendix I.

Variable	N	Mean	SD	P25	Median	P75
Panel A: Firm-Call Level Variables						
DPRisk (Total)	159,066	2.172	2.844	0.000	1.401	3.120
DPRisk (Presentation)	159,043	2.592	3.905	0.000	1.550	3.774
DPRisk (Q&A)	156,310	1.512	4.109	0.000	0.000	0.000
DPSentiment	159,066	1.043	4.486	-1.282	0.732	2.967
Panel B: Firm-Quarter Level Variables						
DPRisk	62,849	2.148	2.582	0.000	1.416	3.135
DPSentiment	62,849	1.704	4.668	-0.997	1.167	3.670
Size	62,849	7.089	1.992	5.696	7.053	8.436
ROA	62,849	-0.005	0.054	-0.013	0.008	0.020
Cash Holding	62,849	0.244	0.218	0.066	0.173	0.373
R&D Intensity	62,849	0.168	0.539	0.012	0.069	0.169
SG&A	62,849	0.061	0.539	0.012	0.045	0.083
COGS	62,849	0.127	0.129	0.039	0.088	0.166
Tobin's Q	62,849	2.284	2.073	1.038	1.594	2.702
Sales	80,363	0.842	0.620	0.436	0.709	1.094
Institutional Holding Ratio	41,666	0.267	0.140	0.158	0.254	0.355
# of Institutional Holders	41,666	3.187	1.539	2.000	3.000	4.000
Cash Flow Volatility (Ind)	41,666	2.626	4.252	0.344	1.459	3.383
GDPR Exposure	106,236	0.350	1.393	0.003	0.020	0.121
GDPR Segment	80,363	0.038	0.190	0.000	0.000	0.000
Previous Violation	41,666	0.044	0.206	0.000	0.000	0.000
Stock Return Volatility	129,438	2.908	2.067	1.685	2.402	3.531
Panel C: Firm-Year Level Variables						
DPRisk	16,264	2.163	2.390	0.573	1.523	2.997
DPSentiment	16,264	0.854	3.449	-0.823	0.637	2.245
PRisk	16,264	127.474	175.425	39.411	79.046	151.776
CCRisk (*10^3)	16,264	0.032	0.179	0.000	0.000	0.000
SCRisk	16,264	415.890	938.716	132.061	223.077	388.170
Risk	16,264	65.359	53.471	33.782	53.436	81.418
Sentiment	16,264	863.371	454.783	572.694	857.770	1146.513
CyberRisk	16,264	0.334	0.205	0.191	0.405	0.492

Table 2: Correlation Matrix of Data Privacy Risk and Related Indicators

This table provides a correlation matrix between data privacy risk (*DPRisk*), data privacy sentiment (*DPSentiment*), and other firm-level, textual-based sentiment and risk indicators from the literature. Political risk (*PRisk*), all-content risk (*Risk)*, and all-content sentiment (*Sentiment*) are from Hassan et al (2019). Climate change risk (*CCRisk*) is from Sautner et al. (2023). Cybersecurity risk (*CyberRisk*) is from Florackis et al. (2023). Supply chain risk (*SCRisk*) is from Ersahin et al. (2024).

	DPRisk	DPSentime	PRisk	CCRisk	CyberRisk	SCRisk	Risk	Sentiment
DPRisk	1.000	-	-	-	-	-	-	-
DPSentiment	-0.004	1.000	-	-	-	-	-	-
PRisk	0.122	-0.093	1.000	-	-	-	-	-
CCRisk	0.012	-0.011	0.070	1.000	-	-	-	-
CyberRisk	-0.039	0.120	-0.018	-0.024	1.000	-	-	-
SCRisk	0.087	-0.019	0.101	0.031	-0.026	1.000	-	-
Risk	0.155	-0.140	0.058	0.058	-0.037	0.160	1.000	-
Sentiment	-0.091	0.392	-0.029	-0.029	0.139	-0.026	-0.264	1.000

Table 3: Data Privacy Risk and GDPR Exposure

This table analyzes the relationship between data privacy risk (*DPRisk*) and GDPR exposure using OLS regressions. *DPRisk*, the dependent variable, is constructed at the earnings call level for different segments (presentation, Q&A, and the full transcript) and scaled by a factor of 1,000 for readability. GDPR exposure is a measure of exposure to GDPR obligations using country-industry trade flows following Frey and Presidente (2024). Detailed variable definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var.	Data Privacy Risk (*DPRisk*)		
Phase	Total	Presentation	Q&A
	(1)	(2)	(3)
GDPR_exposure	61.584**	-22.798	90.159***
	(2.00)	(-0.39)	(2.66)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
# Observations	106,960	106,945	105,280
Adjusted R2	0.366	0.430	0.102

Table 4: Data Privacy Risk and GDPR Adoption

This table reports the relationship between data privacy risk (*DPRisk*) and the adoption of GDPR in an event study framework using OLS regressions. *DPRisk*, the dependent variable, is constructed at the earnings call level for different segments (presentation, Q&A, and the full transcript) and scaled by a factor of 1,000 for readability. *GDPR_Segment* is a firm-level dummy variable equal to one if a firm has at least one segment operating in the European market subject to the GDPR. *Post* is a dummy variable equal to one for calls that occur on or after April 14th, 2016, the date the European Parliament adopted the GDPR proposal. Firm-level controls include firm size, profitability (ROA), cash holding ratio, research and development expenditure, selling general and administration fees, growth opportunities (Tobin's Q), and sales ratio. Detailed variable definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var.	Data Privacy Risk (*DPRisk*)		
Phase	Total	Presentation	Q&A
	(1)	(2)	(3)
GDPR_Segment*Post	0.387*	0.311	0.512**
	(1.65)	(1.18)	(2.35)
GDPR_Segment	-0.014	-0.155	-0.031
	(-0.09)	(-0.79)	(-0.19)
Firm Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Industry*Year FE	Yes	Yes	Yes
# Observations	80,953	80,937	79,768
Adjusted R2	0.431	0.485	0.127

Table 5: Data Privacy Risk and Privacy Violation Incidents

This table reports the relationship between data privacy risk (*DPRisk*) and the privacy violation record using OLS regressions. The dependent variable is *DPRisk* constructed at the earnings call level for different segments (presentation, Q&A, and the full transcript) and scaled by a factor of 1,000 for readability. The main independent variable is the *Previous_violation_dummy* indicating whether a firm has privacy violation records or not. Control variables include *DPSentiment*, size, growth opportunity (Tobin's Q), profitability (ROA), tangibility, research and development expenditure, intuitional holding ratio, number of institutional holders, and cash flow volatility at the industry level. Detailed variable definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var.	Data Privacy Risk (*DPRisk*)		
Phase	Total	Presentation	Q&A
	(1)	(2)	(3)
Previous_violation_dummy	0.424**	0.327	0.380**
	(2.10)	(1.30)	(1.98)
Firm Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Observations	41,993	41,983	41,457
Adjusted R-squared	0.111	0.117	0.041

Table 6: Realized Stock Return Volatility and Data Privacy Risk

This table reports the relationship between realized stock return volatility and data privacy risk (*DPRisk*) using OLS regressions. Realized stock return volatility is measured as the standard deviation of daily stock returns within a given quarter. *DPRisk* is aggregated at the quarterly level using all earnings call transcripts within the corresponding quarter. Control variables include firm size (natural logarithm of total assets), profitability (return on assets), the VIX index, and the Economic Policy Uncertainty (EPU) index. Detailed variable definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var	Realized Volatility					
	(1)	(2)	(3)	(4)	(5)	(6)
DPRisk	0.010*	0.009**	0.009***	0.006**	0.007**	0.007**
	(2.70)	(2.47)	(2.74)	(2.05)	(2.18)	(2.32)
Size	-0.285***	-0.306***	-0.282***	-0.274***	-0.283***	-0.271***
	(-39.30)	(-42.40)	(-11.90)	(-11.57)	(-11.93)	(-11.47)
ROA	-5.884***	-5.275***	-2.275***	-2.17***	-2.267***	-2.17***
	(-20.747)	(-20.21)	(-10.48)	(-10.16)	(-10.44)	(-10.17)
VIX				0.125***		0.143***
				(71.59)		(71.87)
EPU					0.003***	0.004***
					(17.61)	(19.04)
Year FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes	Yes
# Observations	130,762	130,762	130,762	130,762	130,762	130,762
Adjusted R2	0.183	0.278	0.434	0.459	0.436	0.460

Table 7: Data Privacy Risk and Firm Characteristics

This table explores the relationship between quarterly data privacy risk (*DPRisk*) and a range of firm characteristics using contemporaneous OLS regressions in which *DPRisk* is the dependent variable and firm characteristics are the independent variables. *High DPRisk industry* is a dummy variable equal to one if the firm belongs to either Healthcare, Medical Equipment & Drugs, Money Finance, or Business Equipment sectors, based on the Fama-French 12 Industry Portfolios. The sample spans from Q1 2010 to Q4 2023. Detailed definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var.	Data Privacy Risk (*DPRisk*)					
	(1)	(2)	(3)	(4)	(5)	(6)
High DPRisk Industry	0.454***	0.437***	0.338***	0.320***	0.249***	0.260***
	(5.73)	(5.44)	(4.09)	(3.83)	(2.83)	(2.95)
Number of Bigrams	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(-6.74)	(-5.97)	(-6.39)	(-6.21)	(-6.41)	(-6.18)
Size		-0.024	0.017	-0.001	-0.002	0.000
		(-1.10)	(0.68)	(-0.05)	(-0.09)	(-0.01)
ROA			-1.226**	-1.89***	-1.713***	-1.427**
			(-2.36)	(-3.38)	(-3.03)	(-2.44)
Cash Holding			0.676***	0.783***	0.691***	0.824***
			(3.77)	(4.25)	(3.73)	(4.35)
R&D Intensity				-0.115***	-0.125***	-0.124***
				(-2.96)	(-3.21)	(-3.19)
SG&A				-1.547**	-1.047	-0.614
				(-1.98)	(-1.30)	(-0.749)
COGS					-0.759**	-0.816***
					(-2.43)	(-2.61)
Tobin's Q						-0.041***
						(-2.60)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	62,849	62,849	62,849	62,849	62,849	62,849
Adjusted R2	0.019	0.020	0.023	0.024	0.025	0.026

Table 8: Stock Market Reaction and Data Privacy Risk

This table explores the relationship between stock market reactions and data privacy risk (*DPRisk*) using OLS regressions. Stock market reaction is measured by cumulative abnormal return (CAR). Columns (1) to (4) report short-term responses using CAR [-1, 1], and columns (5) to (8) report long-term responses using CAR [2, 60]. CARs are estimated using the Fama-French five-factor model augmented with a momentum factor. Detailed definitions of all variables are provided in Appendix I. Standard errors are heteroscedasticity-robust following Newey and West (1987) and clustered by firm and quarter following Petersen (2009). T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep.Var.	Short-term CAR			Long-term CAR		
	(1)	(2)	(3)	(4)	(5)	(6)
DPRisk	-0.041***	-0.045***	-0.044***	0.023	0.027	0.028
	(-3.32)	(-3.39)	(-3.25)	(0.91)	(0.97)	(1.01)
Size	0.005	-0.032*	0.038	-0.206***	-0.221***	-0.251***
	(0.24)	(-1.66)	(1.41)	(-5.05)	(-5.55)	(-4.95)
Book-to-Market	0.077**	0.118***	0.117***	0.117	0.100	0.054
	(2.05)	(2.95)	(2.71)	(1.34)	(1.19)	(0.61)
Profitability	0.013	0.029***	0.025**	0.026	0.029	0.032
	(1.26)	(2.86)	(2.30)	(1.14)	(1.03)	(1.09)
Leverage	0.0003	0.093	0.045	0.002	0.318	0.292
	(0.18)	(0.83)	(0.41)	(0.51)	(1.60)	(1.49)
DPSentiment		0.108***	0.106***		0.014	0.013
		(13.07)	(12.66)		(0.82)	(0.74)
SUE			7.641***			5.156***
			(12.89)			(3.76)
Volume			-0.000***			0.000
			(-3.33)			(1.60)
Volatility			0.001**			0.000
			(2.08)			(0.50)
Declaration			0.215***			0.212*
			(2.87)			(1.81)
Constant	0.188	0.311**	-0.336	1.677***	1.515***	1.492***
	(1.25)	(2.01)	(-1.50)	(5.43)	(5.21)	(3.38)
Observations	95,109	95,109	95,109	95,109	95,109	95,109
Adjusted R2	0.0002	0.002	0.007	0.001	0.001	0.001

Table 9: Robustness of Stock Market Reaction and Data Privacy Risk

This table reports robustness checks of the relationship between cumulative abnormal returns (CARs) and data privacy risk (*DPRisk*), using alternative event windows and return specifications in OLS regressions. CARs are estimated using stock excess returns in Panel A, and using six-factor (Fama-French five-factor plus momentum) risk-adjusted alphas in Panel B. Columns (1) and (2) report results using CAR[0, 1]; columns (3) and (4) use CAR[0, 3]; and columns (5) and (6) use CAR[0, 5]. Firm-level controls include size, book-to-market ratio, profitability, and leverage. Call-level controls include DPSentiment, SUE, volume, volatility, and a dividend declaration dummy equal to one if a dividend was declared within the [-1,1] window around the earnings call. Detailed definitions of all variables are provided in Appendix I. Standard errors are heteroscedasticity-robust following Newey and West (1987) and clustered by firm and quarter following Petersen (2009). t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: CAR Constructed by Stock Excess Return

Dep.Var.	CAR[0, 1]		CAR[0, 3]		CAR[0, 5]	
	(1)	(2)	(3)	(4)	(5)	(6)
DPRisk	-0.045***	-0.044***	-0.049***	-0.048***	-0.059***	-0.058***
	(-3.53)	(-3.34)	(-3.61)	(-3.48)	(-4.12)	(-3.98)
Firm-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call-level Controls	No	Yes	No	Yes	No	Yes
Observations	95,159	95,159	95,159	95,159	95,159	95,159
Adjusted R2	0.003	0.008	0.003	0.007	0.002	0.008

Panel B: CAR Constructed by Risk-Adjusted Alpha

Dep.Var.	CAR[0, 1]		CAR[0, 3]		CAR[0, 5]	
	(1)	(2)	(3)	(4)	(5)	(6)
DPRisk	-0.044***	-0.043***	-0.045***	-0.043***	-0.051***	-0.050***
	(-3.45)	(-3.22)	(-3.35)	(-3.15)	(-3.53)	(-3.39)
Firm-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call-level Controls	No	Yes	No	Yes	No	Yes
Observations	95,159	95,159	95,159	95,159	95,159	95,159
Adjusted R2	0.003	0.008	0.003	0.007	0.003	0.007

Table 10: Bank Loan Spread and Data Privacy Risk

This table explores the relationship between data privacy risk (*DPRisk*) and the cost of bank loans in OLS regressions. The cost of bank loans is measured by the natural logarithm of the all-in-spread-drawn loan spread in basis points. *DPRisk* is aggregated at the annual level. All continuous variables are winsorized at the 2nd and 98th percentiles. Leverage is restricted to the [0,1] range. Industry classifications are based on four-digit SIC codes. Lender characteristics are assigned based on the lead arranger. Definitions of all variables are provided in Appendix I. Heteroskedasticity-robust t-statistics in parentheses are clustered at the borrower and year level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dep. Var.	Natural Logarithm of Loan Spread				
	(1)	(2)	(3)	(4)	(5)
DPRisk	0.021***	0.021***	0.019**	0.015**	0.016**
	(3.73)	(3.69)	(3.16)	(2.64)	(2.48)
Size	-0.098***	-0.102***	-0.069***	-0.067***	-0.069**
	(-5.81)	(-5.86)	(-3.90)	(-3.51)	(-3.19)
Profitability	-0.149	-0.152	-0.126	-0.200	-0.202
	(-0.43)	(-0.45)	(-0.36)	(-0.71)	(-0.72)
Market-to-Book	-0.060**	-0.060**	-0.055*	-0.043*	-0.041
	(-2.45)	(-2.41)	(-2.17)	(-2.00)	(-1.65)
Tangibility	-0.046	-0.047	0.016	0.003	-0.027
	(-0.252)	(-0.27)	(0.09)	(0.019)	(-0.158)
Leverage	0.482***	0.465***	0.471***	0.499***	0.491***
	(6.78)	(5.42)	(5.16)	(5.97)	(5.94)
Altman Z-score	-0.054*	-0.054*	-0.056*	-0.060**	-0.058**
	(-2.03)	(-2.02)	(-2.24)	(-2.49)	(-2.36)
Cash Holding	0.198	0.197	0.167	0.114	0.104
	(0.88)	(0.87)	(0.81)	(0.60)	(0.50)
Credit Ratings		0.026	0.026	0.026	0.023
		(0.49)	(0.49)	(0.49)	(0.48)
ln(Loan Maturity)			0.134***	0.132***	0.144***
			(3.62)	(3.50)	(4.94)
ln(Loan Size)			-0.057***	-0.056***	-0.057**
			(-4.62)	(-4.41)	(-4.48)
Credit Spread				0.029	
				(1.21)	
Term Spread				-0.071**	
				(-2.74)	
GDP Growth				-6.044**	
				(-2.85)	
Year FE	No	No	No	No	Yes
Borrower Industry FE	Yes	Yes	Yes	Yes	Yes
Bank Lender FE	No	No	Yes	Yes	Yes
Loan Type FE	No	No	Yes	Yes	Yes
Loan Secured FE	No	No	Yes	Yes	Yes
Observations	4,525	4,525	4,525	4,525	4,525
Adjusted R-squared	0.760	0.760	0.775	0.781	0.784