4-27-2021 12:15 PM

Essays on Information Asymmetry and Leakage

Jun Hyun (Joseph) Ryoo, *The University of Western Ontario*

Supervisor: Wang, Xin (Shane), *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Business

Abstract

Since the underlying qualities of products and firms are not readily apparent, information asymmetry exists at the heart of marketing. This dissertation investigates information asymmetry that is present specifically between: (1) firms and consumers, and (2) firms and investors. I advance our knowledge of how information asymmetry can be reduced in beneficial ways for the firm either by voluntary or involuntary means. This dissertation consists of two essays. In Essay 1, I examine involuntary information leakage in the movie industry. I find that spoilers, which prematurely resolve plot uncertainty for those who have yet to see the movie, can increase box office revenues for movie studios. The positive spoiling effect is driven by uncertainty reduction, in which spoilers provide diagnostic information to consumers unsure about the quality of a movie. In Essay 2, I examine voluntary information leakage in the context of firm signaling. As investors do not have access to private information and cannot observe firm activities such as innovation projects and corporate policy changes, firms send signals to investors that provide cues to such information. I find that data breaches previously experienced by firms can serve as information that negatively influences the interpretation of otherwise positive signals. Taken together, this dissertation outlines implications for firms to effectively respond to and manage information asymmetry in the marketplace.

Keywords: Information asymmetry, Spoilers, Signaling, Online word-of-mouth, Data breach

Summary for Lay Audience

This dissertation examines the imbalance of information that is present between two parties within a transaction. Because the true qualities of products and firms are not readily apparent, consumers and investors rely on information in their environment to make better decisions. I specifically examine the information imbalance between: (1) firms and consumers, and (2) firms and investors, in the contexts of movies and signaling respectively. In Essay 1, I find that spoiler reviews can increase box office revenues of movie studios because spoilers provide helpful information to consumers who are unsure about the quality of a movie. In Essay 2, I find that when firms provide cues to investors who are unsure about the quality of firms that are worth investing in, data breaches can serve as information that influences the interpretation of positive cues to have more negative meanings. This dissertation then outlines recommendations for firms to effectively manage these voluntary and involuntary leakages of information in the marketplace.

Co-authorship Statement

The content of Essay 1 in Chapter 2 of this dissertation has been published in Journal of Marketing, with co-authors Prof. Xin (Shane) Wang, and Prof. Shijie Lu. The content of Essay 2 in Chapter 3 has not been submitted to a journal for publication. I certify that I am the leading author and had a major role in the preparation, literature review, data analysis, and writing of both Essay 1 and Essay 2, in accordance with the thesis regulations of Western University.

Reference to the Published Article:

Ryoo, Jun Hyun (Joseph), Xin (Shane) Wang, and Shijie Lu (2021), "Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue," *Journal of Marketing*, 85(2), 70-88.

Acknowledgements

Foremost, I thank my advisor, Prof. Shane Wang, for his kind and unreserved mentorship. I am grateful to him for accepting me as his student during my undergraduate studies, and closely nurturing me in my growth from a doctoral student to a scholar. In addition, I thank Prof. Neil Bendle and Prof. Shijie Lu, both of whom I had the privilege of working closely with during my formative years as a doctoral student. I have learned tremendously from their expertise and supportive critiques. I thank Prof. Mark Vandenbosch for his time and feedback as my dissertation committee member, and thank each of the marketing faculty members at Ivey Business School for their teachings and support throughout my development as a doctoral student.

Finally, I thank my family, especially my parents for their unconditional support and sacrifices, which have made my doctoral pursuit possible in the first place. I dedicate my dissertation to them.

Table of Contents

ABSTRACT II

SUMMARY FOR LAY AUDIENCE III

CO-AUTHORSHIP STATEMENT IV

ACKNOWLEDGEMENTS V

1. ON INFORMATION ASYMMETRY 1

1.1 Introduction 1
1.2 Firms and Consumers in the Movie Industry 3
1.3 Firms and Investors in the Signaling Context 5
References 7

2. DO SPOILERS REALLY SPOIL? USING TOPIC MODELING TO MEASURE THE EFFECT OF SPOILER REVIEWS ON BOX OFFICE REVENUE 8

2.1 Abstract 8
2.2 Introduction 9
2.3 Related Literature 12
2.4 Conceptual Discussion 14
2.5 Setting and Data 19
2.6 Measuring Spoiler Intensity 22
2.7 Empirical Analysis 30
2.8 Underlying Mechanism 38
2.9 Discussion 46
References 51

3. DATA SECURITY AND FIRM SIGNALING: THE DIRECT AND INDIRECT CONSEQUENCES OF DATA BREACH 54

3.1 Abstract 54
3.2 Introduction 55
3.3 Theoretical Framework 59
3.4 Method and Data 70
3.5 Results 80
3.6 Discussion 84
References 88

4. FINAL REMARKS 91

APPENDICES 94

A. Supplementary Tables and Figures 94
B. Complementary Analyses in the Measurement of Spoiler Intensity 98
C. ESG Issues Monitored by RepRisk 102

CURRICULUM VITAE 103

List of Tables

Table 1. Variable Definitions 21

Table 2. Descriptive Statistics 21

Table 3. Topics Associated with Spoiler Reviews 26

Table 4. Estimation Results of the Model of Box Office Revenue 35

Table 5. Estimation Results from Robustness Checks 37

Table 6. Estimation Results of the Model with Interaction Terms with Spoiler Intensity 41

Table 7. Descriptive Statistics of Data in the Event Study 43

Table 8. Impacts of Spoiler Reviews Before and After the IMDb Website Update 45

Table 9. Average Abnormal Returns 74

Table 10. Coding Scheme for Risk Rating 75

Table 11. Coding Scheme for Attribution of Data Breach 76

Table 12. Variable Descriptions and Sources 78

Table 13. Summary Statistics of Data Breach Sample 78

Table 14. Determinants of Returns from Data Breaches 80

Table 15. Determinants of Returns from Patent Announcements 82

Table 16. Determinants of Returns from Executive Hiring Decisions 84

List of Figures

Figure 1. User Review Guidelines on IMDb	20
Figure 2. Cumulative Volume and Proportion of Spoiler Reviews over Time	20
Figure 3. Examples of Spoiler and Non-Spoiler Reviews for the Movie About Time	24
Figure 4. Example of Topic Distributed in Spoiler Reviews Across Movies	27
Figure 5. Screenshot of Spoiler Reviews Before and After the Update	43
Figure 6. Conceptual Framework	63
Figure 7. Distribution of Average Risk Rating	76

1. On Information Asymmetry

1.1 Introduction

Information asymmetry is fundamentally concerned with the imbalance of information between two parties in a transaction. Simply, one party may be aware of things that the other party may not be. This seemingly intuitive statement creates non-obvious complications and costs for both parties within the transaction. Previous research makes the theoretical distinction between information about quality, and information about intent (Connelly et al. 2011; Stiglitz 2000). Information about intent is often conceptualized in a principal-agent relationship, in which the principal may not be aware of the agent's behaviors or intentions (Connelly et al. 2011). This subsequently leads to problems associated with differing incentives and moral hazards. However, the main focus of this dissertation is on information about quality.

Information about quality deals with the fact that it is difficult to observe the true qualities of products and firms. Granted perfect information, individuals could easily optimize their decision-making processes. For example, a consumer could select from an assortment the product that provides the best fit, and an investor could purchase shares of only high-performing firms over low-performing firms to maximize returns. However, the lack of perfect information constrains consumers and investors alike to make decisions based primarily on public information, unable to access private information firms hold that may be diagnostic of quality. The objective of this dissertation is to explore how leakages of private information to the public, either by voluntary or involuntary means, benefit the firm.

Essay 1 examines involuntary information leakage between firms and consumers in the movie industry. Although advertising can be used by movie studios as a direct means to communicate information to consumers about movie quality, online word-of-mouth (WOM) can serve as an alternative trustworthy source of information (Mudambi and Schuff 2010). For example, potential moviegoers can read online reviews written by consumers who have already seen the movie in theaters to make more informed purchase decisions. An industry concern for movie studios is that online WOM often contain spoilers, which give away important plot details. On the one hand, the leakage of plot details can serve to reduce information asymmetry between movie studios and potential moviegoers about movie quality. On the other hand, spoilers can remove the elements of suspense and surprise from the movie, potentially harming box office revenue.

Essay 2 examines voluntary information leakage between firms and investors in the signaling context. Investors seek high-quality firms to maximize their investment returns. However, public information alone may be insufficient to clearly distinguish between high-quality and low-quality firms. In order to maximize shareholder value, high-quality firms are incentivized to voluntarily signal their unobservable characteristics to investors. For example, in order to signal to investors about the progress of an unobservable innovation project, the firm can issue related patent announcements. Since patent announcements are public information, investors may observe the signals and then interpret that the firm will be more competitive in the future when the innovation project is commercialized. This results in positive financial returns from the patent announcements for the firm (Sood and Tellis 2009). This essay examines how a negative event experienced by the firm can influence its investors' interpretations of subsequent signals. Specifically, I examine how the financial returns from patent announcements and executive hiring decisions, which are in themselves positive signals to investors, are negatively impacted by prior data breaches.

The central theme that weaves the two essays in this dissertation together is that reduction of information asymmetry, either by voluntary or involuntary information leakages, can be beneficial for the firm. This dissertation investigates the behavioral mechanism and moderators that can specifically guide how firms can respond to and manage information leakages more effectively. Next, I provide an overview for each essay.

1.2 Firms and Consumers in the Movie Industry

Movies continue to be a popular form of entertainment for consumers. The total box office revenues accrued in North America in 2018 were $11.89 billion USD, a 7.4% increase from the year prior (Statista 2021a). The closure of theaters due to the recent COVID-19 pandemic have merely shifted the mode of movie consumption to streaming services. It is projected that the revenues from video streaming services will reach $35 billion USD by the end of 2021, and experience an annual growth rate of 9.6% moving forward (Statista 2021b).

Essay 1 examines in this context the information asymmetry that is present between movie studios and potential moviegoers. Movies are experiential products that can be characterized by high subjective quality (Holbrook and Hirschman 1982; Wilcox, Roggeveen, and Grewal 2011). As a result, the true quality of movies is difficult to ascertain by consumers prior to actual movie consumption. To reduce the information asymmetry and attract a bigger audience, movie studios that release high-quality movies spend heavily on advertising, which can serve as a credible signal of quality to consumers (Basuroy, Desai, and Talukdar 2006; Milgrom and Roberts 1986). Basuroy, Desai, and Talukdar (2006) argue that it is difficult for movie studios that release low-quality movies to similarly spend heavily on advertising because low-quality movies can generate negative WOM that damages the movie studios' reputation, leading to long-term harms.

In addition to advertising, potential moviegoers have access to online WOM as an additional source of movie information. Accordingly, previous research finds that online WOM has a positive effect on box office revenue by increasing the awareness of the movie, and decreasing the uncertainty related to the movie's quality (Liu 2006; Chintagunta, Gopinath, and Venkataraman 2010). However, it is difficult for movie studios to control the content of online WOM relative to advertising. This highlights an industry concern for movie studios that online WOM often contains spoilers.

Spoilers are defined as information that prematurely resolves plot uncertainty for those who have yet to see the movie. The research question that drives this essay is whether spoiler reviews are beneficial or harmful to the box office revenues of movie studios. Since spoiler reviews reveal plot-related information as justifications when critiquing a movie, potential moviegoers can have access to diagnostic information about movie quality. However, plot uncertainty, which stimulates tension and suspense, serves as an important source of utility in story consumption (Ely, Frankel, and Kamenica 2015). For example, consumers often become emotionally invested in the protagonist as the movie unfolds, and the protagonist's uncertain fate as the movie reaches its climax creates suspense that causes consumers to yearn for its resolution (Zillmann 1995). By prematurely removing the elements of suspense and surprise, spoilers can reduce expected enjoyment and discourage theater visits.

Essay 1 addresses this question by assembling a data set of 140,869 reviews for 993 movies released in the United States between January 2013 and December 2017. A conceptual background of spoilers is developed, along with the properties of metrics necessary to measure the spoiling content of movie reviews. After developing an appropriate spoiler intensity metric, this essay demonstrates a positive association between spoiler reviews and box office revenue, and further provides evidence that uncertainty reduction is responsible for the positive spoiling effect. Managerial implications of how movie marketers can respond to spoiler reviews, and how review platforms that display spoiler reviews can potentially increase consumer welfare are discussed.

1.3 Firms and Investors in the Signaling Context

It is in the interests of high-quality firms to distinguish themselves from low-quality firms to maximize shareholder value. In order to reduce the information asymmetry present between firms and investors, firms can voluntarily send signals that provide information or cues regarding their unobservable characteristics (Connelly et al. 2011). For example, although information regarding an innovation project is private information that is inaccessible by most investors, the firm can use patent announcements to publicly signal the innovation project's progress. This framework based on signaling has been applied by previous research to explain various strategic firm behaviors, from advertising to firm alliance announcements, that voluntarily reduce information asymmetry to increase financial returns (Kim and McAlister 2011; Swaminathan and Moorman 2009).

Essay 2 explores how signals, which otherwise lead to positive financial returns, are affected by unanticipated firm events. This essay specifically examines data breaches, which is defined as the disclosure of private and confidential information to an unauthorized party. A notorious example is Equifax, which in 2017 experienced a data breach perpetrated by a hacker that stole personal data, including the names, addresses, and Social Security numbers, of more than 147 million Americans. A survey of IT professionals reveals that more than 90 percent of firms have experienced some form of threat to their data security (Kaspersky Lab 2015). Furthermore, it is estimated that the average cost of data breaches in the United States is $8.64 million, which is increasing annually (Ponemon Institute 2020; Berinato and Perry 2018). This research demonstrates that such a detrimental event can have both direct and indirect consequences, which negatively influence the interpretation of firm signals and decrease their financial returns. Drawing from signaling theory, this essay proposes that data breaches lead to signal calibration, which can be defined as the change in degree of valance associated with a signal due to information present in the external environment. I construct a matched data set of 135 data breach disclosures, 6,541 patent announcements, and 228 executive hiring decisions in the financial and insurance services industry between January 1, 2010 to December 31, 2018, to demonstrate that recency and attribution of the data breach, and reputation of the firm are responsible for the moderation of financial returns.

This essay contributes to the extant literature by documenting both the direct and indirect financial harm of data breaches. From this, the research provides implications for firms that have previously experienced a data breach to optimize their shareholder value by adjusting their signaling routines.

References

Basuroy, S., Desai, K. K., & Talukdar, D. (2006), "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research*, 43(2), 287-295.

Berinato, S., & Perry, M. (2018), "Security Trends by the Numbers," Harvard Business Review, Retrieved from https://hbr.org/2018/05/security-trends-by-the-numbers.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29(5), 944-957.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011), "Signaling Theory: A Review and Assessment," *Journal of Management*, 37(1), 39-67.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica (2015), "Suspense and Surprise," *Journal of Political Economy*, 123 (1), 215–260.

Holbrook, M. B., & Hirschman, E. C. (1982), "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun," *Journal of Consumer Research*, 9(2), 132-140.

Kaspersky Lab (2015), "Global IT Security Risks Survey 2015," https://media.kaspersky.com/en/business-security/it-security-risks-survey-2015.pdf.

Kim, M., & McAlister, L. M. (2011), "Stock Market Reaction to Unexpected Growth in Marketing Expenditure: Negative for Sales Force, Contingent on Spending Level for Advertising," *Journal of Marketing*, 75(4), 68-85.

Liu, Y. (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70(3), 74-89.

Milgrom, P., & Roberts, J. (1986), "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94(4), 796-821.

Mudambi, Susan M. and David Schuff (2010), "Research Note: What Makes a Helpful Online Review? A Study ofCustomer Reviews on Amazon.com," *MIS Quarterly*, 34 (1), 185–200.

Ponemon Institute (2020), "Cost of a Data Breach Report 2020," *IBM Security.*

Sood, A., & Tellis, G. J. (2009), "Do Innovations Really Pay Off? Total Stock Market Returns to Innovation," *Marketing Science*, 28(3), 442-456.

Statista (2021a), "Box office revenue in the United States and Canada from 1980 to 2020," accessed February 24, 2021.

Statista (2021b), "Video Streaming (SVoD)," accessed February 24, 2021.

Stiglitz, J. E. (2000), "The Contributions of the Economics of Information to Twentieth Century Economics," *The Quarterly Journal of Economics*, 115(4), 1441-1478.

Swaminathan, V., & Moorman, C. (2009), "Marketing Alliances, Firm Networks, and Firm Value Creation," *Journal of Marketing*, 73(5), 52-69.

Wilcox, K., Roggeveen, A. L., & Grewal, D. (2011), "Shall I Tell You Now or Later? Assimilation and Contrast in the Evaluation of Experiential Products," *Journal of Consumer Research*, 38(4), 763-773.

Zillmann, Dolf (1995), "Mechanisms of Emotional Involvement with Drama," *Poetics*, 23 (1-2), 33–51.

2. Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue

2.1 Abstract

A sizable portion of online movie reviews contains spoilers, defined as information that prematurely resolves plot uncertainty. In this research, the authors study the consequences of spoiler reviews using data on box office revenue and online word of mouth for movies released in the United States between January 2013 and December 2017. To capture the degree of information in spoiler review text that reduces plot uncertainty, the authors propose a spoiler intensity metric and measure it using a correlated topic model. Using a dynamic panel model with movie fixed effects and instrumental variables, the authors find a significant and positive relationship between spoiler intensity and box office revenue with an elasticity of .06. The positive effect of spoiler intensity is more prominent for movies with limited release, smaller advertising spending, and moderate user ratings, and is stronger in earlier days after the movie's release. These findings are consistent with the mechanism that more intense spoiler reviews can help consumers reduce their uncertainty about the quality of the movie and therefore encourage theater visits. By studying an exogenous update that changed the display of movie reviews on an online review platform, the authors provide further evidence in support of the uncertainty-reduction mechanism of spoiler reviews. Results from this study suggest that movie studios can benefit from consumers' access to plot-intense reviews, and should actively monitor the content of spoiler reviews to better forecast box office performance.

2.2 Introduction

In April 2019, the directors of *Avengers: Endgame* issued a stern warning to fans about the much-anticipated blockbuster film: “When you see Endgame in the coming weeks, please don’t spoil it for others, the same way you wouldn’t want it spoiled for you” (Kooser 2019). As a marketing tactic, this ploy was successful, generating significant buzz on social media. However, the directors’ true intention behind their statement remains ambiguous. Did they truly want to silence viewers? What is the relationship between spoilers and box office revenue? Should movie studios be concerned about the exchange of spoilers among consumers? Extant marketing research is unequivocal that online word of mouth (WOM) is vital for the financial success of new products such as movies (e.g., Babić Rosario et al. 2016; Kerrigan 2017). However, the understanding of spoilers and how they influence consumer purchase decisions is still limited.

In the context of movies, a *spoiler review* refers to a movie review that contains spoilers, and a *non-spoiler review* refers to a movie review without any spoilers, where a “spoiler” is defined as information that prematurely resolves plot uncertainty for those who have yet to see the movie. According to data from Internet Movie Database (IMDb), approximately 93% of movies released between January 2013 and December 2017 in the United States garnered at least one spoiler review throughout their screenings and approximately 31% of total movie reviews contained spoilers, suggesting the prevalence of spoiler reviews in the movie industry. With the growth of social media, spoiler reviews can spread rapidly throughout the Internet to reach a broad audience. Conventional wisdom suggests a negative relationship between spoiler reviews and consumer demand, as exemplified by the concern raised by the directors of *Avengers: Endgame.* However, previous research has shown either mixed or null effect of spoilers on consumer behavior (Johnson and Rosenbaum 2015; Leavitt and Christenfeld 2011). Thus, the prevalence of spoilers in the movie industry and its unclear ramifications call for a deeper understanding of whether and how spoiler reviews affect consumers' moviegoing decisions—questions we attempt to address in this research.

We provide a conceptual discussion of spoilers which guides the development of *spoiler intensity,* defined as the degree of information in spoiler reviews that reduces plot uncertainty. Although previous marketing research has examined the relationship between consumer demand and various aspects of online WOM, such as volume (Godes and Mayzlin 2004; Liu 2006), valence (Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Moon, Bergey, and Iacobucci 2010), and variance (Sun 2012), most studies have not considered the information within review content beyond the sentiment. Unlike spoiler volume, spoiler intensity is a latent construct that needs to be inferred from review text. In this study, we use a correlated topic model (CTM; Blei and Lafferty 2005) to identify key topics in movie reviews and propose a spoiler intensity metric as a function of these topics.

We assemble a data set of 140,869 reviews for 993 movies released in the United States between January 2013 and December 2017. We collect both spoiler and non-spoiler reviews from IMDb and exploit the review platform's spoiler labels for movie reviews as a training sample to identify topics that are more likely to appear in spoiler than non-spoiler reviews, which we then use in the construction of the spoiler intensity metric. Using a dynamic panel model with movie fixed effects, we quantify the association between spoiler reviews and box office revenue. We alleviate the potential endogeneity concern arising from the inclusion of WOM-related variables and marketing mix variables using instrumental variables (IV). We find that the spoiler intensity of a movie is positively associated with subsequent box office revenue, whereas the association between spoiler volume and subsequent box office revenue is not evident. We also provide evidence that these findings are robust to alternative specifications of spoiler intensity.

We further investigate the behavioral mechanism that may drive the positive relationship between spoiler intensity and demand. Moviegoers often visit online review platforms to seek diagnostic information from their peers and resolve uncertainty about movie quality (Dellarocas 2003; Goh, Heng, and Lin 2013). Unlike non-spoiler reviews, spoiler reviews can reveal plot-related information as justifications when critiquing a movie and therefore tend to be more diagnostic for potential moviegoers. As such, we expect that the diagnostic value of spoiler reviews helps consumers reduce uncertainty about movie quality, which in turn encourages theater visits. To test the uncertainty-reduction mechanism of spoiler reviews, we consider four potential moderators of the effect of spoiler intensity: (i) release type (limited release vs. wide release), (ii) movie age, (iii) advertising, and (iv) average user rating. We find that the positive effect of spoiler intensity is larger for movies characterized by greater uncertainty for moviegoers, such as limited release movies and movies with smaller advertising spending. In addition, the effect of spoiler intensity decays over time, which is consistent with the higher uncertainty at the earlier (rather than later) stages of a movie's life cycle. We also find an inverted-U relationship between average user ratings and the effect of spoiler intensity, which suggests that the positive spoiling effect is stronger for movies that receive moderate or mixed ratings compared to movies that receive extreme ratings (i.e., either very high or low). This finding is likely driven by the fact that user ratings in the middle range tend to convey more ambiguous signals about movie quality than extreme ratings (Tang, Fang, and Wang 2014). Thus, potential consumers of movies with moderate user ratings have greater incentive to seek diagnostic information to reduce their uncertainty about future consumption.

Moreover, we present additional evidence in support of the uncertainty-reduction mechanism of spoiler reviews from an event study. In particular, we examine the change in the effect of spoiler intensity on box office revenues after an exogenous update on the IMDb website which increased both the cost of reading spoiler reviews and the diagnosticity of non-spoiler reviews. If the uncertainty-reduction mechanism is indeed important, we would expect the positive effect of spoiler intensity on demand to be weakened after the website update because of the decrease in the relative diagnostic value, and the increase in the cost of reading spoiler reviews. Our results from the event study are consistent with this expectation and therefore provide additional support for the proposed mechanism.

With this research, we aim to make three contributions. First, we provide a conceptual background of spoilers by formally defining what constitutes spoiling information in a movie review and discussing several key properties that a spoiler intensity metric needs to capture. Second, we make substantive contributions by showing a positive association between spoiler reviews and consumer demand driven by spoiler intensity rather than spoiler volume. Furthermore, we show that the effect of spoiler intensity is more prominent for movies with limited release, smaller advertising spending, and moderate user ratings. The positive effect of spoiler intensity is also stronger in earlier periods of a movie's life cycle. Finally, we present data patterns that support the behavioral mechanism that uncertainty reduction drives the positive effect of spoiler intensity.

2.3 Related Literature

Given our focus on spoiler reviews, this research builds on the literature on online WOM. Extant marketing research conceptualizes the influence of online WOM on demand through two distinct channels (e.g., Babić Rosario et al. 2016; Seiler, Yao, and Wang 2017): the informative effect of online WOM involves increasing the awareness of consumers about the existence of a product and providing information about the product that consumers seek and value; the persuasive effect of online WOM involves increasing consumers' appreciation for a product without delivering specific product information. The informative role of online WOM is supported by the positive relationship found between number of reviews and box office sales (Duan, Gu, and Whinston 2008; Liu 2006) and between the amount of online conversation and television ratings (Godes and Mayzlin 2004). The persuasive effect of online WOM is supported by the positive relationship found between valence (e.g., review ratings, sentiment) and demand (Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Dellarocas, Zhang, and Awad 2007). Regarding the variance of online WOM, measured by the statistical dispersion of ratings, previous findings are less consistent, in part because of the complex ways in which variance may affect sales (Clemons, Gao, and Hitt 2006; Sun 2012).

In addition to the summary statistics of online WOM (e.g., ratings and volume), marketing scholars have explored specific types and patterns of online WOM observed in the movie industry. For example, Hennig-Thurau, Wiertz, and Feldhaus (2015) examine Twitter to study the diagnostic value of microblogging WOM and find that negative tweets are potentially harmful to a movie's early box office revenue. Gelper, Peres, and Eliashberg (2018) note that sporadic volume bursts, or spikes of online WOM prior to a movie's release, are positively associated with opening weekend box office revenue. Recently, a growing academic attention has been paid to online WOM content beyond its overall valence. Gopinath, Thomas, and Krishnamurthi (2014) use human coders to examine the attribute-, emotion-, and recommendation-oriented dimensions of online WOM and find that only the valence of the recommendation-oriented dimensions impacts sales. Liu, Singh, and Srinivasan (2016) use the principal components of words in tweets to show that the content of online WOM can significantly increase the accuracy of predictions about television show ratings.

It is particularly important to account for the WOM content when examining the impact of online WOM in the entertainment industry for at least two reasons. First, summary statistics alone cannot provide a full picture. For instance, previous research has shown that review ratings are subject to inflation (Chevalier and Mayzlin 2006) and selection bias (Dellarocas 2003; Godes and Silva 2012; Li and Hitt 2008), suggesting that ratings can sometimes be misleading in signaling a movie's true quality. Second, to minimize the risk of watching movies of poor quality, potential moviegoers have incentives to read detailed content (Mudambi and Schuff 2010), especially content related to plots, to seek diagnostic information. We contribute to the online WOM literature by presenting the first empirical study of the relationship between plot-related WOM, which often appears in spoiler reviews, and consumer demand in the movie industry.

2.4 Conceptual Discussion

What Are Spoilers?

Previous research in the field of literature finds that consumption of stories involves a prospective orientation in the minds of consumers, related to forming predictions and looking ahead to what will happen next in the plot (Olson, Mack, and Duffy 1981). As a result, plot uncertainty, which stimulates tension and suspense, serves as an important source of utility in story consumption (Ely, Frankel, and Kamenica 2015). For example, consumers often become emotionally invested in the protagonist, who might encounter danger in a story, and the protagonist's uncertain fate creates suspense that causes consumers to yearn for its resolution (Zillmann 1995). In the context of movies, plot uncertainty can be resolved either by watching the movie or by reading reviews that include plot-related information before the movie consumption. We therefore define spoilers as information that prematurely resolves plot uncertainty for those who have yet to see the movie.

Effect of Spoilers

Extant research in psychology and communication has revealed mixed findings regarding the impact of spoilers on story enjoyment. By manipulating the types of short stories read by subjects in laboratory conditions, Leavitt and Christenfeld (2011) find that spoilers can have a *positive* effect on media enjoyment. The authors later explain this effect by the increased ease of understanding the media experience due to spoilers, which frees cognitive resources and allows consumers to enjoy media at a deeper level (Leavitt and Christenfeld 2013). In contrast, Johnson and Rosenbaum (2015) find that spoiled stories are less fun and suspenseful when using a multidimensional approach to measure enjoyment. They explain their findings using excitation-transfer theory (Zillmann, Hay, and Bryant 1975), positing that spoilers have a *negative* effect on media enjoyment because they displace the physiological arousal generated by suspense that should be resolved by media consumption.

The relationship between spoilers and consumer demand is arguably more relevant to marketers. In contrast to the conventional knowledge that spoilers harm demand, Johnson and Rosenbaum (2015) fail to find a significant effect of spoilers on media selection; when subjects were presented with a choice between spoiled and unspoiled short stories, they were just as likely to choose the spoiled stories as those unspoiled stories. However, the relationship between spoiler reviews and movie demand has not yet been examined.

On the one hand, spoiler reviews might discourage theater visits. By prematurely revealing plot-related information, spoiler reviews can ruin the element of surprise in a movie experience and consequently decrease consumption utility. Such a surprise-burst effect can be triggered by different types of plot-related information of movies from different genres. For example, the death of a character could be a surprising event for a dramatic movie, while the proposal and marriage between characters could be the ultimate surprise for a romantic movie.

On the other hand, spoiler reviews might help consumers reduce the uncertainty about product fit. Due to their subjective nature, the quality of experiential products such as movies is difficult to evaluate by consumers prior to consumption (Alba and Williams 2013). By revealing important plot details and increasing the informative value of WOM, spoiler reviews could have a positive effect on movie demand. It is unclear whether this positive uncertainty-reduction effect outweighs the negative surprise-burst effect of spoiler reviews in the movie industry. We seek to extend the literature on spoilers by investigating the net effect of spoiler reviews on movie demand, that is, the sum of the positive effect from uncertainty reduction and the negative effect from the burst of surprise.

Definition and Properties of Spoiler Intensity

Studying the consequences of spoiler reviews requires measuring both the volume and intensity of spoilers, where spoiler intensity is defined as the degree of information in spoiler reviews that reduces plot uncertainty. Consider a movie that receives multiple spoiler reviews. Measuring only the number of spoiler reviews is inadequate at capturing the spoiling effect because these reviews may provide similar plot-related information and therefore do not accumulate in resolving plot uncertainty. As such, spoiler intensity is an important construct that differs from spoiler volume. Below we present and explain several key properties that an adequate measure of spoiler intensity should capture.

Property 1: Spoiler intensity should be a continuous rather than dichotomous variable because the extent to which plot uncertainty is revolved depends on the level of details in a spoiler review. For example, a spoiler review for the movie *Avengers: End Game* can reveal not only the names of characters who died at the end (e.g., "Iron Man dies"), but also the causes and consequences of the deaths (e.g., "Iron Man sacrifices himself to defeat Thanos"), which further resolve plot uncertainty for consumers. Therefore, a dichotomous variable is insufficient to capture the level of plot uncertainty prematurely resolved in a spoiler review.

Property 2: Spoiler intensity should capture a multitude of plot-related topics that are involved in the structure of a story. Previous research suggests that stories in general share similar patterns and plot structures, and stories in movies are no exception (Deighton, Romer, and McQueen 1989). In particular, movie plots typically unfold in a three-act structure: exposition, rising action, and climax (Trottier 1998), where exposition is used to introduce the major characters, the rising action occurs when the protagonist encounters some sort of crisis that creates tension, and the climax features the resolution of the main tensions of the story. For each act, the screenwriter can craft the story using various elements, which we call plot-related topics (e.g., topics related to "fight" often appear in the climax of action movies, while topics related to "emotion" often appear in the climax of romantic movies). Because of the similar patterns and structures of stories in the movie industry, we assume that a discrete number of plot-related topics are conveyed by movie reviews.

Property 3: Spoiler intensity should allow for the degree of uncertainty resolved by the same topic to vary across movies. For example, although both *Avengers: End Game* and *The*

Lego Movie might include the topic of “survival,” the level of suspense resolved by reading plot details related to the topic of “survival” in a spoiler review is likely to be greater for *Avengers: End Game* than for *The Lego Movie* due to the overall storyline and the plot structure. Thus, an adequate measure of spoiler intensity should account for the potential heterogeneity in each topic’s contribution to resolve plot uncertainty across movies.

Property 4: Spoiler intensity should discount the degree of plot uncertainty resolved by a certain topic that has appeared in previous reviews. This property captures the potential dynamics in the spoiling process when a consumer reads multiple reviews. For instance, suppose a consumer has already read several spoiler reviews. Three scenarios might occur when this consumer reads a subsequent spoiler review. First, the new spoiler review includes information of new plot-related topics that have not appeared in previous reviews. Given that a new facet of plot uncertainty can be resolved by reading this new spoiler review, the degree of plot uncertainty resolved by this additional spoiler review should not be discounted when assessing the overall spoiler intensity of multiple reviews. Second, the new spoiler review includes information on plot-related topics that have already appeared in previous reviews but provides additional details for these existing topics. In this case, the degree of plot uncertainty pertaining to existing topics is further resolved by this new spoiler review because of the additional information provided. Third, the new spoiler review includes information on plot-related topics that have appeared in previous reviews but does not provide any new information for these existing topics. The contribution to the reduction of plot uncertainty by this new spoiler review needs to be discounted because consumers’ feeling of suspense is still driven by previous reviews.

2.5 Setting and Data

We obtained a list of movies released in the United States between January 2013 and December 2017 from WildAboutMovies.com. From this list, we sampled 993 movies that have their daily box office revenue data available on BoxOfficeMojo.com. We focus on the first eight weeks of daily box office revenue because 97% of total box office revenue is accrued within the first eight weeks of a movie's release (Liu 2006). We collected daily box office revenue and daily number of theaters, as well as other movie characteristics (e.g., Motion Picture Association of America rating, genre, and release type) from both BoxOfficeMojo.com and IMDb. We matched our movie sample with advertising spending data provided by Kantar Media.

We use IMDb to collect online WOM data for two reasons. First, IMDb is by far the most popular online movie review platform in the United States.[1] Second, IMDb requires users to label their reviews as spoilers if a user believes that his or her review discloses any critical plot elements of a movie. As Figure 1 shows, IMDb penalizes users who do not label spoiler reviews by blacklisting their accounts and deleting their reviews automatically. This institutional feature gives us a data set with a clear classification between spoiler and non-spoiler reviews.

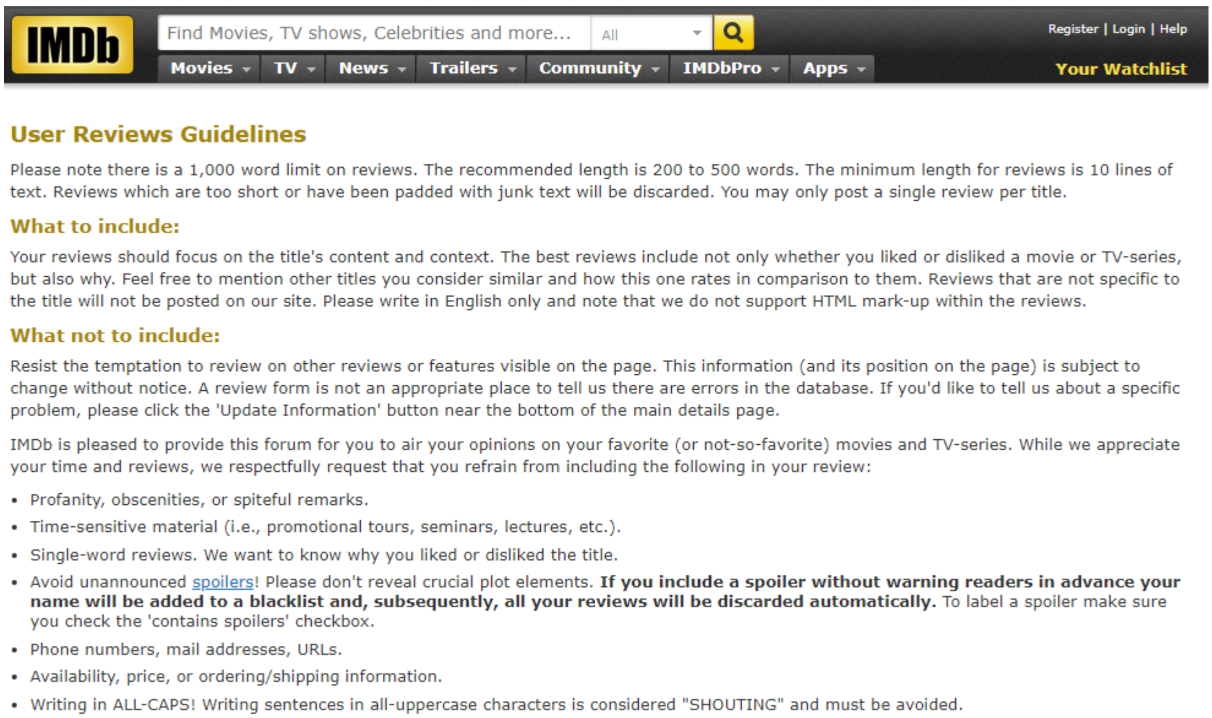
Table 1 lists key time-varying variables in this study, along with their descriptions. Table 2 presents summary statistics of time-varying variables and time-invariant movie characteristics. On average, each movie's daily box office revenue was $1.04 million. Each movie received approximately one spoiler review and two non-spoiler reviews per day.[2] As Figure 2(a) shows, both the volume of spoiler reviews and the volume of total reviews grow over time, though with greater momentum in the earlier than later days after movie release. We also plot the dynamics in

[1] IMDb was ranked 25th, Rotten Tomatoes 322nd, and Metacritic 841st for websites in the United States on Alexa.com, accessed July 2019.

[2] Please see Figure A1 in the Appendix for a Pareto chart of the distribution of spoiler reviews across movies.

the proportion of spoiler reviews in Figure 2(b). The average proportion of spoiler reviews across movies is 26% on day one and gradually increases to 31% by the end of the eighth week.

Figure 1. User Review Guidelines on IMDb


IMDb | Find Movies, TV shows, Celebrities and more... | All | Register | Login | Help

Movies | TV | News | Trailers | Community | IMDbPro | Apps | Your Watchlist

User Reviews Guidelines

Please note there is a 1,000 word limit on reviews. The recommended length is 200 to 500 words. The minimum length for reviews is 10 lines of text. Reviews which are too short or have been padded with junk text will be discarded. You may only post a single review per title.

What to include:

Your reviews should focus on the title's content and context. The best reviews include not only whether you liked or disliked a movie or TV-series, but also why. Feel free to mention other titles you consider similar and how this one rates in comparison to them. Reviews that are not specific to the title will not be posted on our site. Please write in English only and note that we do not support HTML mark-up within the reviews.

What not to include:

Resist the temptation to review on other reviews or features visible on the page. This information (and its position on the page) is subject to change without notice. A review form is not an appropriate place to tell us there are errors in the database. If you'd like to tell us about a specific problem, please click the 'Update Information' button near the bottom of the main details page.

IMDb is pleased to provide this forum for you to air your opinions on your favorite (or not-so-favorite) movies and TV-series. While we appreciate your time and reviews, we respectfully request that you refrain from including the following in your review:

- Profanity, obscenities, or spiteful remarks.
- Time-sensitive material (i.e., promotional tours, seminars, lectures, etc.).
- Single-word reviews. We want to know why you liked or disliked the title.
- Avoid unannounced spoilers! Please don't reveal crucial plot elements. **If you include a spoiler without warning readers in advance your name will be added to a blacklist and, subsequently, all your reviews will be discarded automatically.** To label a spoiler make sure you check the 'contains spoilers' checkbox.
- Phone numbers, mail addresses, URLs.
- Availability, price, or ordering/shipping information.
- Writing in ALL-CAPS! Writing sentences in all-uppercase characters is considered "SHOUTING" and must be avoided.

Figure 2. Cumulative Volume and Proportion of Spoiler Reviews over Time

(a) Dynamics in Cumulative Volume **(b) Dynamics in Proportion of Spoiler Reviews**

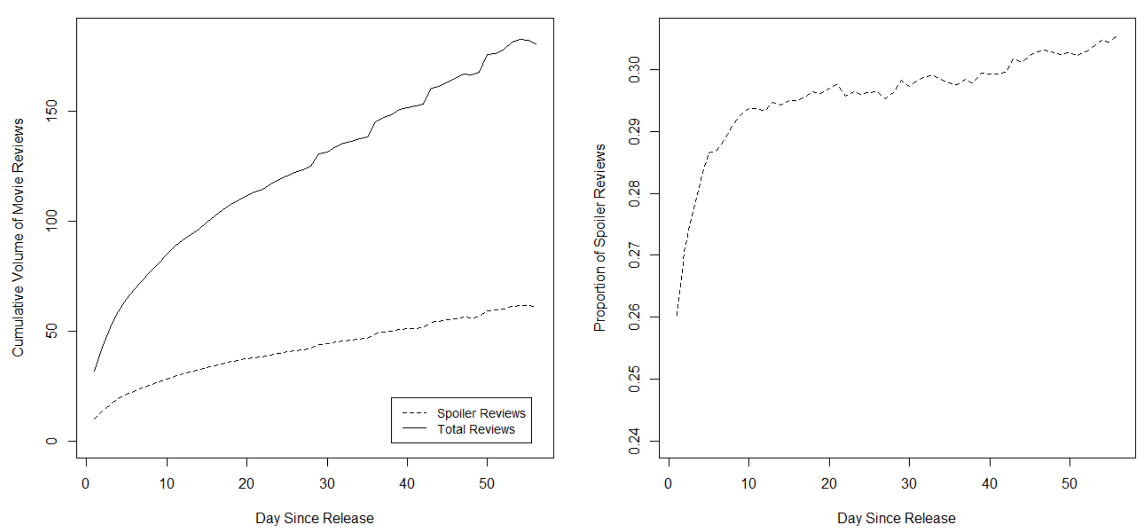


Table 1. Variable Definitions

Variable Name	Description
DAILYREV	Box office revenue on day *t* for movie *i*.
INTENSITY	Spoiler intensity of spoiler reviews within the last 10 days of day *t* for movie *i*.
PROP	Moving average of proportion of spoiler reviews within the last 10 days of day *t* for movie *i*.
CUMRATING	Mean ratings of cumulative movie reviews on day *t* for movie *i*.
CUMVOL	Number of cumulative movie reviews on day *t* for movie *i*.
ADVERT	Average daily advertising expenditure on day *t* for movie *i*.
THEATERS	Number of theaters that screen movie *i* on day *t*.
AGE (t)	Number of days since the release of movie *i* in theaters.
HOLIDAY	Dummy variable for the 10 federal holidays in the United States.
DAYOFWEEK	Indicator variables for each day of the week.

Table 2. Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Daily level				
DAILYREV (in $)	1,039,985	3,265,580	5	119,119,282
INTENSITY	2.48	2.69	0	45.17
PROP	.18	.15	0	1
CUMRATING	6.27	1.48	0	10
CUMVOL	12.87	247.73	0	4,276
ADVERT (in $1,000)	126.3	621.7	0	6,807
THEATERS	1,240	1,309	1	4,535
Movie level				
MPAA ratings				
G & PG	.15	.36	0	1
PG-13	.40	.49	0	1
R	.40	.49	0	1
Unrated	.05	.21	0	1
Genres				
Action	.09	.28	0	1
Adventure/Sci-Fi	.10	.30	0	1
Comedy	.20	.40	0	1
Drama	.32	.47	0	1
Family	.10	.30	0	1
Foreign	.02	.14	0	1
Horror	.06	.24	0	1
Musical	.02	.12	0	1
Romance	.02	.14	0	1
Thriller	.08	.27	0	1
Release type				
Limited Release	.40	.49	0	1

2.6 Measuring Spoiler Intensity

Uncovering Topics from Review Text

The construction of spoiler intensity requires revealing a multitude of plot-related topics from review texts (Property 2). We use text mining—in particular, CTM (Blei and Lafferty 2005)—to uncover the set of topics that generate movie reviews. CTM is an extension of latent Dirichlet allocation (LDA; Blei, Ng, and Jordan 2003), which has been used in previous marketing research to study the emerging topics in scholarly articles (Wang et al. 2015), the dimensions of customer product reviews (Tirunillai and Tellis 2014), and the predictive power of text in peer-to-peer loan applications (Netzer, Lemaire, and Herzenstein 2019). CTM replaces the Dirichlet distribution in LDA with a multinomial distribution in its data generation process. This modification allows flexible correlations between topics and therefore leads to an improved fit with the data (Blei and Lafferty 2005, 2007). Indeed, we find that CTM consistently outperforms LDA in terms of model fit in our empirical context, which provides support for the use of CTM in this study.[3] To apply the CTM, we prepare the textual data by removing stop words, tokenizing each word using a standard stemming algorithm, and removing sparse words that appear in less than 1% of movie reviews. This procedure yields a pre-processed document-term matrix of 140,869 reviews (including both spoiler and non-spoiler reviews) represented by 1,624 unique words.

We refer to a movie review as a *document*, and the collection of movie reviews as a *corpus*. The CTM of each document from the corpus can be described as follows:

1. Draw $\eta|\{\mu, \Sigma\} \sim N(\mu, \Sigma)$.

[3] Please see Figure A2 in the Appendix for details regarding the model fit comparison between CTM and LDA.

2. For each word w contained in the document:

    a. Draw topic assignment variable $z|\eta$ from Multinomial($f(\eta)$),

    b. Draw a word $w|z, \beta$ from Multinomial(β).

where $f(\eta)$ in step 2a maps a natural parameterization of $\eta = (\eta_1, \dots, \eta_K)$ to the vector of topic probabilities $\theta = (\theta_1, \dots, \theta_K)$ expressed below:

$$\theta_k = f(\eta_k) = \frac{\exp(\eta_k)}{\sum_{k=1}^{K} \exp(\eta_k)} \qquad (2.1)$$

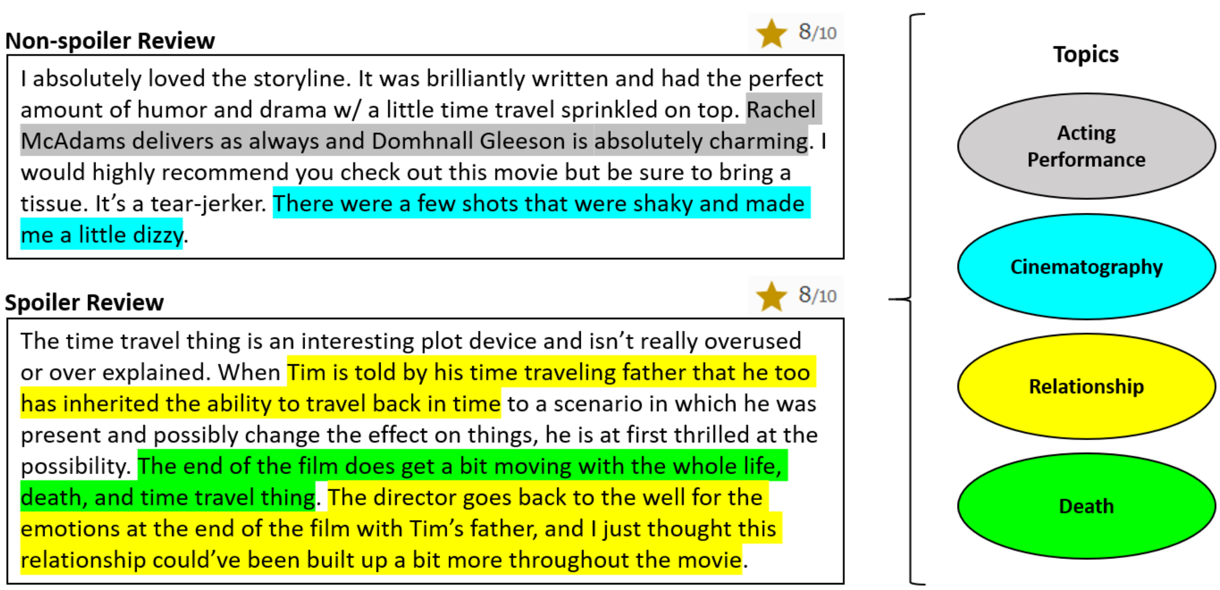
The data generation process of CTM can be interpreted as follows. When a user starts writing a movie review, he or she first decides on the weight of each topic (θ_k) that will appear in the movie review from a fixed number of topics (K). When choosing which word to write, the user selects a topic (z) according to its probabilistic distribution (Multinomial(θ)). Conditional on the topic (z), the user's word choice (w) is then drawn from the associated distribution (Multinomial(β)). The mapping of η to θ in Equation (2.1) allows the K×1 vector of topic probabilities for each document to carry a correlational relationship from Σ. We estimate the posterior distribution of the latent variables using a variational Expectation-Maximization algorithm (Roberts, Stewart, and Airoldi 2016). We refer interested readers to Blei and Lafferty (2007) for the derivation of the posterior distribution for CTM.

The CTM assumes a fixed number of topics K, which is a hyperparameter that must be predetermined by researchers (Chang et al. 2009). We use the algorithm proposed by Lee and Mimno (2017), which estimates the vertices of the convex hull of word co-occurrences using a method of t-distributed stochastic neighbor embedding. Compared to cross-validation, an advantage of this algorithm is the computational efficiency for large data sets like the one in this study. We find that K = 61 is the optimal number of topics for movie reviews (including spoiler and non-spoiler reviews). We name each topic using its representative words and present all topics in Table A1 in the Appendix.

Identifying Spoiling Topics

As not all topics resolve plot uncertainty, we further rely on the difference between texts in spoiler and non-spoiler reviews to identify the set of topics deemed important in resolving plot uncertainty. To better explain the intuition behind the identification strategy, we provide examples of a spoiler review and a non-spoiler review in Figure 3, both of which are real reviews for the movie *About Time*. Notably, the text of each review can be well summarized by its underlying topics. For example, the non-spoiler review includes topics related to "cinematography" and "acting performance," and the spoiler review includes topics related to "relationship" and "death," as evidenced by sentences in their associated colors.

Figure 3. Examples of Spoiler and Non-Spoiler Reviews for the Movie *About Time*



The topics revealed in the non-spoiler review and those revealed in the spoiler review are different in terms of the amount of plot-related information. Two plot-related topics that we clearly observe in the spoiler review are “death” (which occurred at the end of the movie) and “relationship” (between the protagonist Tim and his father). Although the non-spoiler review describes the movie as a “tear-jerker,” the plot details as to why the movie is a “tear-jerker” are not provided. Another observation is that both reviews mentioned “time travel,” suggesting that the topic of time travel is not regarded as spoiling for this movie. Therefore, not all topics that appear in movie reviews are regarded spoiling; a spoiling topic is more likely to occur in spoiler than non-spoiler reviews, whereas a non-spoiling topic has either equal or higher likelihood to appear in non-spoiler reviews.

To identify spoiling topics, we run a logistic regression in which the outcome variable is the review type (i.e., 1 = spoiler, 0 = non-spoiler), and predictors are the number of words in a review associated with each topic. We operationalize the number of words from topic j in review l as $w_{jl} = \theta_{jl} \times n_l$, where θ_{jl} is the weight of topic j in review l from the estimation of CTM, and n_l is the number of words in review l.[4] We also include movie dummies in the regression to account for movie heterogeneity.

We report in Table 3 the 23 topics that have significantly larger weights ($p < .05$) in spoiler reviews than in non-spoiler reviews. The top three spoiler-related topics (i.e., topics that weigh the most in spoiler reviews) are “disappointment,” “kill,” and “death.” Not surprisingly, “kill” and “death” are often involved in critical plot points of movies (e.g., death of the main character). The topic “disappointment” is associated with words “worst,” “ruin,” and “disappoint.” These are common words one might use when expressing one’s unsatisfactory movie experience followed by the reveal of plot information as a justification. Although not presented in Table 3, the top three topics related to non-spoiler reviews (i.e., topics that weigh

[4] We report in Appendix B.1 more details about the predictive power of topics from CTM.

the most in non-spoiler reviews) are: "cinematography," "expectation," and "acting performance." The topic "cinematography" is associated with the words "beautiful," "visual," and "set," which are related to the visual appeal of the movie and therefore unrelated to movie plot. Similarly, "expectation" (associated with the words "time," "expect," and "watch") and "acting performance" (associated with the words "actor," "perform," and "role") are not directly associated with the plot of the movie. This comparison between the top topics related to spoiler and non-spoiler reviews provides some face validity to our identification of spoiling topics.

Table 3. Topics Associated with Spoiler Reviews

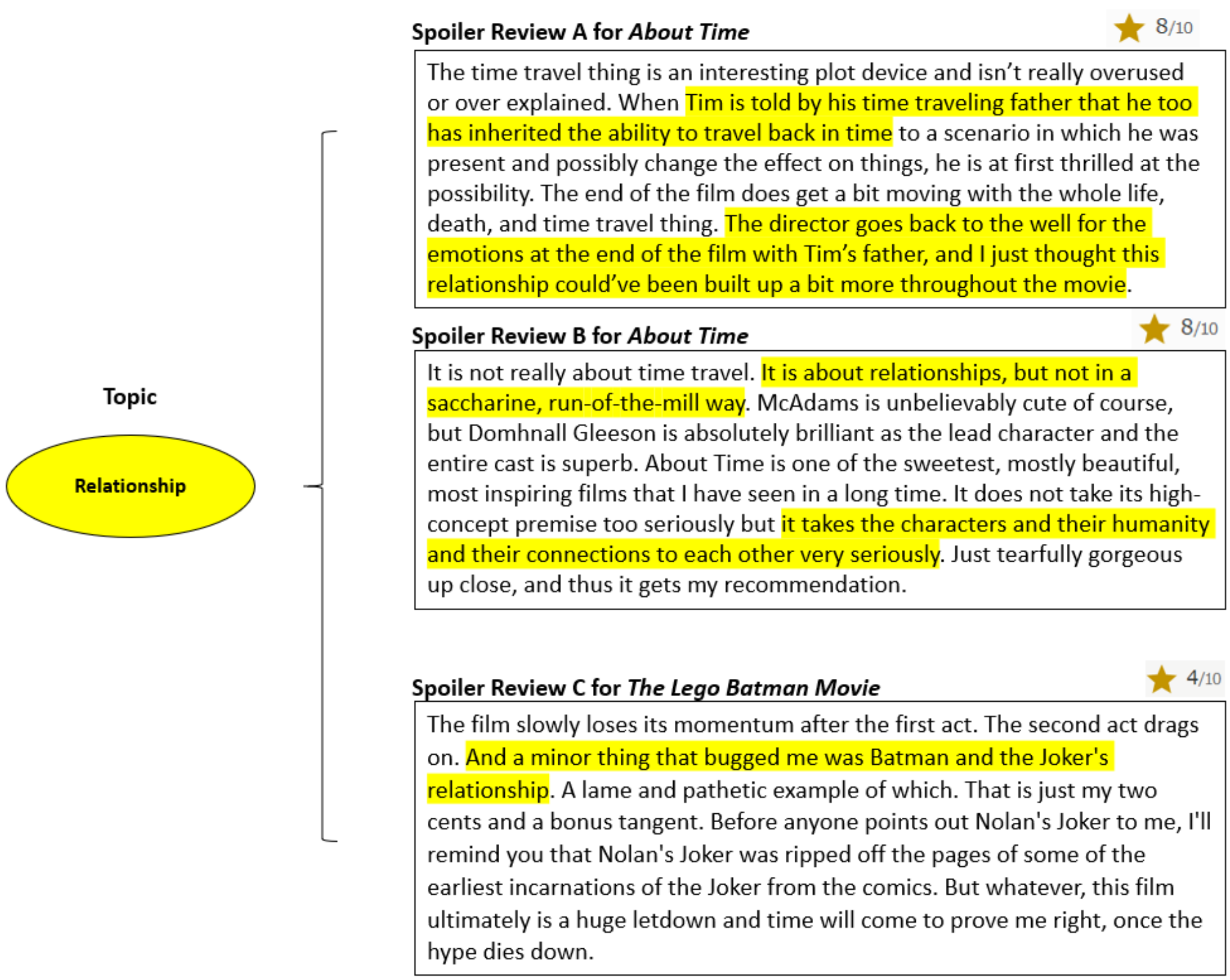
Topic Name	Coefficient	Std. Error	Significance
America	1.582e-02	5.973e-03	**
Book	2.796e-02	3.489e-03	***
Character Development	1.021e-02	2.610e-03	***
Death	5.816e-02	5.282e-03	***
Disappointment	4.457e-01	6.030e-02	***
Emotion	3.205e-02	3.718e-03	***
Fight	1.472e-02	2.448e-03	***
Ghost	9.521e-03	3.969e-03	*
Historical	7.063e-03	2.646e-03	**
Humans and Robots	1.657e-02	5.245e-03	**
Kill	1.889e-01	4.407e-03	***
Length of Movie	1.349e-02	2.329e-03	***
Lesson	7.624e-03	2.867e-03	**
Office	2.797e-02	3.144e-03	***
Overall Evaluation	5.410e-02	2.892e-03	***
Relationship	3.160e-02	2.848e-03	***
Romance	2.939e-02	4.327e-03	***
Science Fiction	8.301e-03	3.566e-03	*
Soundtrack	2.186e-02	3.471e-03	***
Space Travel	2.850e-02	2.932e-03	***
Star Wars Characters	3.177e-02	2.074e-03	***
Survival	1.425e-02	3.644e-03	***
Western	9.939e-03	4.860e-03	*

$*p < .05$; $**p < .01$; $***p < .001$.

Constructing the Spoiler Intensity Metric

With the uncovered set of topics that constitute spoiling information, we further construct the spoiler intensity metric following the guidance of the remaining three properties previously discussed (i.e., Property 1, 3, and 4). To better illustrate these properties, we provide two spoiler reviews for the movie *About Time* (one of which is the same review as in Figure 3), and one spoiler review for *The Lego Batman Movie* in Figure 4.

Figure 4. Example of Topic Distributed in Spoiler Reviews Across Movies



Comparing the two spoiler reviews for *About Time*, we notice that spoiling topics may receive different degrees of elaboration. Spoiler review A provides more details for the topic of "relationship" than spoiler review B. In particular, spoiler review A reveals that the protagonist's father can time travel as the ability is heritable, and that the father is involved in the movie's emotional ending. Spoiler review B provides a relatively limited description that the topic of relationship is not saccharine, and that it is taken seriously by the movie. As the degree of elaboration is often associated with the length of description, we use the number of words related to each topic (w_{jl}) as a proxy for the amount of plot-related information revealed in a spoiler review. This specification renders the spoiler intensity variable continuous and therefore satisfies Property 1.

Property 3 suggests that the degree of spoiling per topic might vary across movies. For example, in addition to spoiler reviews A and B, spoiler review C for *The Lego Batman Movie* in Figure 4 also discusses the topic of relationship (between Batman and the Joker). However, since *The Lego Batman Movie* is a comedy, the degree of spoiling from reading the topic of relationship is potentially less than that for a romantic movie like *About Time.* As such, for each *J* = 23 plot-related topic, we quantify the degree of spoiling of topic *j* for movie *i*. Recall that the probability of review *l* associated with movie *i* being a spoiler review is predicted by the logistic model as follows:

$$y_{il} = \frac{\exp(\gamma w_l + \omega_i)}{1 + \exp(\gamma w_l + \omega_i)} \qquad (2.2)$$

where $w_l = (w_{1l}, \dots, w_{Kl})$, and ω_i is the fixed effect of movie *i*.

We calculate the contribution of spoiling information from topic *j* in review *l* as follows:

$$c_{ijl} = \frac{\exp\big(\gamma(w_{jl} + 1) + \delta w_{-jl} + \omega_i\big)}{1 + \exp\big(\gamma(w_{jl} + 1) + \delta w_{-jl} + \omega_i\big)} - \frac{\exp\big(\gamma w_{jl} + \delta w_{-jl} + \omega_i\big)}{1 + \exp\big(\gamma w_{jl} + \delta w_{-jl} + \omega_i\big)} \qquad (2.3)$$

where w_{-jl} is a vector of the number of words from other topics. The difference of the two terms on the right-hand side of Equation (2.3) measures the change in the likelihood (y_{il}) in response to a topic share increase in θ_{jl} (i.e., $(\theta_{jl} + \Delta\theta) \times n_l = w_{jl} + 1$, where $\Delta\theta = \frac{1}{n_l}$). A greater c_{ijl} suggests a higher degree of spoiling from topic *j* in review *l* for movie *i*.

We aggregate the degree of spoiling of topic *j* for movie *i*, denoted by α_{ij}, using the normalized sum of c_{ijl} across all reviews of movie *i* as follows:

$$\alpha_{ij} = \frac{\sum_{l=1}^{L_i} c_{ijl}}{\sum_{j=1}^{J}\left(\sum_{l=1}^{L_i} c_{ijl}\right)} \qquad (2.4)$$

where L_i represents the set of all reviews associated with movie *i*, and $\sum_{j=1}^{J} \alpha_{ij} = 1$. The parameter α_{ij} in Equation (2.4) measures the spoiling effect of topic *j* for movie *i*, suggesting that the inclusion of α_{ij} in spoiler intensity metric will satisfy Property 3.

Let S_{it} denote the set of spoiler reviews for movie *i* generated within a lagged time-window of day *t.* We operationalize spoiler intensity of movie *i* on day *t* using all spoiler reviews from S_{it} as follows:

$$INTENSITY_{it} = \sum_{j=1}^{J} \alpha_{ij} \times \operatorname*{Max}_{l \in S_{it}} w_{jl} \qquad (2.5)$$

Property 4 suggests that once the spoiling information related to a certain topic has been revealed, information from the same topic does not further reduce plot uncertainty when it reappears in subsequent reviews, unless additional information is provided. Consider again the two spoiler reviews for *About Time* in Figure 4. If an individual reads spoiler review A after spoiler review B, this individual can further reduce plot uncertainty because spoiler review A contains more specific plot details regarding the topic of relationship (e.g., with the protagonist's father, his involvement in the emotional ending) than spoiler review B, which only indicates that the movie treats relationships between characters seriously. However, if the order is reversed (i.e., reading spoiler review B after spoiler review A), it is unlikely for the individual to reduce plot uncertainty by spoiler review B because much of the plot-related information has been covered by spoiler review A. As such, we use the maximum function in Equation (2.5) to capture Property 4 of the spoiler intensity metric. The maximum function ensures that once a piece of information has been spoiled, it cannot spoil again. We provide evidence for the validity of the proposed spoiler intensity metric in capturing the level of spoiling information perceived by real people in Appendix B.2.

We choose the lag window that we use to construct spoiler-related variables (i.e., spoiler intensity and spoiler volume) to be 10 days based on a separate panel data set of movie reviews that we collected for 45 movies released in the United States in April 2019. For each movie, we tracked first-page spoiler reviews on IMDb daily in April 2019. The recency of spoiler reviews on the first page has a mean of 9.53 days, where we calculate the recency of each spoiler review by the difference between the date of observation and the date of creation. Therefore, we assume that consumers typically read spoiler reviews generated within the last 10 days.

2.7 Empirical Analysis

Model of Box Office Revenue

Let *i* denote movies and *t* the days after release. The dependent variable is $\ln(DAILYREV)_{it}$, which represents the log-transformed daily box office revenue for movie *i* on day *t.* To examine the relationship between spoiler reviews and box office revenue, we consider the following model specification:

$$\begin{aligned}\ln(DAILYREV)_{it} &= \beta_1 \ln(DAILYREV)_{i,t-1} + \beta_2 \ln(INTENSITY)_{i,t-1} + \beta_3 PROP_{i,t-1} \\ &+ \beta_4 \ln(CUMRATING)_{i,t-1} + \beta_5 \ln(CUMVOL)_{i,t-1} + \beta_6 \ln(ADVERT)_{i,t-1} \\ &+ \beta_7 \ln(THEATERS)_{it} + \beta_8 t + \beta_9 HOLIDAY_{it} + \sum_{d=1}^{6} \gamma_j I\{DAYOFWEEK_{it} = d\} \\ &+ \omega_i + \varepsilon_{it} \qquad (2.6)\end{aligned}$$

We include the lagged dependent variable, $\ln(DAILYREV)_{i,t-1}$, on the right-hand side of Equation (2.6) to better capture the dynamics and indirectly control for past realizations of independent variables (e.g., WOM-related variables), which can persist to influence contemporaneous box office revenue (Keele and Kelly 2006). *INTENSITY* denotes the spoiler intensity described in Equation (2.5), and *PROP* measures the proportion of spoiler volume, defined as the moving average of the proportion of spoiler reviews to total movie reviews within the last 10 days (i.e., from *t*-10 to *t*-1).[5]

For controls, we include the mean rating (*CUMRATING*) and volume (*CUMVOL*) of cumulative movie reviews because IMDb presents these summary statistics on the main page of each movie. We also include marketing mix variables, which comprise log-transformed advertising expenditure (*ADVERT*) and theater release count (*THEATERS*); and time-related variables, which comprise days after movie release (*t*), a dummy variable for federal holidays in the United States (*HOLIDAY),* and indicator variables ($I\{\cdot\}$) for each day of the week (*DAYOFWEEK*).

In line with previous research (e.g., Duan, Gu, and Whinston 2008; Liu 2006), we lag WOM-related and marketing mix variables except for the number of theaters to alleviate simultaneity concerns. We include ω, the movie fixed effect, to control for time-invariant heterogeneity of movies that include observable factors (e.g., budget, genre, star power) and unobservable factors (e.g., quality of the script, plot). Finally, ε is the idiosyncratic error term with a mean of zero.

Endogeneity Issues

[5] We log-transform *INTENSITY* using $\ln(x+1)$. *PROP* is not in log because it is bounded between 0 and 1.

It is well known that the inclusion of the lagged dependent variable as a predictor leads to a specific endogeneity issue known as the dynamic panel bias (Nickell 1981). As such, we estimate Equation (2.6) using the generalized method of moments (GMM) proposed by Blundell and Bond (1998). This estimation approach involves instrumenting the lagged dependent variable using both of its lagged levels and lagged differences. Our panel data allows the use of multiple lags (i.e., lags 2 and up) as GMM-type instruments to increase the efficiency of our estimation (Blundell and Bond 1998).

Unobserved time-variant characteristics of movies can induce a correlation between the regressors and the error term. The first potential source of endogeneity stems from the WOM-related variables. For example, unobserved offline WOM may increase both the demand for movies and the number of movie reviews. In addition, a user's interest in writing a spoiler review may also be associated with unobserved demand factors. Our solution follows Anderson and Hsiao (1981, 1982) to instrument the endogenous variable (*CUMVOL, PROP,* and *INTENSITY*) using its lagged level. The lagged levels of the endogenous variables are valid instruments under zero second-order autocorrelation (Anderson and Hsiao 1981, 1982), an assumption we empirically checked and confirmed.

Moreover, strategic information held by movie studios may also be a potential source of endogeneity. After a movie's release, private market information may be obtained by studio managers, allowing for adjustments of *THEATERS* and *ADVERT*. We follow previous research (e.g., Chintagunta, Gopinath, and Venkataraman 2010; Lu, Wang, and Bendle 2020) to use the means of *THEATERS* and *ADVERT* of other movies from the same genre as movie *i* and the same number of days *t* from the release as instruments for $THEATERS_{it}$ and $ADVERT_{it}$. The rationale for the relevance of these instruments is similar to that provided by Chintagunta, Gopinath, and Venkataraman (2010): movies of the same genre are likely to share similar release patterns and promotional strategies. The exclusion restrictions of these instruments stem from the fact that the means of marketing mixes set by other movies at different times are unlikely to be correlated with the current demand shock of the focal movie.

Empirical Findings

We begin with standard ordinary least squares (OLS) regression of the model of box office revenue without the lagged dependent variable.[6] We report the results in column 1 of Table 4, which provides preliminary evidence that the association between spoiler intensity and box office revenue is positive and significant (.180, $p < .001$). We find that the association between spoiler volume and box office revenue is also positive and significant (.564, $p < .001$). Estimates for the control variables are of expected signs. For example, both *CUMRATING* and *CUMVOL* have positive associations with box office revenue. In addition, box office revenue is greater for movies that played in a larger number of theaters and spent more on advertising.

In column 2 of Table 4, we present model estimates using standard fixed effects regression and report robust standard errors clustered at the movie level. After controlling for time-invariant heterogeneity of movies, the association between spoiler intensity and box office revenue remains positive and significant (.045, $p < .001$), whereas the association between spoiler volume and box office revenue becomes nonsignificant (-.016, $p > .05$).

We report estimates using the GMM method (Blundell and Bond 1998) in column 3 and column 4 of Table 4, where robust standard errors clustered at the movie level are reported. We show the results with endogeneity correction for only the lagged dependent variable in column 3, and endogeneity corrections for the lagged dependent variable, WOM-related variables, and

[6] Including the lagged dependent variable in OLS leads to an almost perfect linear relationship (adjusted R-square of 1); therefore, the results are uninformative.

marketing mix variables in column 4. We conduct Hansen's *J*-test and the Arellano-Bond test for AR(2) to check the validity of over-identifying restrictions and second-order autocorrelation, respectively. The *p*-values of the *J*-test and the test for AR(2) are .362 and .147 for column 3, and .488 and .134 for column 4, supporting the validity of proposed instruments and providing no evidence of second-order autocorrelation. Results from both GMM specifications show that the coefficient of *INTENSITY* is positive and significant. Although the estimate of *PROP* is positive and significant in column 3, it becomes nonsignificant in column 4 after the endogeneity corrections for WOM-related and marketing mix variables. We focus on the results in column 4 in the rest of the article because of the more careful endogeneity corrections. The log-log model indicates that one percentage increase in spoiler intensity for movie *i* on day *t* is associated with a .06 percentage increase in box office revenue on the following day.

Table 4. Estimation Results of the Model of Box Office Revenue

	OLS	FE	GMM with IVs for Lagged DV	GMM with IVs for Lagged DV, WOM, & Marketing Mix
	(1)	(2)	(3)	(4)
$Intercept$	5.958***	-	-	-
	(.025)			
$\ln(DAILYREV)_{i,t-1}$	-	.474***	.606***	.638***
		(.011)	(.013)	(.017)
$\ln(INTENSITY)_{i,t-1}$	.180***	.045***	.077***	.060***
	(.010)	(.008)	(.014)	(.014)
$PROP_{i,t-1}$	.564***	-.016	.170**	.075
	(.041)	(.031)	(.055)	(.062)
$\ln(CUMRATING)_{i,t-1}$	.457***	-.004	.202***	.171***
	(.012)	(.020)	(.029)	(.027)
$\ln(CUMVOL)_{i,t-1}$	.140***	-.192***	.048***	.037***
	(.004)	(.014)	(.008)	(.010)
$\ln(ADVERT)_{i,t-1}$	.123***	.018***	.051***	.097***
	(.002)	(.002)	(.003)	(.007)
$\ln(THEATERS)_{it}$	.894***	.431***	.356***	.337***
	(.002)	(.010)	(.012)	(.020)
$AGE\ (t)$	-.033***	-.019***	-.012***	-.008***
	(3.33e-4)	(.001)	(.001)	(.001)
$HOLIDAY_{it}$	.759***	.572***	.558***	.533***
	(.023)	(.016)	(.018)	(.018)
DAYOFWEEK Dummies	Yes	Yes	Yes	Yes
Movie Fixed Effects	No	Yes	Yes	Yes
Adjusted R-Squared	.899	.952	-	-
Cluster-Robust Standard Error	No	Yes	Yes	Yes
Number of Observations	49,057			

*$p < .05$; **$p < .01$; ***$p < .001$.

Robustness Checks

We check the robustness of our findings against alternative measures of spoiler reviews and report estimation results from GMM with IV corrections in Table 5.[7] We first re-estimate the model in Equation (2.6) using simpler measures of spoiler reviews. In column 1, we consider a benchmark model to include $SPOILER$, a dummy variable that equals one if there is at least one spoiler review within the last 10 days, to capture the relationship between the availability of spoiler reviews and box office revenues. In column 2, we replace *INTENSITY* in Equation (2.6) with $NWORDS$, the total count of words associated with spoiling topics in spoiler reviews within the last 10 days. Consistent with the main findings, both $SPOILER$ and $\ln(NWORDS)$ have positive and significant associations with box office revenue.

We further check the sensitivity of our results against various aspects in the spoiler intensity specification. In particular, we consider an alternative spoiler intensity metric denoted by $INTENSITY^A$, which assumes equal weight (i.e., $\alpha_{ij} = \frac{1}{J}$ in Equation 2.5) among spoiling topics (column 3), or uses average function for aggregation (column 4), or uses sum function for aggregation (column 5), or uses a longer lag window of three weeks (column 6)[8], which covers 92.1% of first-page spoiler reviews according to the data collected in April 2019. Across column 3 to column 6, the coefficient of $\ln(INTENSITY^A)$ is positive and significant, whereas the coefficient of $PROP$ is nonsignificant, supporting the robustness of our findings.

Lastly, we consider the possibility that high-quality movies can attract more intense spoiler reviews over time, creating the risk that the cross-sectional differences in box office dynamics can load onto the spoiler intensity variable. To test this possibility, we allow for

[7] We conduct additional robustness checks to spoiler intensity from non-spoiler reviews in Appendix B.3.
[8] We update the variable *PROP* using the three-week window accordingly.

heterogeneous time trend across movies by including an interaction term between *CUMRATING* and *AGE*, in which *CUMRATING* serves as a proxy for movie quality. Results in column 7 confirm that the positive effect of spoiler intensity still holds.

Table 5. Estimation Results from Robustness Checks

	Simpler Measures of Spoiler Reviews		Alternative Specifications of Spoiler Intensity				Heterogeneous Trend
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\ln(DAILYREV)_{i,t-1}$	.637*** (.017)	.625*** (.018)	.647*** (.017)	.649*** (.017)	.661*** (.016)	.652*** (.017)	.622*** (.018)
$\ln(INTENSITY)_{i,t-1}$	-	-	-	-	-	-	.067*** (.015)
$\ln(INTENSITY^{A})_{i,t-1}$	-	-	.072*** (.012)	.102** (.038)	.064*** (.009)	.060* (.026)	-
$SPOILER_{i,t-1}$	.082*** (.017)	-	-	-	-	-	-
$\ln(NWORDS)_{i,t-1}$	-	.036*** (.005)	-	-	-	-	-
$PROP_{i,t-1}$	-	-.072 (.066)	4.68e-4 (.056)	.100 (.061)	-.021 (.055)	.052 (.087)	.097 (.069)
$\ln(CUMRATING)_{i,t-1}$	.161*** (.028)	.168*** (.027)	.166*** (.027)	.171*** (.027)	.171*** (.025)	.164*** (.027)	.115** (.037)
$\ln(CUMVOL)_{i,t-1}$	.050*** (.010)	.033*** (.010)	.028** (.009)	.037*** (.010)	.023* (.010)	.040*** (.011)	.043*** (.011)
$\ln(ADVERT)_{i,t-1}$	.101*** (.007)	.101*** (.007)	.089*** (.006)	.085*** (.006)	.080*** (.006)	.094*** (.006)	.103*** (.007)
$\ln(THEATERS)_{it}$	.338*** (.020)	.340*** (.020)	.335*** (.019)	.332*** (.019)	.309*** (.017)	.317*** (.018)	.344*** (.020)
$AGE\ (t)$	-.008*** (.001)	-.007*** (.001)	-.007*** (.001)	-.008*** (.001)	-.007*** (.001)	-.008*** (.001)	-.013*** (.003)
$HOLIDAY_{it}$	.534*** (.018)	.535*** (.018)	.531*** (.018)	.532*** (.018)	.528*** (.018)	.528*** (.018)	.536*** (.018)
$\ln(CUMRATING)_{i,t-1} \times t$	-	-	-	-	-	-	.003* (.001)
DAYOFWEEK Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Movie Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Endogeneity Corrections	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster-Robust Standard Error	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	49,057						

*$p < .05$; **$p < .01$; ***$p < .001$.

2.8 Underlying Mechanism

We further investigate the behavioral mechanism that may drive the positive effect of spoiler reviews on demand. For experiential products like movies, potential consumers often visit online review platforms to seek diagnostic information to resolve uncertainty (Dellarocas 2003; Goh, Heng, and Lin 2013). Compared to non-spoiler reviews, spoiler reviews are more diagnostic in reducing uncertainty because spoiler reviews can reveal important plot-related information as justification when critiquing a movie, while non-spoiler reviews cannot. The reduction in potential moviegoers' uncertainty about movie quality due to spoiler reviews might lead to higher demand.

Moderator Analysis

To indirectly test the uncertainty-reduction mechanism of spoiler reviews, we consider four potential moderators of the effect of spoiler intensity: (i) release type (limited release vs. wide release), (ii) movie age, (iii) advertising, and (iv) average user rating. If uncertainty reduction is important, we expect the positive effect of spoiler intensity to be stronger under greater movie uncertainty.

We first consider whether the positive effect of spoiler intensity varies by the release type of a movie. Intuitively, it is in the movie studio's financial interest to play the movie in as many theaters as possible. However, a wide release strategy typically requires significant marketing investment and substantial negotiating power on behalf of the distributor (Kerrigan 2017). As a result, this strategy is often reserved for mainstream and potential blockbuster movies, whereas independent movies often employ a limited release strategy. Compared to mainstream movies (typically developed to appeal to the masses and thus more predictable or formulaic), independent movies are generally avant-garde and associated with higher uncertainty in terms of artistic quality (Holbrook 1999). As such, we anticipate the positive effect of spoiler intensity to be stronger for independent movies — which, as noted, often use limited release.[9] We follow the literature to define a limited release movie, denoted by a dummy variable *LIMITED,* as a movie that plays in less than 700 theaters on its opening day, and a wide release movie (i.e., *LIMITED* = 0) as a movie that plays in more than 700 theaters (Fellman 2006; Kerrigan 2017).

We also examine the moderating role of movie age. We expect that consumers have higher movie uncertainty in the earlier (vs. later) period of a movie's life cycle since more quality signals (e.g., online WOM) become available as time goes by. For instance, past box office revenue can serve as a quality signal for potential moviegoers because high-quality movies tend to accrue greater ticket sales over time than low-quality movies (Moon, Bergey, and Iacobucci 2010). Following this rationale, we anticipate the positive effect of spoiler intensity to be greater in the earlier period after the movie release due to the higher movie uncertainty.

It is well known that the informative function of advertising can reduce product uncertainty for potential buyers (Bagwell 2007; Hoch and Ha 1986). For example, Kim and Krishnan (2015) find that product descriptions and video commercials provided by online market platforms have a significant effect in reducing product uncertainty for intangible products. Moreover, Basuroy, Desai, and Talukdar (2006) suggest that advertising can serve as a credible signal of quality in the movie industry because any upward deviation of true quality in advertising content (i.e., overselling) can result in negative WOM and long-term harms, and therefore will not be adopted by movie studios. Based on these findings, we expect that the positive effect of spoiler intensity is more salient for movies that spend less on advertising.

[9] Release type is more objective and well-defined in the industry than movie type, which is subjective in nature.

The last moderator we consider is average user rating. Compared to extreme ratings (either very high or low), ratings in the middle range tend to convey more ambiguous signals about movie quality (Tang, Fang, and Wang 2014). Thus, we expect that for movies with moderate or mixed ratings, consumers are more likely to seek additional information to reduce movie uncertainty. Following this line of thought, we hypothesize an inverted-U relationship between the effect of spoiler intensity and average user ratings. To test this relationship, we classify our movie sample into quartiles based on the average user ratings, and then include the first and fourth quartile dummies—denoted as *QUART1* and *QUART4*, respectively—as moderators for spoiler intensity.[10]

We examine the moderators by re-estimating Equation (2.6) with additional interaction terms with spoiler intensity using GMM. We report estimation results in Table 6, where column 1 presents the results without the interactions between spoiler intensity and quartile dummies of average user ratings, and column 2 presents results using the complete set of moderators.[11] Given the consistency of estimates, we summarize findings by focusing on the results in column 2. In line with our hypotheses, the positive and significant coefficient of the interaction between *INTENSITY* and *LIMITED* suggests that the positive effect of spoiler intensity is greater for limited release movies than for wide release movies. The negative and significant coefficient of the interaction between *INTENSITY* and *t* indicates a decay in the effect of spoiler reviews over time. Furthermore, the positive effect of spoiler intensity is negatively associated with advertising spending and is stronger for movies with moderate user ratings. The negative and significant coefficients of the interaction terms between *INTENSITY* and the two quartile

[10] We do not include the interaction terms between spoiler intensity and average user rating and its squared term directly because of multicollinearity: the variance inflation factor has a mean of 13.44 and a maximum of 69.10.
[11] P-values of Hansen's *J*-test and the test for AR(2) are .713 and .306 for column 1, and .645 and .498 for column 2.

dummies (i.e., *QUART1* and *QUART4*) reveal an inverted-U relationship between average user rating and the effect of spoiler intensity on box office revenue. In sum, the results from the moderator analysis are consistent with the uncertainty-reduction mechanism of spoiler reviews.

Table 6. Estimation Results of the Model with Interaction Terms with Spoiler Intensity

	Excluding Interaction with Quartile Dummies of Average User Ratings (1)	Including Interaction with Quartile Dummies of Average User Ratings (2)
$\ln(INTENSITY)^{MC}{}_{i,t-1} \times LIMITED_i$	.351*** (.067)	.439*** (.071)
$\ln(INTENSITY)^{MC}{}_{i,t-1} \times t$	-.005*** (.001)	-.002** (.001)
$\ln(INTENSITY)^{MC}{}_{i,t-1} \times \ln(ADVERT)^{MC}_{i,t-1}$	-.186*** (.026)	-.163*** (.022)
$\ln(INTENSITY)^{MC}{}_{i,t-1} \times QUART1_i$	-	-.409*** (.079)
$\ln(INTENSITY)^{MC}{}_{i,t-1} \times QUART4_i$	-	-.250*** (.052)
$\ln(DAILYREV)_{i,t-1}$	.529*** (.019)	.499*** (.018)
$\ln(INTENSITY)^{MC}{}_{i,t-1}$	.085* (.041)	.154** (.050)
$PROP_{i,t-1}$	-.119 (.101)	-.136 (.109)
$\ln(CUMRATING)_{i,t-1}$	.214*** (.052)	.194*** (.054)
$\ln(CUMVOL)_{i,t-1}$	.058*** (.016)	.036* (.017)
$\ln(ADVERT)^{MC}{}_{i,t-1}$	.193*** (.018)	.215*** (.018)
$\ln(THEATERS)_{it}$	.430*** (.021)	.484*** (.020)
$AGE\ (t)$	-.012*** (.001)	-.010*** (.001)
$HOLIDAY_{it}$	.552*** (.018)	.560*** (.018)
DAYOFWEEK Dummies	Yes	Yes
Movie Fixed Effects	Yes	Yes
Endogeneity Corrections	Yes	Yes
Cluster-Robust Standard Error	Yes	Yes
Number of Observations	49,057	

*p < .05; **p < .01; ***p < .001. "MC" denotes mean-centered.

Event Study

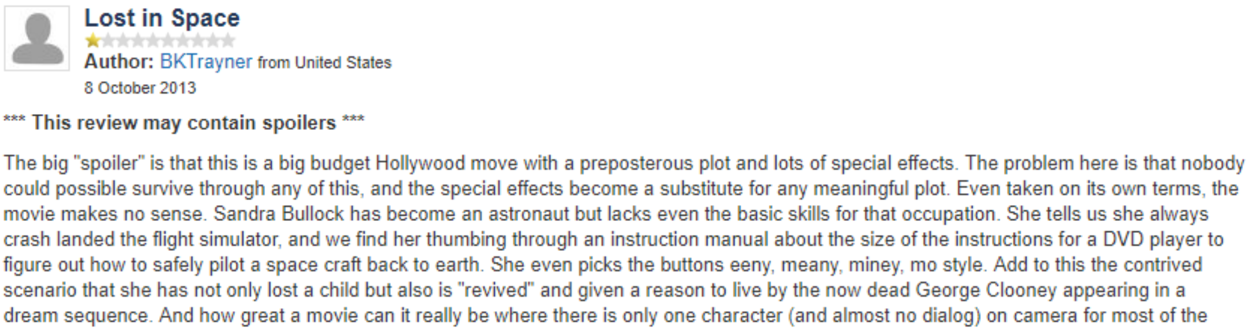
We provide additional support for the uncertainty-reduction mechanism using an event study, which focuses on an exogenous IMDb website update on December 11, 2017 (IMDb 2017). This update made two major changes to the way movie reviews are displayed on IMDb: (i) reviews are displayed *only* in the order of "helpfulness" (from most helpful to least helpful), while reviews could be sorted in a variety of ways (e.g., by date, most positive/negative, etc.) prior to the update; and (ii) the content of spoiler reviews is hidden by default and IMDb requires users to manually click on the spoiler review to see the content. The ability to sort reviews by methods other than helpfulness was restored after a subsequent update on February 10, 2018.

Theoretically, displaying all reviews according to their helpfulness should increase the diagnostic value of non-spoiler reviews and therefore decrease the relative usefulness of spoiler reviews in reducing movie uncertainty. In addition, hiding spoiler reviews by default increases the consumers' cost of reading spoiler reviews. Because of the decrease in relative benefit and the increase in cost of reading, we expect the positive effect of spoiler reviews on demand to be smaller after the IMDb update.

To compare the effect of spoiler reviews before and after the IMDb update (see Figure 5 for an example), we collected the daily box office revenue data for all 47 movies that were screened in the United States both before and after December 11, 2017. For these 47 movies, we further collected a total of 16,742 movie reviews, 4,309 of which are spoiler reviews. We report the descriptive statistics in Table 7.

Figure 5. Screenshot of Spoiler Reviews Before and After the Update

a. Before December 11, 2017



Lost in Space

Author: BKTrayner from United States

8 October 2013

***** This review may contain spoilers *****

The big "spoiler" is that this is a big budget Hollywood move with a preposterous plot and lots of special effects. The problem here is that nobody could possible survive through any of this, and the special effects become a substitute for any meaningful plot. Even taken on its own terms, the movie makes no sense. Sandra Bullock has become an astronaut but lacks even the basic skills for that occupation. She tells us she always crash landed the flight simulator, and we find her thumbing through an instruction manual about the size of the instructions for a DVD player to figure out how to safely pilot a space craft back to earth. She even picks the buttons eeny, meany, miney, mo style. Add to this the contrived scenario that she has not only lost a child but also is "revived" and given a reason to live by the now dead George Clooney appearing in a dream sequence. And how great a movie can it really be where there is only one character (and almost no dialog) on camera for most of the

b. After December 11, 2017

9/10

Magnificent portrait of organized crime

ks4 25 December 2002

Warning: Spoilers

519 out of 736 found this helpful. Was this review helpful? Yes No | Report this

Table 7. Descriptive Statistics of Data in the Event Study

Variable	Mean	Standard Deviation	Minimum	Maximum
DAILYREV (in $1,000)	664	2,536	.005	50,426
THEATERS	839	1,162	1	4,535
CUMRATING	6.46	1.65	0	10
CUMVOL	347.4	438.3	1	2,506
PROP	.11	.14	0	1
INTENSITY	3.66	4.00	0	23.08

We apply the same method to measure spoiler intensity, and estimate the following model:

$$
\begin{aligned}
\ln(DAILYREV)_{it} &= \delta_1 \ln(DAILYREV)_{i,t-1} + \delta_2 \ln(INTENSITY)_{i,t-1} + \delta_3 PROP_{i,t-1} \\
&+ \delta_4 \ln(CUMRATING)_{i,t-1} + \delta_5 \ln(CUMVOL)_{i,t-1} \\
&+ \delta_6 \ln(ADVERT)_{i,t-1} + \delta_7 \ln(THEATERS)_{it} + \delta_8 t + \delta_9 HOLIDAY_{it} \\
&+ \delta_{10} EVENT_{it} + \delta_{11} \ln(INTENSITY)_{i,t-1} \times EVENT_{it} \\
&+ \delta_{12} PROP_{i,t-1} \times EVENT_{it} + \sum_{d=1}^{6} \gamma_j I\{DAYOFWEEK_{it} = d\} + \omega_i \\
&+ \varepsilon_{it} \qquad (2.7)
\end{aligned}
$$

where $EVENT_{it}$ is a dummy variable that equals one if the observation was after December 11, 2017. We allow the effects of spoiler-related variables to be different before and after the update. In particular, δ_2 (δ_3) and δ_{11} (δ_{12}) measure the effect of spoiler intensity (spoiler volume) on box office revenue before and after the update, respectively.

We use the GMM approach with panel instruments to estimate Equation (2.7). We do not use instruments for the marketing mix variables, mainly because of the relatively small sample associated with the event (i.e., 47 movies), which prevents us from creating IVs based on other movies from the same genre. We report the estimation results of Equation (2.7) in Table 8, where only the event dummy is added to the model in column 1, and additional interactions between the event dummy and spoiler-related variables are added to the model in column 2.

We see results consistent with our main findings: *INTENSITY* is a significant and positive predictor of box office revenue, while the effect of spoiler volume captured by *PROP* is nonsignificant. The coefficient of *EVENT* is also nonsignificant, suggesting that the IMDb update is not directly associated with the box office revenues of movies that were screened around this time. This result is not surprising, as the update only affected the display—not the availability—of movie reviews on the platform. Our main parameter of interest is the coefficient of the interaction between *EVENT* and *INTENSITY*. If the uncertainty-reduction mechanism is the key driver of the positive effect of spoiler reviews, the positive effect of *INTENSITY* should be attenuated after the update. We indeed see that the coefficient of the interaction between *EVENT* and *INTENSITY* is significant and negative.

Is it possible that the decrease in the effect of spoiler intensity is driven by behavioral changes after the update? Specifically, might users be less willing to write detailed spoiler reviews after the update because of the unappealing changes regarding the display of spoiler reviews? To investigate this possibility, we regress *INTENSITY* on *AGE, HOLIDAY, DAYOFWEEK, EVENT*, and movie fixed effects, and find that *EVENT* is negative but statistically nonsignificant. Thus, we failed to find evidence that the IMDb update affected the intensity of new spoiler reviews posted on the platform.

Table 8. Impacts of Spoiler Reviews Before and After the IMDb Website Update

	Excluding Interactions with Event Dummy (1)	Including Interactions with Event Dummy (2)
$\ln(DAILYREV)_{i,t-1}$	.720***	.717***
	(.027)	(.029)
$\ln(INTENSITY)_{i,t-1}$	.195***	.293***
	(.043)	(.085)
$PROP_{i,t-1}$	-.133	-.604
	(.310)	(.434)
$\ln(CUMRATING)_{i,t-1}$	.042	.050
	(.067)	(.064)
$\ln(CUMVOL)_{i,t-1}$	.027	.025
	(.028)	(.027)
$\ln(ADVERT)_{i,t-1}$	.043***	.040***
	(.010)	(.011)
$\ln(THEATERS)_{it}$	.232***	.240***
	(.026)	(.029)
$AGE\ (t)$	-.004**	-.004**
	(.001)	(.001)
$HOLIDAY_{it}$	.628***	.622***
	(.048)	(.047)
$EVENT_{it}$	-.034	.108
	(.066)	(.082)
$\ln(INTENSITY)_{i,t-1} \times EVENT_{it}$	-	-.155*
		(.072)
$PROP_{i,t-1} \times EVENT_{it}$	-	.564
		(.490)
DAYOFWEEK Dummies	Yes	Yes
Movie Fixed Effects	Yes	Yes
Cluster-Robust Standard Error	Yes	Yes
Number of Observations	4,290	

* < .05; ** < .01; *** < .001.

2.9 Discussion

Although the relationship between spoilers and media enjoyment has received some academic attention, the relationship between spoilers and demand remains a knowledge gap in the literature. In this research, we show that the degree of plot uncertainty resolved by movie reviews (i.e., spoiler intensity) has a positive and significant association with box office revenue with an elasticity of .06. In addition, we provide evidence that uncertainty reduction is the behavioral mechanism that drives the positive effect of spoiler intensity using moderator analysis and an event study. Our finding of the positive association is novel in the movie industry, where the conventional knowledge is that spoilers hurt box office revenues. Moreover, our conceptual framework of spoilers can be generalized to other product categories (e.g., television shows, role-playing games, novels, etc.). Although we find a positive net effect of spoiler reviews in the movie context, the relative importance of the positive uncertainty-reduction effect and the negative surprise-burst effect of spoilers may vary across product categories, and therefore warrants further investigation.

Managerial Implications

Our findings provide important managerial implications for movie studios, theaters, and review platforms. Foremost, our results suggest that online review platforms can potentially increase consumer welfare in the entertainment industry. The uncertainty-reduction mechanism that we have uncovered suggests a spoiler-friendly review platform can provide diagnostic plot-related information through spoiler reviews to help consumers make purchase decisions. Accordingly, we recommend online review platforms to maintain the availability of spoiler reviews, especially plot-intense spoiler reviews for potential consumers. We also recommend review platforms to keep the warning labels of spoiler reviews because of the benefit of allowing consumers to selfselect into the exposure to spoilers. These spoiler-alert warnings reduce the search cost for consumers who seek to reduce movie uncertainty, while shield consumers who care about movie enjoyment from the unfavorable effects of spoiler reviews. Furthermore, with advances in information technology, online review platforms can even go one step further to customize the number of displayed spoiler reviews and adjust the prominence of warning labels catering to an individual consumer's preference revealed by his or her historical spoiler reading behavior.

Second, movie studios and theaters should actively monitor the content of spoiler reviews to better forecast future box office revenue. To demonstrate the predictive power of spoiler intensity, we randomly split the data into quarters, and then used three-quarters of the data as a training sample and the remaining quarter as a hold-out sample. By adding WOM-related variables individually to the benchmark model without any WOM-related variables, we calculated the predictive power of each WOM-related variable using the lift in the model's R-squared on the hold-out sample. We find that the lift in R-squared is .010 for spoiler intensity, .007 for spoiler volume, .011 for WOM volume, and .004 for WOM valence, suggesting that spoiler intensity explains 1% of data variation. More importantly, the predictive power of spoiler intensity is slightly below that of WOM volume and more than twice that of WOM valence. Given the industrial routine of monitoring WOM volume and valence in forecasting, we recommend that movie studios and theaters also actively monitor the content of spoiler reviews to improve forecasting performance.

Third, the benefit of monitoring the spoiler intensity of movie reviews, a particular act of social listening, is greater for movies with less advertising spending. To support this claim, we conducted a spotlight analysis to examine the elasticity of spoiler intensity at different levels of advertising. Specifically, we calculated the elasticity of spoiler intensity for advertising at the

25th ($1,243 per day) and 75th percentiles ($3,452 per day), respectively. We find that for movies with low levels of advertising (25th percentile), the elasticity of spoiler intensity is significant and large (.234, $p < .001$)—almost four times the magnitude of the elasticity for an average movie (.060, $p < .001$). However, the elasticity for movies with high levels of advertising is statistically nonsignificant (.067, $p > .05$). These findings suggest that movies with relatively small advertising budgets (e.g., most movies released by independent and arthouse studios) benefit the most from monitoring the content of spoiler reviews.

Fourth, the decay of the positive effect of spoiler intensity over time suggests that managers should make greater monitoring efforts in the earlier, rather than later, period of a movie's life cycle. To identify the specific window in which it is most beneficial to monitor spoiler reviews, we conducted a spotlight analysis for the elasticity of spoiler intensity at different days after the movie release. We find that the elasticity is the greatest on the opening day (.149, $p < .01$), and then steadily declines (i.e., week 1: .129, week 2: .110, week 3: .093, all with $p < .05$) until it becomes statistically nonsignificant at the end of the fourth week (.077, $p > .05$).

Finally, we highlight the boundary conditions under which movie studios might benefit from encouraging more intense spoiler reviews that can help reduce the uncertainty of movie quality. In particular, our findings suggest that for movies with small advertising budget and mixed user ratings, the marketing managers should place great emphasis on stimulating online WOM, including those that might spoil the movie plot. However, for movies with large advertising budget and extreme user ratings, we do not recommend movie managers to encourage consumers to generate spoiler reviews because of the lack of a significant effect on sales. In addition, the creation of spoiler reviews after three weeks of movie release does not seem to generate an economically meaningful impact on sales either. Although a no-spoiler policy is not recommended, we also caution movie studios that the dissemination of spoilers is sometimes uncontrollable. For example, spoiler reviews on IMDb can spread via social media where warning labels do not exist, which makes consumers more subject to the unfavorable effect of spoilers.

Limitations and Directions for Future Research

We note several limitations of this study, all of which provide promising directions for future research. Although we focus on online movie review platforms as the main source of online WOM, these platforms represent only one source of online WOM—one that consumers must actively seek out. Future research could explore whether our findings can be generalized to spoilers on social media platforms, where users are more likely to read spoilers by chance. Furthermore, we focus on the net effect of spoilers in this research. Future research could test for a parallel mediation of spoilers on movie demand, with a positive path via uncertainty reduction and a negative path via the burst of surprise.

We also note that this research focuses on spoilers that are generated by consumers. While we find a positive net effect of spoiler reviews, our results may not generalize to "leaks." Leaks, unlike spoilers, refer to information that is typically released from the supply side (e.g., movie producers, staff), either accidentally or maliciously prior to a movie's release. Leaks can take many forms but are often disseminated via images (e.g., camera shots on set, posters, etc.) and videos (e.g., production footage, unedited clips, etc.). Although the effects from leaks would be controlled by movie fixed effects in our model, conceptual questions remain as to how spoilers and leaks differ in affecting ticket sales and whether they operate by the same behavioral mechanism. We leave these questions to future researchers.

Spoilers may also appear in other media, such as images on Pinterest and videos on YouTube. A notable feature of IMDb is that it offers expressive freedom to consumers at a relatively low cost of content generation (i.e., it is free to create an account and write reviews). In contrast, the creation of images and videos often requires skills of artistic design, content editing, and so on, suggesting a high cost of content generation. Consequently, we suspect that the generation of spoilers on platforms that focus on images and videos is centralized to professional content creators. With the advances in machine learning and unstructured data analysis, future research could examine how user-generated spoilers delivered via media other than text affect consumer demand.

Finally, we use the max function in the specification of spoiler intensity to capture the discount of spoiling information that has appeared in previous spoiler reviews without the knowledge of individual review-viewing behavior. Because detailed review-viewing data are typically unavailable to movie studios, the proposed spoiler intensity metric should be useful to managers in the movie industry and therefore serves as a first step. Should individual-level data become available, future studies could relax the assumptions we made and extend the spoiler intensity metrics for both academics and practitioners.

References

Alba, Joseph W., and Elanor F. Williams (2013), "Pleasure Principles: A Review of Research on Hedonic Consumption," Journal of Consumer Psychology, 23 (1), 2–18.

Anderson, Theodore W., and Cheng Hsiao (1981), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76 (375), 598–606.

Anderson, Theodore W., and Cheng Hsiao (1982), "Formulation and Estimation of Dynamic Models using Panel Data," *Journal of Econometrics*, 18 (1), 47–82.

Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo HA Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297–318.

Bagwell, Kyle (2007), "The Economic Analysis of Advertising," in *Handbook of Industrial Organization*, 3, 1701–1844.

Basuroy, Suman, Kalpesh K. Desai, and Debabrata Talukdar (2006), "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research*, 43 (2), 287–295.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3 (Jan), 993–1022.

Blei, David M., and John Lafferty (2005), "Correlated Topic Models," *In Proceedings of the 18th International Conference on Neural Information Processing Systems* (pp. 147–154). MIT Press.

Blei, David M., and John Lafferty (2007), "A Correlated Topic Model of Science," *The Annals of Applied Statistics*, 1 (1), 17–35.

Blundell, Richard, and Stephen Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87 (1), 115–143.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L. Boyd-Graber, and David M. Blei (2009), "Reading Tea Leaves: How Humans Interpret Topic Models," *In Advances in Neural Information Processing Systems* (pp. 288–296).

Chevalier, Judith A., and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–354.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944–957.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (2), 149–171.

Dellarocas, Chrysanthos (2003), "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49 (10), 1407–1424.

Dellarocas, Chrysanthos, Xiaoquan Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Deighton, John, Daniel Romer, and Josh McQueen (1989), "Using Drama to Persuade," *Journal of Consumer Research*, 16 (3), 335–343.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007–1016.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica (2015), "Suspense and Surprise," *Journal of Political Economy*, 123 (1), 215–260.
Fellman, Daniel R. (2006), "Theatrical Distribution," in *The Movie Business Book*, 362–374.
Gelper, Sarah, Renana Peres, and Jehoshua Eliashberg (2018), "Talk Bursts: The Role of Spikes in Prerelease Word-of-Mouth Dynamics," *Journal of Marketing Research*, 55 (6), 801–817.
Godes, David, and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–560.
Godes, David, and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (3), 448–473.
Goh, Khim-Yong, Cheng-Suang Heng, and Zhijie Lin (2013), "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-and Marketer-generated Content," *Information Systems Research*, 24 (1), 88–107.
Gopinath, Shyam, Jacquelyn S. Thomas, and Lakshman Krishnamurthi (2014), "Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance," *Marketing Science*, 33 (2), 241–258.
Hennig-Thurau, Thorsten, Caroline Wiertz, and Fabian Feldhaus (2015), "Does Twitter Matter? The Impact of Microblogging Word of Mouth on Consumers' Adoption of New Movies," *Journal of the Academy of Marketing Science*, 43 (3), 375–394.
Hoch, Stephen J., and Young-Won Ha (1986), "Consumer Learning: Advertising and the Ambiguity of Product Experience," *Journal of Consumer Research*, 13 (2), 221–233.
Holbrook, Morris B. (1999), "Popular Appeal versus Expert Judgments of Motion Pictures," *Journal of Consumer Research*, 26 (2), 144–155.
IMDb (2017), Updates to User Reviews: IMDb.com Customer Community. Retrieved July 17, 2019, from https://getsatisfaction.com/imdb/topics/update-to-user-reviews.
Johnson, Benjamin K., and Judith E. Rosenbaum (2015), "Spoiler Alert: Consequences of Narrative Spoilers for Dimensions of Enjoyment, Appreciation, and Transportation," *Communication Research*, 42 (8), 1068–1088.
Keele, Luke, and Nathan J. Kelly (2006), "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables," *Political Analysis*, 14 (2), 186–205.
Kerrigan, Finola (2017), *Film Marketing*. Routledge.
Kim, Youngsoo, and Ramayya Krishnan (2015), "On Product-level Uncertainty and Online Purchase Behavior: An Empirical Analysis," *Management Science*, 61 (10), 2449–2467.
Kooser, Amanda (2019), "Avengers: Endgame spoilers? Just don't, Russo Brothers say," Retrieved July 17, 2019, from https://www.cnet.com/news/avengers-endgame-spoilers-just-dont-russo-brothers-say.
Leavitt, Jonathan, and Nicholas Christenfeld (2011), "Story Spoilers Don't Spoil Stories," *Psychological Science*, 22 (9), 1152–1154.
Leavitt, Jonathan, and Nicholas Christenfeld (2013), "The Fluency of Spoilers: Why Giving Away Endings Improves Stories," *Scientific Study of Literature*, 3 (1), 93–104.
Lee, Moontae, and David Mimno (2017), "Low-Dimensional Embeddings for Interpretable Anchor-based Topic Inference," arXiv preprint arXiv:1711.06826.
Li, Xinxin, and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456–474.

Liu, Xiao, Param Vir Singh, and Kannan Srinivasan (2016), “A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing,” *Marketing Science*, 35 (3), 363–388.
Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (3), 74–89.
Lu, Shijie, Xin Wang, and Neil Bendle (2020), “Does Piracy Create Online Word of Mouth? An Empirical Analysis in the Movie Industry,” *Management Science*, 66 (5), 2140–2162.
Moon, Sangkil, Paul K. Bergey, and Dawn Iacobucci (2010), “Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction,” *Journal of Marketing*, 74 (1), 108–121.
Mudambi, Susan M., and David Schuff (2010), “Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon. com,” *MIS Quarterly*, 34 (1), 185–200.
Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), “When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications,” *Journal of Marketing Research*, 56 (6), 960–980.
Nickell, Stephen (1981), “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49 (6), 1417–1426.
Olson, Gary M., Robert L. Mack, and Susan A. Duffy (1981), “Cognitive Aspects of Genre,” *Poetics*, 10 (2-3), 283–315.
Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi (2016), “A Model of Text for Experimentation in the Social Sciences,” *Journal of the American Statistical Association*, 111 (515), 988–1003.
Seiler, Stephan, Song Yao, and Wenbo Wang (2017), “Does Online Word of Mouth Increase Demand? (and How?) Evidence From a Natural Experiment,” *Marketing Science*, 36 (6), 838–861.
Sun, Monic (2012), “How Does the Variance of Product Ratings Matter?” *Management Science*, 58 (4), 696–707.
Tang, Tanya, Eric Fang, and Feng Wang (2014), “Is Neutral Really Neutral? The Effects of Neutral User-generated Content on Product Sales,” *Journal of Marketing*, 78 (4), 41–58.
Tirunillai, Seshadri, and Gerard J. Tellis (2014), “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data using Latent Dirichlet Allocation,” *Journal of Marketing Research*, 51 (4), 463–479.
Trottier, David (1998), *The Screenwriter's Bible: A Complete Guide to Writing, Formatting,* Los Angeles: Silman-James Press.
Wang, Xin, Neil T. Bendle, Feng Mai, and June Cotte (2015), “The Journal of Consumer Research at 40: A Historical Analysis,” *Journal of Consumer Research*, 42 (1), 5–18.
Zillmann, Dolf, T. Alan Hay, and Jennings Bryant (1975), “The Effect of Suspense and its Resolution on the Appreciation of Dramatic Presentations,” *Journal of Research in Personality*, 9 (4), 307–323.
Zillmann, Dolf (1995), “Mechanisms of Emotional Involvement with Drama,” *Poetics*, 23 (1-2), 33–51.

3. Data Security and Firm Signaling: The Direct and Indirect Consequences of Data Breach

3.1 Abstract

This research investigates the direct and indirect consequences of data breaches that moderate the financial returns of subsequent firm signals. Drawing from signaling theory, I hypothesize that data breaches serve as information to investors that leads to signal calibration. I examine 135 data breach disclosures from 72 firms in the financial and insurance services industry between January 1, 2010 and December 31, 2018. I find that the direct financial harm from data breaches is moderated by the firm's reputation, and whether the perpetrator of the data breach is internal to the firm. To study how firm signals subsequent to data breaches are affected, I examine 6,541 patent announcements and 228 executive hiring decisions made after data breaches. I find evidence that recency and attribution of data breaches, and reputation of firms partake in the signal calibration, subsequently moderating the signals' financial returns. This research contributes to a more holistic understanding of the financial consequences of data breaches. Additionally, this research highlights the importance of firms' internal data security, as well as the planning involved in the delivery of firm announcements to investors following data breaches.

3.2 Introduction

In 2017, Equifax experienced a data breach perpetrated by a hacker that stole personal data, including the names, addresses, and Social Security numbers, of more than 147 million Americans. Despite the company’s chief executive officer issuing a video apology after the official disclosure, consumers and lawmakers across the country were nonetheless appalled by the sheer scale of the data breach (Ng 2018). Experiencing such a detrimental event can have wide-reaching consequences, both direct and indirect to the firm.

I define a data breach as disclosure of private and confidential information to an unauthorized party. Although a subset of instances, such as that experienced by Equifax, can garner disproportionately large amounts of media attention, data breaches are not rare events that affect a minority of firms. An international survey of IT professionals reveals that more than 90 percent of firms have experienced some form of threat to their data security (Kaspersky Lab 2015). Further, it is estimated that the average cost of a data breach in the United States is $8.64 million, which is increasing annually (Ponemon Institute 2020; Berinato and Perry 2018). This highlights the importance for firms to not only prevent potential data breaches, but also plan how to adapt when a data breach inevitably does occur.

Previous literature on data breaches has conceptualized the financial consequences primarily from a consumer perspective. It has been argued that when a firm notifies its customers of a data breach, customers’ perceived data vulnerability is heightened and leads to negative psychological responses (Martin et al. 2017). Feelings associated with heightened data vulnerability has been linked to negative word-of-mouth, decline in stock price, and decrease in customer spending (Martin et al. 2017; Janakiraman et al. 2018).

In this research, I examine the financial consequences of data breaches from the investor perspective. I find that in addition to the direct financial consequences, data breaches can spillover to affect the financial returns from subsequent firm actions. I draw from signaling theory, which posits that investors, in order to optimize their trading decisions, seek information from their environment to reduce the information asymmetry with firms. Strategic signals, such as patent announcements and executive hiring decisions, can then serve as cues to investors about the firm's unobservable qualities, such as an innovation project or an impending policy change respectively. Against this theoretical background, I hypothesize that a prior data breach that had affected the firm can remain in the environment as information to investors, which then influences their interpretation of the firm's subsequent signals.

I purposefully examine patent announcements and executive hiring decisions as signals for two reasons. First, patent announcements and executive hiring decisions each signal firm initiatives that are long-term oriented. As multiple signals are required to be transmitted throughout the life of a long-term initiative, my research provides possible solutions to firms that have experienced a data breach and require adjustments to their signaling routine. Second, previous research documents that both signals per se are positively interpreted by investors. By studying how otherwise positive signals can be negatively affected by a data breach, I capture the full extent of the financial consequences of data breaches, and highlight the importance of sending correct signals with respect to data breach-related and firm-related characteristics.

To empirically test my predictions, I construct a sample of data breach disclosures in the financial and insurance services industry between January 1, 2010, to December 31, 2018 in the United States. For the 72 firms in my sample of 135 data breach disclosures, I further collect their patent announcements, and executive hiring decisions made within the same time window.

As my research focuses on shareholder value, I use the event study methodology to calculate the abnormal returns from data breaches, patent announcements, and executive hiring decisions, to examine each of their determinants.

Analysis of the abnormal returns from data breaches reveals that in accordance with attribution, data breaches caused by internal perpetrators lead to worse financial outcomes compared to data breaches caused by external perpetrators. In addition, positive firm reputation attenuates the negative abnormal returns from data breaches, indicating an insurance-like effect that buffers financial harm. Independent analyses of returns from patent announcements and executive hiring decisions reveal that positive firm reputation, which is associated with signal credibility, can overall increase the signals' financial returns. Due to differences in signal interpretation, I find that the recency of prior data breaches increases the returns from executive hiring decisions, but decreases the returns from patent announcements.

It is important to note that I do not assert patent announcements and executive hiring decisions are made specifically in response to a data breach. Rather, the objective is to demonstrate that firms should consider the information in their environment prior to signaling to investors. As my sample consists of firms that have specifically experienced a data breach, I generalize my findings and implications to the corresponding population of firms.

This work contributes to the extant literature in three ways. First, previous research has primarily focused on the direct financial harm from data breaches. I contribute by additionally studying the indirect consequences from the perspective of investors. I theorize that due to signal calibration, data breaches can affect the financial returns of the afflicted firm's signals. This suggests that the total consequences from data breaches are much larger than previously anticipated, and that the relevant mitigation plans for firms should involve more than addressing technical problems directly related to the data breach.

Second, I extend the literature on signal calibration to the context of data breaches. I provide the theoretical importance of firm-related and data breach-related characteristics that affect the interpretation of firm signals by investors. From this, I provide managerial implications to optimize shareholder value from signaling routines, such as patent announcements and executive hiring decisions, for firms that have previously experienced a data breach.

Finally, I contribute to the literature on firm reputation. With respect to public crises and scandals, previous research posits two potential effects of positive reputation. The first is an insurance-like effect, in which the accrued goodwill serves as a buffer to attenuate the negative financial harm. The second is a spotlight effect, in which positive reputation leads to greater expectations from the firm's stakeholders. Consequently, a public crisis, such as a data breach can lead to a greater sense of violation and worsen its associated financial consequences. In this research, I find that positive firm reputation not only serves as a buffer to the direct financial harm from data breaches, but also increases the financial returns from subsequent signals by strengthening the firm's signal credibility.

The remainder of this essay is structured as follows. First, I review the previous literature to develop a theoretical framework based on signaling theory, conceptualizing the consequences of data breaches from the perspective of investors. Next, I theorize the link between data breaches and firm signaling to formulate my hypotheses. I then describe my empirical setting and data, along with the relevant measures of independent variables. I then discuss the event study methodology and model specifications, followed by the results. I conclude the essay with a discussion of managerial implications and limitations.

3.3 Theoretical Framework

Previous research in marketing theorizes the consequences of data breaches primarily from the perspective of consumers. Particularly, Martin et al. (2017) use gossip theory to explain that consumers' psychological responses to a data breach is similar to that of being a target of an evaluative communication by others. Becoming a target of gossip elicits negative feelings associated with violation of trust and betrayal (Martin et al. 2017). Likewise, receiving a notification that an unauthorized party has breached the firm's data increases the perception of vulnerability in the firm's customers, subsequently leading to behaviors that generate negative word-of-mouth and financial harm (Martin et al. 2017). In accordance with this customer perspective, Janakiraman et al. (2018) find in the retailing context that customers' heightened perception of data vulnerability leads to decrease in spending and increase in channel-switching behaviors.

In this research, I take the perspective of investors when conceptualizing the financial consequences of data breaches. My objective is to examine whether the financial returns from strategic firm actions, specifically patent announcements and executive hiring decisions, are affected by data breaches experienced by the corresponding firm. I draw from signaling theory to argue that the knowledge that a firm has previously experienced a data breach remains in the environment as information for investors, subsequently calibrating the valence of signals that firms may strategically release.

Signaling Theory

To optimize the return on investment, investors seek from their environment diagnostic information that can distinguish between high-quality and low-quality firms. However, information asymmetry is at the heart of the relationship between investors and firms, as investors are unable to access private or insider information held by firms (Myers and Majluf 1984). Information asymmetry poses problems for both parties, as it is also in the interests of high-quality firms to distinguish themselves from low-quality firms, in order to maximize their shareholder value. In this theoretical setting, high-quality firms then become signalers that attempt to communicate information or cues about their unobservable qualities to investors, who then become receivers of the signals (Connelly et al. 2011).

This framework based on signaling has been applied by previous research to explain various strategic firm behaviors. For example, Eliashberg and Robertson (1988) examine preannouncements of new products as a signal to consumers who might consider adopting the new product in the near future. Swaminathan and Moorman (2009) examine reputation in firm networks as a signal to investors about the firm's knowledge and skills in managing firm alliances. Joshi and Hanssens (2010) examine advertising as a signal to investors about the firm's financial well-being and competitive capability. Kim and McAlister (2011) extend the previous work by proposing that advertising can also strategically signal the firm's future cashflows to investors.

In the context of data breaches, I examine two particular firm signals: patent announcements, and executive hiring decisions. My focus on these signals is driven by two reasons. First, patent announcements and executive hiring decisions each signal firm initiatives that are long-term oriented. For firms to realize the full financial returns from their investment in either innovation or a newly hired executive requires a significant amount of time, in which the returns are slowly accrued (Sood and Tellis 2009; Schwartz and Menon 1985). This highlights the importance for the firm to develop an effective signaling routine that maximizes shareholder value, implemented until the full returns are realized. I argue that an effective signaling routine must consider the impact of a prior data breach. Second, previous research documents that both firm actions are interpreted as positive signals of firm quality by investors. By examining how a prior data breach can negatively calibrate these otherwise positive signals, I capture the full extent of the financial consequences of data breaches. This also highlights the importance for firms to consider prior data breaches when designing and transmitting their signals to investors. Below, I discuss each of the two firm signals independently, and then how a prior data breach can change their degree of valence.

Patent announcements. Innovation births new products, fuels growth, and is often necessary for firms to simply remain as incumbents in a market. Sood and Tellis (2009) define an innovation project as "the total of a firm's activities in researching, developing, and introducing any new product based on a new technology." Various firm activities constitute an innovation project, and they can each be categorized into one of three phases: *initiation*, which includes activities such as obtaining grants and new manufacturing facilities, *development*, which includes patents and product prototypes, and *commercialization*, which includes product launches (Sood and Tellis 2009). Innovation projects require a substantial amount of time and resources to initiate and develop (McGrath and Nerkar 2004). As the progress of an innovation project is unobservable to investors, firms with high-quality innovation projects can rely on signaling to reduce the information asymmetry. In this research, I focus on patent announcements because it belongs in the development phase of an innovation project, which is documented to have the largest effect on financial returns (Sood and Tellis 2009). This implies that patent announcements as a signal are unambiguously interpreted as positive by investors.

Although withholding the progress of innovation activities from the public, which competitors have access, can be beneficial for firms to obtain a competitive advantage, disclosure can also be beneficial in providing legitimacy and reducing transaction costs for the firm (Simeth and Raffo 2013). Furthermore, specific investments and capabilities are essential for the firm for scientific disclosure. For example, trained researchers with knowledge of the patent filing process, as well as balancing the benefits and drawbacks of disclosing an innovation project are necessary (Simeth and Lhuillery 2015). Therefore, I conceptualize patent announcements as voluntary signals, the content and timing of which can be designed by the firm.

Executive hiring decisions. I broadly define executives as top-level management roles. Previous research suggests that investors pay close attention to the appointments of executives, such as Chairman, Presidents, and C-suite, as they are responsible for oversight and formulation of strategies with respect to the overarching goals of the firm (Varadarajan and Clark 1994). Because of the significant resources and expertise that each executive brings to the firm, a new appointment signals to investors an impending policy change or a new corporate direction (Nath and Mahajan 2011). This signal is positively interpreted by investors, as Davidson et al. (1990) suggest that investors perceive the new executive appointment will lead to a positive difference for the firm. Accordingly, the new appointments of Chairman, Presidents, and Chief Executives have a positive effect on financial returns (Davidson et al. 1990; Furtado and Rozeff 1987).

I do not assert that executive hiring decisions are made as a direct response to data breaches. The appointments of new executives are made routinely by firms to alter their strategy, internal structures or processes, as well as in response to problems that require special skills (Davidson et al. 1990; Schwartz and Menon 1985). My argument is rather that their financial returns are moderated by a prior data breach, and I provide implications with regards to the design and timing of the announcements related to the hiring decision.

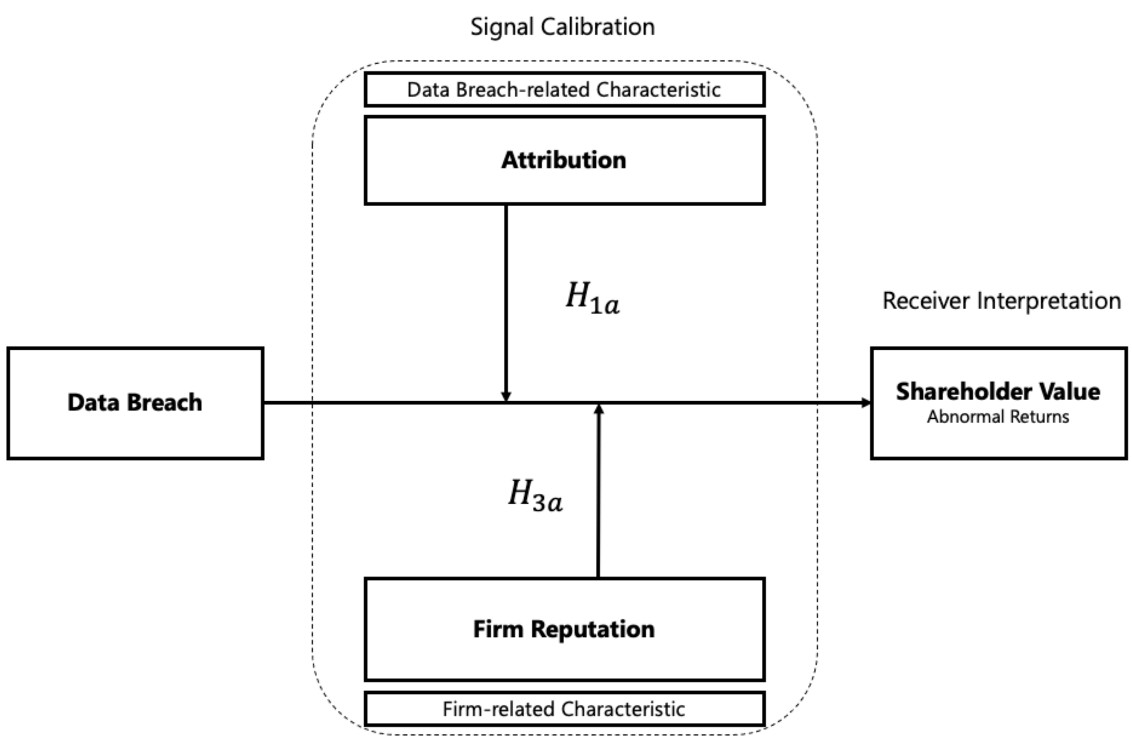
I next discuss how these two positive signals are calibrated by a prior data breach, and then formulate hypotheses for empirical testing.

Hypothesis Development

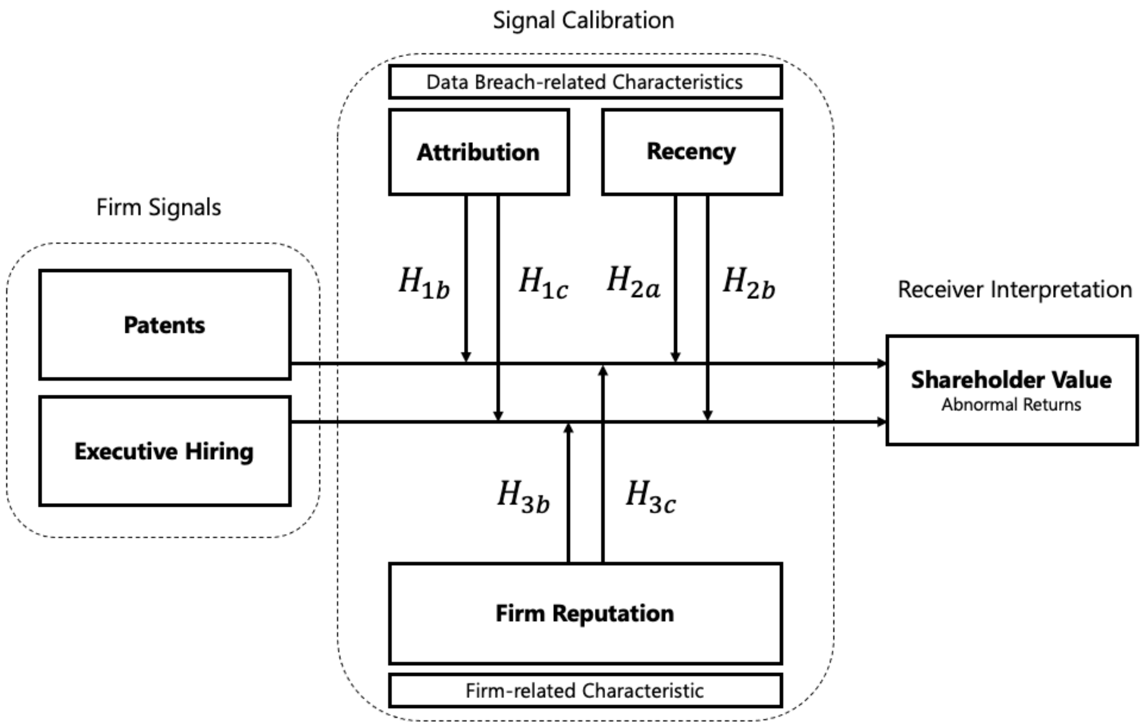
Signal calibration. It is important to note that meaning is not intrinsic to a signal—signals have to be observed and interpreted by the receiver. Consequently, the environment in which the signal is transmitted can contain other information that either strengthens or weakens the signal for receivers. In this research, I define *signal calibration* as the change in degree of valance associated with a signal due to information present in the external environment. Against this theoretical background, I formulate my hypotheses, which I present visually in Figure 6 below.

Figure 6. Conceptual Framework

A. Effect of Data Breach on Shareholder Value



B. Effect of Firm Signals on Shareholder Value



Signal calibration occurs when information in the environment is combined with the transmitted signal to affect the signal's interpretation in the minds of receivers. For example, university rankings may serve as a signal for prospective students about the quality of a potential university, but opinions from student peers may alter the valence of the signal (Connelly et al. 2011). Rynes et al. (1991) examine signal calibration in the context of job recruitment, finding that timing is an important factor for recruiters to manage. Specifically, delays in the recruitment process can generate negative inferences for the job candidate that calibrate the initial, otherwise positive signals that the recruiter may have previously sent to the candidate (Rynes et al. 1991).

Significant events that impact the firm may also serve as information for investors. For example, Park and Mezias (2005) examine firm alliance announcements made before and after the technology sector crash in 2000, revealing that the financial returns from announcements made after the crash were lower. The authors explain that firm alliances signal capability at times of high environmental munificence (i.e., before the crash), but weakness at times of low environmental munificence (i.e., after the crash; Park and Mezias 2005). Gao et al. (2015) examine advertising expenditure as a signal specifically before a product recall announcement, finding that advertising, which serves as a positive signal of future cashflows, serves as a buffer to the impending negative news. Data breaches are also detrimental events, previously conceptualized as infractions of the social contract and as service failures to the afflicted firm's customers (Janakiraman et al. 2018; Malhotra and Malhotra 2011). I hypothesize that data breaches can calibrate signals that firms transmit to investors, and that the calibration depends on data breach-related and firm-related characteristics: attribution and timing associated with the data breach, and reputation of the firm.

Attribution of breach. Previous research has viewed firms merely as victims of a data breach (Janakiraman et al. 2018). I argue by drawing from attribution theory that investors can attribute the cause of a data breach to the firm, rather than just to an external factor beyond the firm's control. Attribution theory explains how an observer gathers and combines information to arrive at a causal explanation for an event (Fiske and Taylor 1991). In this line of research, a distinction is made between internal and external causes for events. For example, Phares (1957) examines attribution in the context of achievement with respect to skill and chance. When experimental subjects were told that their success on a given task was due to skill, the subjects forecasted higher achievement in the future task, compared to subjects who were told that their success on the task was due to luck (Phares 1957). Because skill is an attribute that is internal to subjects, successes on the task were attributed to the subjects themselves. On the other hand, since luck is an attribute that is external to subjects, successes were attributed to factors deemed beyond the subjects' control (Kelley and Michela 1980).

In accordance with attribution theory, I hypothesize that investors can attribute the cause of a data breach based on whether the perpetrator is internal or external to the afflicted firm. If the perpetrator is internal to the firm (e.g., a rogue employee), the cause of the data breach can be associated with suspect firm characteristics, such as poor employee training and absence of appropriate protocols. Conversely, a data breach caused by an external perpetrator (e.g., a hacker) lacks association with characteristics that are internal to the afflicted firm, and is likely to be attributed to factors beyond the firm's control. Therefore, I hypothesize that a data breach caused by an internal perpetrator will lead to greater financial harm than data breaches caused by an external perpetrator.

H_{1a}: Data breach caused by an internal perpetrator leads to greater financial harm than data breach caused by an external perpetrator.

I propose that transmitting a positive signal (e.g., patent announcement, executive hiring decision) after the disclosure of an attributable data breach leads to signal incongruence. Signal incongruence occurs when positive and negative signals associated with the same subject come into conflict. In such an instance, previous research predicts that the negative signal will receive greater attention in the minds of receivers because negative information is perceived as more diagnostic than positive information about a subject (Skowronski and Carlston 1989; Vergne et al. 2018). For example, Vergne et al. (2018) find that when firms were revealed to have overcompensated their executives after engaging in corporate philanthropy, the firms received greater disapproval from the media. In accordance with signal incongruence, I hypothesize that a data breach caused by an internal perpetrator will lead to the attenuation of financial returns from subsequent patent announcements, or executive hiring decisions.

***H_{1b}*: Prior data breach caused by an internal perpetrator leads to attenuation in the financial returns from patent announcements.**

***H_{1c}*: Prior data breach caused by an internal perpetrator leads to attenuation in the financial returns from executive hiring decisions.**

Recency of breach. The temporal information of how far back in the past a data breach had occurred relative to the focal firm signal can influence signal calibration. Previous research documents the prevalence of recency bias, which is a cognitive bias that places greater importance for recent events when forecasting the future (Kunreuther et al. 2002). Accordingly, recency can make the prior data breach more salient in the minds of investors, and affect the degree of signal calibration. For patent announcements and executive hiring decisions, I hypothesize two different effects of data breach recency, which I discuss individually below.

As previously discussed, innovation projects incur substantial amounts of risks and resources to complete (Sood and Tellis 2009). Signaling the progress of an innovation project soon after a data breach may not be positively received because an inadequate amount of time had been made available for the firm to recover. The recovery from a data breach involves various costs from identifying and repairing the damages, monitoring the servers for future breaches, acquiring new customers, and resolving legal fines (Martin et al. 2017). It is estimated that 39 percent of the total costs are incurred more than a year after the data breach had taken place (Ponemon Institute 2020). Thus, by announcing a patent soon after a data breach, investors may perceive that the firm is pursuing untimely risks by advancing its innovation project. Furthermore, due to recency bias, investors may forecast that pursuing such risks may result in the firm increasing the likelihood of another data breach in the future, and the innovation project experiencing less than anticipated cashflows when it is later commercialized as a consequence of the data breach. In sum, I hypothesize that the more recent a data breach is, the smaller the financial returns from the subsequent patent announcement.

***H_{2a}*: The more recent a data breach is, the smaller the financial returns from the subsequent patent announcement.**

A data breach is a firm crisis that requires a comprehensive response from management to address (Malhotra and Malhotra 2011). In responding to a data breach, firms can bring together short-term and long-term remedies. For example, the identification and containment of a data breach can be viewed as short-term remedies, whereas the improvement of employee training and implementation of new policies to securely handle customer data can be viewed as long-term remedies.

Investors are prone to assigning blame to executives for organizational problems (Arthaud-Day et al. 2006). Since a new executive appointment signals to investors an impending policy change, a recent data breach can subsequently strengthen the signal when interpreted as a long-term remedy. In accordance with this conjecture, previous research indicates that changes in top management are perceived as critical determinants of an organization's adaptive behaviors, and potential solutions to organizational problems (Schwartz and Menon 1985). Moreover, I argue based on recency bias that the more recent a prior data breach is to the executive hiring decision, the more salient the data breach is in the minds of investors. This leads to an easier formation of the link between the data breach as a crisis, and the newly hired executive as a potential remedy. Thus, I hypothesize that for executive hiring decisions, the recency of data breach will increase their financial returns.

***H_{2b}*: The more recent a data breach is, the greater the financial returns from the subsequent executive hiring decision.**

Firm reputation. There are numerous determinants of firm reputation, such as financial performance, corporate social responsibility, and successful delivery of promises (Deephouse and Carter 2005; Christensen and Raynor 2013; Balmer 1997). Previous research conceptualizes reputation as an intangible asset or goodwill that consists of two dimensions: the firm's perceived capability to produce and deliver high quality products, and the firm's prominence in the minds of its stakeholders (Rindova et al. 2005). In this research, my primary focus is on the latter dimension, and define reputation as "the public recognition and social approval of an organization" (Zavyalova et al. 2016). Consequently, to manage high firm reputation, it is critical to safeguard how society and stakeholders perceive the firm (Hogarth et al. 2018).

As an intangible asset, reputation can have an insurance-like effect for firms. For example, Love and Kraatz (2009) find that high reputation can benefit the firm after corporate downsizing, which in itself signals opportunism and lack of communal commitment, by buffering these negative perceptions. Pfarrer et al. (2010) find that firms with high reputation experienced lighter financial consequences after disclosing negative earnings. Luo and Bhattacharya (2009) find that building a positive perception of the firm through corporate social responsibility leads to moral capital that lowers firm-idiosyncratic risk. In accordance with this stream of research, I hypothesize that firm reputation will serve to attenuate the direct financial harm from a data breach.

H_{3a}*: *Positive firm reputation attenuates the direct financial consequences from a data breach.

In the signaling context, positive reputation of the signaler can strengthen signal credibility. Signal credibility can be broadly defined as "whether the signal can be trusted" by its receivers (Connelly et al. 2016). Positive reputation is accrued over time through repeated interactions between the firm and its stakeholders, where the firm manages to consistently satisfy its stakeholders' expectations (Zavyalova et al. 2016). As a result, signals from firms with a positive reputation are more likely to be perceived as trustworthy by investors, in that the signaler will likely meet the signaled expectations. When risk and uncertainty are particularly salient for investors after a data breach, I argue that investors will rely on firm reputation to gauge signal credibility. Thus, I hypothesize that positive firm reputation increases the financial returns from firm signals made subsequent to a data breach.

***H_{3b}*: Positive firm reputation increases the financial returns from patent announcements made by firms that have experienced a data breach.**

***H_{3c}*: Positive firm reputation increases the financial returns from executive hiring decisions made by firms that have experienced a data breach.**

3.4 Method and Data

Event Study

As my research takes the perspective of investors and focuses on shareholder value, I examine financial abnormal returns as the dependent variable. The derivation of abnormal returns is based on the event study methodology, widely used in financial and accounting research (Sorescu et al. 2017). The theoretical underpinning for event studies is the efficient market hypothesis, which posits that stock prices at a certain point in time reflect all publicly available information that are relevant to investors' trading decisions (Fama et al. 1969). As such, fluctuations of stock price for a given firm under the efficient market hypothesis reflect investors interpreting newly revealed information and adjusting their expectations about the firm's discounted future cashflows. The event study allows me to examine how patent announcements and executive hiring decisions made after a data breach are interpreted by investors, which in turn affect shareholder value.

I use the market model to estimate abnormal returns, following conventional practice in finance and methodological recommendations, which has shown the consistency of the market model using simulations (Sorescu et al. 2017; Brown and Warner 1985). Thus, abnormal returns from an event (i.e., either a data breach, patent announcement, or executive hiring decision) can be expressed as follows:

$$A_{it} = R_{it} - (\hat{\alpha}_i + \hat{\beta}_i R_{mt}) \qquad (3.1)$$

where A_{it} is the abnormal return, and R_{it} the observed price for stock *i* on day *t.* The event date is set as $t = 0$. R_{mt} is the return from the CRSP equally weighted index for day *t* (Brown and Warner 1985), and $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the parameters from the estimation period, the length of which I specify as maximum of 255 days, and ends 46 days before the event.

A concern associated with event studies is known as event uncertainty (Flammer 2013). Event uncertainty casts doubt on whether the fluctuation of a stock price on a given date can be attributed to the event that is disclosed to the public. For example, a patent announcement could be made on day *t,* but in the evening after the stock market had closed. This would result in the information of the announcement to be reflected in the stock price the following day (i.e., $t + 1$, not *t*). Alternatively, information leakage may occur where a number of investors could receive information about an announcement earlier than the general public via unobservable channels. To address these concerns, I estimate the abnormal returns from three separate windows for a given event. In addition to the returns from the single event day $t = 0$ (i.e., $[0, 0]$), I estimate the cumulative returns from the two-day window (i.e., $[-1, 0]$), and the three-day window (i.e., $[-1, 1]$). I calculate the cumulative abnormal returns CAR_i for stock *i* as follows:

$$CAR_i[-m, n] = \sum_{t=-m}^{n} A_{it} \qquad (3.2)$$

where -m and n denote the $[-m, n]$ event window, encompassing t = 0.

Sample

Data breach announcements. To study the impact of data breaches on subsequent firm signals, I foremost require a sample of firms that have experienced a data breach, and the dates that each of the breaches were disclosed to the public. I obtain the necessary data from Privacy Rights Clearing House (PRCH), which is a non-profit organization that monitors data breach disclosures in the United States. I focus on firms in the financial and insurance services industry as classified by PRCH for two reasons. First, according to the *2019 Ninth Annual Cost of Cybercrime Study* by Accenture, the financial services industry incurred the highest costs related to cybercrime among all industries.[12] Second, the nature of customer data in this industry is particularly sensitive, but plays an important role in the day-to-day operations for firms from customer relationship management to new product innovation (Gomber et al. 2018).

From PRCH, I collected the data breach disclosures from publicly-traded firms in the United States between January 1, 2010 to December 31, 2018. I also collected the descriptions that accompany each of the breach disclosures, which detail the nature and cause of the data breach if known at the time. This resulted in a tentative sample of 174 data breach disclosures from 94 firms.

I then matched this tentative sample with the risk ratings collected from RepRisk. RepRisk is a business intelligence service that specializes in assessing environmental, social, and governance (ESG) risks of firms. By aggregating over 80,000 news sources, RepRisk monitors

[12] https://www.accenture.com/_acnmedia/PDF-96/Accenture-2019-Cost-of-Cybercrime-Study-Final.pdf#zoom=50

negative media coverage related to 28 ESG issues for over 160,000 international firms, assigning each firm a proprietary letter rating (see Appendix A for the list of issues monitored by RepRisk). These ratings resemble standard credit ratings, where AAA reflects the safest, while D the riskiest. I subsequently use these ratings as a measure of firm reputation. Firms in my sample that were not monitored by RepRisk were dropped, yielding the final sample of 135 data breach disclosures from 72 firms.

Patent announcements. Patent announcements made between January 1, 2010 to December 31, 2018 by firms in the data breach sample were collected from Factiva, a major database that aggregates news articles from numerous sources. On Factiva, the search terms “Patent Issued” were used for each company name and subject set to “patent” when searching for all the patent announcements made within the sample time window. When multiple media publications existed for the same patent, only the one with the earliest date was collected. Given my focus on patent announcements subsequent to a data breach, I only collected patent announcements that were made after the earliest data breach disclosure for each company, resulting in a total of 6,541 patent announcement dates for 47 firms.

Executive hiring decisions. Executive hiring decisions made between January 1, 2010 to December 31, 2018 by firms in the data breach sample were collected from BoardEx, which is a data provider that consolidates information regarding board members and senior management of international firms. Similar to the procedure for patent announcements, I focus on hiring decisions that were made after the earliest data breach disclosure for each company. As my focus is on executive management roles in the United States, I removed hiring announcements related to divisional and regional positions outside the United States. Moreover, I removed announcements related to Emeritus roles, and dates when multiple hiring decisions were made to avoid confounding with unobservable corporate events, such as firm acquisitions, divestitures, and restructuring. Thus, I obtained a total of 228 executive hiring decisions made by 47 firms.

Dependent Variables

For each of the three events in my sample, I estimate the abnormal returns that will serve as dependent variables for my multivariate analyses. Table 9 shows the average *CAR* for the prespecified event windows of each event.

Table 9. Average Abnormal Returns

Event Window	Mean Cumulative Abnormal Returns	Standardized Cross-Sectional Z-statistic	*N*	Multivariate Analysis F-statistic
A: Data Breach				
$[0, 0]$	-0.13%	-1.538*	135	0.950
$[-1, 0]$	-0.34%	-2.280**	135	1.366
$[-1, 1]$	-0.52%	-2.037**	135	1.446*
B: Patent				
$[0, 0]$	0.02%	1.676*	6,541	1.202
$[-1, 0]$	-0.01%	-0.184	6,541	1.167
$[-1, 1]$	0.03%	1.123	6,541	1.387**
C: Executive Hiring				
$[0, 0]$	0.08%	0.573	228	1.151
$[-1, 0]$	0.08%	0.070	228	1.160
$[-1, 1]$	-0.07%	-1.020	228	1.451**

* $p < 0.10$; ** $p < .05$; *** $p < .01$.

I find in Section A of Table 1 that the average *CAR* of data breaches is negative and statistically significant, consistent with prior research (Martin et al. 2017). However, the statistical significance of *CAR* from patent announcements is inconsistent, and that of executive hiring decisions is nonsignificant. I note that in my context, the statistical significance of *CAR* for patent announcements and executive hiring decisions per se are relatively less meaningful. Whereas previous research finds that the two events from a random sample of firms have a positive and statistically significant effect on abnormal returns (e.g., Boyd et al. 2010; Sood and Tellis 2009; Austin 1993), my sample of events is non-random, consisting of those from firms that have experienced a data breach. Given the hidden moderators that I subsequently examine, I attribute greater theoretical meaning to the subsequent multivariate analyses where I examine in-depth how the *CAR* of these events are affected by prior data breaches. My sample allows me to generalize my results not to the average firm, but to the specific population of firms that have experienced a data breach. From the three specified event-windows in Table 9, I select for multivariate analysis the window that yields a statistically significant *F*-statistic with my model specification.

Independent Variables

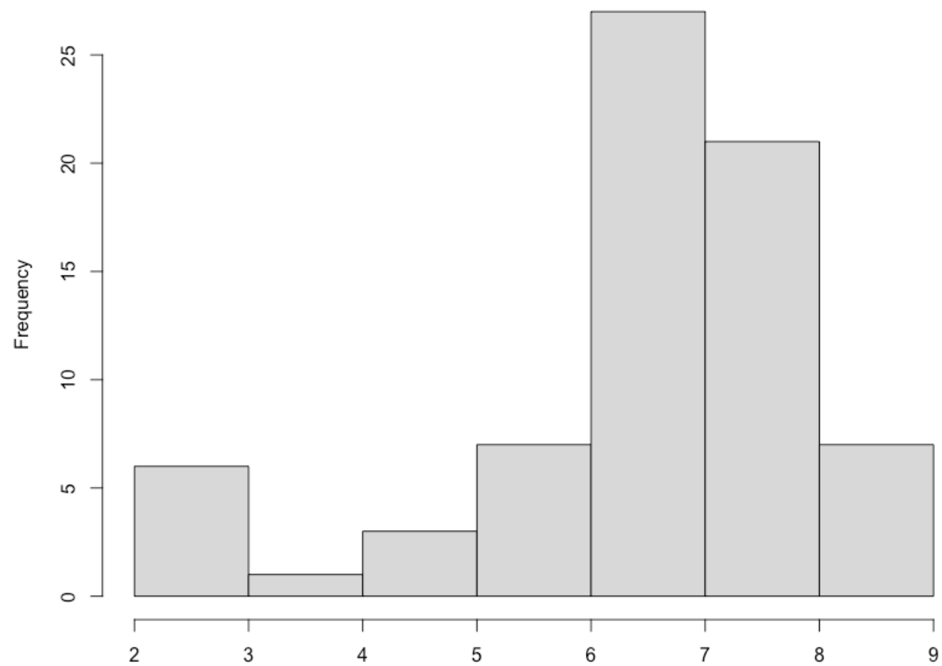
Firm reputation. I measure firm reputation using the monthly-level letter ratings collected from RepRisk. To transform the letter ratings into an ordinal variable *RiskRating*, I use the coding scheme shown in Table 10. I find that the distribution of average *RiskRating* in my data breach sample, as visualized in Figure 7, provides adequate variance to examine the moderating role of firm reputation with respect to data breaches.

Table 10. Coding Scheme for Risk Rating

Rating	**Coding**
AAA	9
AA	8
A	7
BBB	6
BB	5
B	4
CCC	3
CC	2
C	1

Note: AAA reflects the safest, while C the riskiest firm with respect to ESG factors.
"C" is the lowest rating in my sample.

Figure 7. Distribution of Average Risk Rating



Attribution of breach. Based on the descriptions of data breaches provided by PRCH, I manually code the causes of data breaches based on the coding scheme shown in Table 11.

Table 11. Coding Scheme for Attribution of Data Breach

Internal Perpetrator	External Perpetrator
Unintended Disclosures	Third-party Hacker
Rogue Employee or Insider	Skimming Devices at POS Terminals
Unidentified Cause	Theft of Physical Storage Devices

I create *Internal,* which is an indicator variable that equals one if the perpetrator of the data breach is internal to the firm. As discussed in my theoretical framework, I argue that the cause of a data breach is attributable to the firm if the perpetrator is internal (e.g., insider, employee) rather than external (e.g., third-party hacker) to the firm. Of the 135 data breaches in my sample, I find that the cause of seven data breaches were described as unknown. I code these breaches as attributable to the firm because they signal to investors flaws or incompetence in identifying and rectifying the source of the data breach.

When analyzing the abnormal returns of patent announcements and executive hiring decisions, I use *InternalPast,* which is an indicator variable that equals one if the most recent data breach prior to the focal event (i.e., patent or executive hiring announcement) is attributable to the firm (i.e., *Internal* = 1).

Recency of breach. Particularly for patent announcements and executive hiring decisions, I measure the recency of the prior data breach using *DaysSinceBreach,* which is the number of days between the focal event and the most recent data breach. Therefore, an increase in *DaysSinceBreach* indicates that the most recent data breach had occurred much further back in time, relative to the focal event. I find in my sample that the average *DaysSinceBreach* for patent announcements is 1,004 days, and that for executive hiring decisions is 848 days.

Finally, I measure the number of data breaches that a firm has experienced using *BreachCount,* the number of patent announcements made by a firm using *PatentCount,* and the number of executive hiring decisions made by a firm using *HiringCount.*

Control variables. I control for firm size (*FirmSize*) using the natural logarithm of the number of employees. Furthermore, I control for firm profitability (*Profitability*) using return on assets, which is the ratio of net income to total assets of a firm, and control for innovation intensity (*InnovationIntensity*) using the ratio of R&D expenditure to sales. The data for each of these control variables were collected from COMPUSTAT. In accordance with previous research, missing values were coded as 0 (Borah and Tellis 2014). Finally, I control for potential unobserved heterogeneity by using firm- and year-specific dummy variables. The complete list of independent variables, descriptions, and sources are presented in Table 12, and the summary statistics of the data breach sample in Table 13.

Table 12: Variable Descriptions and Sources

Variable	Description	Source
RiskRating	Ordinal variable of firm reputation based on negative media coverage related to environmental, social, and governance issues associated with the firm.	RepRisk
Internal	Indicator variable that equals one if the cause of the data breach is attributable to the firm.	Privacy Rights Clearing House
InternalPast	Indicator variable that equals one if the cause of the most recent data breach to an event of interest is attributable to the firm.	Privacy Rights Clearing House
DaysSinceBreach	Number of days since the last data breach prior to an event of interest.	Privacy Rights Clearing House
BreachCount	Number of data breaches experienced by the affected firm up to the event of interest.	Privacy Rights Clearing House
PatentCount	Number of patent announcements released by the focal firm up to the event of interest.	Factiva
HiringCount	Number of appointments to executive or corporate board by the focal firm up to the event of interest.	BoardEx
FirmSize	Natural logarithm of the number of employees in the firm.	COMPUSTAT
Profitability	Return on assets (i.e., net income/sales).	COMPUSTAT
InnovationIntensity	Proportion of R&D expenditure to sales.	COMPUSTAT

Table 13. Summary Statistics of Data Breach Sample

	Mean	Std. Dev	Min	Max
RiskScore	6.09	2.08	2	9
Internal	0.54	0.50	0	1
BreachCount	2.24	1.95	1	10
FirmSize	3.23	1.77	0	5.67
InnovationIntensity	0.01	0.04	0	0.20
Profitability	0.06	0.08	-0.24	0.40

Model Specification

To test my hypotheses, I specify separate models of *CAR* for data breaches, patent announcements, and executive hiring decisions. First, to examine whether firm reputation and attribution of data breach moderate the direct financial consequences of data breaches, I estimate the following equation:

$$CAR_i = \beta_1 + \beta_2 RiskRating_i + \beta_3 Internal_i + \beta_4 BreachCount_i + \boldsymbol{\beta'X_i} + \varepsilon_i \qquad (3.3)$$

where *i* denotes data breach disclosures as the unit of analysis. The dependent variable is CAR_i over the $[-1, 1]$ event window, which has a statistically significant F-statistic with the above model specification. $\boldsymbol{X}$ is the matrix of control variables, and ε_i is the idiosyncratic error term. Coefficients β are estimated using ordinary-least squares.

To estimate whether a prior data breach affects the *CAR* from subsequent patent announcements and executive hiring decisions, I estimate the following separate equations:

$$CAR_j = \theta_1 + \theta_2 RiskRating_j + \theta_3 InternalPast_j + \theta_4 BreachCount_j + \theta_5 PatentCount_j + \theta_6 DaysSinceBreach_j + \boldsymbol{\theta'X}_j + \epsilon_j \qquad (3.4)$$

$$CAR_m = \gamma_1 + \gamma_2 RiskRating_m + \gamma_3 InternalPast_m + \gamma_4 BreachCount_m + \gamma_5 HiringCount_m + \gamma_6 DaysSinceBreach_m + \boldsymbol{\gamma'X}_m + \pi_m \qquad (3.5)$$

where *j* denotes patent announcements, and *m* denotes executive hiring decisions as the unit of analysis. Coefficients θ and γ are also estimated using ordinary-least squares.

3.5 Results

I begin with the results of Equation (3.3) to examine the determinants of abnormal returns from data breaches. I report the main effects in column 1, and the estimates with an interaction term in column 2 of Table 14.

Table 14. Determinants of Returns from Data Breaches

	(1)	(2)
Intercept	0.356	0.128
	(0.878)	(0.858)
RiskRating × *BreachCount*	-	-0.002**
		(0.001)
RiskRating	0.010**	0.017***
	(0.005)	(0.006)
Internal	-0.016*	-0.015*
	(0.009)	(0.008)
BreachCount	-0.004	0.004
	(0.004)	(0.005)
FirmSize	-0.010	0.008
	(0.035)	(0.035)
InnovationIntensity	-2.684	-1.784
	(7.229)	(7.017)
Profitability	-0.098	-0.133
	(0.355)	(0.344)
N	135	
Adjusted R-squared	0.220	0.268
Firm Dummies	Yes	Yes
Year Dummies	Yes	Yes

* p < 0.10; ** *p* < .05; *** *p* < .01.

The main parameters of interest are *Internal* and *RiskRating.* In column 1, I find that the main effect of *Internal* is negative and statistically significant ($p < 0.10$), which provides support for H_{1a}, that data breaches attributable to the firm leads to greater financial harm. I additionally find that the effect of *RiskRating* is positive and significant ($p < 0.05$), which provides support for H_{3a}, that positive firm reputation attenuates the financial harm from data breaches.

In column 2, I examine the interaction between *RiskRating* and *BreachCount* to further explore the insurance-like effect of firm reputation. I find that the interaction term is negative and significant ($p < 0.05$), indicating that additional data breaches experienced by the firm reduces the positive buffer effect from firm reputation.

I next examine the results of Equation (3.4), which models the relationship between *CAR* of patent announcements and prior data breaches. The main effects are report in column 1, and the estimates with interaction terms in column 2 of Table 15. In column 1, I find that *DaysSinceBreach* is positive and significant ($p < 0.05$), providing support for H_{2a}, that the greater the amount of time between the patent announcement and the data breach, the larger the financial returns. Since the average *CAR* from patent announcements in my sample is 0.0003, I use the coefficient of *DaysSinceBreach* to calculate that the percentage increase of *CAR* by delaying a patent announcement by 30 days is approximately 30 percent. In addition, I find that *RiskRating is* also positive and significant ($p < 0.10$), providing support for H_{3b}, that firm reputation increases the returns from signals subsequent to a data breach. The main effect of *InternalPast* is statistically nonsignificant, which does not provide support to H_{1b}.

Table 15. Determinants of Returns from Patent Announcements

	(1)	(2)
Intercept	-0.022	-0.034
	(0.021)	(0.021)
RiskRating × *BreachCount*	-	-0.001**
		(2.25e-4)
PatentCount × *InternalPast*	-	-7.81e-6***
		(2.19e-6)
DaysSinceBreach	2.96e-6**	2.71e-6**
	(1.25e-6)	(1.28e-6)
RiskRating	0.001*	0.001**
	(4.12e-4)	(5.33e-4)
InternalPast	-0.003	-0.004
	(0.002)	(0.002)
BreachCount	0.001	0.003***
	(0.001)	(0.001)
PatentCount	4.15e-6	8.42e-6***
	(2.56e-6)	(2.97e-6)
FirmSize	0.003	0.005
	(0.004)	(0.004)
InnovationIntensity	0.063	0.050
	(0.040)	(0.041)
Profitability	0.004	0.005
	(0.013)	(0.014)
N	6,541	
Adjusted R-squared	0.004	0.006
Firm Dummies	Yes	Yes
Year Dummies	Yes	Yes

* $p < 0.10$; ** $p < .05$; *** $p < .01$.

In column 2, I find that the interaction between *RiskRating* and *BreachCount* is negative and significant ($p < 0.05$). This is consistent with the results from Equation (3.3), which indicates depletion of the buffer from positive reputation with additional data breaches experienced by the firm. I also examine the interaction between *InternalPast* and *PatentCount* to explore whether the frequency of patent announcements is affected by the attribution of data breach. I find that the interaction is negative and significant ($p < 0.01$), indicating that continuously announcing new patents after an attributable data breach leads to smaller financial returns.

I finally examine the results of Equation (3.5) that models the determinants of *CAR* from executive hiring decisions made after a data breach. I present the main effects in column 1, and the model with interaction terms in column 2 of Table 16. In column 1, I find that the main effect of *RiskRating* is positive and significant ($p < 0.05$), which provides support for H_{3c}, that positive firm reputation increases the returns from executive hiring decisions after a data breach. As in Equation (3.4), I find that the main effect of *InternalPast* is statistically nonsignificant, which does not support H_{1c}. I also find that the main effect of *DaysSinceBreach* is statistically nonsignificant.

In column 2, I examine the interaction between *DaysSinceBreach* and *InternalPast,* finding that the effect is negative and significant ($p < 0.10$). This provides conditional support for H_{2b}; the more recent a data breach that is specifically attributable to the firm, the greater the returns from executive hiring decisions. Finally, I find consistent results with Equation (3.3) and Equation (3.4), in that the interaction between *RiskRating* and *BreachCount* is negative and significant ($p < 0.01$).

Table 16. Determinants of Returns from Executive Hiring Decisions

	(1)	(2)
Intercept	-0.066	-0.108**
	(0.041)	(0.043)
RiskRating × *BreachCount*	-	-0.002***
		(0.001)
DaysSinceBreach × *InternalPast*	-	-1.10e-5*
		(5.71e-6)
DaysSinceBreach	-6.88e-6	-1.03e-6
	(4.93e-6)	(5.63e-6)
RiskRating	0.006**	0.009***
	(0.002)	(0.003)
InternalPast	0.006	0.013
	(0.008)	(0.008)
BreachCount	-0.005**	0.002
	(0.002)	(0.004)
HireCount	2.18e-4	1.00e-4
	(0.001)	(0.001)
FirmSize	-0.010	-0.010
	(0.008)	(0.008)
InnovationIntensity	-0.520	-0.417
	(0.730)	(0.712)
Profitability	0.126	0.201*
	(0.104)	(0.106)
N	228	
Adjusted R-squared	0.110	0.153
Firm Dummies	Yes	Yes
Year Dummies	Yes	Yes

* $p < 0.10$; ** $p < .05$; *** $p < .01$.

3.6 Discussion

Whereas previous research primarily focuses on the direct financial harm of data breaches, I additionally study the harm that can indirectly affect the financial returns from subsequent firm signals. I draw from signaling theory to propose that prior data breaches serve as information to investors that subsequently lead to signal calibration. To understand the signal calibration in greater detail, I theorize and test the recency and attribution of data breaches, and reputation of firms as moderators to the abnormal returns from data breach disclosures, patent announcements, and executive hiring decisions.

I foremost find that data breaches caused by perpetrators internal to the afflicted firm lead to greater financial harm relative to those caused by external perpetrators. If the perpetrator is internal to the firm, I theorize using attribution theory that the cause of the data breach can be associated with unobservable firm characteristics by investors. On the other hand, external perpetrators lack clear associations with the firm and the data breach is likely to be attributed to factors beyond the firm's control.

I also find that firm reputation is an important asset in the context of data breaches. Positive reputation not only provides an insurance-like buffer to the direct negative harm from data breaches, but also serves to increase the financial returns from subsequent firm signals. As positive reputation is accrued over time by the firm consistently meeting its stakeholders' expectations, I theorize that positive reputation can increase the firm's signal credibility.

My research also indicates that timing of firm signals after a data breach is important. I find that the recency of data breach decreases the financial returns from patent announcements, but increases the returns from executive hiring decisions. Since innovation projects require a substantial amount of time and resources, signaling its progress soon after a data breach may lead to negative perceptions from investors. Specifically, investors may perceive that the firm is pursuing untimely risks by advancing its innovation project. On the other hand, a new executive appointment after a data breach signals to investors a long-term remedy, as investors are prone to assigning blame to executives for organizational problems. Accordingly, I find that the more recent an attributable data breach (i.e., caused by an internal perpetrator) is to the executive hiring decision, the greater the financial returns.

Managerial Implications

This research provides several implications for firms with respect to data security and signaling. First, in addition to investments in data security from external threats, firms should invest in security from internal threats. Since I find that the financial harm from data breaches caused by internal perpetrators is greater, firms may benefit from strengthening their internal data security. This may involve implementing new policies regarding the handling of data, and improving employee training to minimize accidental data breaches.

Second, firms should consider the timing of their signals to investors with respect to the prior data breaches that they have experienced. Signaling firm initiatives that involve a substantial amount of risk, such as an innovation project, may financially benefit when it is timed further after a data breach. On the other hand, appointments of new executives, which signal firm initiatives related to long-term remedies for organizational problems, may financially benefit when it is timed closer to a data breach.

Third, firms should invest and safeguard their reputation. A positive reputation provides an insurance for the firm that can buffer the harm from a future data breach. As well, positive reputation increases signal credibility, which can generate greater financial returns from signaling after a data breach. Although firm reputation is a complex construct that is influenced by various factors, I have measured reputation in this research using risk ratings derived from ESG issues. Therefore, firms concerned with their reputation can monitor their ESG performance, and ensure that they do not receive negative coverage by the media on such issues.

Limitations and Directions for Future Research

I note limitations of this research, and outline potential directions for future research. The first limitation stems from my measurement of firm reputation using ESG risks. Previous research conceptualizes firm reputation as consisting of two dimensions: the perceived capability to produce high quality products, and the firm's prominence in the minds of its stakeholders (Rindova et al. 2005). The ESG risk ratings obtained from RepRisk capture the latter dimension, but not the former. Thus, firm reputation that results from high-quality products and services are not considered in this research. Future research can apply a more comprehensive measure of firm reputation, and extend this research with respect to data breaches.

Second, this research examines how the financial returns of signals made subsequent to a data breach are affected. Future research can examine potential signals that are available to the firm prior to a data breach disclosure. Similar to the research of Gao et al. (2015), which examines advertising as a signal before a product recall announcement, there may exist potential signals that can soften the negative returns directly from a data breach disclosure. This would broaden the number of signals available to the firm before and after a data breach disclosure.

Third, I only examine firms in the financial and insurance services industry. While this restriction provides a strong test of the proposed theoretical framework because it reduces the unobserved heterogeneity between firms, it may hinder the generalizability of results to other industries. Future research may examine alternative industries based on the proposed signaling framework.

References

Arthaud-Day, M. L., Certo, S. T., Dalton, C. M., & Dalton, D. R. (2006), "A Changing of the Guard: Executive and Director Turnover Following Corporate Financial Restatements," *Academy of Management Journal*, 49(6), 1119-1136.

Austin, D. H. (1993), "An Event-study Approach to Measuring Innovative Output: The Case of Biotechnology," *The American Economic Review*, 83(2), 253-258.

Balmer, J. M., van Riel, C. B., Markwick, N., & Fill, C. (1997), "Towards a Framework for Managing Corporate Identity," *European Journal of Marketing*.

Berinato, S., & Perry, M. (2018), "Security Trends by the Numbers," *Harvard Business Review,* Retrieved from https://hbr.org/2018/05/security-trends-by-the-numbers.

Borah, A., & Tellis, G. J. (2014), "Make, Buy, or Ally? Choice of and Payoff from Announcements of Alternate Strategies for Innovations," *Marketing Science*, 33(1), 114-133.

Boyd, D. E., Chandy, R. K., & Cunha Jr, M. (2010), "When Do Chief Marketing Officers Affect Firm Value? A Customer Power Explanation," *Journal of Marketing Research*, 47(6), 1162-1176.

Brown, S. J., & Warner, J. B. (1985), "Using Daily Stock Returns: The Case of Event Studies," *Journal of Financial Economics*, 14(1), 3-31.

Christensen, C., & Raynor, M. (2013), *The Innovator's Solution: Creating and Sustaining Successful Growth*, Harvard Business Review Press.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011), "Signaling Theory: A Review and Assessment," *Journal of Management*, 37(1), 39-67.

Connelly, B. L., Ketchen Jr, D. J., Gangloff, K. A., & Shook, C. L. (2016), "Investor Perceptions of CEO Successor Selection in the Wake of Integrity and Competence Failures: A Policy Capturing Study," *Strategic Management Journal*, 37(10), 2135-2151.

Davidson III, W. N., Worrell, D. L., & Cheng, L. (1990), "Key Executive Succession and Stockholder Wealth: The Influence of Successor's Origin, Position, and Age," *Journal of Management*, 16(3), 647-664.

Deephouse, D. L., & Carter, S. M. (2005), "An Examination of Differences Between Organizational Legitimacy and Organizational Reputation," *Journal of Management Studies*, 42(2), 329-360.

Eliashberg, J., & Robertson, T. S. (1988), "New Product Preannouncing Behavior: A Market Signaling Study," *Journal of Marketing Research,* 25(3), 282-292.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969), "The Adjustment of Stock Prices to New Information," *International Economic Review*, 10(1), 1-21.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Flammer, C. (2013), "Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors," *Academy of Management Journal*, 56(3), 758-781.

Furtado, E. P., & Rozeff, M. S. (1987), "The Wealth Effects of Company Initiated Management Changes," *Journal of Financial Economics*, 18(1), 147-160.

Gao, H., Xie, J., Wang, Q., & Wilbur, K. C. (2015), "Should Ad Spending Increase or Decrease Before a Recall Announcement? The Marketing–Finance Interface in Product-Harm Crisis Management," *Journal of Marketing*, 79(5), 80-99.

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018), "On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption, and Transformation in Financial Services," *Journal of Management Information Systems*, 35(1), 220-265.

Hogarth, K., Hutchinson, M., & Scaife, W. (2018), "Corporate Philanthropy, Reputation Risk Management and Shareholder Value: A Study of Australian Corporate Giving," *Journal of Business Ethics*, 151(2), 375-390.

Homburg, C., Vollmayr, J., & Hahn, A. (2014), "Firm Value Creation Through Major Channel Expansions: Evidence from an Event Study in the United States, Germany, and China," *Journal of Marketing*, 78(3), 38-61.

Janakiraman, R., Lim, J. H., & Rishika, R. (2018), "The Effect of a Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer," *Journal of Marketing*, 82(2), 85-105.

Joshi, A., & Hanssens, D. M. (2010), "The Direct and Indirect Effects of Advertising Spending on Firm Value," *Journal of Marketing,* 74(1), 20-33.

Kaspersky Lab (2015), "Global IT Security Risks Survey 2015," https://media.kaspersky.com/en/business-security/it-security-risks-survey-2015.pdf.

Kelley, H. H., & Michela, J. L. (1980), "Attribution Theory and Research," *Annual Review of Psychology*, 31(1), 457-501.

Kim, M., & McAlister, L. M. (2011), "Stock Market Reaction to Unexpected Growth in Marketing Expenditure: Negative for Sales Force, Contingent on Spending Level for Advertising," *Journal of Marketing*, 75(4), 68-85.

Kunreuther, H., Meyer, R., Zeckhauser, R., Slovic, P., Schwartz, B., Schade, C., ... & Hogarth, R. (2002), "High Stakes Decision Making: Normative, Descriptive and Prescriptive Considerations," *Marketing Letters*, 13(3), 259-268.

Love, E. G., & Kraatz, M. (2009), "Character, Conformity, or the Bottom Line? How and Why Downsizing Affected Corporate Reputation," *Academy of Management Journal*, 52(2), 314-335.

Luo, X., & Bhattacharya, C. B. (2009), "The Debate Over Doing Good: Corporate Social Performance, Strategic Marketing Levers, and Firm-Idiosyncratic Risk," *Journal of Marketing*, 73(6), 198-213.

Malhotra, A., & Kubowicz Malhotra, C. (2011), "Evaluating Customer Information Breaches as Service Failures: An Event Study Approach," *Journal of Service Research*, 14(1), 44-59.

Martin, K. D., Borah, A., & Palmatier, R. W. (2017), "Data Privacy: Effects on Customer and Firm Performance," *Journal of Marketing*, 81(1), 36-58.

McGrath, R. G., & Nerkar, A. (2004), "Real Options Reasoning and a New Look at the R&D Investment Strategies of Pharmaceutical Firms," *Strategic Management Journal*, 25(1), 1-21.

Myers, S. C., & Majluf, N. S. (1984), "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have," *Journal of Financial Economics*, 13(2), 187-221.

Nath, P., & Mahajan, V. (2011), "Marketing in the C-suite: A Study of Chief Marketing Officer Power in Firms' Top Management Teams," *Journal of Marketing*, 75(1), 60-77.

Ng, A. (2018, September 07), "How the Equifax Hack Happened, and What Still Needs to be Done," *CNet*, Retrieved from https://www.cnet.com/news/equifaxs-hack-one-year-later-a-look-back-at-how-it-happened-and-whats-changed.

Park, N. K., & Mezias, J. M. (2005), "Before and After the Technology Sector Crash: The Effect of Environmental Munificence on Stock Market Response to Alliances of e-Commerce Firms," *Strategic Management Journal*, 26(11), 987-1007.

Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. (2010), "A Tale of Two Assets: The Effects of Firm Reputation and Celebrity on Earnings Surprises and Investors' Reactions," *Academy of Management Journal*, 53(5), 1131-1152.

Phares, E. J. (1957), "Expectancy Changes in Skill and Chance Situations," *The Journal of Abnormal and Social Psychology*, 54(3), 339.

Ponemon Institute (2020), "Cost of a Data Breach Report 2020," *IBM Security.*

Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005), "Being Good or Being Known: An Empirical Examination of the Dimensions, Antecedents, and Consequences of Organizational Reputation," *Academy of Management Journal*, 48(6), 1033-1049.

Rynes, S. L., Bretz Jr, R. D., & Gerhart, B. (1991), "The Importance of Recruitment in Job Choice: A Different Way of Looking," *Personnel Psychology*, 44(3), 487-521.

Schwartz, K. B., & Menon, K. (1985), "Executive Succession in Failing Firms," *Academy of Management Journal*, 28(3), 680-686.

Simeth, M., & Lhuillery, S. (2015), "How Do Firms Develop Capabilities for Scientific Disclosure?" *Research Policy*, 44(7), 1283-1295.

Simeth, M., & Raffo, J. D. (2013), "What Makes Companies Pursue an Open Science Strategy?" *Research Policy*, 42(9), 1531-1543.

Skowronski, J. J., & Carlston, D. E. (1989), "Negativity and Extremity Biases in Impression Formation: A Review of Explanations," *Psychological Bulletin*, 105(1), 131.

Sood, A., & Tellis, G. J. (2009), "Do Innovations Really Pay Off? Total Stock Market Returns to Innovation," *Marketing Science*, 28(3), 442-456.

Sorescu, A., Warren, N. L., & Ertekin, L. (2017), "Event Study Methodology in the Marketing Literature: An Overview," *Journal of the Academy of Marketing Science*, 45(2), 186-207.

Swaminathan, V., & Moorman, C. (2009), "Marketing Alliances, Firm Networks, and Firm Value Creation," *Journal of Marketing*, 73(5), 52-69.

Varadarajan, P. R., & Clark, T. (1994), "Delineating the Scope of Corporate, Business, and Marketing Strategy," *Journal of Business Research*, 31(2-3), 93-105.

Vergne, J. P., Wernicke, G., & Brenner, S. (2018), "Signal Incongruence and its Consequences: A Study of Media Disapproval and CEO Overcompensation," *Organization Science*, 29(5), 796-817.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Hubbard, T. D. (2016), "Reputation as a Benefit and a Burden? How Stakeholders' Organizational Identification Affects the Role of Reputation Following a Negative Event," *Academy of Management Journal*, 59(1), 253-276.

4. Final Remarks

With perfect information, consumers and investors could easily optimize their decision-making processes. However, information asymmetry that arises due to the inaccessibility of information regarding true quality constrains decision making to be based primarily on public information. The main focus of this dissertation is to examine how leakages of information regarding quality, either by voluntary or involuntary means, can benefit firms.

In Essay 1, I study involuntary information leakage between firms and consumers in the movie industry. I find that the spoiling content of movie reviews, measured using the spoiler intensity metric, has a positive association with box office revenue. Furthermore, this positive spoiling effect is more prominent for movies with limited release, smaller advertising spending, and moderate user ratings, and is stronger in earlier days after the movie's release. The results indicate that uncertainty reduction is the behavioral mechanism that is driving the positive effect—potential moviegoers who are unsure about the quality of a movie can use spoilers to reduce their uncertainty.

In Essay 2, I study voluntary information leakage between firms and investors in the signaling context. In an investment setting, public information alone may be insufficient for investors to clearly differentiate between high-quality and low-quality firms. High-quality firms can then voluntarily signal their unobservable characteristics to investors using cues, such as patent announcements and executive hiring decisions that signal the progress of an innovation project and an impending policy change, respectively. I find that data breaches can alter the investors' interpretations of these signals. Specifically, the recency and attribution of data breaches, and the reputation of firms moderate the financial returns from these signals by changing the degree of the signals' valence. Thus, the research proposes that firms can carefully design their signaling routine considering these moderators to maximize their shareholder value.

With the increasing digitalization that allows consumers to freely discuss their opinions on online platforms and widely share news on social media, the information imbalance between firms and consumers is becoming easier to bypass. For example, consumers can not only use firm-generated advertisements to obtain product information, but also read online reviews generated by other consumers to make more informed purchase decisions. The theme that is shared between the two essays in this dissertation is that the reduction of information asymmetry can be beneficial for firms. Even in the context of the movie industry, where conventional wisdom suggests movies are better left unspoiled, I find that spoilers have a positive effect on box office revenue due to the uncertainty reduction mechanism.

I hope to advance with this dissertation how in this environment, firms can effectively respond to and manage information leakages. Essay 1 highlights that movies that instil greater quality uncertainty, such as limited release movies, benefit the most from spoilers. This provides recommendations for movie studios to forecast future box office revenue and adjust their marketing mix in accordance with the amount of spoiling information online. Essay 2 highlights that the environment in which signals are transmitted is important in shaping the signals' meaning. This provides recommendations for firms that have experienced a data breach to consider their reputation, timing of the signal, and attribution of the data breach prior to signaling to investors. Taken together, firms stand to benefit by unambiguously addressing the quality uncertainty in their stakeholders, be it consumers or investors.

Appendices

A. Supplementary Tables and Figures

Figure A1. Pareto Chart of Spoiler Reviews for 993 Movies

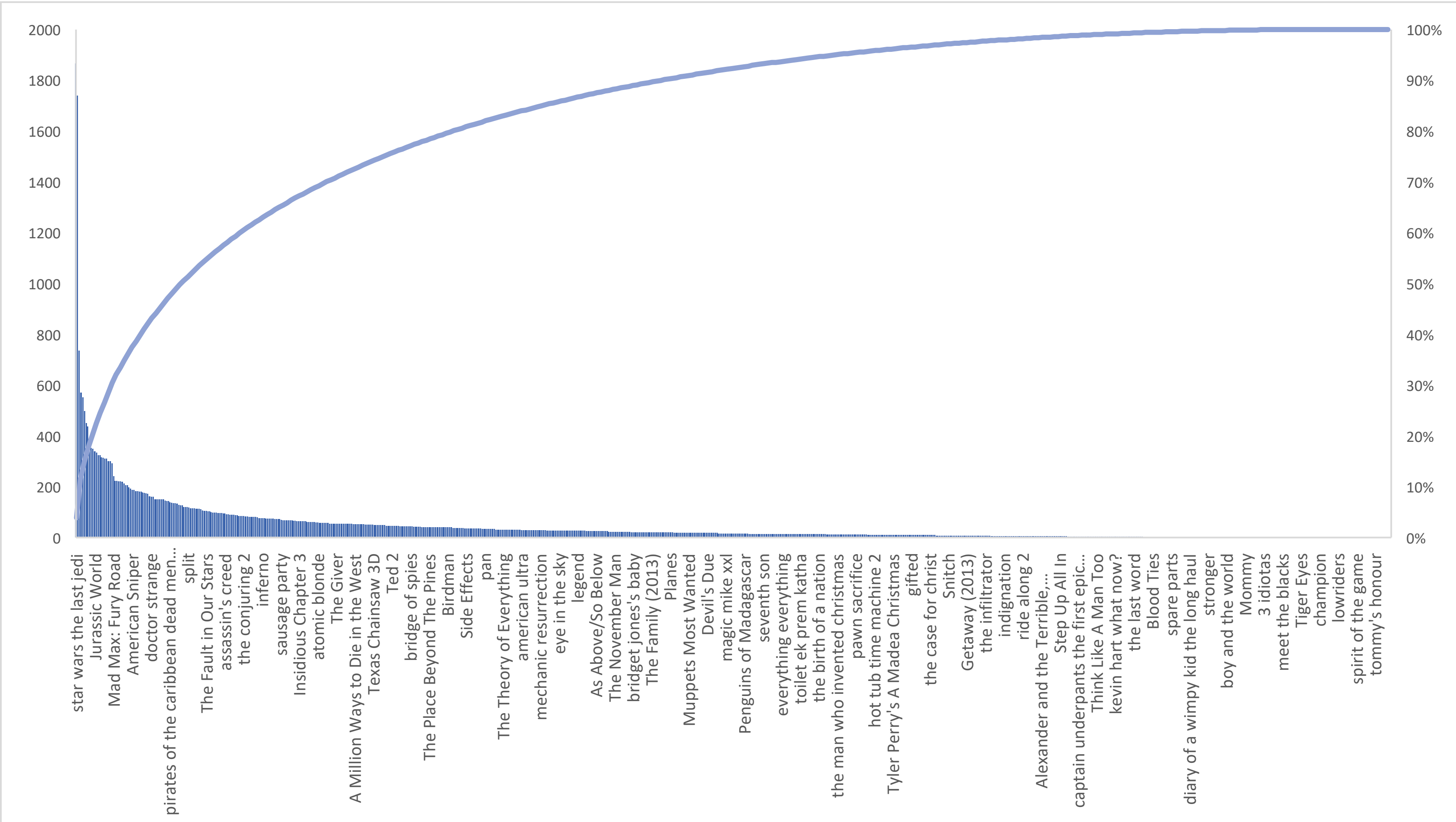
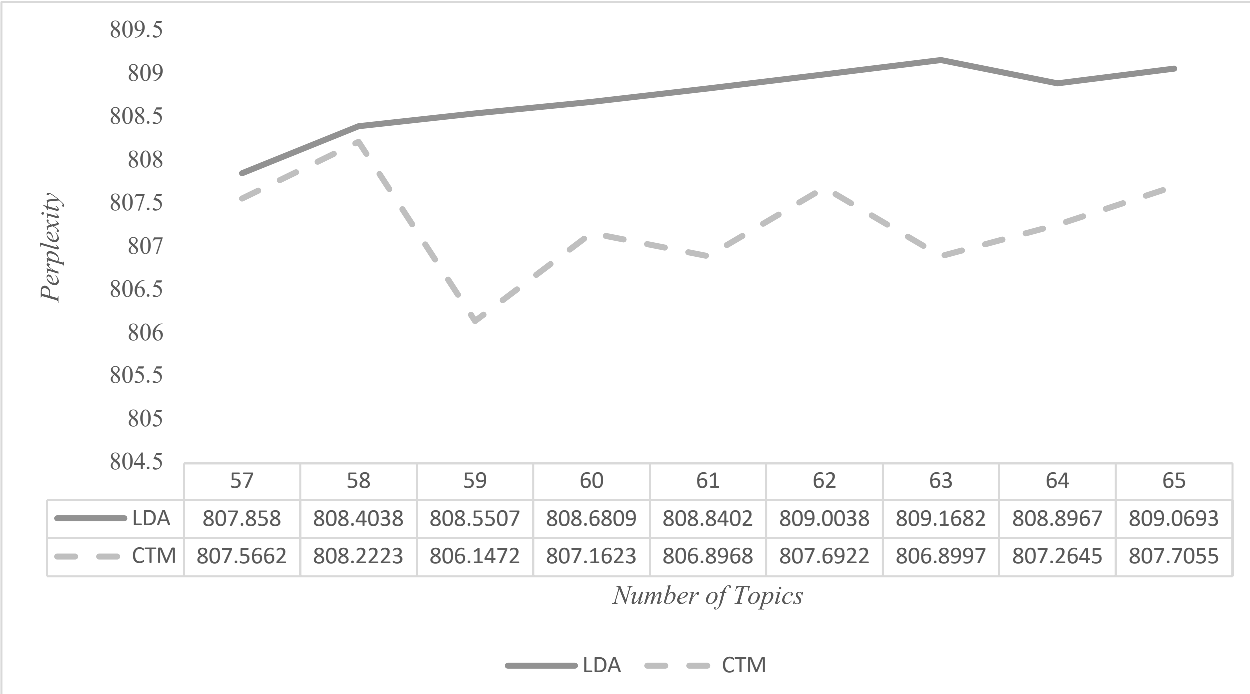


Figure A2. Model Performance of CTM and LDA



	57	58	59	60	61	62	63	64	65
LDA	807.858	808.4038	808.5507	808.6809	808.8402	809.0038	809.1682	808.8967	809.0693
CTM	807.5662	808.2223	806.1472	807.1623	806.8968	807.6922	806.8997	807.2645	807.7055

Notes: Perplexity is a measure used in natural language processing to evaluate the performance of probability models. Lower perplexity scores indicate better generalization performance (Blei, Ng, and Jordan 2003). We fit both CTM and LDA on our corpus of movie reviews across a nine-topic window surrounding $K = 61$, which is the optimal number of topics for our data based on the algorithm proposed in Lee and Mimno (2017). The comparison of perplexity scores shows that CTM outperforms LDA, providing support for the use of CTM in this study.

Table A1. Names and Representative Words of 61 Topics

Topic Name	Representative Words					
Acting Performance	Perform	Actor	Role	Cast	Play	Act	Oscar
America	Black	America	Civil	Govern	Call	Reason	Side
Animation	Kid	Anim	Voic	Children	Disney	Cartoon	Magic
Atmosphere	Feel	Create	Light	Sens	Tone	Misteri	Dark
Book	Book	Read	Adapt	Base	Materi	Sourc	Chang
Brother	Play	Brother	Race	Sport	Role	Work	Cast
Character Arc	Stori	Interest	Focus	Charact	Line	Engag	Tell
Character Development	Charact	Plot	Main	Develop	Lack	Problem	Scene
Cinematography	Visual	Beauti	Work	Score	Set	Product	Cinematographi
Comedic Effect	Funni	Comedi	Laugh	Fun	Joke	Humor	Moment
D.C. Cinematic Universe	Woman	Justic	Ben	Batman	Affleck	Leagu	Flash
Death	Girl	Evil	Dead	Blood	Bodi	Murder	Killer
Disappointment	Worst	Made	Make	Ruin	Give	Bad	Disappoint
Emotion	Love	Feel	Heart	Emot	Wonder	Perfect	Happi
Enjoyable	Great	Amaz	Watch	Enjoy	Job	Awesom	Recommend
Expectation	Movi	Watch	Theater	Made	Make	Expect	Time
Fan	Seri	Fan	Previous	Trek	Instal	Star	Reboot
Fantasy	Battl	Epic	King	Fantasi	Lord	Trilog	Legend
Fight	Superman	Batman	Fight	Destruct	Build	Power	Save
Franchise	Franchis	Time	Back	Termin	Chapter	Travel	Machin
Ghost	Hous	Mother	Ghost	Visit	Haunt	Disturb	Activ
Hero	Man	Iron	Steel	Age	Hero	Dark	Super
Historical	American	Women	Histori	White	Peopl	Polit	Female
Hollywood	Watch	Cinema	Hollywood	Shown	Express	Worth	Director
Horror	Horror	Jump	Scare	Scari	Genr	Creepi	Gore
Human and Robots	Effect	Human	Special	Transform	Robot	Brain	Intellig
Interview	Room	Interview	Split	Show	Produc	Make	Artist
Kill	Kill	Die	Death	Destroy	Fight	Power	End
Length of Movie	Time	Minut	Bad	Hour	Bore	Money	Make
Lesson	Life	Live	Father	Inspir	Child	Messag	Learn
Marvel Cinematic Universe	Marvel	Villain	Comic	Superhero	Univers	Hero	Charact
Monster	Monster	Creatur	Godzilla	Giant	Disast	Island	Cgi
Movie	Film	Make	Found	Made	Work	Enjoy	Feel
Movie Characters	Adam	David	Cast	Big	Line	Work	Direct
Nature	Attack	Wild	Bear	Warm	Tree	Natur	Mile
Office	Team	Citi	Offic	Group	Escap	Plan	Crime
Opinion	Peopl	Review	Rate	Critic	Give	Understand	Opinion
Overall Evaluation	Good	Thing	Lot	Bad	Part	Pretti	Kinda

Past and Future	Men	Futur	Past	Final	Year	Back	Hugh
Power	Power	Remain	Ultim	Attempt	World	Result	Abil
Relationship	Famili	Friend	Home	Wife	Day	Night	Daughter
Road Action	Tom	Max	Mad	Road	Cruis	Insan	Vehicl
Role	Play	Player	Shadow	Screen	Year	Role	Charisma
Romance	Relationship	Chemistri	Romanc	Sex	Babi	Romant	Sexual
Scenes	Scene	End	Pace	Moment	Slow	Intens	Build
School	School	Spiderman	Peter	High	Teenag	Friend	Teen
Science Fiction	Alien	Question	Scifi	Answer	Creation	Technolog	Engin
Secret Mission	Bond	Chase	Secret	Agent	Servic	Spi	Mission
Sequel	Origin	Sequel	Version	Classic	Remake	Predecessor	Fresh
Sequence of Action	Action	Sequenc	Scene	Fast	Fight	Kick	Explos
Soundtrack	Music	Song	Number	Danc	Sing	Band	Soundtrack
Space	Nolan	Space	Fiction	Scienc	Time	Interstellar	Graviti
Space Travel	Earth	Planet	Crew	Space	Ship	Mission	Engin
Star Wars Characters	Luke	Jedi	Ren	Solo	Charact	Order	Vader
Star Wars Franchise	Star	War	Episod	Forc	Fan	Prequel	Trilog
Survival	Surviv	Water	Shot	Reveng	Brutal	Cold	Aliv
Terror	Fear	Chill	Stephen	Terrifi	Psycholog	Terror	Afraid
Video Game	Game	Video	Catch	Runner	Part	Hunger	Fire
Wall Street	Job	Wall	Drug	Street	Busi	Play	Work
War	War	Soldier	Enemi	Hero	Fight	World	Furi
Western	Jack	Jason	Western	Play	Magnific	Town	Rock

B. Complementary Analyses in the Measurement of Spoiler Intensity

B.1 Evidence on the Difference in Topics between Spoiler and Non-Spoiler Reviews

One presumption about IMDb's review data is that the distribution of topics in spoiler reviews is different from that in non-spoiler reviews because spoiler reviews allow sensitive plot-related information while non-spoiler reviews do not. We test this presumption by running a logistic regression, in which the outcome variable is the review type (i.e., 1 = spoiler, 0 = non-spoiler), and predictors are the number of words in a review associated with each topic. We use a 10-fold cross-validation to assess the predictive performance. The area under the ROC curve (AUC) is .70, which is similar to the average AUC reported by previous research (Netzer, Lemaire, and Herzenstein 2019). Thus, the topics from CTM include sufficient information to distinguish between spoiler and non-spoiler reviews.

B.2 Validation of the Proposed Spoiler Intensity Metric

We test whether the proposed spoiler intensity metric is capable of capturing what we theorized by examining whether it adequately agrees with human judgment in determining which review spoils more for a movie. For simplicity, we focus on spoiler intensity at the review level, defined as $\sum_{j=1}^{J} \alpha_{ij} w_{jl}$. We randomly sample 100 movies from our sample that have at least four spoiler reviews (approximately 84% of our movies). For each movie, we order all spoiler reviews based on their spoiler intensity and then randomly sample one review from each quartile. This procedure samples four spoiler reviews per movie. Finally, we use the sampled reviews to randomly form two pairs of spoiler reviews for each movie, resulting in a total of 200 pairs of spoiler reviews.

Using MTurk, we recruited three human coders for each pair of spoiler reviews. We instructed each coder to read the spoiler reviews and indicate "Which of the two movie reviews reveal more of the plot?" We present only the text of reviews, omitting other review information such as username and rating to avoid potential confounders. We use the majority voting rule to determine the review with more spoiling information as evaluated by human judges, and compare the result with that determined by the spoiler intensity metric. We present one coded pair of spoiler reviews from the movie *The Wolverine* as an example in Figure B1. We find an overall agreement rate of 82.5%, which provides evidence for the validity of the proposed spoiler intensity metric.

B.3 Additional Robustness Checks to Spoiler Intensity from Non-Spoiler Reviews

We construct spoiler intensity using spoiler reviews in our main analyses under the assumption that the majority of plot-related information is provided by spoiler reviews rather than non-spoiler reviews. As a robustness check, we relax this assumption to see whether results still hold after controlling for the additional spoiler intensity from non-spoiler reviews.

We begin by considering an alternative intensity metric that aggregates spoiling information from both spoiler and non-spoiler reviews. We denote the aggregate intensity metric by ${INTENSITY_{it}}^{AG}$ and calculate ${INTENSITY_{it}}^{AG}$ using Equation (2.5) summing over all reviews rather than spoiler reviews only. The mean of ${INTENSITY_{it}}^{AG}$ and that of $INTENSITY_{it}$ (i.e., intensity of spoiler reviews) are 3.29 and 2.48 respectively, suggesting that spoiler reviews contribute to the majority of aggregate intensity (2.48/3.29 = 75.4%). Column 1 of Table B1 reports the estimation results with ${INTENSITY_{it}}^{AG}$. Consistent with the uncertainty-reduction mechanism, we find that ${INTENSITY_{it}}^{AG}$ has a positive and statistically significant association with box office revenue. The elasticity of aggregate intensity is .073, which is slightly larger than the elasticity of spoiler intensity (.06), suggesting that our estimate of elasticity of spoiler intensity is conservative.

As a further exploration, we examine whether the effect of intensity from spoiler reviews and that from non-spoiler reviews are different. To do so, we include the log of intensity from non-spoiler reviews, denoted by $\ln(INTENSITY^{NS})$, into the regression, where ${INTENSITY_{it}}^{NS} = {INTENSITY_{it}}^{AG} - INTENSITY_{it}$. As column 2 of Table B1 shows, the elasticities of intensity from spoiler reviews and non-spoiler reviews are .074 and .058, and the difference is not statistically significant (p = .428). These results indicate that the positive association between spoiler intensity from spoiler reviews and box office revenue still holds after controlling for the effect of intensity from non-spoiler reviews. In addition, the positive coefficients of both $INTENSITY$ and $INTENSITY^{NS}$ suggest that the uncertainty-reduction mechanism is in play by the plot-related information from both the spoiler and non-spoiler reviews, despite the fact that spoiler reviews contain the majority of plot-related information.

Figure B1. Coded Pair of Spoiler Reviews for *The Wolverine*

	Spoiler Review 1		Spoiler Review 2
Text	This is the best Wolverine to date. It's fast paced full of action and the story keeps flowing with a few good twists and turns. Hugh Jackman is back to his best with a story following his past present and future. It follows on from an event in Logans past where he is given the chance to have a taste of mortality but obviously at a price. It is mostly set in Japan with stunning sequences and breathtaking scenery. Without making it too complicated the movie flows through the gears and gives a few flashbacks to previous movies. Whereas previous versions have lacked in seriousness this movie maintains a good rhythm throughout. With a good blend of martial arts, samurais, mutants and robots yes that's right robots its an altogether fun action packed film with well directed fight scenes that you'll enjoy immensely. There is a bonus scene for the die hard fans after the credits well worth the wait. Easily watched without having seen any of the previous movies its very enjoyable. Most probably better watched in 2D but I've not had a chance with the 3D yet.		This film was awesome and blows "X-men Origins Wolverine" out of the water, everything about this film was amazing, I loved the fight scenes the one on the Bullet train was brilliant, so was the one at Yashida's funeral. One of the things that made this film so good for me was how for the most part of the film Wolverine/Logan (Hugh Jackman) doesn't have his power of regeneration so he's just a mortal with awesome claws. One thing I wasn't keen on with this film was Jean Grey (Famke Janssen) presence in this film I thought she was stupidly pointless and if anything brought unnecessary complexity to the plot. I loved the end credit scene this just wrapped it up as Logan was leaving with Yukio (Rila Fukushima) on a plane and then he's in an airport in the end credit scene. The end credit scene has left me feeling really excited for "Days of Future past" and I can't wait for the 23rd of May 2014 to see that. 10/10 Thanks for reading :)
Spoiler Intensity	.466	<	.972
Human Votes	0/3	<	3/3

Table B1. Results of Models Controlling for Intensity from Non-Spoiler Reviews

	Homogenous Effect of Intensity from Spoiler and Non-Spoiler Reviews (1)	Separate Effects of Intensity from Spoiler and Non-Spoiler Reviews (2)
$\ln(DAILYREV)_{i,t-1}$	.642***	.642***
	(.017)	(.017)
$\ln(INTENSITY)_{i,t-1}$	-	.074***
		(.015)
$\ln(INTENSITY^{AG})_{i,t-1}$	.073***	-
	(.016)	
$\ln(INTENSITY^{NS})_{i,t-1}$	-	.058***
		(.014)
$PROP_{i,t-1}$	.122*	.128
	(.051)	(.066)
$\ln(CUMRATING)_{i,t-1}$	.163***	.163***
	(.027)	(.026)
$\ln(CUMVOL)_{i,t-1}$	.029**	.034***
	(.010)	(.009)
$\ln(ADVERT)_{i,t-1}$	.093***	.091***
	(.006)	(.006)
$\ln(THEATERS)_{it}$	.336***	.324***
	(.019)	(.019)
$AGE\ (t)$	-.007***	-.007***
	(.001)	(.001)
$HOLIDAY_{it}$	.532***	.532***
	(.018)	(.018)
DAYOFWEEK Dummies	Yes	Yes
Movie Fixed Effects	Yes	Yes
Endogeneity Corrections	Yes	Yes
Cluster-Robust Standard Error	Yes	Yes
Number of Observations	49,057	

* < .05; ** < .01; *** < .001.

C. ESG Issues Monitored by RepRisk

Environment	Social	Governance
Animal mistreatment	Child labor	Anti-competitive practices
Climate change, GHG emissions, and global pollution	Discrimination in employment	Corruption, bribery, extortion, money laundering
Impacts on landscapes, ecosystems, and biodiversity	Forced labor	Executive compensation issues
Local pollution	Freedom of association and collective bargaining	Fraud
Overuse and wasting of resources	Human rights abuses, corporate complicity	Misleading communication
Waste issues	Impacts on communities	Tax evasion
	Local participation issues	Tax optimization
	Occupational health and safety issues	
	Poor employment conditions	
	Social discrimination	

Cross-cutting Issues
Controversial products and services
Products (health and environmental issues)
Supply chain issues
Violation of international standards
Violation of national legislation

Information in this table was obtained from: https://www.reprisk.com/content/static/reprisk-esg-issues-definitions.pdf (accessed February 5, 2021).

Curriculum Vitae

Name:	Jun Hyun (Joseph) Ryoo
Post-secondary Education and Degrees:	Western University London, Ontario, Canada 2012-2016 B.A. 2016-2021 Ph.D.
Honours and Awards:	Province of Ontario Graduate Scholarship 2019-2021

Publications:

Wang, Xin (Shane), Jun Hyun (Joseph) Ryoo, Neil Bendle, and Praveen K. Kopalle, “The Role of Machine Learning Analytics and Metrics in Retailing Research,” *Journal of Retailing,* forthcoming.

Ryoo, Jun Hyun (Joseph), Xin (Shane) Wang, and Shijie Lu (2021), "Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue," *Journal of Marketing*, 85(2), 70-88.

Borah, Abhishek, Xin (Shane) Wang, and Jun Hyun (Joseph) Ryoo (2018), "Understanding Influence of Marketing Thought on Practice: An Analysis of Business Journals Using Textual and Latent Dirichlet Allocation (LDA) Analysis," *Customer Needs and Solutions*, 5 (3-4), 146.

Ryoo, Jun Hyun (Joseph), and Neil Bendle (2017), “Understanding the Social Media Strategies of U.S. Primary Candidates,” *Journal of Political Marketing*, 16 (3-4), 244.