The Effect of Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer

Ramkumar Janakiraman[1]
Associate Professor of Marketing
Business Partnership Foundation Research Fellow
Darla Moore School of Business, University of South Carolina
Columbia, SC (USA) 29208
Phone: (803) 777-0534
Email: Janakiraman.Ramkumar@gmail.com

Joon Ho Lim
Assistant Professor of Marketing
College of Business
Illinois State University
Normal, IL (USA) 61790-5590
Phone: (309) 438-8429
Email: jlim12@ilstu.edu

Rishika Rishika
Clinical Assistant Professor of Marketing
Darla Moore School of Business, University of South Carolina
Columbia, SC (USA) 29208
Phone: (803) 777-8778
Email: rishika@moore.sc.edu

[1]The first author would like to acknowledge an anonymous retailer for sharing the data. The three authors contributed equally. All errors are our own.

The Effect of Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer

Abstract

Data breach incidents have become increasingly common and businesses incur enormous costs to recover from such events. In this study, the authors assess the effects of a data breach announcement by a multichannel retailer on customer behavior. They exploit a *natural* experiment and use *individual* customer transaction data from the retailer to conduct a detailed and systematic empirical examination of the effects of a data breach announcement on customer spending and channel migration behavior. For the identification of the data breach announcement effects, the authors compare the change in customer behavior *before* and *after* the data breach announcement *between* two groups of customers—a treatment group (customers whose information is breached) and a control group (customers whose information is not breached)—using the difference-in-differences modeling framework. They find that the data breach results in a significant decrease in customer spending. However, there is a silver lining for the focal multichannel retailer as they find that customers of the firm migrate from the breached to the unbreached channels of the retailer. The findings further reflect that customers with a higher retailer patronage are more forgiving as the negative effects of the data breach announcement are *lower* for customers with a higher level of patronage. By leveraging data on individual customers' access to email communication from the retailer after the breach, the authors propose and empirically test for the role of customer data vulnerability as the behavioral mechanism that drives customer behavior subsequent to a data breach announcement. They perform a series of robustness checks (with alternative operationalization of variables, model specifications, and study time periods) and falsification tests to validate the findings. Based on the results, the authors offer prescriptions for managers on how to engage with customers following data breach announcements and elaborate on the role of multichannel strategy in absorbing negative demand shocks.

Keywords: customer data breach, data vulnerability, multichannel retailing, customer channel choice, natural experiment, difference-in-differences, propensity score matching

In his first major television interview, the former director of the Federal Bureau of Investigation noted that "*There's only two types of corporations -- big corporations -- in America. There are those who've been hacked.....and those who don't know they've been hacked…*" (*Bloomberg Business* 2014). Cybercrimes related to data breaches have exploded in recent times and more and more businesses—ranging from health and retail (e.g., Anthem and Target) to financial sector (e.g., JPMorgan Chase)—have reported data breach incidents (*New York Times* 2015). One of the nation's three major credit reporting agencies, Equifax, recently reported a data breach that could potentially affect the personal information of 143 million consumers (*Wall Street Journal* 2017). The economic impact of data breach events for business entities could be staggering. As per the Center for Strategic and International Studies estimates, cyber security issues could lead to a loss of $445 billion and 200,000 jobs for businesses annually (Intel Security-McAfee 2014). In 2014, Target Corporation's CEO resigned in the aftermath of the retailer's massive data breach during the Thanksgiving holiday season in 2013 in which the personal information of about 100 million customers was stolen (*Forbes* 2014). It was widely covered in the business reports that the acquisition of Yahoo by Verizon was delayed and subsequently the acquisition price was lowered because of breach liability faced by Yahoo (*Reuters* 2017).

From a customer's perspective, breach of data containing personal and financial information can be perceived as a violation of social contract and a service failure (Malhotra and Malhotra 2011) negatively impacting the customer-firm relationship. Thus, from a retailer's perspective, understanding the impact of data breach announcements on customer behavior is vital in developing strategies that can help mitigate long-term negative consequences if the retailer were to experience such an event. Often companies, both large and small, are under-

prepared for these incidents and they may not accurately estimate the impact of a data breach (*USA Today* 2013). Yet, there is virtually no study that uses *actual* customer transaction data and empirically examines the effect of a data breach announcement on customer behavior. While a few but limited number of studies have examined the impact of security breaches on firm value (e.g., Schatz and Bashroush 2016), there is no study that examines the impact of a data breach announcement on customer behavior. We attempt to fill this gap in the literature. Thus, our *first* objective is to leverage *a natural experiment* and use a unique customer transaction dataset from a multichannel retailer both *before* and *after* a data (customer related) breach announcement by the retailer to examine the effect of the data breach announcement on individual customer *spending level*.

Many retailers operate via multiple channels and rarely does a data breach involve all of their channels. For example, in the data breach reported by Home Depot (*USA Today* 2014b), the breach was limited to the payment systems at the physical stores' self-checkout lanes. In such a case, a retailer's multichannel strategy (à la Venkatesan, Kumar, and Ravishanker 2007) can prove to be advantageous and help mitigate negative consequences of a data breach in the breached channel by facilitating channel migration for customers who may feel vulnerable shopping from the affected channel. It is also possible, however, that perceptions of vulnerability (stemming from potential misuse of data) following a data breach may spillover to other unaffected channels of the firm multiplying the negative outcomes for the firm. We examine this critical multichannel issue with the aim to shed light on the effect of negative demand shocks for a multichannel retailer. Thus, our *second* objective is to examine customers' channel migration (from the breached channel to other unbreached channels) behavior in response to a data breach announcement.

Studies have documented that firm related negative outcomes are likely to be dampened for customers with greater brand commitment and familiarity (e.g., Ahluwalia 2002). Some customers may perceive greater anxiety about the threat of harm from a data breach and may exhibit a greater response to such an event. Therefore, our *third* objective is to examine whether customer characteristics moderate the effect of a data breach announcement on customer behavior. In particular, we study how customers with a high level of retailer patronage change their spending and channel migration behavior in response to a data breach announcement as compared to customers with a low level of retailer patronage.

In a digital era where firms create large databases with different types of customer information not limiting to their transaction history and extending to payment information, data breaches and security protocols to prevent data breaches are increasingly garnering attention. Indeed, a single data breach incident has the potential to immediately heighten a customer's perception of susceptibility to harm (Martin, Borah, and Palmatier 2017) reducing trust and negatively affecting subsequent behavior. Firms often follow a data breach announcement with email communication to customers who are affected by the data breach which can make the data breach more salient to customers. Research suggests that customers inclined towards processing information more deeply will engage in a more thoughtful assessment and evaluation of communication messages thus leading to stronger attitudes (MacInnis, Moorman, and Jaworski 1991). Thus, email communications to the affected customers that follow a public data breach announcement can enhance customers' perceptions of harm or data vulnerability and can further reduce customer trust in the customer-firm relationship. A violation of customer trust is linked to a host of negative outcomes for the firm such as negative word-of-mouth, reduced customer satisfaction and repurchase intentions (e.g., Wang and Huff 2007). Thus, it becomes crucial to

gain a better understanding of the underlying mechanism that drives the undesirable customer outcomes in the aftermath of a data breach. Recent studies (e.g., Goldfarb and Tucker 2014) also argue for the importance of a mechanism check in making causal inference claims. Accordingly, our *final* objective is to establish the role of customer data vulnerability as the underlying mechanism that drives customer behavior subsequent to a data breach announcement.

Examining customer response to a data breach announcement at the individual customer level has both data and econometric challenges. Firms are often reluctant to disclose data breach events and may not want to reveal the damage or the extent of the damage to avoid negative publicity and potential legal liabilities (*CNBC* 2013). It becomes even more difficult to obtain individual customer level transaction data after a data breach event as firms strive to install more security protocols and become hesitant to share data with other entities. In addition, firms also desire to avoid publicizing the negative effects of the data breach as they are taking steps to undo the damage incurred. This makes it nearly impossible to obtain data from a firm to study the effects of a data breach. We have, however, been fortunate to obtain a unique panel customer transaction dataset from a multichannel retailer that reported a data breach in one of its channels. From an econometric perspective, issues of reverse causality and endogeneity can also plague the identification of the data breach effect. To address these issues and to accomplish our objectives, we use transaction data of the *same panel* of customers *before* and *after* a data breach announcement by our focal multichannel retailer.

A key feature of our dataset is that while the focal retailer operates via multiple sales channels, the data breach was limited to only one channel, i.e., the customer payment card information from only one channel was compromised. We exploit this "*natural experiment*" and use the customer transaction data that spans *pre*- and *post*-data breach announcement time

periods to cast our empirical analyses in the *difference-in-differences* modeling framework (Angrist and Pischke 2009) that has been employed in recent studies in marketing (Kumar et al. 2016; Shi et al. 2017). In our context, we track the behavior of two groups of customers, namely the treatment group customers whose information was reported as breached and the control group customers whose data was not breached. Another highlight of our data is that we have information on individual customers' opening of emails that they received from the retailer following the public data breach announcement. We leverage this information to uncover the role of customer data vulnerability in customers' response to data breach announcement.

Our results show that following the announcement of a data breach by the focal multichannel retailer, the affected customers decrease their spending level by 32.45%. We find support for customer data vulnerability as the behavioral mechanism through which a data breach announcement affects customer behavior. We also find significant evidence of customer migration from the *breached channel* to the non-breached channels following the data breach announcement by our focal multichannel retailer. However, the negative effects of the data breach announcement are *lower* for customers with a higher level of retailer patronage as compared to customers with lower patronage. We find that the customers who received and opened emails about the breach respond more negatively to the data breach as compared to customers who did not open the emails. We perform a series of robustness checks with alternative operationalization of variables, model specifications, and study time periods and conduct a battery of falsification tests to validate our findings.

Our study contributes to the literature in the following four ways. *First*, our study contributes to the emerging literature on data breach and data privacy related issues. While studies in marketing (for e.g., Martin, Borah, and Palmatier 2017) and other fields such as

information systems (for e.g., Cavusoglu, Mishra, and Raghunathan 2004) have focused on how *Wall Street* responds to data breach announcements, no study to our knowledge, perhaps due to data challenges that we described earlier, has examined how the *Main Street*, i.e., the customers of a firm respond to data breach announcements. Our study attempts to fill this gap. *Second*, our findings related to customer channel migration prove valuable in furthering our understanding of the data breach effects. We find that it is not prudent to assume that customers would stop using both the breached and the unbreached channels. Our results suggest that a multichannel retailing strategy can help absorb some of the negative fallout after a data breach event. Following recent studies that highlight the benefits of a multichannel strategy (Venkatesan, Kumar, and Ravishanker 2007), we believe the results of our study provide a new justification for using multiple channels as a strategic tool for firms in creating a sustainable long-term advantage. *Third*, we show that customer data vulnerability is at play behind the undesirable customer outcomes after a data breach suggesting that firms must invest in allaying consumer concerns regarding their digital data security features. *Finally*, as many studies in the area of customer relationship management have extolled the benefits of deep customer relationships (Reinartz, Krafft, and Hoyer 2004; Reinartz and Kumar 2003), our study demonstrates that customers with a stronger retailer patronage are more forgiving and exhibit a weaker negative response to the data breach event thus highlighting the role of investing in customer relationship management initiatives.

We structure the rest of the article as follows: First, we provide a brief background on data breaches and a review of the existing set of studies that have examined the consequences of a data breach announcement and delineate the contributions of our study. Next, we develop a set of testable hypotheses on the effects of a data breach announcement on customer behavior. Then

we describe our research setting and data and present our econometric modeling approach followed by the results of our proposed models. We then present supplementary analyses, a series of robustness checks and falsification tests and conclude the paper with a discussion on the implications of our study.

Conceptual Background

We begin this section by providing a brief background on data breach announcements and highlight how our study is different from related studies on brand scandals and data privacy issues. We then present our conceptual background and a set of testable hypotheses related to the effects of data breach announcement (henceforth DBA) on customer behavior.

Research Background

DBAs often involve firms informing their customers that their security systems that protect customers' payment and other personally identifiable information have been breached by people or entities that may have intentions to use this information in an unlawful way. Although the number of data breaches reported by firms in the United States and across the world has been on the rise (*USA Today* 2014a), currently there are no laws that mandate public companies to disclose cyber security in their SEC (the U.S. Securities and Exchange Commission) filings. In 2011, the SEC issued guidance advising companies to report cyber threat and security issues.[1] However, this guidance was simply advice, not regulation.

Much of the cost associated with a data breach is due to the recovery costs in the form of fines, lost revenue, hiring people to fix the problem, paying for credit monitoring services for

[1] http://www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm

customers, public relations, etc. Studies in the area of information systems have investigated the effect of DBAs from the viewpoint of stock market reaction. For example, Cavusoglu, Mishra, and Raghunathan (2004) analyzed the effect of 66 distinct security breaches of public firms and concluded that the breached firms, on average, lost 2.1 percent of their market value within two days of the announcement of the data breach. Based on the average market value of the analyzed firms, the authors find that this amounts to a $1.65 billion loss in market capitalization. More recent studies by Malhotra and Malhotra (2011) and Martin, Borah, and Palmatier (2017) employ event study methodology to understand the effect of privacy breaches on firms' abnormal stock returns and find that data breaches have a significant negative effect on firm performance. Cavusoglu, Mishra, and Raghunathan (2004) argue that the costs faced by firms due to security breaches include the transitory costs due to lost business, decreased productivity, stolen information, litigation costs of containing the damages of data breach and the costs incurred by firms due to the loss of current customers to competitors. Although these studies document that the stock market looks at DBAs unfavorably, the impact of such data breaches on customer behavior using objective individual customer level data has not been examined.

Effect of Data Breach Announcement on Customer Spending and Channel Migration

Brand crises and negative news of any type about a brand can erode a brand's image and trust in the eyes of consumers (Dawar and Pillutla 2000). Based on attribution theory, scholars have classified crisis into three types of crisis responsibility: (1) victim crisis, in which the firm has a weak attribution of responsibility; (2) accidental crisis which is an unintended accident in which the firm has minimal attributions of crisis responsibility; (3) intentional crisis which includes accidents and errors caused by human error and violation of law and other misdeeds by the

organization (Coombs and Holladay 2002; Pick et al. 2016). Given the prevalence of digital security and data breach issues, in the absence of employee caused errors or misdeeds, we reckon that customers of a firm would typically view customer data breach as a victim crisis or as an accidental crisis.[2]

Prior studies in marketing have established that brand crisis in the form of product-harm crisis lowers brand equity and consumers' perceptions of brand quality (Van Heerde, Helsen, and Dekimpe 2007; Zhao, Zhao, and Helsen 2011). Whereas in the case of product harm crisis, product consumption may lead to serious and direct harm to an individual consumer, in the case of a data breach, consumers are more likely to form and suffer from perceptions of harm or data vulnerability that stems from a risk of impersonation, fraud or identity theft (Acquisti, Friedman, and Telang 2006).

Consumers share personal and financial information (for e.g., credit card details, mailing address, etc.) with retailers for a better shopping experience and they are likely to view DBA as a violation of their trust and a breach of the psychological contract that they perceive to have with retailers (Malhotra and Malhotra 2011). From this point of view, data breaches are likely to be construed as a firm service failure. Levitt (1981) argues that a product purchase from a retailer includes not just the tangible product but also the *augmented product* which comprises of all the other benefits a consumer derives from product purchase and consumption. Consumers expect retailers to safeguard their personal information and consider this service as part of a retailer's augmented product. Congressional Research Service report (Cashell et al. 2004) suggests that the

[2] In other words, while customers are more likely to attribute product harm crisis to the actions of managers, firms that suffer from data breach are seen as victims of cybercrimes. In particular, loyal customers or customers who are more familiar with a retailer are less likely to attribute a data breach to the actions of managers and can be more forgiving of the retailer. We explore the differential response to data breach announcement across customers later in this section.

negative publicity associated with cyberattack announcements can lead to an erosion of confidence among customers and firms may experience lost sales in the process. Other studies in law (Fisher 2013) and marketing (Martin, Borah, and Palmatier 2017) also argue that customers will experience anxiety and data vulnerability at the moment of the breach, irrespective of whether data was subsequently misused. All these taken together, we expect that a DBA would lead to decreased customer spending. Thus, we formulate the following hypothesis:

H_1: Following a data breach announcement, individual-customer spending will decrease.

Multichannel retailers operate via multiple channels and sometimes only one of the channels suffers a data breach at a given point of time. For example, in the case of the data breach reported by Home Depot, the breach of data was restricted to customer self-checkout point-of-sale terminals (*USA Today* 2014b). In such cases, retailers typically emphasize in their marketing communication that only one of their channels was breached in order to minimize the fallout to other unaffected channels. When a DBA is made, while some customers may cease their relationship with the retailer, other customers may choose to modify their behavior and look for alternative ways to shop from the retailer until their trust in the retailer is restored. For such customers, shopping from an alternative channel is a viable option especially if the alternative channel is not compromised. In recent years, multi-channel retailing and customer management have gained increased prominence with retailers investing in building a seamless transition experience for its customers across different channels of the firm (Neslin et al. 2006). This has created a fluid multi-channel environment for customers where they can easily transition from shopping from one channel to another. Studies also suggest that retailers invest in a multichannel strategy in order to enhance customers' experience, increase customer satisfaction and build customer loyalty (Wallace, Giese, and Johnson 2004). Thus, in case of a DBA, if a retailer has an

efficient multichannel environment, many customers will choose to simply migrate from the breached channel to the unbreached channel(s).

From a consumer behavior perspective, theories in consumer psychology suggest that consumers tend to discount negative information that is inconsistent with their preferences (Klein and Ahluwalia 2005). This suggests that when encountered with negative information related to only one retail channel, customers may not immediately transfer the negative associations onto the other channel(s) and end their relationship with the retailer. Some customers may indeed choose to transition to other retail channels of the retailer instead of terminating their relationship with the retailer. In our context, these arguments lead us to expect that many customers will migrate from the breached to the non-breached channel(s) following a DBA and continue their patronage with the retailer. Based on these arguments, we propose the following hypothesis:

H_2: Following a data breach announcement, customers will migrate from the breached channel to the unbreached channel(s).

Role of Patronage in Customer Response to a Data Breach Announcement

Several studies have linked retail patronage to favorable attitudes towards the firm (e.g., Eastlick and Liu 1997; Korgaonkar, Lund, and Price 1985). Customers with a high level of retailer patronage have stronger positive attitudes towards the retailer which can introduce skepticism towards negative information regarding the retailer. Customers with a stronger relationship with the retailer tend to be more familiar with the retailer's products, prices and customer service which can increase their switching costs as well. In the light of new negative information which is disconfirmatory with their prior positive experience with the retailer, such patrons are more likely to discount the newly encountered information (Dawar and Pillutla 2000). Studies further

argue that customers with favorable prior attitudes will assign a lower weight to new information that may reflect poorly on the firm (Ahluwalia 2002) and such customers may not consider the preference-inconsistent negative news as relevant as they selectively avoid inconsistent information (Xiong and Bharadwaj 2013).

Based on the above arguments, we posit that the negative effect of a data breach announcement on customer spending will be less pronounced for customers with a higher level of retailer patronage as compared to customers with a lower level of retailer patronage. We thus propose the following hypothesis:

H_3: The impact of a data breach announcement on customer spending will be weaker for customers with a higher level of patronage.

Customers' channel choice decision has been characterized as a moving target and one that can evolve over a customer's lifetime with the firm (Valentini, Montaguti, and Neslin 2011). Among the factors that can induce a customer to migrate to other channels of the firm, an unsatisfying experience has been cited as a critical driving force. It has been long known that attitudes are a precursor to forming behavioral intentions (Ajzen and Fishbein 1980) and they have been shown to affect channel migration behavior (Verhoef, Neslin, and Vroomen 2007). Hence, an exogenous negative shock in the form of data breach announcement that affects one particular channel of a firm, while inducing channel migration across all customers, may affect customers differentially based on their prior attitudes towards the firm. Customers with a strong patronage with the firm harbor a deeper relationship with the firm and have built more positive and favorable associations with the firm over time. Therefore, they are more likely to continue to shop from their preferred channel and would be less likely to migrate to non-preferred channels subsequent to a data breach announcement. Customers with a weaker patronage with the firm, on the other hand,

are more likely to exploit alternative channels of the current firm which may include transitioning to unaffected channels.

Thus, we expect that the effect of a data breach announcement on customer channel migration (from the breached channel to the unbreached channels) will be weaker for customers with a higher level of retail patronage and propose the following hypothesis:

H_4: The impact of a data breach announcement on customer channel migration will be weaker for customers with a higher level of patronage.

Customer Information Processing and Data Breach Announcement: Role of Customer Data Vulnerability

Subsequent to the public announcement of a data breach, firms and retailers typically communicate with individual customers—who are affected by the breach—via email outlining the steps that will be undertaken in response to the incident. Retailers' objectives in sending these communication messages are to notify their customers about the data breach and to lay out the steps customers need to follow to minimize the potential harm that they can experience as a result of the data breach. Such communication messages from the firms in the aftermath of a data breach incident can amplify customers' concerns regarding the breach and misuse of their personal information. In particular, customers who open and read these email communication messages may particularly internalize the DBA and perceive a serious imminent threat to their personal information and may react strongly to the DBA. Thus, customers who are exposed to email communication messages from the firm may feel an amplified sense of data vulnerability after a data breach incident has occurred.

Martin, Borah, and Palmatier (2017, p. 37) describe customer data vulnerability as a "*customer's perception of his or her susceptibility to being harmed as a result of various uses of his or her personal data.*" The authors argue that customers' perception of their susceptibility to being harmed can lead to negative outcomes even if they do not become actual victims. We argue that customers' exposure to email communication messages from a retailer about a data breach would make the data breach more salient triggering an enhanced perception of harm resulting from the misuse of personal information. This would further invoke a customer's sense of data vulnerability that can exacerbate customers' response to a DBA.[3]

A rich stream of literature in the area of consumer information processing helps us understand and predict consumers' attitudinal response to marketing communication (see MacInnis and Jaworski 1989 for a review). Specifically, we build on the arguments of the elaboration likelihood model (henceforth, ELM) to understand how customers who receive and open email marketing communication (from the retailer about the data breach) subsequent to a data breach announcement will respond to the DBA. The basic tenet of ELM is that marketing communication information can change consumers' attitudes towards messages via two different information processing routes, the central and the peripheral routes (Petty and Cacioppo 1986). Whereas the central route requires diligent processing of information on the part of consumers, the peripheral route requires less cognitive effort. Prior research in consumer psychology has established that attitudes that are formed via the central route are highly accessible and are more predictive of subsequent behavior (Petty and Krosnick 1995). It is also argued that individuals with greater motivation and ability are more likely to process marketing communication via the

[3] We do not know if a customer faced real harm because of the data breach. Furthermore, our data is not conducive to disentangling attitudinal vulnerability from actual harm or vulnerability. Thus, in our context, customers' perception of data vulnerability is attitudinal or psychological.

central route of information processing. A recent study by Sahni, Wheeler, and Chintagunta (2018) suggests that when people have both the motivation and the ability, they are more likely to carefully process email marketing communication. The authors conduct a randomized field experiment and find that customers with a greater level of motivation and ability are likely to engage in deeper processing of messages shared via email communications. In our context, following a DBA, customers who received and opened emails from the breached retailer will use the central route for processing information which will lead to a more thoughtful and cognitive process resulting in an increase in customer data vulnerability that will lead to a stronger response to the DBA.

While the public announcement of data breach by the retailer is likely to affect all customers whose data is potentially breached, information processing theory based on the ELM suggests that the impact of the DBA would be greater for customers who have more motivation and ability to process information. Such affected customers would be more motivated to open the emails from the retailer due to their heightened perceptions of harm resulting from the DBA, will process information via the central route (more thoughtfully) which will further increase data breach vulnerability and will therefore, exhibit a stronger response in terms of customer spending and customer channel migration (from the breached to the unaffected channels). Based on the above arguments, we propose the following hypotheses:

H_{5a}: The impact of a data breach announcement on customer spending will be stronger for customers who perceive higher data vulnerability.

H_{5b}: The impact of a data breach announcement on customer channel migration will be stronger for customers who perceive higher data vulnerability.

Methods

Research Setting

The data set for this study comes from a publicly owned department store retailer headquartered in the United States. The product categories that the retailer carries include men's apparel, women's apparel, footwear, accessories for men and women, and kids' apparel. The retailer operated via three channels: the physical store (multiple brick and mortar stores), the Internet and the catalog channels during the data time period. The brick and mortar stores and the Internet channel are the two dominant channels for the multichannel retailer. During the timespan of the dataset, the focal multichannel retailer suffered a data breach and subsequently announced publicly that customer payment data related to its physical store channel during a specific transaction time period was breached. In particular, customer debit and/or credit card information was compromised for customers who purchased via the physical stores (henceforth referred to as the "breached" channel) during a certain time window (the "breached" or the "affected" time period).[4] Customer information related to other channels of the multichannel firm, the Internet and the catalog channels (henceforth, we refer to these as the "unbreached" channels), was not affected or compromised in any way. It is also worth noting that customers who shopped via the breached channel but outside of the breached time period were also not affected. The retailer specifically mentioned and clarified these issues in its public announcement and subsequent email communication with the affected customers.

Following the public announcement of the data breach, the retailer also contacted all of the customers who purchased via the breached channel during the affected time period and for whom mailing and/or email addresses were available. The retailer also posted details of the DBA

[4] Due to confidentiality agreements, we are not able to disclose the name of the retailer, the specific time period during which the breach occurred and other detailed and technical information about the nature of the data breach.

on its main website. We have transaction data of the *same panel* of customers both *before* and *after* the DBA. The data specifically includes detailed customer level information on the items purchased, the dollar amount of the purchases and the channel used for purchasing. Besides the individual customer level transaction data, we also have access to data on individual customers' opening of emails from the retailer following the DBA.

Identification Strategy

Difference-in-differences (DD) approach. In this sub-section, we discuss the econometric challenges and the identification strategy that we undertake in examining the effect of DBA. Examining the effect of DBA using aggregate level sales data across a panel of firms with and without DBAs would be limiting as such data may mask important patterns at the individual customer level. To resolve this, we work with customer level transaction data that spans pre and post DBA by the focal multichannel retailer. However, a simple *before-and-after* comparison, even if one were to get access to customer level data, would not be able to rule out the effect of temporal factors (for e.g., competitors' actions) that are unobserved to us as researchers. Such factors can make the identification of the DBA challenging. To rule out the effect of such factors, we rely on a *natural experiment* in which we examine the effect of an exogenous treatment (in our context, DBA) on customer behavior using the *difference-in-differences* (DD) approach. The DD approach examines the behavior of customers in the treatment group (the affected customers whose data was breached) and the control group (the unaffected customers whose data was not breached) before-and-after the treatment (DBA). DD modeling approach helps account for both time invariant customer characteristics and any time trend effects and establish the causal effect of DBA on customer behavior (Angrist and Pischke 2009; Huang et al. 2012; Shi et al. 2017;

Wooldridge 2002).

The first step in implementing the DD approach involves the construction of the two groups of customers—the treatment and the control group. We define the treatment group as the set of customers whose information was breached, i.e., customers who transacted with the retailer via the "breached" channel at least once *during* the "affected" time period (T_{1b} in Figure W1.1 in the Web Appendix W1). The control group comprises of customers whose data was not breached as they did not transact via the breached channel during the affected time period and hence their payment information was not compromised. It is worth noting that at the time of purchasing via the *breached channel* during the *affected time period*, customers had no extra information that the channel was going to get hacked subsequently. The firm also did not have any information about the timing of the data breach until after the breach was discovered. In other words, to the extent that the timing of the data breach was exogenous and to the extent that customers and the retailer did not know that a particular channel is going to get hacked at a particular time, the classification of customers into the treatment and the control group is exogenous. In the robustness checks section, we test if the treatment and the control group customers have similar trends in behavior over time before the DBA. In the Web Appendix W2, we elaborate on the motivation and the setup of the DD modeling framework.

DD approach with propensity score matching. We argued above that since the timing and the channel of breach is exogenous, the assignment of customers into the treatment and the control groups is likely to be random.[5] However, we recognize that the construction or definition of control group in a natural experiment setting is not always apparent. To rule out any customer

[5] We also find that there is no significant difference between the treatment and the control group customers in their purchase behavior during the pre-DBA time period. See the sub-section on descriptive statistics and model-free evidence of the results section.

self-selection issue and to ensure similarity of the treatment and the control group customers, following recent studies in marketing (e.g., Huang et al. 2012; Kumar et al. 2016; Shi et al. 2017), operations management (Bell, Gallino, and Moreno 2016), and economics (O'Keefe 2004), we apply a quasi-experimental approach and use the DD approach with propensity score matching (PSM).[6] PSM helps mimic a randomized experimental study design (Rubin 2006) by creating matched pairs of treatment and control group customers who are similar on a set of observed characteristics thus addressing the customer self-selection effect (à la Rishika et al. 2013). The DD modeling approach helps account for "common shocks" or time trend effects that affect all customers and any inherent differences (not observed by the researcher) between the two groups. In sum, the combination of PSM and DD helps control for selection due to both observed and (time invariant) unobserved factors that could confound the effect of DBA (Gertler et al. 2011).

To sum, we rely on two identification strategies: (1) the difference-in-differences and (2) the combination of difference-in-differences and propensity score matching to establish the impact of DBA on customer behavior.

Data Sample

For our analysis, we use customer transaction data over a time period that spans seven months before and after the DBA. From our background research on multichannel retailing and after our discussions with managers of the focal retailer, we found that this time period is long enough to pick noticeable changes (if any) in customer behavior. In order to construct our estimation dataset, we applied the following data filtering steps: we began with a sample of randomly

[6] We thank the Area Editor for the suggestion.

selected 15,000 customers. We then removed observations with missing values or errors. In order to keep the treatment and the control group customers comparable, we work with a sample of customers who are multi-channel customers, i.e., those who shopped via more than one channel in the pre-DBA period. For the customer spending model, we did not apply any filters in the post-DBA period as we want to include customers who cut back on their spending and those who left the retailer altogether post DBA.[7] However, for the channel migration model, we did not include those customers who did not purchase in the post-DBA time period as we cannot model their channel choice post-DBA if the customers did not shop at the retailer.[8] The data cleaning and filtering procedure yielded two separate datasets, one for the customer spending model and the other for the channel migration model. The dataset for the spending model consists of 5,004 treatment and 6,455 control customers and the dataset for the channel migration model comprises of 1,572 treatment and 2,883 control customers.

Operationalization of Dependent Variables

Our dependent variables of interest are customer *spending* and customer *channel migration* behavior. A focal customer's spending level (denoted by $Spending_{it}$) is operationalized as the total amount of spending (in United States dollars) by the focal customer i at time period t. We operationalize customer migration from the breached channel to the set of non-breached channels (CM_Trip_{it}) by the proportion of purchase trips undertaken by a focal customer to the unaffected channels (i.e., the Internet and the catalog channels) to the total number of purchase trips (across the three channels of the retailer). In other words, CM_Trip_{it} is defined as follows:

[7] We thank the Area Editor and the anonymous reviewers for the suggestion.
[8] In other words, our channel migration model is conditional on a customer shopping with the retailer. As a result, the composition of the sample for the channel migration model is a subset of the sample for the spending behavior model.

$$\text{CM\_Trip}_{it} = \frac{\text{Trip}_{it}^{\text{unaffected}}}{\text{Trip}_{it}}, \quad (1)$$

where $\text{Trip}_{it}^{\text{unaffected}}$ is the number of purchase trips undertaken by a focal customer i via the unaffected channels at time period t. Trip_{it} denotes the total number of purchase trips taken by the customer i at time period t.[9]

We note that we first aggregate customers' transaction data over seven months in the pre-DBA period and seven months in the post-DBA period and analyze their spending and channel migration behavior with the balanced panel data. The reason for doing so is similar to the reasons expounded in recent studies that use similar modeling techniques (Shi et al. 2017). First, customers do not purchase products from a department store every week or every month. The average number of purchase trips for a seven-month time period in our sample is 3.96 (with a standard deviation of 3.39). Second, by collapsing the customer transaction data into two time periods, we can circumvent the inconsistent estimation of standard errors due to a serial correlation in the DD models over multiple time periods (Bertrand, Duflo, and Mullainathan 2004; Shi et al. 2017). We also note that we collapse the Internet and the catalog channels into one "non-breached" channel for the operationalization of CM_Trip_{it} as customers' use of the catalog channel is extremely low as compared to the other two channels, physical stores and the Internet channel.[10]

[9] We note that the operationalization does not restrict customers' choice of channels and allows for customers' switching back and forth from the breached channel to the unbreached channels. If customers were to prefer or migrate to the unbreached channels over the breached channel, CM_Trip would approach 1. If customers were to prefer or stay with the breached channel (there is no migration from the breached channel to the unbreached channels), CM_Trip would be closer to 0.

[10] In our sample, purchase trips to the breached channel (i.e., physical stores) and the Internet channel account for 76.16% and 20.32% of the total purchase trips respectively. Purchase trips to the catalog channel account for only about 3.51% of the total purchase trips.

Econometric Analyses

Effect of the data breach announcement on customer spending behavior. Following prior studies that have used the DD modeling approach (Danaher et al. 2010; Goldfarb and Tucker 2011b), we model the impact of the DBA on customers' spending behavior as follows:

$$\ln(\text{Spending}_{it}) = \alpha^{sp}\text{DBA}_t + \beta^{sp}\text{HACKED}_i \times \text{DBA}_t + \mu_i + \varepsilon_{it}^{sp}, \quad (2)$$

where HACKED_i is a treatment group indicator that is equal to 1 if the customer i is in the treatment group and 0 if the customer i is in the control group. DBA_t is a time period indicator that is equal to 1 if time period t is the post-DBA period and 0 if time period t is the pre-DBA period. We also include customer fixed effects (μ_i) that help control for unobserved customer heterogeneity and remove endogeneity bias (Rossi 2014). Since the customer specific fixed effects would be perfectly collinear with the treatment group indicator (denoted by HACKED_i), we do not include the main effect of the treatment group indicator (see Goldfarb and Tucker 2011b for details).[11] ε_{it}^{sp} denotes the error term of the model proposed in Equation 2. We note that we transformed the dependent variable, Spending_{it}, by taking a natural logarithm in order to reduce right skewness.[12] Our primary coefficient of interest is the interaction coefficient, β^{sp}, which captures the effect of the retailer's DBA on the change in the spending of the treatment group customers and the spending of unaffected customers (across the pre-DBA and the post-DBA time periods). Based on the model presented in Equation 2, β^{sp} can be interpreted as the causal effect of DBA on customer spending behavior (subject to the identifying assumptions of

[11] There is no intercept in the model as we have fixed effects for all the customers.

[12] Before applying the log-transformation, we rescaled Spending_{it} by dividing it by the average spending amount across all customers in both the pre- and the post-DBA time periods for reasons of confidentiality. We also added a small number (0.001) to the rescaled dependent variable to handle zero values (Collett 2002).

the DD modeling approach). We conduct tests to check for the identifying assumptions in a later section.

Effect of the data breach announcement on customer channel migration behavior. We now turn our attention to customers' channel migration behavior. We specify the DD model of customers' channel migration as follows:

$$\text{(3)} \qquad \text{logit}(CM\_Trip_{it}) = \alpha^{cm} DBA_t + \beta^{cm} HACKED_i \times DBA_t + \omega_i + \varepsilon_{it}^{cm}.$$

The independent variables in Equation 3 are the same as the ones in Equation 2. ω_i refers to the customer fixed effects. As before, the focal coefficient of interest is β^{cm}, which captures the impact of the DBA on customer channel migration behavior. We note that given our operationalization of channel migration, CM_Trip_{it} falls in the interval [0, 1]. Since the normality assumption of Ordinary Least Squares (OLS) regression does not hold, we cannot estimate the model in Equation 3 via the OLS technique. We thus work with a logit transformation of CM_Trip_{it}.[13]

Effect of DBA on high vs. low patronage customers. Using the pre-DBA time period transaction data, we segment customers into two groups of high and low patronage customers by using a median split of their spending level (Tucker, Zhang, and Zhu 2012). We use the first three months of the pre-DBA period to calibrate the pre-DBA customer patronage level; thus our customer segmentation analysis spans four months before and after DBA.[14] This helps ensure that customer classification does not confound with the estimation time period and also aids in easy interpretation. To investigate how the effect of DBA on customer spending and channel

[13] We add a small number (0.001) to both the numerator and the denominator of the variable (Collett 2002).
[14] As a result, the number of observations for the DD analyses and the customer segmentation (DDD based) analyses are different.

migration varies across high versus low levels of customer patronage, following recent studies (Danaher et al. 2010; Goldfarb and Tucker 2011b; Shi et al. 2017), we extend our DD models to *difference-in-difference-in-differences* (DDD) modeling framework. The DDD models for customer spending and customer channel migration are as follows:

$$\ln(\text{Spending}_{it}) = \alpha^{sp}\text{DBA}_t + \beta^{sp}\text{HACKED}_i \times \text{DBA}_t + \gamma^{sp}\text{DBA}_t \times \text{HIGH}_i + \delta^{sp}\text{HACKED}_i \times \text{DBA}_t \times \text{HIGH}_i + \mu_i + \varepsilon_{it}^{sp}, \quad (4)$$

$$\text{logit}(\text{CM\_Trip}_{it}) = \alpha^{cm}\text{DBA}_t + \beta^{cm}\text{HACKED}_i \times \text{DBA}_t + \gamma^{cm}\text{DBA}_t \times \text{HIGH}_i + \delta^{cm}\text{HACKED}_i \times \text{DBA}_t \times \text{HIGH}_i + \omega_i + \varepsilon_{it}^{cm}, \quad (5)$$

where HIGH_i takes the value 1 if a customer i is a high patronage customer and 0 if the customer i is a low patronage customer. In Equations 4 and 5, all the other variables are the same as the ones used in Equations 2 and 3, respectively. In Equations 4 and 5, the main coefficients of interest are δ^{sp} and δ^{cm}. The three-way interaction model—presented in Equations 4 and 5—is commonly referred to as the *difference-in-difference-in-differences* (DDD) model or the triple-difference model as it helps examine the variation in the outcome variable specific to the high patronage customers (relative to low patronage customers) in the treatment group (as compared to the control group) in the time period following DBA (relative to time period prior to DBA). For example, δ^{sp} captures the effect of the DBA on the spending behavior of *high patronage* customers (relative to low patronage customers) in the *treatment group* (relative to the control group) in the *post-DBA* period (relative to the pre-DBA period).

Effect of customer email opening following DBA. We had built on the arguments of ELM and suggested that the affected customers who receive email marketing communication following a data breach would perceive a greater sense of data vulnerability. Thus, while all the affected customers whose data was announced as breached may feel apprehensive, we argue that customers who receive and open emails will exhibit a stronger response to the DBA. To

empirically test this proposition, we leverage access to unique individual customer level marketing communication data. More specifically, we have information on the customers who received and opened the retailer's email within one week of the public announcement regarding the data breach. Customers who opened the retailer's email (with information on the data breach) immediately following the announcement of the breach may become more acutely aware of the data breach and feel more threatened that the data breach would affect them personally. We use a customer's email communication exposure within a week of the public DBA as a measure for checking the proposed underlying mechanism of customer data vulnerability.

To empirically examine how customer response to DBA varies with customer email opening behavior, we leveraged our customer level email communication data and classified the *treatment* customers into two groups—treatment customers who received and opened the retailer's email (within one week) after the DBA and treatment customers who received the email but did *not* open it. Following the technique in recent DD modeling literature (Levine and Toffel 2010), we extended the DD models of spending and channel migration presented in Equations 2 and 3 respectively to examine the differential response of the two types of treatment group customers:

$$(6)\quad \ln(\text{Spending}_{it}) = \alpha_1^{sp}\text{DBA}_t + \alpha_2^{sp}\text{HACKED}_i^{\text{EmailOpen}} \times \text{DBA}_t + \alpha_3^{sp}\text{HACKED}_i^{\text{EmailNotOpen}} \times \text{DBA}_t + \mu_i + \varepsilon_{it}^{sp},$$

$$(7)\quad \text{logit}(\text{CM\_Trip}_{it}) = \alpha_1^{cm}\text{DBA}_t + \alpha_2^{cm}\text{HACKED}_i^{\text{EmailOpen}} \times \text{DBA}_t + \alpha_3^{cm}\text{HACKED}_i^{\text{EmailNotOpen}} \times \text{DBA}_t + \omega_i + \varepsilon_{it}^{cm},$$

where $\text{HACKED}_i^{\text{EmailOpen}}$ ($\text{HACKED}_i^{\text{EmailNotOpen}}$) is an indicator variable that equals 1 if a treatment customer i received and opened the retailer's email within a week of the DBA

(received the email but did not open it) and 0 otherwise.[15] In Equations 6 and 7, all the other variables are the same as the ones used in Equations 2 and 3, respectively. Comparison of α_2^{sp} to α_3^{sp} would help compare the effect of the DBA on the spending level of the treatment group customers who received and opened emails as compared to the treatment group customers who did not open the emails. Likewise, comparison of α_2^{cm} to α_3^{cm} would help quantify the differential response in terms of channel migration behavior.

Propensity score matching in combination with difference-in-differences. We note that we follow the procedure in Caliendo and Kopeinig (2008) and recent studies in marketing (e.g., Huang et al. 2012; Kumar et al. 2016; Shi et al. 2017) to implement propensity score matching (PSM henceforth). PSM involves the calculation of the propensity score, the probability that a customer purchased via the breached channel at least once during the affected time period, obtained from a logit model of a set of customer specific factors.[16] We used the following customer-specific observables as the matching variables: the total number of purchase trips to the breached channel prior to the breach period (TRIP_HACKED_CHANNEL), whether a customer used multiple channels or only a single channel to purchase prior to the breach period (MULTICHANNEL_CUSTOMER), age (AGE), gender (FEMALE), marital status (SINGLE) and income level (INCOME). Table 1 presents the descriptions and the summary statistics of these matching variables and Table W3.1 (see the Web Appendix W3) provides the estimation results of the logistic regression models.[17]

[Insert Table 1 about here]

[15] We note that the retailer sent emails to all the affected customers. Thus, the retailer's decision to send emails is exogenous to individual customers' perceptions of vulnerability. We thank the Area Editor for pointing this out.
[16] We note that since the sample for the spending and the channel migration models are different, we run two logit models for matching, one for spending analysis and the other for channel migration analysis.
[17] In the interest of space, we do not discuss the results of the logistic regression models.

We matched the treatment to the control customers with the closest propensity score (which is equal to the estimated probability) using the *optimal full matching* algorithm. We employ this algorithm as it allows for a more general type of matching of one treatment unit to one or more comparison units and vice-versa and thus does not require discarding of any unmatched observations (Hansen and Klopfer 2006; Rosenbaum 2002). Furthermore, the optimal full matching is more robust against violations of the common support region assumption as compared to the conventional greedy matching algorithms (Guo and Fraser 2015).

After matching, we conducted a detailed examination to assess the quality of matching by checking if the matching variables are well balanced between the treatment and the control groups where balance refers to the similarity of their covariate distributions. In the Web Appendix W3, we present the standardized differences between the treatment and the control groups on the matching variables before and after matching. As can be seen in Table W3.2 (Panel A), whereas most of the standardized difference measures are statistically significant prior to matching, the measures are *not* significant *after* matching, which implies that PSM helps achieve covariate balance between the treatment and the control group customers. We obtain significant bias reduction after conducting PSM thus attesting to the appropriateness of our matching technique. In addition, we follow Hansen and Bowers (2008) and conduct an *omnibus test* for balance on all of the matching variables simultaneously (as opposed to comparing the treatment and the control groups on each matching variable separately). In Table W3.2 (Panel A), we present the results of the omnibus balance test. Large *p*-value (.8698) of the combined baseline difference statistic (d^2) after matching suggests that the null hypothesis of well-balanced matched sets cannot be rejected. The balance between the treatment and the control groups on the matching variables also holds for the PSM conducted for the channel migration analysis (see

Panel B of Table W3.2 in the Web Appendix W3). All of these results taken together suggest that we are able to achieve statistical balance between the treatment and the control customers for both the spending and the channel migration analyses. In Figure W3.1 (see the Web Appendix W3), we present the graphical representations of covariate balance before and after matching shown in Table W3.2.

The final step of our effort in reducing potential selection biases and capturing the causal impact of DBA is to combine propensity score matching with the difference-in-differences model. We follow the *propensity score weighting* procedure that has been expounded in statistics and program evaluation literature (e.g., Hirano and Imbens 2001; Hirano, Imbens, and Ridder 2003; McCaffrey, Ridgeway, and Morral 2004; Rosenbaum 1987). Similar procedure has been implemented in recent management science studies as well (e.g., Bell, Gallino, and Moreno 2016). We obtained the propensity scores from the logit model estimation results and used them as sampling weights in the DD model estimation. Weighted regression (with the weights being the customers' propensity scores) involves assigning propensity scores as weights to the treatment and the control customers and helps make the two customer groups as similar as possible on their observed characteristics (matching variables). In Figure W3.2 (see the Web Appendix W3), we summarize the steps of PSM in combination with the DD modeling approach. To facilitate a better understanding of the structure of our analyses, we present the overview of our study timeline and how we construct key variables based on specific time windows in Figure W1.1 (see the Web Appendix W1).

Results

Descriptive Statistics and Model-Free Evidence

Table 1 summarizes the main dependent variables and the matching variables that we use in the customer spending and migration models. In Table 2, we present model-free evidence of the effect of DBA on the two groups—the treatment and the control group—of customers across the two time periods—the pre-DBA and the post-DBA periods. The key takeaway is that the treatment customers, i.e., those customers whose data was reported as breached, reduced their spending level more after the DBA as compared to the control customers. More specifically, the average spending amounts for the treatment customers are 1.0739 and .7398 in the pre- and post-DBA periods respectively. The average spending level decreased by .3341, which is statistically significant ($t = 32.41$, $p < .01$).

The average spending amounts of the control group customers are 1.0758 and 1.0687 in the pre- and post-DBA periods respectively, and the difference between the two periods is *not* statistically significant ($t = .40$). This suggests that the DBA did *not* have a significant effect on the control group customers. All these results taken together suggest that the "*difference-in-differences*" for spending is negative and significant ($-.3270$, $p < .01$) thus providing prima facie evidence of the negative effect of DBA on customers' spending behavior. We find similar results for customers' channel migration behavior. Specifically, we find that the treatment customers show a significant change in their channel preference and prefer the non-breached channels in the post-DBA period as compared to the control customers. The "*difference-in-differences*" value for channel migration is positive and significant ($.0581$, $p < .01$) suggesting a positive effect of DBA on customers' channel migration behavior. We also find that there is no significant difference between the treatment and the control group customers in their spending and channel

migration behavior in the pre-DBA time period.[18] This suggests that the two groups are similar (in terms of the outcome variables) prior to the DBA which in turn implies that customer self-selection is not a major concern in our context. Figure 1 visually presents the model-free evidence of the DBA effects.

We also present the customer spending level for the treatment and control group customers that spans 100 days before and after the DBA (see Figure W4.1 in the Web Appendix W4). The plot suggests that there is no clear difference in the spending level of the control group customers (whose data was not breached) before and after DBA which supports the earlier finding that the DBA did not have a noticeable effect on the control group customers. However, we can see a large drop in the spending level of the treatment group customers right after the DBA. We also note that there is *no* significant difference in the spending patterns between treatment and control customers during the *pre-DBA* period. In other words, the visual analytics based on time series plot are in conformance with the *raw* difference-in-differences results we discussed earlier (in Table 2). Taken together, the model-free evidence suggests that customers cut back on their spending and increasingly migrate to the non-breached channels of the focal retailer in response to the DBA. We now present the results of the effect of DBA on customer behavior based on a series of DD models.

[Insert Table 2 about here]

[Insert Figure 1 about here]

[18] The difference between the mean of Spending for the treatment group prior to DBA and that of the control group prior to DBA (−.0019) is *not* statistically significant. In addition, the difference between the mean of CM_Trip for the treatment group prior to DBA and that of the control group prior to DBA (−.0067) is also *not* statistically significant.

Main Effects of DBA on Spending and Channel Migration

In Table 3, we present the estimation results of the DD model of customer spending. We note that the standard errors reported in the table are clustered at the customer level and are heteroskedasticity robust.[19] We find that β^{sp} (from Equation 2) is negative and significant which suggests that the DBA has a negative and significant impact on customer spending behavior. More specifically, on average, we find that the DBA leads to about 32.45% decrease in customer spending (over a period of seven months). We thus find support for H_1. With respect to channel migration, we find that β^{cm} (from Equation 3) is positive and significant which suggests that customers prefer to purchase via the non-breached channels after the DBA (see Table 4). More specifically, on average, we find that the ratio of the number of purchase trips to the non-breached channels to the number of purchase trips to the breached channel for the treatment group customers is 167.54%[20] greater than the corresponding ratio for the control group customers. This suggests evidence of channel migration from the breached channel to the non-breached channels; we thus find support for H_2.

[Insert Tables 3 and 4 about here]

Effect of DBA on High vs. Low Patronage Customers

Hypotheses H_3 and H_4 propose that the effect of the DBA will differ across high and low patronage customers. In Table 5, we present the parameter estimates of the DDD models of customer spending and channel migration behavior respectively.[21] The DDD models show that δ^{sp} (from Equation 4) is positive and significant (.3474, $p < .05$) suggesting that the negative

[19] The reported standard errors in all of the DD and DDD models are clustered at the customer level.
[20] $100 \times [\{\exp(\beta^{cm})\} - 1]$.
[21] We computed the variance inflation factor (VIF) for each of the independent variable and found that none of the VIFs are larger than 10. Based on VIF diagnostics (Hair et al. 2010; Kutner, Nachtsheim, and Neter 2004), we note that multicollinearity is not a concern in our context. We thank an anonymous reviewer for the suggestion regarding this multicollinearity check.

impact of the DBA on customer spending is lower for customers with high patronage as compared to customers with low patronage. With respect to channel migration, we find that δ^{cm} (from Equation 5) is negative and significant (−1.8220, $p < .01$) suggesting that migration from the breached channel to the unbreached channels is less pronounced for customers with high patronage as compared to customers with low patronage. We thus find support for both H_3 and H_4. In sum, while the DD model results suggest that the DBA leads to customer spending reduction and channel migration (from the breached channel) to the unaffected channels in the aftermath of the crisis, the DDD results suggest that the DBA has a weaker impact on customers with a high level of patronage as compared to those with a low level of patronage.

[Insert Table 5 about here]

Effect of Emails and Customer Data Vulnerability

In Table 6, we present the results of the models related to the role of opening of email on customer response. We find that both α_2^{sp} and α_3^{sp} (from Equation 6) are negative and significant. This suggests that DBA has a negative impact on the buying behavior of the two groups of treatment customers, the group who received and opened the retailer's email within a week of the DBA and the group who received the email but did not open it. However, the Wald test suggests that the negative effect of DBA on spending level is greater (more negative) for the treatment customers who opened emails from the retailer as compared to customers who did not open emails from the retailer. We find similar results for channel migration behavior. We had argued that customer data vulnerability is the mechanism that drives customers' response to the DBA. These results support our arguments and hence support H_{5a} and H_{5b}. We thus find support for all of our proposed hypotheses.

[Insert Table 6 about here]

We note that ours is the first study to empirically test for and document the role of customer data vulnerability using actual individual customer level transaction data and breached retailer's email communication information. Goldfarb and Tucker (2014, p. 32) argue that "*mechanism check is important because it helps support claims of causal inference and because it enhances the likelihood that a paper is remembered.*" They also suggest that "*if the effect is larger when theory suggests it should be, then this helps identify the mechanism*" (Goldfarb and Tucker 2014, p. 31). In our context, our results that the DBA effect is larger for customers with greater customer data vulnerability highlight the potential role of customer data vulnerability as the underlying mechanism behind customer behavior following DBA by a retailer.

Results of the DD Model with Propensity Score Matching

To account for possible customer self-selection driven confounding factors, we re-estimated our proposed DD model using the matched sets of treatment and control customers by using propensity scores as weights (Hirano, Imbens, and Ridder 2003; Khandker, Koolwal, and Samad 2010). Propensity score is the probability that a customer would be in the treatment group (in our context, the group of customers whose data was breached) given a set of covariates (also known as the matching variables). We added matched-set fixed effects to the DD models in order to account for unobserved heterogeneity at the *matched-set* level. We present these results of the revised DD models—based on matched samples of treatment and control customers—in Table 7. The key takeaway from the table is that the results of the DD models based on matched samples are in conformance with the results of the proposed DD models presented earlier. This suggests that the core set of results related to the effect of DBA on customer behavior are robust to potential customer self-selection issues.

[Insert Table 7 about here]

Robustness Checks

In this section, we discuss various robustness checks and a series of falsification tests we conducted to examine potential concerns about our empirical strategy and any spurious correlations.

Alternative Variable Operationalization and Model Specifications

We perform various checks to ascertain that our core results related to the DD models are robust to alternative operationalization of dependent variables and model specifications. We first checked if core results of the effect of DBA on customer spending behavior would hold if we analyze customer behavior in terms of the number of purchase trips (denoted by $Trip_{it}$) and the number of products purchased (denoted by $Quantity_{it}$). We examine these two outcomes using a Poisson model specification (Anderson et al. 2010). We present the results of these two alternative dependent variables in Columns 1 and 2 of Table 8 respectively. We find that the pattern of results for the two alternative outcome variables is consistent with the results of the effect of DBA on customer spending. Specifically, we find that the DBA leads to a decrease in the number of purchase trips undertaken and the number of products purchased by 20.28%[22] and 22.31% respectively (over a period of seven months).

We further check the robustness of our results by using a DD model of customer purchase incidence. We use and estimate a logistic regression specification of purchase incidence to model the probability of purchase from the retailer in a given month (Goldfarb and Tucker

[22] $100 \times [\{\exp(-.2267)\} - 1]$.

2011a). In this model of purchase incidence, Buy_{it} is the alternative dependent variable which is equal to 1 if a customer i makes a purchase in a year-month t and 0 otherwise. The negative and significant DD estimate (see Column 3 of Table 8) of the proposed model supports our main finding that the DBA negatively affects customers' purchase behavior.

For the customer channel migration analysis, we had worked with the ratio of a focal customer's number of purchase trips to the unaffected channels to the total number of purchase trips (see Equation 1). As a robustness check, we operationalized channel migration in two alternative ways: (1) the ratio of spending at the unaffected channels to the total spending (denoted by $CM\_Spending_{it}$) and (2) the ratio of number of items purchased via the unaffected channels to the total number of items bought by a focal customer (denoted by $CM\_Quantity_{it}$). We present the results of these robustness checks in Columns 4 and 5 of Table 8 respectively. We find that the results of models with these alternative channel migration variables are consistent with the main results.

Finally, we develop a simple channel choice model that models the probability that a customer would purchase via one of the unbreached channels. More specifically, we use a logistic regression specification with a dependent variable, $Channel_{it}$, that is equal to 1 if a customer i shops via the unbreached channels at time (purchase trip date) t and 0 otherwise. As can be seen from the results in Column 6 of Table 8, we find that the customers are more likely to choose the unbreached channels over the breached channel subsequent to the DBA. To sum, all the results from the alternative operationalization of dependent variables and model specifications are consistent with the results of the main DD models.

[Insert Table 8 about here]

Our analyses related to customer patronage were based on the median split of customers' spending level during the calibration period (T_{1a} in Figure W1.1 in the Web Appendix W1). To check the robustness of the results, we operationalized customer patronage based on the median split of the number of purchase trips (denoted by $Trip_{it}$) and the number of products purchased (denoted by $Quantity_{it}$). Table 9 shows that the DDD estimates based on these alternative measures of customer patronage are consistent with the results of the main DDD models for both spending level and channel migration analyses.

[Insert Table 9 about here]

With respect to the role of the email communication on customer response, we defined the treatment group as the customers who *not only* shopped in the breached channel during the breach period *but also* received and opened the retailer's email within a week of the DBA. We then estimated the models of customer spending and customer channel migration presented in Equations 2 and 3 on this redefined treatment group customers and the control group customers. We find that the DBA effects based on the subsample with the redefined treatment customers are stronger than those from the full sample with the original treatment customers.[23] This highlights the role of customer data vulnerability as the underlying mechanism behind customer behavior following DBA by a retailer.

Short-Term and Long-Term Effects of Data Breach Announcement

In our main DD analyses, we used customer transaction data that spans seven months pre and post the focal retailer's DBA. To check the robustness of our DD results to alternative time periods and examine how the effects of DBA change over time, we estimate a series of DD

[23] The results are available from the authors upon request.

models (Equations 2 and 3) with one month pre and post, two months pre and post, three months pre and post, four months pre and post, five months pre and post and six months pre and post DBA.[24] This lets us work with balanced data and helps resolve any inconsistent estimation of standard errors (Bertrand, Duflo, and Mullainathan 2004; Shi et al. 2017). We present the results of these models of customer spending and customer channel migration in Tables 10 and 11 respectively. We find that while the direction of the DD estimates of the six models of both customer spending and channel migration is consistent with that of the original DD estimates, the effects of the data breach announcement attenuate over time. The good news is that the negative effect of DBA on customer spending seems to wane over time indicating that breached firms need to take immediate action to minimize negative publicity at the early stage of the data breach crisis.

[Insert Tables 10 and 11 about here]

Falsification Test 1: Placebo Effect

The identifying assumption behind the DD modeling approach is that the treatment group and the control group customers have similar trends in behavior over time before the intervention (in our context, data breach announcement). To check the validity of this identifying assumption, we conduct a "placebo" test that has been used in the economics literature (Carrieri, D'Amato, and Zotti 2015; Kim, Urpelainen, and Cooper 2015; Puri, Rocholl, and Steffen 2011) by using only the pre-DBA period data and treating the data from the first half of the pre-DBA period as the new pre-DBA period data and the second half of it as the *fake* post-DBA period data. That is,

[24] We set the same time window for pre- and post-DBA periods in order to make the shopping behavior of interest observed in the two time periods comparable. All of the data filtering criteria we applied for the sample of the main analysis holds for these additional analyses. Our analyses of short-term and long-term effects are based on the balanced data construction.

there is no real treatment between the new pre-DBA period and the fake post-DBA period. Given this setting, we estimate the following DD model:

$$(8) \qquad \ln(\text{Spending}_{it}) = \alpha^{sp}\text{DBA}_t^{\text{Placebo}} + \beta^{sp}\text{HACKED}_i \times \text{DBA}_t^{\text{Placebo}} + \mu_i + \varepsilon_{it}^{sp},$$

$$(9) \qquad \text{logit}(\text{CM\_Trip}_{it}) = \alpha^{cm}\text{DBA}_t^{\text{Placebo}} + \beta^{cm}\text{HACKED}_i \times \text{DBA}_t^{\text{Placebo}} + \omega_i + \varepsilon_{it}^{cm},$$

where HACKED_i is a treatment group indicator that is equal to 1 if the customer i is in the treatment group and 0 otherwise. $\text{DBA}_t^{\text{Placebo}}$ is the placebo indicator variable that distinguishes between the first half of the original pre-DBA period (the new pre-DBA period) and the second half of the original pre-DBA period (the fake post-DBA period). The DD estimates (β^{sp} and β^{cm}) will not be statistically significant if the pre-treatment parallel trend assumption holds. Table 12 indicates that the DD estimates for spending and channel migration models are *not* statistically significant suggesting that the treatment and control customers follow the same trend in the pre-DBA period and thus the assumption of parallel trend holds in our study.

[Insert Table 12 about here]

Falsification Test 2: Fake Treatment Group

In our analysis, the treatment group consists of customers who used the breached channel during the affected time period and thus suffered a possible breach of their personal information. The control group customers did not use the breached channel during the affected time period and thus are not affected by the data breach. The DD modeling framework builds on this assumption and compares the behavior of the two groups of customers—*the treatment* and *the control group*—across the two time periods—*before* and *after the data breach announcement*. To assess the validity of the construction of our treatment group, we conduct a falsification test by

randomly treating half of our control customers as "fake treatment customers" and the other half as control customers. If the DD estimate based on this fake treatment group is different from zero, the construction of the treatment group would be questionable. To confirm the stability of the estimation results, we repeat the random sampling 5,000 times and report the bootstrap DD estimates in Table 13. We find that the bootstrap DD estimates are *not* statistically significant implying that our original construction of the treatment group is valid (see Table 13).

[Insert Table 13 about here]

To summarize, our main results regarding the effects of the DBA on customer spending and channel migration behavior survive a battery of additional analyses with alternative operationalization of variables, alternative model specifications, alternative treatment group, different study time periods, falsification tests, and different estimation strategies such as combining the propensity score matching and the difference-in-differences approach. Therefore, we conclude that we find compelling evidence of the causal effect of the data breach announcement on customer behavior.

Discussion

Firms spend vast resources to build reputation and create brand equity and yet a single data security incident can inflict serious damage to the firm's reputation, lead to significant customer churn and increase customer acquisition costs. According to a study by IBM and the Ponemon Institute, a public policy think tank dedicated to privacy and information security policy, the average total cost of a data breach registered an increase of 23% from 2013-2015 to $3.79 million (Ponemon Institute 2015). In the light of such huge financial impact of data breach events, our study becomes the first one to document the impact of a data breach announcement

on customer behavior and thereby makes significant contributions to the understanding of the phenomenon from both theoretical and managerial perspectives.

Theoretical Implications

A recent white paper suggests that one quarter of all security breaches occur in a retail environment (Retail Perceptions 2014). The report uses survey data and suggests that customers are likely to spend and visit the retail stores less following a security violation. While several such business reports suggest the possibility of reduced customer spending following a security incident, there is virtually no study that uses actual customer transaction data to study the effect. Thus, from a theoretical perspective, our study helps make inroads into understanding the impact of data breach announcements at an individual customer level. Broadly, this study contributes to the growing literature on brand crises (such as product recalls) that can cause revenue and share losses and harm brand equity (e.g., Cleeren, Van Heerde, and Dekimpe 2013; Dawar and Pillutla 2000).

While product harm crises negatively affect product quality and customer purchase behavior, we find that the mechanism through which a data breach operates is different. Data breach announcements typically contain information about the number of individuals affected through the data breach and specify the time period during which a particular breach took place. Such announcements are followed by email communications to the affected customers in which an individual customer is informed that their information is stolen and that it can be potentially misused in an unauthorized manner in the future. We argue that a data breach announcement would immediately heighten affected customers' perceptions of data vulnerability resulting in negative customer outcomes. We argue further that the email communications initiated by the

firm to the affected customers would make the incident even more salient in customers' minds who would process the information more deeply (via the central processing route à la ELM) thus strengthening the negative effects of the breach. Thus, we propose and test for customer data vulnerability as the underlying behavioral mechanism that would help explain the effect of data breach announcements on customer behavior. We find that customers' perception of the severity of the harm enhances their data vulnerability and negatively affects their subsequent behavior. This is a novel finding that makes new contributions to the emerging literature on data breaches and data vulnerabilities (e.g., Martin, Borah, and Palmatier 2017; Ransbotham et al. 2016).

Increased customer profitability of multichannel shoppers has generated a fair bit of interest in marketing literature and is often cited as a key reason for firms to engage in multichannel strategy (e.g., Montaguti, Neslin, and Valentini 2016; Thomas and Sullivan 2005; Venkatesan, Kumar, and Ravishanker 2007). Our study adds a new dimension to the current multichannel marketing literature by showing that investing in a multichannel strategy is beneficial for firms as it can help absorb the negative impact of data breach announcements. Using customers' actual transaction data and exploiting a unique natural experiment where an exogenous shock (i.e., data breach) occurs at only one channel of the firm, we are able to compare *pre* and *post* buying and channel usage behavior of treatment group versus control group customers. Our findings suggest that while customer spending substantially decreased, significant customer migration to the unaffected channels also occurred at the same time. Thus, from a strategic perspective, shocks that affect customer trust can be mitigated to an extent by diverting resources to the unaffected channels while the firm is recovering from the crises. While existing literature has advocated several benefits of multichannel strategy such as increasing customer engagement and loyalty leading to increased customer profitability (Venkatesan,

Kumar, and Ravishanker 2007), our findings are the first to demonstrate that multiple channels can have a broader strategic purpose in an era of cybercrimes where data breaches are threatening to disrupt the pace of business at a much larger scale.

Our findings also contribute to the literature on relationship marketing and customer relationship management. Customer relationship management literature suggests that the relationship value to the firm is not generally homogeneous across customers and thus, it is important to invest in the right relationships (Reinartz, Krafft, and Hoyer 2004). Customer loyalty has occupied a key place in this literature stream and studies suggest that cultivating attitudinal loyalty is important in leading to the creation of more profitable customer-firm relationships (Reinartz and Kumar 2003). Our results build on this stream of literature and help furnish a new justification for investing in the right customers. We find that firms definitely have a lot to gain from nurturing relationships with more loyal customers as they tend to support a firm through a crisis. The willingness of loyal customers to withstand negative shocks has rarely been examined before and therefore, this study takes important strides in adding to the customer relationship management literature.

Managerial Implications

Data breach events are on the rise and most firms today face an unprecedented security risk that managers must actively manage. Based on the results from our study, we offer the following prescriptions for managers.

Engage actively in damage control and address customer data vulnerability. When online clothing and shoe retailer Zappos suffered a data breach incident that affected 24 million customers in 2012, it took assertive and remedial steps immediately following the discovery of

the breach. For example, Zappos reset the customers' passwords promptly and encouraged them to alter their usernames and passwords used in Zappos for any other websites (*CNNMoney* 2012). Such active damage control strategy can be crucial in preventing loss of customer trust. Our results show that the announcement of a data breach leads to a 32.45% reduction in customer spending, 20.28% decrease in the number of purchase trips and a 22.31% decrease in the number of products purchased by customers (over a period of seven months). These findings suggest that how a company responds to a data breach will determine how well it survives in the wake of a data breach announcement. While prevention is a key element in cyberattacks, new threats keep evolving in the dynamic marketplace and marketing managers must be prepared to engage aggressively in damage control once a threat to the company's cybersecurity is realized. A data breach response plan is quintessential in surviving and successfully managing a data breach incident.

It is also crucial not only to work on mitigation of the negative fallout from a data breach incident but also to invest in consumer trust building initiatives. In particular, managers must address customers' perception of data vulnerability as it influences how they respond to data breaches. To the extent possible, retailers must communicate the steps that they plan to take to assuage consumers' perceptions of vulnerability. It is also worth noting that we find that the negative effect of DBA decreases over time. However, this would imply that the breached firms need to take immediate and concrete actions at the early stage of the crisis. In the months following a DBA, since the traffic comes mainly from high patronage customers, we suggest that retailers initiate measures that serve to maintain customer loyalty while the retailer is dealing with the crises.

Invest in multiple channels. An important result of our study, from a managerial perspective, is that customers migrate from the breached channel to the unbreached channels subsequent to the data breach announcement. We find that the number of purchase trips by the affected customers to the non-breached channels of the firm increases substantially subsequent to a data breach announcement as compared to the trips by the unaffected customers. This finding provides a significant justification for pursuing a multichannel strategy. This would also suggest that a multichannel retailer should be prepared for increased traffic to the unbreached channels following the DBA. Since several data breaches involve only one channel, operating via multiple channels can help absorb external shocks that affect consumer attitudes and behavior in only one channel. However, different channels have different operational and logistical challenges and therefore, firms must invest in multichannel development and management strategies simultaneously such that the unaffected channels can seamlessly integrate excess demand from the affected channel.

Conclusion and Limitations

Although our study is the first to examine the effect of a data breach announcement using actual customer behavioral data, it is not without its limitations. Our results are based on data from only one multichannel retailer that experienced a data breach and followed the breach with a public announcement. While we leverage the data breach affecting only one of the channels of the multichannel retailer as a natural experiment and hence control for other firms' actions, we are not able to specifically examine competitor actions. Although we leveraged the natural experimental based research design and supplemented our core results with analyses based on matching techniques, we caution any causal interpretation based on the results that we report is

subject to the identifying assumptions. We believe there is a lot of scope for future research in this area and future studies could examine how different retailers react to a data breach announcement and how their customers respond to negative publicity associated with different forms of data breaches. Future studies could also examine the role of the severity of data breaches on customer behavior, which we are unable to do due to data limitations. We leveraged data on emails that individual affected customers received following the data breach to shed light on customer data vulnerability. However, future research can examine the role of marketing communication efforts in getting customers back to the stores in the days following a data breach announcement. Despite these limitations, we hope that our study helps convey the direct costs to firms in the form of lost business due to a data breach announcement and spurs more studies in the area of business implications of cyberattacks and data breaches.

REFERENCES

Acquisti, Alessandro, Allan Friedman, and Rahul Telang (2006), "Is There a Cost to Privacy Breaches? An Event Study," Proceedings of the Twenty Seventh International Conference on Information System, Milwaukee, WI, (accessed March 6, 2017), [available at http://www.heinz.cmu.edu/~acquisti/papers/acquisti-friedman-telang-privacy-breaches.pdf].

Ahluwalia, Rohini (2002), "How Prevalent is the Negativity Effect in Consumer Environments?" *Journal of Consumer Research*, 29 (2), 270−79.

Ajzen, Icek and Martin Fishbein (1980), *Understanding Attitudes and Predicting Social Behaviour*. Englewood Cliffs, NJ: Prentice Hall.

Anderson, Eric T., Nathan M. Fong, Duncan I. Simester, and Catherine E. Tucker (2010), "How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet," *Journal of Marketing Research*, 47 (2), 229−39.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bell, David R., Santiago Gallino, and Antonio Moreno (2016), "Offline Showrooms in Omni-Channel Retail: Demand and Operational Benefits," working paper, (December 24), (accessed April 25, 2017), [available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2370535].

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (1), 249–75.

Bloomberg Business (2014), "Chinese Hackers Like a 'Drunk Burglar,' 'Kicking Down the Door,' Says FBI Director," (October 6), (accessed February 28, 2015), [available at http://www.bloomberg.com/bw/articles/2014-10-06/fbi-chief-james-comey-lambasts-chinese-hackers].

Caliendo, Marco and Sabine Kopeinig (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 22 (1), 31−72.

Carrieri, Vincenzo, Marcello D'Amato, Roberto Zotti (2015), "On the Causal Effects of Selective Admission Policies on Students' Performances: Evidence from a Quasi-Experiment in a Large Italian University," *Oxford Economic Papers*, 67 (4), 1034−56.

Cashell, Brian, William D. Jackson, Mark Jickling, and Baird Webel (2004), "The Economic Impact of Cyber-Attacks," Congressional Research Service report for congress, (April 1), (accessed January 16, 2016), [available at http://archive.nyu.edu/bitstream/2451/14999/2/Infosec_ISR_Congress.pdf].

Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan (2004), "The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers," *International Journal of Electronic Commerce*, 9 (1), 69–104.

Cleeren, Kathleen, Harald J. Van Heerde, and Marnik G. Dekimpe (2013), "Rising from the Ashes: How Brands and Categories Can Overcome Product-Harm Crises," *Journal of Marketing*, 77 (2), 58–77.

CNBC (2013), "Cyberattacks: Why Companies Keep Quiet," (February 25), (accessed January 16, 2016), [available at http://www.cnbc.com/id/100491610].

CNNMoney (2012), "Zappos Hacked, 24 Million Accounts Accessed," (January 16), (accessed October 31, 2017), [available at http://money.cnn.com/2012/01/16/technology/zappos_hack/index.htm].

Collett, David (2002), *Modelling Binary Data*. 2nd ed. Boca Raton, FL: Chapman & Hall.

Coombs, Timothy W. and Sherry J. Holladay (2002), "Helping Crisis Managers Protect Reputational Assets: Initial Tests of the Situational Crisis Communication Theory," *Management Communication Quarterly*, 16 (2), 165–86.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang (2010), "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science*, 29 (6), 1138–51.

Dawar, Niraj and Madan M. Pillutla (2000), "Impact of Product-Harm Crises on Brand Equity: The Moderating Role of Consumer Expectations," *Journal of Marketing Research*, 37 (2), 215–26.

Eastlick, Mary Ann and Mengmeng Liu (1997), "The Influence of Store Attitudes and Other Nonstore Shopping Patterns on Patronage of Television Shopping Programs," *Journal of Interactive Marketing*, 11 (3), 14–24.

Fisher, John A. (2013), "Secure My Data or Pay the Price: Consumer Remedy for the Negligent Enablement of Data Breach," *William & Mary Business Law Review*, 4 (1), 215–39.

Forbes (2014), "Target's CEO Steps Down Following the Massive Data Breach and Canadian Debacle," (May 8), (accessed April 3, 2016), [available at http://www.forbes.com/sites/greatspeculations/2014/05/08/targets-ceo-steps-down-following-the-massive-data-breach-and-canadian-debacle/#28cb80763f56].

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch (2011), *Impact Evaluation in Practice*, 1st ed. Washington, DC: World Bank.

Goldfarb, Avi and Catherine E. Tucker (2011a), "Advertising Bans and the Substitutability of Online and Offline Advertising," *Journal of Marketing Research*, 48 (2), 207–27.

Goldfarb, Avi and Catherine E. Tucker (2011b), "Privacy Regulation and Online Advertising," *Management Science*, 57 (1), 57–71.

Goldfarb, Avi and Catherine E. Tucker (2014), "Conducting Research with Quasi-Experiments: A Guide for Marketers," working paper, (March 28), (accessed February 21, 2017), [available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2420920].

Guo, Shenyang Y. and Mark W. Fraser (2015), *Propensity Score Analysis: Statistical Methods and Applications*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Hair, Joseph F. Jr., William C. Black, Barry J. Babin, and Rolph E. Anderson (2010), *Multivariate data analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hansen, Ben B. and Jake Bowers (2008), "Covariate Balance in Simple, Stratified and Clustered Comparative Studies," *Statistical Science*, 23 (2), 219−36.

Hansen, Ben B. and Stephanie Olsen Klopfer (2006), "Optimal Full Matching and Related Designs via Network Flows," *Journal of Computational and Graphical Statistics*, 15 (3), 609–27.

Hirano, Keisuke and Guido W. Imbens (2001), "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2 (3-4), 259–78.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder (2003), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71 (4), 1161–89.

Huang, Qingyi, Vincent R. Nijs, Karsten Hansen, and Eric T. Anderson (2012), "Wal-Mart's Impact on Supplier Profits," *Journal of Marketing Research*, 49 (2), 131−43.

Intel Security-McAfee (2014), "Net Losses: Estimating the Global Cost of Cybercrime," (June 1), (accessed February 1, 2015), [available at http://globalinitiative.net/wp-content/uploads/2017/01/csis-estimating-the-global-cost-of-cybercrime-june-2014.pdf].

Khandker, Shahidur R., Gayatri B. Koolwal, and Hussain A. Samad (2010), *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, DC: World Bank.

Kim, Sung Eun, Johannes Urpelainen, and Jasper Jack Cooper (2015), "The American Resource Curse: Quasi-Experimental Evidence for the Impact of the Shale Gas Revolution on Politics," working paper, (December 14), (accessed August 8, 2016), [available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703603].

Klein, Jill G. and Rohini Ahluwalia (2005), "Negativity in the Evaluation of Political Candidates," *Journal of Marketing*, 69 (1), 131–42.

Korgaonkar, Pradeep K., Daulat Lund, and Barbara Price (1985), "A Structural Equations Approach Toward Examination of Store Attitude and Store Patronage Behavior," *Journal of Retailing*, 61 (2), 39–60.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior," *Journal of Marketing*, 80 (1), 7–25.

Kutner, Michael H., Christopher J. Nachtsheim, and John Neter (2004), *Applied Linear Regression Models*, 4th ed. New York, NY: McGraw-Hill Irwin.

Levine, David I. and Michael W. Toffel (2010), "Quality Management and Job Quality: How the ISO 9001 Standard for Quality Management Systems Affects Employees and Employers," *Management Science*, 56 (6), 978−96.

Levitt, Theodore M. (1981), "Marketing Intangible Products and Product Intangibles," *Harvard Business Review*, 59 (3), 94–102.

MacInnis, Deborah J. and Bernard J. Jaworski (1989), "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of Marketing*, 53 (4), 1–23.

MacInnis, Deborah J., Christine Moorman, and Bernard J. Jaworski (1991), "Enhancing and Measuring Consumers' Motivation, Opportunity, and Ability to Process Brand Information from Ads," *Journal of Marketing*, 55 (4), 32–53.

Malhotra, Arvind and Claudia K. Malhotra (2011), "Evaluating Customer Information Breaches as Service Failures: An Event Study Approach," *Journal of Service Research*, 14 (1), 44–59.

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier (2017), "Data Privacy: Effects on Customer and Firm Performance," *Journal of Marketing*, 81 (1), 36–58.

McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral (2004), "Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies," *Psychological Methods*, 9 (4), 403–25.

Montaguti, Elisa, Scott A. Neslin, and Sara Valentini (2016), "Can Marketing Campaigns Induce Multichannel Buying and More Profitable Customers? A Field Experiment," *Marketing Science*, 35 (2), 201–17.

Neslin, Scott A., Dhruv Grewal, Robert Leghorn, Venkatesh Shankar, Marije L. Teerling, Jacquelyn S. Thomas, and Peter C. Verhoef (2006), "Challenges and Opportunities in Multichannel Customer Management," *Journal of Service Research*, 9 (2), 95–112.

New York Times (2015), "Millions of Anthem Customers Targeted in Cyberattack," (February 5), (accessed March 1, 2015), [available at http://www.nytimes.com/2015/02/05/business/hackers-breached-data-of-millions-insurer-says.html?_r=2].

O'Keefe, Suzanne (2004), "Job Creation in California's Enterprise Zones: A Comparison Using a Propensity Score Matching Model," *Journal of Urban Economics*, 55 (1), 131–50.

Petty, Richard E. and John T. Cacioppo (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York, NY: Springer-Verlag.

Petty, Richard E. and Jon A. Krosnick (Eds.) (1995), *Attitude Strength: Antecedents and Consequences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pick, Doreén, Jacquelyn S. Thomas, Sebastian Tillmanns, and Manfred Krafft (2016), "Customer Win-Back: The Role of Attributions and Perceptions in Customers' Willingness to Return," *Journal of the Academy of Marketing Science*, 44 (2), 218–40.

Ponemon Institute (2015), "2015 Cost of Data Breach Study: Global Analysis," (accessed March 28, 2016), [available at https://nhlearningsolutions.com/Portals/0/Documents/2015-Cost-of-Data-Breach-Study.PDF].

Puri, Manju, Jörg Rocholl, and Sascha Steffen (2011), "Global Retail Lending in the Aftermath of the US Financial Crisis: Distinguishing between Supply and Demand Effects," *Journal of Financial Economics*, 100 (3), 556–78.

Ransbotham, Sam, Robert G. Fichman, Ram Gopal, and Alok Gupta (2016), "Special Section Introduction—Ubiquitous IT and Digital Vulnerabilities," *Information Systems Research*, 27 (4), 834–47.

Reinartz, Werner J., Manfred Krafft, and Wayne D. Hoyer (2004), "The Customer Relationship Management Process: Its Measurement and Impact on Performance," *Journal of Marketing Research*, 41 (3), 293–305.

Reinartz, Werner and V. Kumar (2003), "The impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (1), 77–99.

Retail Perceptions (2014), "Retail's Reality: Shopping Behavior After Security Breaches," (accessed February 14, 2015), [available at http://www.interactionsmarketing.com/retailperceptions/2014/06/retails-reality-shopping-behavior-after-security-breaches/].

Reuters (2017), "Verizon, Yahoo Agree to Lowered $4.48 Billion Deal Following Cyber Attacks," (February 21), (accessed February 27, 2017), [available at http://www.reuters.com/article/us-yahoo-m-a-verizon-idUSKBN1601EK].

Rishika, Rishika, Ashish Kumar, Ramkumar Janakiraman, and Ram Bezawada (2013), "The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation," *Information Systems Research*, 24 (1), 108−27.

Rosenbaum, Paul R. (1987), "Model-Based Direct Adjustment," *Journal of the American Statistical Association*, 82 (398), 387–94.

Rosenbaum, Paul R. (2002), "Covariance Adjustment in Randomized Experiments and Observational Studies," *Statistical Science*, 17 (3), 286−327.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33 (5), 655–72.

Rubin, Donald B. (2006), *Matched Sampling for Causal Effects*. New York, NY: Cambridge University Press.

Sahni, Navdeep S., S. Christian Wheeler, and Pradeep K. Chintagunta (2018), "Personalization in Email Marketing: The Role of Non-Informative Advertising Content," *Marketing Science*, forthcoming, (accessed September 23, 2017), [available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2725251].

Schatz, Daniel and Rabih Bashroush (2016), "The Impact of Repeated Data Breach Events on Organisations' Market Value," *Information & Computer Security*, 24 (1), 73–92.

Shi, Huanhuan, Shrihari Sridhar, Rajdeep Grewal, and Gary Lilien (2017), "Sales Representative Departures and Customer Reassignment Strategies in Business-to-Business Markets," *Journal of Marketing*, 81 (2), 25–44.

Thomas, Jacquelyn S. and Ursula Y. Sullivan (2005), "Managing Marketing Communications with Multichannel Customers," *Journal of Marketing*, 69 (4), 239–51.

Tucker, Catherine, Juanjuan Zhang, and Ting Zhu (2012), "Days on Market and Home Sales," working paper, (March 21), (accessed January 16, 2016), [available at http://ssrn.com/abstract=1481321].

USA Today (2013), "Target Confirms Massive Credit-Card Data Breach," (December 18), (accessed August 22, 2016), [available at https://www.usatoday.com/story/news/nation/2013/12/18/secret-service-target-data-breach/4119337/].

USA Today (2014a), "43% of Companies Had a Data Breach in the Past Year," (September 24), (accessed August 27, 2016), [available at https://www.usatoday.com/story/tech/2014/09/24/data-breach-companies-60/16106197/].

USA Today (2014b), "Home Depot Hackers Used Vendor Log-On to Steal Data, E-mails," (November 7), (accessed January 16, 2016), [available at http://www.usatoday.com/story/money/business/2014/11/06/home-depot-hackers-stolen-data/18613167/].

Valentini, Sara, Elisa Montaguti, and Scott A. Neslin (2011), "Decision Process Evolution in Customer Channel Choice," *Journal of Marketing*, 75 (6), 72–86.

Van Heerde, Harald, Kristiaan Helsen, and Marnik G. Dekimpe (2007), "The Impact of a Product-Harm Crisis on Marketing Effectiveness," *Marketing Science*, 26 (2), 230–45.

Venkatesan, Rajkumar, V. Kumar, and Nalini Ravishanker (2007), "Multichannel Shopping: Causes and Consequences," *Journal of Marketing*, 71 (2), 114–32.

Verhoef, Peter C., Scott A. Neslin, and Björn Vroomen (2007), "Multichannel Customer Management: Understanding the Research-Shopper Phenomenon," *International Journal of Research in Marketing*, 24 (2), 129–48.

Wall Street Journal (2017), "Equifax Breach Could Cost Billions," (September 15), (accessed October 22, 2017), [available at https://www.wsj.com/articles/equifax-breach-could-cost-billions-1505474692].

Wallace, David W., Joan L. Giese, and Jean L. Johnson (2004), "Customer Retailer Loyalty in the Context of Multiple Channel Strategies," *Journal of Retailing*, 80 (4), 249–63.

Wang, Sijun and Lenard C. Huff (2007), "Explaining Buyers' Responses to Sellers' Violation of Trust," *European Journal of Marketing*, 41 (9/10), 1033–52.

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: Massachusetts Institute of Technology Press.

Xiong, Guiyang and Sundar Bharadwaj (2013), "Asymmetric Roles of Advertising and Marketing Capability in Financial Returns to News: Turning Bad into Good and Good into Great," *Journal of Marketing Research*, 50 (6), 706–24.

Zhao, Yi, Ying Zhao, and Kristiaan Helsen (2011), "Consumer Learning in a Turbulent Market Environment: Modeling Consumer Choice Dynamics After a Product-Harm Crisis," *Journal of Marketing Research*, 48 (2), 255–67.

TABLE 1
Summary Statistics

Variable	Description	Spending Model				Channel Migration Model			
		Treatment Group (N = 5,004)		Control Group (N = 6,455)		Treatment Group (N = 1,572)		Control Group (N = 2,883)	
		M	SD	M	SD	M	SD	M	SD
Spending	Spending amount (in USD) over a period of seven months	.91	1.05	1.07	1.40	–	–	–	–
CM_Trip	Proportion of the number of purchase trips undertaken by a customer in the unaffected channels to the total number of purchase trips over a period of seven months	–	–	–	–	.42	.18	.40	.23
TRIP_HACKED_CHANNEL	Number of purchase trips to the breached channel prior to the breach time period	2.29	2.06	2.92	2.71	2.95	2.09	3.25	3.00
MULTICHANNEL_CUSTOMER	Whether a customer used multiple channels prior to the breach time period	.76	.43	.93	.26	.87	.33	.91	.28
AGE	Customer age	51.79	13.79	51.01	13.48	53.05	13.57	51.62	13.32
FEMALE	Whether a customer is female	.84	.37	.82	.38	.89	.31	.85	.35
SINGLE	Whether a customer is single	.29	.45	.29	.45	.28	.45	.28	.45
INCOME	Customer income level that is coded as an ordinal variable with 36 categories: 1 indicates under $15,000, 2 indicates $15,000-$19,999, 3 indicates $20,000-$24,999 and so on.	26.31	9.34	26.39	9.31	26.59	9.23	26.76	9.20

Notes: We rescale Spending by dividing it by the average of spending amount across all customers in both pre- and post-DBA periods for reasons of confidentiality.

TABLE 2
Raw Difference-in-Differences

	Treatment Customers			Control Customers			Difference between Treatment and Control Customers in Pre-DBA Period	Difference-in-Differences
	(1)	(2)	(2) − (1)	(3)	(4)	(4) − (3)	(1) − (3)	{(2) − (1)} − {(4) − (3)}
	Pre-DBA	Post-DBA	Difference	Pre-DBA	Post-DBA	Difference		
Spending	1.0739 (1.0535)	.7398 (1.0291)	**−.3341*****	1.0758 (1.3134)	1.0687 (1.4868)	−.0071	−.0019	**−.3270*****
CM_Trip	.3951 (.1529)	.4509 (.1913)	**.0558*****	.4019 (.1348)	.3995 (.3030)	−.0024	−.0067	**.0581*****

Notes: The table compares the means of our focal dependent variables, Spending and CM_Trip, between treatment and control customers during pre- and post-data breach announcement periods. We calculate a "difference" that indicates the change in outcome variable pre- and post-data breach announcement for each group and a "difference-in-differences" measure by subtracting "difference" for control customers from "difference" for treatment customers. In addition, we perform t-tests to confirm whether they are statistically significant. Standard deviations are in parentheses.
*$p < .10$; **$p < .05$; ***$p < .01$.

TABLE 3
Impact of Data Breach Announcement on Spending

	DV: ln(Spending)
HACKED × DBA	**−.3245***** **(.0674)**
DBA	−1.3292*** (.0459)
Customer fixed effects	Yes
# of observations	22,918
# of customers	11,459
R^2	.6993

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold.
*$p < .10$; **$p < .05$; ***$p < .01$.

TABLE 4
Impact of Data Breach Announcement on Channel Migration

	DV: logit(CM_Trip)
HACKED × DBA	**.9841*** (.1030)**
DBA	−.7561*** (.0977)
Customer fixed effects	Yes
# of observations	8,910
# of customers	4,455
R^2	.5019

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold.
*p < .10; **p < .05; ***p < .01.

TABLE 5
Effect of Data Breach Announcement on High vs. Low Patronage Customers

	(1)	(2)
	DV: ln(Spending)	DV: logit(CM_Trip)
HACKED × DBA × HIGH	**.3474** (.1381)**	**−1.8220*** (.7053)**
DBA	−1.2942*** (.0839)	−.6096 (.5044)
HACKED × DBA	−1.1138*** (.1043)	1.6405*** (.5783)
DBA × HIGH	.5748*** (.1057)	1.6854*** (.5914)
Customer fixed effects	Yes	Yes
# of observations	14,700	5,052
# of customers	7,350	2,526
R^2	.6589	.4281

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DDD estimate) that is statistically significant is highlighted in bold.
*p < .10; **p < .05; ***p < .01.

TABLE 6
Customer Data Vulnerability and Response to Data Breach Announcement: Effect of Email Communication

	(1)	(2)
	DV: ln(Spending)	DV: logit(CM_Trip)
$HACKED^{EmailOpen}$ × DBA	**−1.4147*** (.5417)**	**1.2668*** (.2022)**
$HACKED^{EmailNotOpen}$ × DBA	**−.2333** (.0916)**	**.6052*** (.1119)**
DBA	−1.3132*** (.0455)	−.7561*** (.0977)
Customer fixed effects	Yes	Yes
# of observations	16,212	6,460
# of customers	8,106	3,230
R^2	.7074	.4916
Wald test (H_0: $\alpha_2 = \alpha_3$)	28.5642***	4.1249**

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold. Wald tests report F statistics.
*p < .10; **p < .05; ***p < .01.

TABLE 7
Impact of Data Breach Announcement with Propensity Score Matching

	(1)	(2)
	DV: ln(Spending)	DV: logit(CM_Trip)
HACKED × DBA	**−.3007** (.0684)**	**.9741*** (.0740)**
DBA	−1.3597*** (.0471)	−.7561** (.0702)
Customer fixed effects	Yes	Yes
Matched-set fixed effects	Yes	Yes
# of observations	22,572	8,656
# of customers	11,286	4,328
R^2	.6961	.5085

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold. We note that the number of customers in the PSM based model is different from that of the main model because of missing matching variables for some of the customers.
*p < .10; **p < .05; ***p < .01.

TABLE 8
Robustness to Alternative Dependent Variables and Model Specifications

	Original DV: ln(Spending)			Original DV: logit(CM_Trip)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Alternative DV: Trip	Alternative DV: Quantity	Alternative DV: Buy	Alternative DV: logit(CM_Spending)	Alternative DV: logit(CM_Quantity)	Alternative DV: Channel
	Model: Poisson	Model: Poisson	Model: Logistic	Model: OLS	Model: OLS	Model: Logistic
HACKED × DBA	**−.2267*** (.0163)**	**−.2525*** (.0218)**	**−.3676*** (.0256)**	**.7656*** (.1501)**	**.6806*** (.1425)**	**.1621*** (.0494)**
DBA	−.2019*** (.0112)	−.1640*** (.0157)	−.4514*** (.0187)	−.4821*** (.1195)	−.3601*** (.1147)	−.4563*** (.0334)
Customer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	No	No	Yes	No	No	Yes
Day of week fixed effects	No	No	No	No	No	Yes
# of observations	22,918	22,918	183,344	8,910	8,910	135,134
# of customers	11,459	11,459	11,459	4,455	4,455	4,455
Log-likelihood	−19,428.535	−43,767.021	−75,754.300	–	–	−61,794.035
R^2	–	–	–	.5179	.5172	–

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold.
*p < .10; **p < .05; ***p < .01.

TABLE 9
Effect of Data Breach Announcement on High vs. Low Patronage Customers: Robustness to Alternative Measures of Customer Patronage Level

	DV: ln(Spending)		DV: logit(CM_Trip)	
	Alternative Measure of Patronage Level		Alternative Measure of Patronage Level	
	(1)	(2)	(3)	(4)
	Trip	Quantity	Trip	Quantity
HACKED × DBA × HIGH	**.3963** **(.1898)**	**.4636** **(.1901)**	**−1.6738** **(.6641)**	**−1.7574***** **(.6716)**
DBA	−1.2828*** (.0994)	−1.3270*** (.1111)	−.1840 (.4168)	−.0923 (.4232)
HACKED × DBA	−1.0914*** (.1270)	−1.1834*** (.1420)	1.3254*** (.4923)	1.4251*** (.5083)
DBA × HIGH	.8231*** (.1416)	.7274*** (.1449)	1.2275** (.5435)	.9400* (.5546)
Customer fixed effects	Yes	Yes	Yes	Yes
# of observations	14,700	14,700	5,052	5,052
# of customers	7,350	7,350	2,526	2,526
R^2	.6627	.6628	.4258	.4255

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DDD estimate) that is statistically significant is highlighted in bold.
*p < .10; **p < .05; ***p < .01.

TABLE 10
Robustness to Alternative Study Time Period: Short-Term and Long-Term Effects of Data Breach Announcement on Spending

	Time Period under Study					
	(1)	(2)	(3)	(4)	(5)	(6)
	1 month pre- and post-DBA	2 months pre- and post-DBA	3 months pre- and post-DBA	4 months pre- and post-DBA	5 months pre- and post-DBA	6 months pre- and post-DBA
	DV: ln(Spending)	DV: ln(Spending)	DV: ln(Spending)	DV: ln(Spending)	DV: ln(Spending)	DV: ln(Spending)
HACKED × DBA	**−4.2349*** (.1627)**	**−2.5095*** (.1200)**	**−1.6572*** (.1013)**	**−1.0933*** (.0881)**	**−.7840*** (.0788)**	**−.5597*** (.0725)**
DBA	−.8477*** (.1510)	−1.2954*** (.1003)	−1.4291*** (.0787)	−1.4090*** (.0649)	−1.3032*** (.0557)	−1.2555*** (.0503)
Customer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	16,048	19,558	21,114	21,960	22,458	22,720
# of customers	8,024	9,779	10,557	10,980	11,229	11,360
R^2	.5929	.5623	.5955	.6335	.6612	.6819

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its estimate (i.e., DD estimate) are highlighted in bold. *p < .10; **p < .05; ***p < .01.

TABLE 11
Robustness to Alternative Study Time Period: Short-Term and Long-Term Effects of Data Breach Announcement on Channel Migration

	Time Period under Study					
	(1)	(2)	(3)	(4)	(5)	(6)
	1 month pre- and post-DBA	2 months pre- and post-DBA	3 months pre- and post-DBA	4 months pre- and post-DBA	5 months pre- and post-DBA	6 months pre- and post-DBA
	DV: logit(CM_Trip)	DV: logit(CM_Trip)	DV: logit(CM_Trip)	DV: logit(CM_Trip)	DV: logit(CM_Trip)	DV: logit(CM_Trip)
HACKED × DBA	**4.0944*** (.6571)**	**2.6687*** (.4111)**	**1.9037*** (.3110)**	**1.4641*** (.2377)**	**1.3821*** (.1900)**	**1.2895*** (.1444)**
DBA	−3.9969*** (.4510)	−3.3169*** (.2923)	−2.5933*** (.2259)	−1.9456*** (.1743)	−1.5056*** (.1448)	−1.1016*** (.1192)
Customer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	1,448	3,120	4,680	5,984	7,026	7,904
# of customers	724	1,560	2,340	2,992	3,513	3,952
R^2	.5941	.5591	.5333	.5267	.5027	.5039

Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its estimate (i.e., DD estimate) are highlighted in bold. *p < .10; **p < .05; ***p < .01.

TABLE 12
Falsification Test 1: Placebo Effect

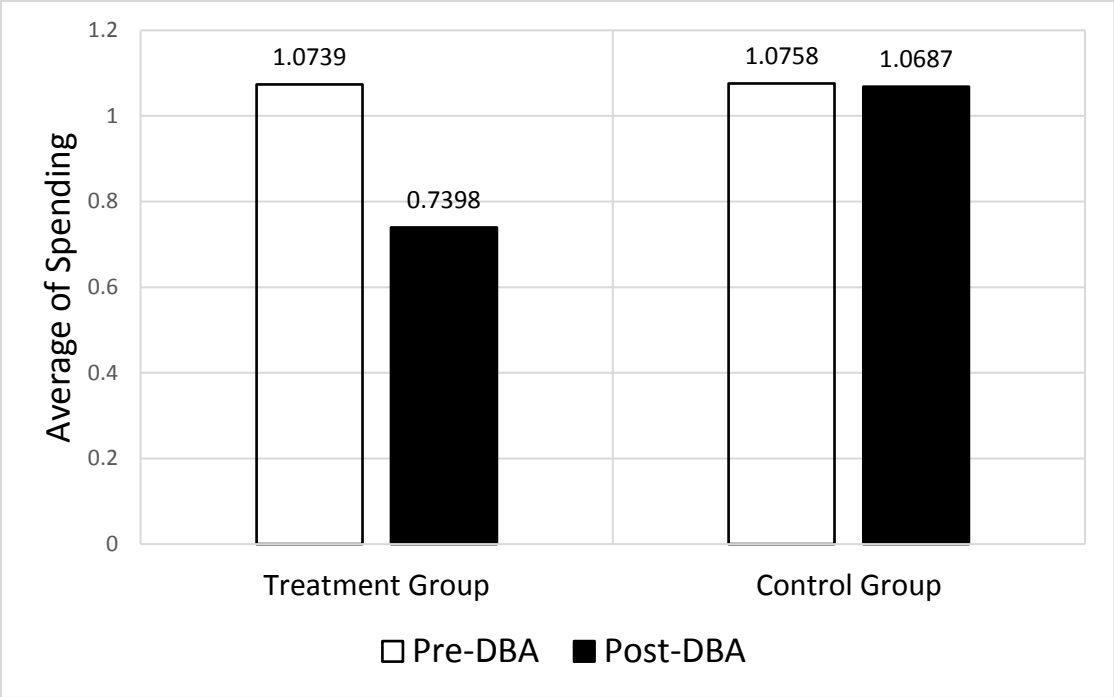
	(1)	(2)
	DV: ln(Spending)	DV: logit(CM_Trip)
HACKED × DBA^{Placebo}	.0358 (.0715)	−.1325 (.2015)
DBA^{Placebo}	.6275*** (.0546)	.3753* (.1944)
Customer fixed effects	Yes	Yes
# of observations	19,020	4,078
# of customers	9,510	2,039
R^2	.6184	.4361
Notes: Robust standard errors that are clustered at the customer level are in parentheses. The focal variable of interest and its coefficient estimate (i.e., DD estimate) that is statistically significant is highlighted in bold. *p < .10; **p < .05; ***p < .01.		

TABLE 13
Falsification Test 2: Fake Treatment Group

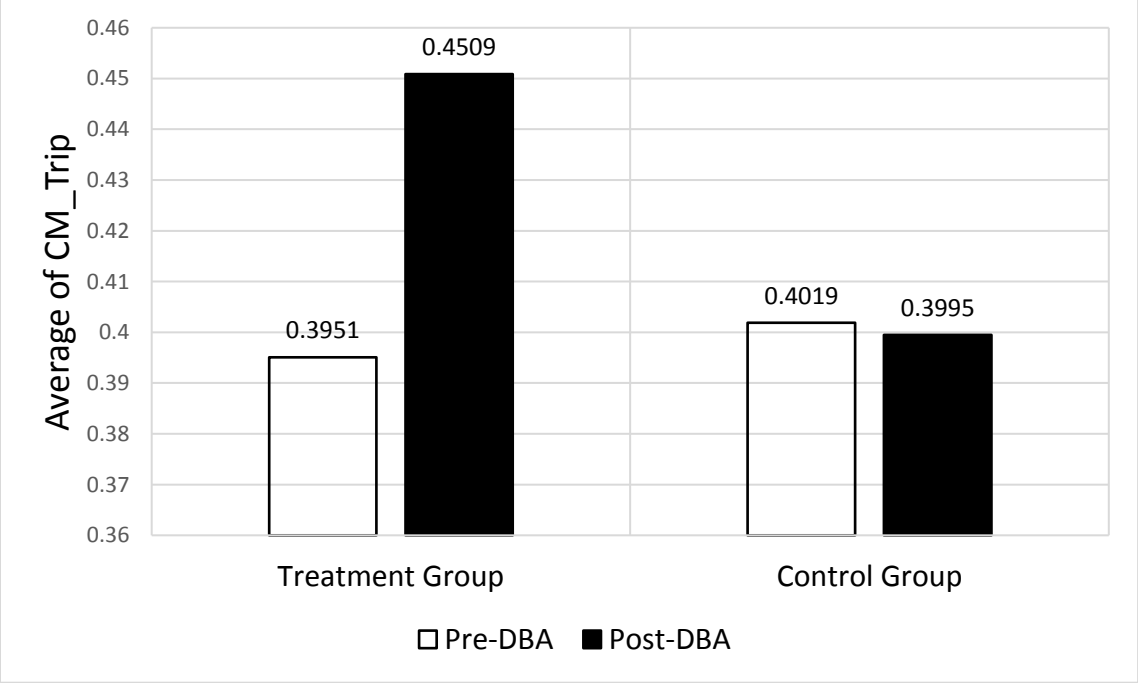
	(1)	(2)
	DV: ln(Spending)	DV: logit(CM_Trip)
HACKED^{Fake} × DBA	−.0010 (.0655)	.0012 (.1371)
DBA	−1.3286*** (.0327)	−.7567*** (.0685)
Customer fixed effects	Yes	Yes
# of observations	12,910	5,766
# of customers	6,455	2,883
R^2	.7019	.4874
Notes: To confirm the stability of the estimation results, we report the bootstrap coefficient estimates and bootstrap robust standard errors that are clustered at the customer level (provided in parentheses) based on the randomly selected 5,000 bootstrap samples. We also provide the bootstrap R-squared. *p < .10; **p < .05; ***p < .01.		

FIGURE 1
Comparison of Shopping Behavior in the Pre- and Post-DBA Periods

(a) Spending level



(b) Channel migration



The Effect of Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer

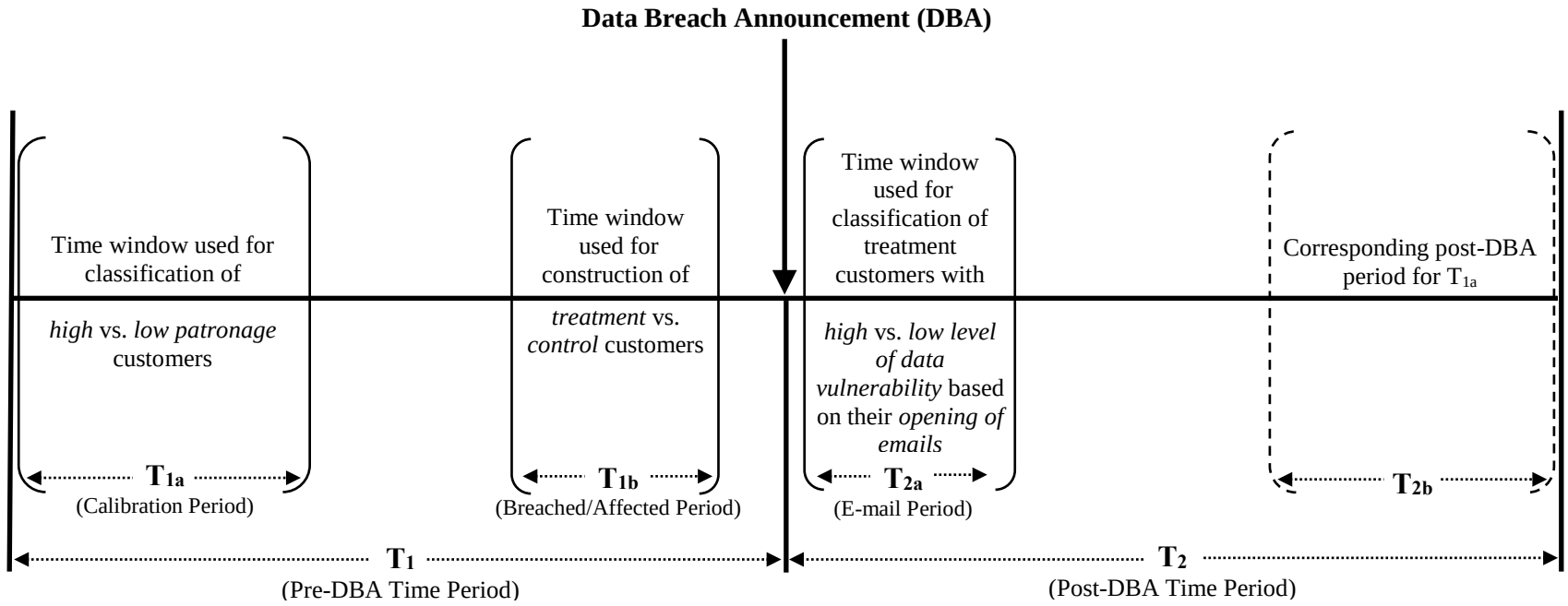
Ramkumar Janakiraman, Joon Ho Lim, and Rishika Rishika

Web Appendix

OVERVIEW

WEB APPENDIX W1. TIMELINE OF STUDY AND CONSTRUCTION OF VARIABLES

Figure W1.1: Illustration of Time Windows for Key Variable Construction and Model Estimation



Notes: This figure illustrates the time windows used to construct the key variables and examine the effect of data breach announcement (DBA) on customer behavior. We use the seven months in the pre-DBA time period (T_1) and the seven months in the post-DBA time period (T_2) for our main difference-in-differences (DD) models of spending level and channel migration. We define our treatment customers as those who transacted with the retailer via the "breached" channel at least once during the "affected" period (T_{1b}). To construct an indicator variable that identifies whether a customer is a high or low patronage customer, we use a median split of customers' spending level (Tucker, Zhang, and Zhu 2012) during the first three months of the pre-DBA period (T_{1a}). By using this indicator variable, we estimate the difference-in-difference-in-differences (DDD) model over the four months before and after the DBA ($T_1 - T_{1a}$ and $T_2 - T_{2b}$) to ensure that the calibration period (T_{1a}) does not overlap with the estimation period of DDD model. By using treatment customers' email opening behavior, we segment treatment customers into two groups, a treatment group with high-level of data vunerability (i.e., treatment group of customers who received and opened the retailer's email within a week following the DBA (T_{2a})) and a treatment group with low-level of data vulnerability (i.e., treatment group of customers who received but did not open the retailer's email). We construct the matching variables based on customers' shopping behavior and demographic information prior to the breached period within the pre-DBA period ($T_1 - T_{1b}$) in order to run a logit model of the propensity score matching (PSM).

WEB APPENDIX W2. DIFFERENCE-IN-DIFFERENCES MODEL: MOTIVATION AND MODEL SETUP

Difference-in-Differences

Randomized controlled trial is widely considered as the gold standard for establishing a cause-and-effect relationship. A successful randomization washes out all differences between the treatment and control groups with the exception of the "treatment effect" and thus the difference in outcomes of treatment and control group can be attributed to the treatment effect. However, it is very common that a controlled experiment is not feasible in observational study settings due to ethical, legal, or practical barriers. To overcome this limitation, researchers who work with observational data to establish the causal effect of a specific treatment or intervention (e.g., policy change, program implementation, enactment of regulation, etc.) employ the *difference-in-differences* (DD) modeling approach (one type of quasi-experimental designs).

To understand how the DD approach works in general and why it is appropriate for our study, it is worth considering two naïve approaches for making a causal inference. The first approach is a simple *before-and-after* comparison of outcomes for a *treatment group alone*. However, this approach does not control for other confounding factors that coincide with the treatment. The second simple approach is to compare outcomes of the treatment and control groups in the post-treatment period, *ignoring pre-treatment outcomes*. However, this approach does not rule out any inherent differences between the treatment and control groups that are invariant over time.

DD approach "combines" these two naïve approaches as it takes into account (1) the difference between post-intervention outcome and pre-intervention outcome (the first difference) and (2) the difference between a treatment group's outcome and a control group's outcome (the second difference). The *first difference* (pre-treatment vs. post-treatment) controls for time-invariant characteristics of each group and *the second difference* (treatment group vs. control group) controls for any time trends or time-varying factors. In other words, the DD approach compares the changes in the outcome variables before and after the treatment between two groups of units or individuals, those that received the treatment (a treatment group) and a group that is not affected by the intervention (a control group) and helps remove the potential confounding factors that are the main sources of biases in the causal effect estimation (Angrist and Pischke 2009; Wooldridge 2002; Gertler et al. 2011).

In our context, the DD approach compares the behavior of customers in the treatment group (the affected customers whose data was breached) and the control group (the unaffected customers) before and after the focal retailer's data breach announcement (DBA). We are thus able to account for both time invariant customer characteristics and for any time trend effects and establish the causal effect of DBA on customer behavior.

Difference-in-Differences in Regression Modeling Framework

The DD method can be implemented by the following simple regression model:

$$\text{(W2.1)} \qquad Y_{it} = \beta_0 + \beta_1 TRT_i + \beta_2 POST_t + \beta_3 TRT_i \times POST_t + \varepsilon_{it},$$

where TRT_i is a treatment group indicator that is equal to 1 if the individual i is in the treatment group and 0 if the individual i is in the control group; $POST_t$ is a time period indicator that is equal to 1 if time period t is the post-treatment period and 0 if time period t is the pre-treatment period; $TRT_i \times POST_t$ is an interaction term of the two indicator variables. The primary coefficient of interest is the interaction coefficient, β_3, which captures the causal effect of the treatment (i.e., changes in Y of the treatment group and Y of the control group across the pre- and post-treatment periods and hence the name *difference-in-differences estimate*). Table W2.1 explains how the different components (regression coefficients) make up the DD estimate.

Table W2.1: Difference-in-Differences Estimate

	Pre-Treatment Period	Post-Treatment Period	Difference
Control Group	β_0	$\beta_0 + \beta_2$	$(\beta_0 + \beta_2) - \beta_0 = \beta_2$
Treatment Group	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) = \beta_2 + \beta_3$
Difference	$(\beta_0 + \beta_1) - \beta_0 = \beta_1$	$(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) = \beta_1 + \beta_3$	**DD** $= (\beta_2 + \beta_3) - \beta_2 = (\beta_1 + \beta_3) - \beta_1 = \boldsymbol{\beta_3}$

WEB APPENDIX W3. PROPENSITY SCORE MATCHING AND DIFFERENCE-IN-DIFFERENCES

Table W3.1: Logit Estimates for Propensity Score Matching

Matching Variable	Analysis	
	(1)	(2)
	Spending Level	Channel Migration
TRIP_HACKED_CHANNEL	.1286***	.3207***
	(.0263)	(.0454)
TRIP_HACKED_CHANNEL^2	−.0201***	−.0381***
	(.0029)	(.0051)
MULTICHANNEL_CUSTOMER	−1.4731***	−.6468***
	(.0656)	(.1087)
AGE	.0063***	.0082***
	(.0015)	(.0025)
FEMALE	.1238**	.3186***
	(.0527)	(.0986)
SINGLE	.0227	.0536
	(.0482)	(.0794)
INCOME	.0014	−.0004
	(.0023)	(.0038)
Constant	.4384***	−1.1505***
	(.1306)	(.2307)
N	11,459	4,455
Log-likelihood	−7,461.995	−2,824.333
Nagelkerke's R^2	.0879	.0413
*p < .10; **p < .05; ***p < .01.		

Table W3.2: Covariate Balance Checks

Panel A. Covariate balance before and after matching for analysis of spending level

Separate balance test of each matching variable

Matching Variable	Before Matching		After Matching		Bias Reduction (%)
	Standardized Difference	z-score	Standardized Difference	z-score	
TRIP_HACKED_CHANNEL	−.2570***	−13.5337	.0000	.0000	100
TRIP_HACKED_CHANNEL^2	−.2036***	−10.7543	.0000	.0000	100
MULTICHANNEL_CUSTOMER	−.4900***	−25.2789	.0000	.0000	100
AGE	.0573***	3.0392	−.0016	−.5120	97.13
FEMALE	.0344*	1.8241	.0000	.0000	100
SINGLE	−.0007	−.0356	−.0000	.0000	100
INCOME	−.0094	−.5012	.0085	.4921	9.84

Omnibus balance test

Before Matching			After Matching		
Combined Baseline Difference Statistic (d^2)	df	*p*-value	Combined Baseline Difference Statistic (d^2)	df	*p*-value
724.5822	7	.0000	.2790	2	.8698

Panel B. Covariate balance before and after matching for analysis of channel migration

Separate balance test of each matching variable

Matching Variable	Before Matching		After Matching		Bias Reduction (%)
	Standardized Difference	z-score	Standardized Difference	z-score	
TRIP_HACKED_CHANNEL	−.1104***	−3.5180	.0000	.0000	100
TRIP_HACKED_CHANNEL^2	−.1654***	−5.2583	−.0000	.0000	100
MULTICHANNEL_CUSTOMER	−.1353***	−4.3062	.0000	.0000	100
AGE	.1064***	3.3894	.0068	1.2224	93.61
FEMALE	.1079***	3.4374	.0000	.0000	100
SINGLE	−.0058	−.1862	.0000	.0000	100
INCOME	−.0187	−.5976	.0147	.6134	21.39

Omnibus balance test

Before Matching			After Matching		
Combined Baseline Difference Statistic (d^2)	df	*p*-value	Combined Baseline Difference Statistic (d^2)	df	*p*-value
77.3050	7	.0000	1.7712	2	.4125

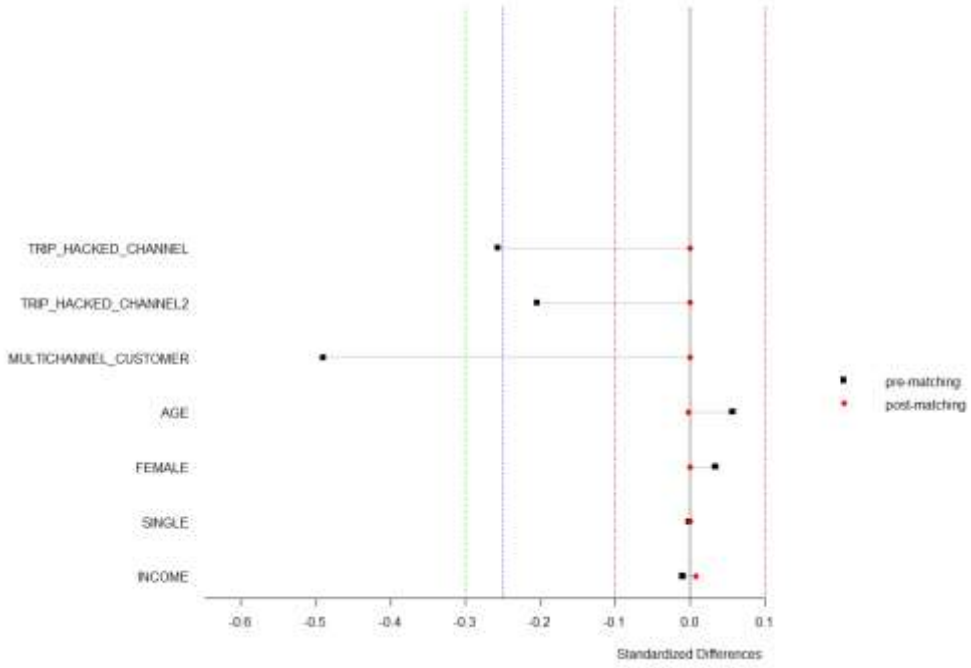
Notes: Refer to Hansen and Bowers (2008) for full details about standardized difference, z-score, and d^2 calculation. Bias reduction is computed by:

$$100 \times \frac{\left(|\text{Standardized difference before matching}| - |\text{Standardized difference after matching}|\right)}{|\text{Standardized difference before matching}|}$$

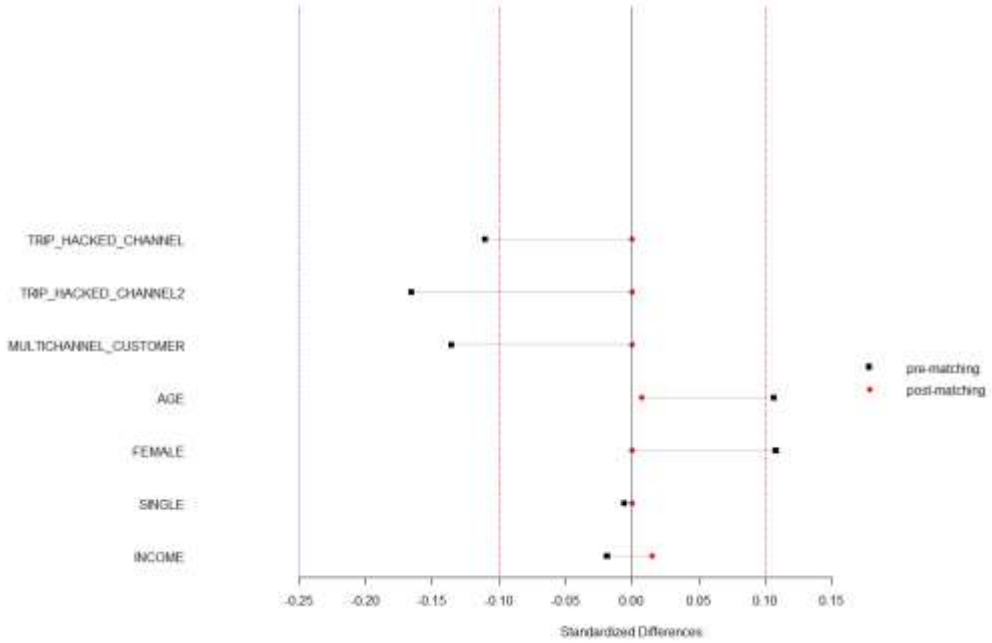
*p < .10; **p < .05; ***p < .01.

Figure W3.1: Balance Assessment Plots

(a) Analysis of spending level



(b) Analysis of channel migration



Notes: Three types of vertical line pairs are presented in the plot. Each type stands on a specific standardized difference value that serves as the decision criterion for achieving covariate balance after matching. We use .10 (D'Agostino 1998), .25 (Ho et al. 2007) and .30 as decision criteria for covariate balance that are presented by red (short dash dotted), blue (dotted), and green (short dash) vertical lines respectively. A covariate balance is accepted when a standardized difference after matching lies between the two vertical lines for each type of criterion. The smaller a decision criterion value (i.e., |Standardized Difference|), the stricter the condition for achieving covariate balance.

Figure W3.2: Steps for Propensity Score Matching and Difference-in-Differences

Step 1. Selection of matching variables

- Find a proper set of matching variables (i.e., observed characteristics) that can lead to an imbalance between treatment and control groups.
- Use theories and prior empirical studies to choose relevant matching variables.
- Check whether treatment and control groups are similar each other on the matching variables.

Step 2. Estimation of propensity scores using a logit model

- Regress a binary variable (1 for treatment group and 0 for control group) on the set of matching variables.
- Calculate the propensity score (i.e., the probability of receiving treatment) for each unit in both treatment and control groups based on the estimation results of the logit model.

Step 3. Matching

- Select a matching method that suits the research setting best. Examples of available matching methods include optimal matching (e.g., pair matching, matching with a variable number of controls), greedy matching (e.g., caliper matching) and Mahalanobis distance matching.
- Match treated units and control units who have similar propensity scores by using the selected matching algorithm and create a group of matched-sets.

Step 4. Assessment of the quality of matching

- Check the balance between the treatment and control groups on the matching variables after matching by comparing it with the balance prior to matching.
- Use standardized differences measure or omnibus test to evaluate the balance.
- If there is still imbalance between treatment and control group, try to find new matching variables or change the logit model specification by including higher order terms of matching variables. That is, repeat steps 1 - 3 until treatment and control groups are well balanced on the observed characteristics.

Step 5. Combination of propensity score weighting and difference-in-differences model

- Transform the propensity scores into weights as follows:

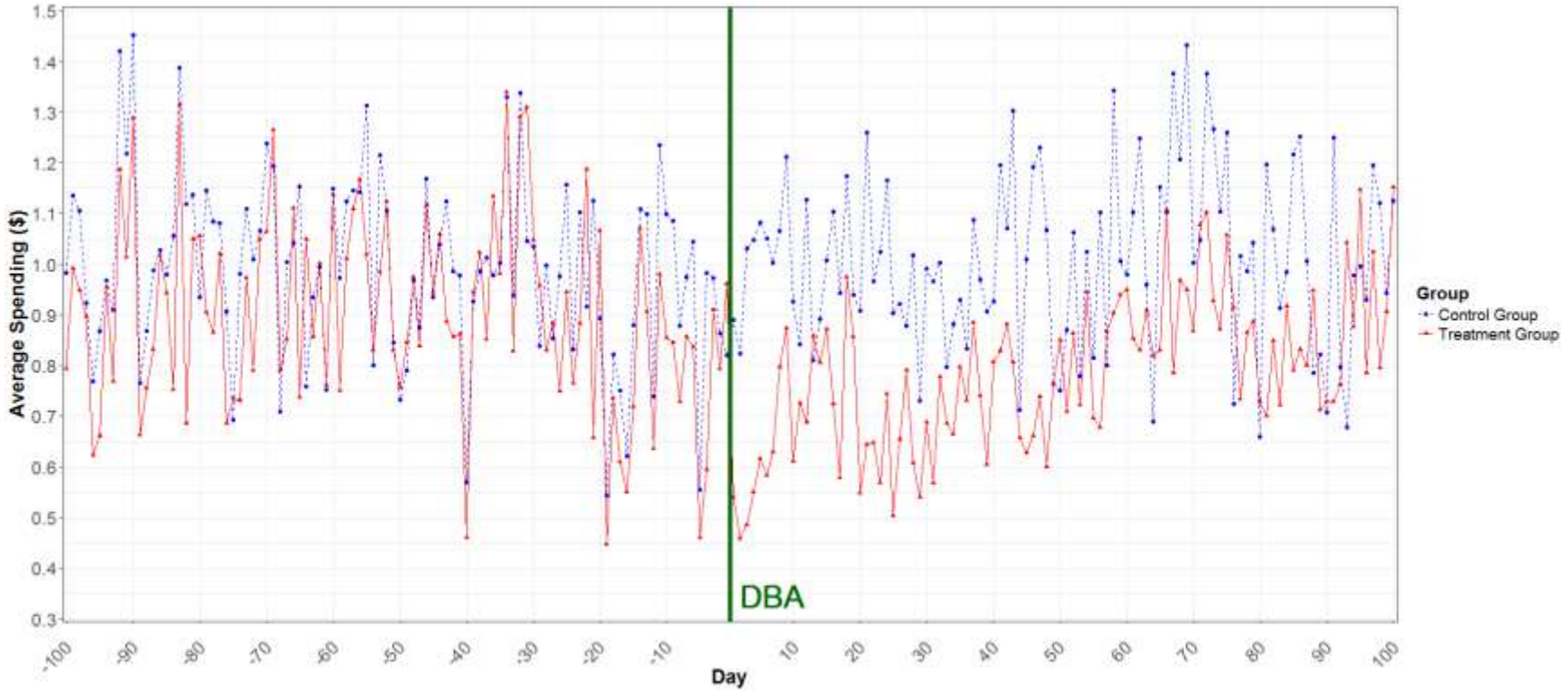
(W3.1)
$$\omega(X,z) = \frac{X}{\hat{p}(z)} + \frac{1-X}{1-\hat{p}(z)},$$

where X = 1 for a treated unit (= 0 for a control unit) and $\hat{p}(z)$ is the estimated probability of being treated from the logit model with a set of matching variables z (i.e., propensity score).

- Include the weights in the difference-in-differences model estimation.

Web Appendix W4. Trend in Customer Spending Level

Figure W4.1: Customer Spending Before and After the Data Breach Announcement



Notes: The time series plot shows the variation in average daily spending level of the treatment and control group customers 100 days before and after the data breach announcement (DBA). For reasons of confidentiality, we rescale the original spending amount (in the United States dollars) by dividing it by the average of spending amount across all customers in both pre- and post-DBA periods.

REFERENCES (WEB APPENDIX)

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

D'Agostino, Ralph B. Jr. (1998), "Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group," *Statistics in Medicine*, 17 (19), 2265−81.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch (2011), *Impact Evaluation in Practice*, 1st ed. Washington, DC: World Bank.

Hansen, Ben B. and Jake Bowers (2008), "Covariate Balance in Simple, Stratified and Clustered Comparative Studies," *Statistical Science*, 23 (2), 219−36.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart (2007), "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15 (3), 199−236.

Tucker, Catherine, Juanjuan Zhang, and Ting Zhu (2012), "Days on Market and Home Sales," working paper, (March 21), (accessed January 16, 2016), [available at http://ssrn.com/abstract=1481321].

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: Massachusetts Institute of Technology Press.