The University of British Columbia
Vancouver, April 2019

Market Effects of Data Breaches

An Empirical Analysis of Disciplinary Market Forces

By: Emmet Hall-Hoffarth

Supervisors: Erik Snowberg and Marit Rehavi

Submitted to the Vancouver School of Economics in partial fulfillment of the requirements for an Honours BA in Economics

April 5, 2019

THE UNIVERSITY OF BRITISH COLUMBIA

VANCOUVER SCHOOL OF ECONOMICS

Abstract

The high costs and privacy implications of data breaches have been widely publicised, but through what mechanisms are these costs imposed on firms, and to what extent do firms internalise them? Do consumers value their privacy enough to punish firms for leaking their private information? This paper uses a data set of over 750 data breaches and an event study methodology to evaluate changes in revenues, expenses, profit, and stock returns that are caused by data breaches. I find evidence that consumers will indeed punish some firms for leaking their data, however, the response is highly heterogeneous and temporary. This punishment does translate into temporarily reduced profits. Furthermore, consistent with previous findings, a firm's stock returns fall significantly in the wake of a data breach. On the other hand there is no evidence that firms change their behaviour as a response to this punishment by increasing spending on security or advertising. In fact, expenses may fall, and this seems to reflect down-scaling or cutbacks which come as a result of data breaches. There is no evidence that any of the identified negative effects persist. The findings of this paper have implications about the sufficiency of market mechanisms in imposing a socially optimal outcome with regards to data security.

Contents

List of Figures

List of Tables

1 Introduction

On September 28, 2018, Facebook (2018) announced a data breach in which the information of more than 50 million of its users was exposed to hackers.[1] In recent years events like this one have become increasingly prominent in the mainstream media as commentators grapple with the societal implications of a so-called, "post-privacy world." Broad philosophical considerations aside, given their scale and cost, data breaches also raise many interesting and important economic questions.

While the frequency of such events is not necessarily increasing, the total annual costs of these events lies in the tens of billions of dollars in the United States alone (Edwards et al., 2016). One study by the Ponemon Institute (2018) found that the average cost of each individual data breach to the victim firm is approximately $4 million. Of particular economic concern is the extent to which various stakeholders internalise these costs. The severity of data breaches provides suggestive evidence that firms may systematically under-invest (relative to socially optimal levels) in preventative and remediation strategies due to insufficient corrective action from market forces. Indeed, Roberds and Schreft (2009) develop a theoretical model which predicts exactly this outcome in equilibrium. If this prediction is true, it would have the clear policy implication that market intervention in order to reduce the risk of data breaches would be advisable, not only in order to protect individual privacy, but also to improve economic outcomes. In order to evaluate this claim from the empirical perspective, it is necessary to understand the costs of data breaches that firms face and the extent to

[1] I would like to acknowledge Anders Wettergreen Gundersen for providing the template used to typeset this thesis, as well as Marek Hlavac for creating the Stargazer package for R which was used to create the tables.

which they are internalised, the total social costs, and how effective preventative strategies actually are. Evaluating all of these would be outside of the scope of this paper, which aims only to open this line of academic inquiry by addressing the types and scale of costs of data breaches faced by firms. The remaining issues are left as topics for future research.

In order to address this question, this paper employs an event study methodology which can identify the effect of data breaches on a number of firm specific and stock market variables. I find some evidence that consumers do indeed punish firms for leaking their information, as a discontinuous drop in revenue can be observed after a data breach. However, this effect is short-lived and highly variable. This drop in revenue does indeed translate to an observable temporary loss in net income[2]. The amount punishment that a firm faces seems to be increasing in the number of data breaches a firm has. Furthermore, consistent with other studies conducted in this area, I find that stock returns fall precipitously in the wake of a data breach, however, there is no evidence that this effect persists. Despite these costs, there is no evidence that firms increase spending on either security or advertising efforts after a data breach — providing suggestive evidence that market forces are not enough to cause firms to change their behaviour and achieve the socially optimal outcome. This is perhaps hardly surprising given that all of the negative effects of data breaches appear to be temporary.

The remainder of this paper will proceed as follows: Section 2 provides an overview of relevant past research. Section 3 describes the empirical estimation strategy used, and without giving a full theoretical model, outlines some intuition that guides the expected results. Section 4 discusses the data used. Section 5 analyses the empirical results, all of which are given in the appendix. Finally, Section 6 summarises the conclusions of this study.

[2] I use profit and net income interchangeably in the context of this paper.

2 Literature Review

There are many aspects that require study in order to understand the costs associated with data breaches. Thus far, most of the academic literature in this area has focused on the impact of data breaches on the stock market. For example, Goel and Shawky (2009) conducted an event study on the stock returns for firms that fell victim to data breaches in the period 2004-2008 and found that these firms lost on average 1% of their market value in the days following the event. Rosati et al. (2019) similarly find significant negative effects on stock returns. Whether or not these effects persist is unclear. A study by Morse et al. (2011) found significant negative returns, and that these negative returns persist for at least two years after the data breach. However, Acquisti et al. (2006) found that these effects do not persist for long; perhaps only a few days. Another study by Gordon et al. (2011) investigated how the effects of data breaches on stock prices have changed over time and found broadly[3] that the effects have been weakening over time. These studies all lend inspiration to the sources of heterogeneity which can be studied when considering how other outcomes respond to data breaches.

However, how might *consumers* be expected to respond to being informed of a data breach, and for what reasons? One aspect to consider is media coverage that is generated in the wake of a data breach. Since these events generate primarily negative coverage, the most obvious conclusion is that sales would be harmed. Indeed, negative publicity can damage reputations, and reputation has been shown to be a strong predictor of sales (Livingston, 2005). However, a study by Berger et al. (2010) found that the effects of negative publicity on sales can be

[3] By comparing pre and post 9/11 periods.

positive for relatively unknown products while being negative for well-known products. This is explained through the competing mechanisms of increased product awareness yet increased negative sentiment caused by negative publicity. Of particular interest is whether — given information about a data breach — consumers will try to punish firms for leaking their information. The limited empirical work which has been carried out in this area seems to suggest that this is not the case. Davis et al. (2009) found no significant impact on the web traffic of online businesses as a result of data breaches. They provide a number of possible explanations for this including the fact that consumers are somewhat insured against damages from a data breach[4] and thus do not sustain substantial losses, and a collective action problem which arises as individual consumers do not believe that they can send a sufficiently strong signal to firms by ceasing their purchases, even if they have sustained substantial losses.

Given that consumer reactions to data breaches are both theoretically ambiguous and empirically small in other studies, the results of this paper to make a significant contribution to existing literature. Furthermore, this paper also considers the internal costs of data breaches, and area in which there is to my knowledge no existing academic literature.

3 Methodology

In order to evaluate all of the potential costs of a data breach as outlined in the introduction, this paper considers two separate models applied to two separate data sets. The first model examines both the internal and external costs that a firm may face. This will be referred to as the *firm study* from the remainder of this paper. The second model examines stock market

[4] For example, in the case of credit card theft consumers are usually not liable.

responses to data breaches. This will be referred to as the *stock market study* for now on. Both studies involve an event study methodology, which attempts to identify a discontinuous change in outcomes before and after the data breach. However, the exact implementations differ greatly, so I will discuss them separately. I begin now by outlining the *firm study.*

3.1 Firm Study

The *firm study* employs an event study methodology to identify the causal impact of data breaches on affected firms. The model considers both whether there is any immediate effect, and if so, whether those effects persist. This model uses the following regression equation:

$$Y_{iq} = \beta_0 + \beta_1 A_{iq} + \beta_2 (Q \times A_{iq}) + \vec{\beta_3} X_{iq} + \vec{\beta_4}(X_{iq} \times \vec{A_{iq}}) + \alpha_i + \delta_q + \epsilon_{iq} \tag{3.1}$$

Where Y_{iq} is the outcome of interest, A_{iq} is a dummy variable which is 1 if a firm i has already announced a data breach in quarter q, Q is the number of quarters since the start of the data, X_{iq} is a matrix of controls, $\vec{A_{iq}}$ is a column vector of A_{iq}, and α_i and δ_q are firm and quarter fixed effects, respectively. In this equation there are two primary coefficients of interest: β_1 and β_2. β_1 measures the immediate discontinuous effect of the data breach on the outcome. Meanwhile, β_2 is a linear time trend interaction which indicates the persistence of any effect. Mathematically, β_1 amounts to a change in intercept of the regression function after the data breach, and β_2 a change in slope. For example, if β_1 is negative, and β_2 is positive, this means that there is some immediate negative effect, but that that effect decays over time. If β_1 is negative, but β_2 is zero, then the effect persists over the event window. I also include specifications where β_2 is excluded. Here β_1 captures the average effect over the event window, so it will be different from zero if there is an effect which persists over the event window, and zero otherwise. Clearly, whether or not a given effect is considered to be persistent depends on the length of the event window considered, since, for example, it is possible for an effect to persist for one year but not two years. Therefore, specifications are given for the primary outcomes with event windows of six months, one year, two years and, three years after the data breach.

In the *firm study* there are three primary outcomes for which this model is evaluated, which I will now discuss in turn. The first is the revenues of firms. In this case, there are theoretical reasons to believe that β_1 may be negative or zero. On one hand, if the representative consumer either believes that data breaches occur randomly, and thus the occurrence of a data breach conveys no information about the firm and its chances of being breached again, or they do not suffer disutility from the firm leaking data, then it would be irrational for the consumers to boycott that firm because their expected utility from continuing to buy from that firm has not changed. If this is true then β_1 and β_2 are expected to be zero. On the other hand, if the representative consumer internalises information from the data breach such that they believe that the firm is more likely to leak information again in the future, and if leaks cause consumers disutility, then it would be entirely rational for consumers (if they behave as a cohesive unit) to reduce future purchases from that firm as they expect that another costly data breach may occur in the future. Furthermore, consumers (again as a cohesive unit) may wish to boycott as a form of punishment since the credible threat of this punishment if large enough will be enough to discipline all firms into better protecting customer data. If this is true β_1 is expected to be negative. However, even if it would benefit consumers as a whole to punish firms, they may face a collective action problem, because each individual consumer has an incentive to continue buying from a firm, in the hopes that other consumers will carry out the punishment. Despite this slight theoretical ambiguity, in light of the findings of Davis et al. (2009) and the likelihood of a collective action problem it seems more likely *a priori* that β_1 in the case of firm revenue will not be significantly different from zero.

The other primary outcomes that are considered are costs to and profits of firms as measured by both operating and non-operating expenses, and net income respectively. The interpretation of the coefficients are the same as for revenue, so what remains is to discuss the theoretical effect of data breaches on these outcomes. It seems likely *ex ante* that expenses, particularly non-operating expenses which encompass unexpected and unusual costs such as legal settlements will increase in the wake of a data breach. Other potential sources of increased costs — which are captured by operating expenses — include security upgrades, and increased advertising to recover public goodwill. However, these expenses may be transitory in nature — when considering a longer time-frame expenses may decrease because the data breach causes a general decline in the firm which leads to downsizing. Given lost sales and potentially increased expenses, profits are likely to be adversely impacted. Including this outcome will check the robustness of the revenue and operating expense specifications, as the sum of the effect on these two should roughly equal the effect on profit.

In addition to considering the magnitude of effects on these primary outcomes, this study also investigates some potential sources of heterogeneity among the responses to data breaches by interacting various outcomes with the A_{iq} dummy. These variables are included in the matrix X of controls, and therefore, their marginal impact on the total effect of a data breach is given by the coefficients in $\vec{\beta_4}$[5]. In particular, I test whether the size of the data breach (as measured by the number of records leaked), media coverage (as proxied for by Google Trends), the size of the firm, or the type of data leaked result in a heterogeneous response. Finally, the study also considers whether the response in outcomes depends on the number of times a firm leaks data by breaking results out into subsets based on the number of data breaches a firm has had in the sample. I will not now express a prior as to the expected effect of these variables, and instead leave it as a strictly empirical exercise. The observed marginal impact of these factors will be discussed in the results section.

3.2 Stock Market Study

The *stock market study* will follow standard procedures for conducting an event study in financial economics. At least two models will be applied in order to test robustness, which are the capital asset pricing (hereafter CAPM) and Fama-French 5 factor models (Fama and French, 2015). These models will be fit to the data during the *prediction window*[6] before the data breach and be used to calculate predicted returns during the *event window* after the data breach, which when subtracted from the actual returns give abnormal returns (AR). These abnormal returns can be summed over the *event window* to generate cumulative abnormal return (CAR). The general logic of this test is that if AR tend to be very large (and thus as is CAR) in absolute value then this means that the event – in this case data breaches – has had a strong effect on returns. There are a number of non-parametric methods to test the significance of these abnormal returns as their distribution can be heavily skewed, however,

[5] Note that for time invariant features such as characteristics of the data breach itself only the interaction term is included because the linear term is co-linear with the firm fixed effect for firms with one data breach.

[6] I use the period 180 days before the data breach

in order to maintain consistency with the *firm study* and broader common practice in the economics literature, this study will employ a simple t-test. The CAPM has the following functional form:

$$R_{it} = \beta_0 + \beta_1 MRP_t + \mu_{it} \tag{3.2}$$

Where R_{it} is the daily excess (greater than risk free) stock return for firm i on day t, and MRP_t is value weighted excess market return as given by CRSP. The Fama-French model expands on this with some additional factors to improve fit. It has the following functional form:

$$R_{it} = \beta_0 + \beta_1 MRP_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \mu_{it} \tag{3.3}$$

Where SMB, HML, RMW, and CMA are controls for firm size, book-to-market value, profitability and, investment respectively. The coefficients on these terms are not of importance for this study. The theoretical advantage of this model is that it may predict returns over the event period more precisely, potentially allowing for the effect of the data breach to be identified more strongly against noise. For robustness both of these models will be used to calculated AR and subsequently CAR.

The t-test on CAR will determine whether data breaches have a significant effect on stock returns. If the t-statistic is significantly negative then there is evidence of punishment on the stock market in the form of reduced returns as a result of the data breach. Alternatively, if the t-statistic is zero, the firms do not face stock market consequences for the data breach. Furthermore, this study will consider whether these effects persist. If there is a significant effect with a short window, there may or may not be a significant effect with a longer window. Therefore, this study considers event windows of 2, 5 10, 90, and 180 days. Longer event periods are possible, however, over a very long period it becomes difficult to rule out other significant shocks which may occur during the event period.

It is also interesting to consider the possible determinants of any stock market response. This is done by regressing CARs on various features such as number of records leaked via simple OLS. It should be noted however, that these results are merely correlations and do not identify the causal determinants of stock market reactions to data breaches.

This study will compare two mechanisms that can potentially explain stock market responses to data breaches. Firstly, data breaches may cause investors to lower their expectations of future profits as a result of lost sales, and higher expenses — a change in expectation which is not entirely unwarranted according to the results of the *firm study.* On the other hand, the occurrence of the data breach could signal new and negative information about the quality of a firm's infrastructure and management. This information could cause investors to reassess the fundamental value of the firm. These are both reasons to expect that negative abnormal returns will be observed.

Given the findings of previous studies (Goel and Shawky, 2009; Morse et al., 2011; Acquisti et al., 2006) it seems likely that some significant negative abnormal stock returns will be observed, at least with short post-event windows. Findings on the persistence of these effects are mixed, and will depend on whether or not data breaches have long-lasting effects, and whether or not the market internalises this information.

3.3 Identification

The regressions in both the *firm model* and the *stock market model* employ an event study methodology and thus have the same identifying assumptions, as well as the same strengths and weaknesses. This approach is in fact quite similar to a difference in difference approach, with the key difference being the absence of an explicit control group — instead the firms in the sample form their own control for the period before the data breach. Therefore, the identifying assumptions are similar to a difference in difference strategy, with some additional caveats. Firstly, the analogue to the parallel trends assumption relates to the trend for each firm during the *prediction window.* Since each firm is its own control in the *prediction window*, once the firm fixed effects are removed there should be no trend in the residuals before the data breach. Secondly, exactly as with a difference in difference approach, there must be no changes which are simultaneous with the data breach, or any significant exogenous shocks during the event window in question. Thirdly, the information about the data breach must not leak before it is publicly announced, as this is how the event date is defined. Otherwise the coefficients may be biased towards zero because the effect actually started before the regression coefficients can pick it up.

The first assumption has to do with the validity of the control group. There should be no trend in the outcomes for a firm in the *prediction window.* If this is not the case then there are two distinct issues. Firstly, if there is some trend in an outcome for a firm in the prediction window, then the coefficients in Equation 3.1 are biased because this functional form implicitly assumes the slope and intercept of the regression line are zero before the data breach, and then change afterwards. Therefore, the effects of any trend can be wrongfully attributed to the data breach. This is precisely what happens when the parallel trends assumption is not met in a difference in difference approach. Secondly, consider a firm that has declining profits before a data breach. This firm is experiencing structural decline and may have been forced to cut back on spending on key security infrastructure, thus making a data breach more likely. In this case we are concerned that the data breach is endogenously determined and that the effect that is measured is not truly the causal effect of a data breach, but rather the causal effect of a firm being in decline, and thus being more prone to a data breach.

Given these serious concerns it is fortunate that it is possible to test and verify the first identifying assumption. One way to do so is to simply inspect the plotted residuals for any signs of a trend during the *prediction window.* The other way to do so is to try placebo event dates before the actual data breach and confirm that the coefficients pick up no effect. Both of these tests can be employed and are discussed in the results section.

On the other hand, the second assumption will be difficult to verify, particularly when considering long event windows. Over a multiple year time period, many omitted factors may change that could introduce bias such as changes of key personnel, changes in competitiveness in industries and so on. In the *firm study* the violation of this assumption could contribute to much of the noise in the observed coefficients. However, this does not necessarily bias estimates in any particular direction, because exogenous shocks could push coefficients in either direction, and the effects of follow up shocks which are endogenous to data breaches (for example the CEO being fired) are — I would argue fairly — lumped into the coefficient measuring the effect of the data breach itself. Nevertheless, for the above reasons the time trend coefficients — especially for long event windows — will not be as strongly identified as the immediate effect as being the direct causal effect of the data breach.

The third assumption seems plausible, although it is possible that information about a data breach leaks out before the public announcement. This is likely not an issue for the *firm study* because if information about a data breach leaks early it likely does so by a number of days, rather than months, which would not change the fiscal quarter that the true event occurs in. However, this could be significant for the *stock study* which uses daily data. There are two methods used to address this concern. Firstly, a *buffer window* of 5 days between the end of the *prediction window* is included, meaning that the CAR begins to accumulate 5 days before the event date. This means that if the data breach was in fact leaked up to 5 days before it was publicly announced the resulting CAR will still be valid. Secondly, a second specification is offered (Tables 8.3 and 8.4) in which events are excluded if the largest change in CAR did not occur within two days of the event date, suggesting that some other significant event took place[7]. This criteria is harder to satisfy for longer event windows, which is why in this table sample size is decreasing in event window length. The other specification (Tables 8.1 and 8.2) does not impose this restriction.

As presented earlier it would be interesting to compare the reaction of consumers and investors to data breach events because the mechanisms through which these parties are affected are vey different. When put on the same scale the investor reaction is expected to be stronger than the consumer reaction, because stock returns price in expectations about consumer behaviour as well as a number of other negative effects of data breaches. Furthermore, if there were no consumer response and yet a stock market response is observed,

[7] This is an (admittedly imperfect) way to automate a common practice which is to manually exclude observations where other significant exogenous shocks occur during the event window

then this implies that the stock market revaluation is based on factors other than lost sales, such as damage to reputation or lower expectations about infrastructure and management.

4 Data

Broadly speaking the primary data used in this paper are data on data breaches themselves as well as data on outcomes for affected firms. Data on data breaches were obtained from Privacy Rights Clearinghouse (2018). This database lists more than 8000 data breaches since the year 2005, and provides for each of these information on the public announcement date, number of records leaked, type of breach (hacking, lost device, etc.), as well as a small description of the event which appears to come from news headlines and press releases. This list is likely to be very close to a comprehensive list of all events, at least in the United States — the area that the proposed study focuses on — because all 50 states have laws that mandate that firms publicly announce the occurrence of a data breach (NCSL, 2018). In other words, the occurrence of a data breach is public information, thus a similar or identical list should be possible to obtain from any other source.

This paper considers the subset of the Privacy Rights Clearinghouse database for which the data breach occurred in the United States and the affected firm is publicly traded[8], for the time period January 01, 2005 to December 31, 2017. This results in a final relevant sample size of 759. The number of records leaked is highly skewed to the right as summarised in Table 1.3 and shown in Figure 1.1, therefore, the natural logarithm of this variable is always used in regression specifications. It is also noteworthy that a large number of data breaches

[8] Those firms for which financial information is readily available

(358) actually result in zero records being exposed. Data on the type of information leaked is summarised in Table 1.2. These variables were created by scraping the event description for the given key words. Therefore, this variable is not perfect and contains a considerable amount of noise, however, it does allow for some novel comparisons.

The primary dependent variables in the *firm study* are firm revenue, profit and expensese[9]. These variables were collected for all publicly traded companies listed in the Privacy Rights Clearinghouse database from the quarterly financial statements as reported by COMPUSTAT (2019) for the duration of the sample period. These data are summarised at the top of Table 1.1. In addition to the primary outcomes, COMPUSTAT also provides a number of general firm specific variables that may be relevant such as number of employees and shareholders' equity. Another supply side variable that is potentially relevant is prices, as firms may lower prices to incentivise consumers to come back after the data breach. However, these data proved difficult to obtain and are not essential to include as lower demand will still be reflected in lower revenues and profits, regardless if it is quantity sold or price that actually falls. For the *stock market study*, comprehensive daily stock return data were also obtained through CRSP 2019, and data for the Fama-French factors were obtained from the personal web-page of Kenneth French (2018). These data are summarised in Table 1.4.

Further data on the monthly amount of Google Searches of a firm were collected using an API for Google Trends (2018). In particular, data were collected on an index of searches for the firm's name[10] and its stock ticker. These data are summarised at the bottom of Table 1.1. These data are a proxy for the amount of media attention that a firm receives and is

[9] COMPUSTAT also provides a variable called net quarterly sales, but using this instead of revenue results in nearly identical results.

[10] After discarding various uninformative strings such as 'INC'

used to test to what extent data breaches cause increased media coverage, as well as what effect on outcomes any such change in media coverage may have.

5 Results and Discussion

This section will begin with discussion of the results of the *firm study*, and will then conclude with discussion of the *stock market study*.

5.1 Firm Study

5.1.1 Revenue

Table 2.1 shows the primary specifications for the *firm study* with quarterly revenues as the outcome. Specification (1) shows the effect when no time trend is included. This is the average effect if the data breach on revenues during the entire event period, which in this table is one year (robustness to other event periods will be discussed soon). This coefficient is positive, but close to zero and insignificant. Specification (2) allows for a time trend. In this specification we observe that there is a significant immediate drop in revenues after the data breach as evidenced by the negative coefficient on β_1, but the positive value for β_2 implies that this loss in revenues recovers over the event period which is consistent with the conclusion of specification (1) that the average effect over the event window is close to zero. These effects are visualised in Figure 2.1. To put these effects into perspective, the immediate drop in revenues is about 20% of the average revenue in this sample. Since the unit of revenue is millions USD, this effect is about $1 billion which is much larger than estimate of total average costs as given by Ponemon-Institute (2018). This is likely because the effect is driven by large outliers, for example, firms that go out of business as a result of the data breach. As evidence of this, when the logarithm of revenue[11] is used as shown in Table 3.1 and Figure 3.1, no significant effect is observed.

In specification (3) controls are added for both the natural logarithm of the number of records leaked in a data breach and the Google searches index. Neither appears to have a strong impact on responses to data breaches, although the coefficient for records leaked is significantly positive at 10%[12]. Specification (4) includes interactions with dummies for which quartile a firm's revenue falls in in the first period in which it appears in the data. This is a proxy for the size of the firm. The coefficient on these interactions is decreasingly negative in the quartile of the firm's revenue which suggests that smaller firms suffer more punishment in terms of lost sales than larger firms. Finally, specification (5) includes interactions with dummies for the type of data lost in the data breach. These show that lost revenues are larger when "customer data" is lost — consistent with the consumer punishment mechanism — and in particular when social security numbers are lost. This seems reasonable because this is generally considered very sensitive information, the loss of which could be very costly for victims.

Table 2.2 shows the basic specification equivalent to specification (2) of Table 2.1 (no controls) for various lengths of event windows. The negative immediate effect on revenues appears to be robust to these various specifications. Note that the slope of the time trend is decreasing in the length of the event period. This is because firms seem to quickly return to

[11] Which gives a better estimate of the average effect

[12] However, I hesitate to conclude therefore that leaking more records results in less loss in revenue.

their previous level of revenue and stay there, so extending the event window simply causes the regression line to have a more shallow slope.

Within the *firm study* data there are a number of firms which have more than one data breach. Tables 2.3 and 2.4 consider whether a firm having multiple data breaches results in a heterogeneous response. Table 2.3 is the same as Table 2.2 however the sample is restricted to the first data breach for every firm. The effects are not significant, and the point estimates for the immediate effect (β_1) are much smaller. Table 2.4 shows the main regression for subsets based on the number of data breaches that a firm has in the sample period. While none of the coefficients are significant, the point estimates are increasingly negative in the number of breaches. These two tables provide some suggestive evidence that any consumer punishment effect is increasing in the number of data breaches a firm has. This is an intuitive result: consumers may readily forgive the first data breach as a random case of misfortune, but if it occurs many times consumers will start to believe that the firm is systematically failing to put in place mechanisms to protect their data.

Finally, Table 2.5 shows results when the event date is offset from the actual public announcement date. In this placebo test the event periods are one year and therefore do not overlap, so (as discussed in the methodology section) the coefficients are expected to not be significant for the offset event dates. However, this does not appear to be the case, so there may be issues with the first identifying assumption. Of particular concern is the significant coefficient for β_1 two years before the data breach. This suggests that there may be a negative trend before the actual data breach. However, when considering Figure 2.1 there seems to be little trend before the event, so the "no-trend" assumption is not obviously violated. Of less concern is the coefficient for β_1 for the event date three years after the actual announcement, which is even larger in magnitude than that for the actual event date. This is most likely to reflect the presence of lagged or "knock-on" effects of data breaches such as the example given earlier which was the CEO being fired.

Overall there seems to be some, but not definitive evidence that firm revenues are adversely affected by data breaches. This is the consumer boycott mechanism for market punishment. However, the effect is not robust to all specifications and there is a large degree of heterogeneity in these results based on the size of the firm, and the number of data breaches a firm has.

5.1.2 Expenses

Having discussed the interpretation of results from the *firm study* with firm revenues as the outcome, we now turn to the remaining outcomes which are expenses and net income. For these I will not go into depth in interpreting all of the coefficients as the logic is exactly the same as in the previous section. The specifications shown for these outcomes are the same and are shown in the same order in the appendix, and therefore the reader should now be equipped to interpret these. Therefore, I will simply draw attention to some important results without exhaustively discussing every table provided.

Turning first to non-operating income/expenses, Table 4.2 shows surprisingly that for at least short time periods there is a positive coefficient on the immediate effect of the data breach — implying in contrast to theoretical predictions that the data breach either causes an increase in non-operating income or a decrease in non-operating expenses[13]. However, these results are only significant at the 10% significance level. Even more surprisingly, this positive

[13] The outcome here is non-operating income net of non-operating expenses, so a positive coefficient actually implies that expenses are falling.

effect on non-operating income seems to be strongest for firms that have had many data breaches as shown in Table 4.4. In any case, no specifications imply a strong negative effect, so there seems to be no evidence for the legal expenses mechanism of market punishment. It is still possible that firms on average do indeed face substantial legal costs, however, they are transitory in nature, which makes them difficult to pick up given the temporal resolution available to this study.

For operating expenses, Table 5.2 shows that for a number of event window lengths there is some negative immediate effect which although insignificant, is consistent across specifications. Figure 5.1 shows no trend before the data breach, and clearly shows the discontinuous drop after. These findings support the conclusion that firms may be forced to downsize or otherwise scale back operations somewhat in the aftermath of a data breach — but only temporarily. In table 5.5, the only significant coefficient in this table is for the three years after offset suggesting that any such downsizing either comes as the result of "knock-on" or lagged effects rather than immediate effects of the data breach. In any case, there is certainly no evidence that firms make an effort to improve security or advertise more after a data breach, which would result in increased operating expenses. In fact, specification (1) of Table 7.1 shows that a proxy for advertising expenditure displays a significant *negative* discontinuous change of about half of its mean after the data breach.

5.1.3 Profits

The final outcome to consider for the *firm study* is net income. Negative immediate effects on both revenues and operating expenses have already been discussed, however, the effect on revenue seems to be larger in magnitude and more significant, so it is expected that there will be a net negative effect of data breaches on profits. This is indeed what is observed. Table 6.2 shows that there is a negative discontinuous effect of data breaches on profit which is significant and robust to different event window lengths. The effect is large, at around 90% of the average quarterly profit in the sample. However, note that this effect is not permanent, both because of the zero effect when no time trend is included, as well as the positive coefficient on the time trend when it is included. We can verify, by comparing Tables 2.2, 5.2, and 6.2 that the immediate effect on net income roughly equals the sum of the immediate effect on revenue and that on operating expenses. Table 6.1 mirrors the results for revenues in many ways: the loss in profits seems to not depend on media coverage as proxied by Google searches or the number of records leaked, but does depend on "consumer data" — especially social security numbers — being leaked. However in contrast to the effect on revenue, there is no strong evidence that the effect on profit varies heavily based on the size of the firm. When considering only the first data breach as shown in Table 6.3 the effect is much weaker, and Table 6.4 shows roughly that the effect is increasingly negative in the number of data breaches. This is consistent with the theoretical prediction, as well as the empirical observation for other outcomes that market punishment is stronger for firms that have multiple data breaches.

There are unfortunately some issues with these results. Table 6.5 shows that the placebo test may fail. The coefficients for event date before the data breach have much smaller point estimates than the actual event date, but they are still significant. Furthermore, Figure 6.1 fails to visually depict the same effect that is implied by the regression results — we should be able to see a lower intercept after the data breach. Nonetheless, this figure also depicts no negative trend before the data breach which will assuage concerns that profits were declining before the data breach, so the "no-trend" assumption is not obviously violated. As with other outcomes, the three years after event date gives a large coefficient, suggesting lagged or "knock-on" effects.

5.1.4 Other Outcomes

In addition to the main outcomes, I also ran the basic regression on a number of additional outcomes available in the COMPUSTAT database. Results are reported in Table 7.1. Here there is evidence that shareholders' equity fall in the wake of a data breach. This is consistent with the results of the *stock market study*. Interestingly, β_1 is negative with Google searches as the outcome. This is likely because any media effects are short lived, and have dissipated by the end of the month (the resolution of this data). Instead, what the regression picks up is a medium term decline in interest the firm consistent with the downsizing hypothesis and reduced advertising expenditure. The fact that data breaches do not appear to cause an increase in Google searches may explain why Google searches do not seem to be an important source of variation in the responses of other variables.

5.2 Stock Market Study

Results for the *stock market study* are given in section 8 of the appendix. Figures 8.1 and 8.2 plot the average CAR for data breaches over a 10 day event period for the CAPM and Fama-French models respectively. While the discontinuity is not completely clean at the event date, it is clear that CAR drops significantly over the event period. This conclusion is supported by the t-test shown in Tables 8.1, 8.2, 8.3 and, 8.4. The former two tables show the effect for the full sample, while the latter two show a sub-sample for which the largest change in CAR over the event period happened within two days of the data breach as discussed in the methodology section. For all specifications with event periods under 10 days CAR is significantly negative at at least the 10% level with most significant at the 5% level. So the finding that firms' stock returns fall in the wake of a data breach is very robust and consistent with previous literature on the subject. The Fama-French models provide lower p-values than the CAPM models which suggests that for these data the additional factors succeed in reducing noise.

When considering the longer event windows (90 and 180 days) the effects become less clear. The specifications using the *two day restriction* do not show significant results for these periods, however, this could well be the result of lack of power from the small sample size that they have access to (51 and 24 respectively). The unrestricted sample for both models shows a significantly negative CAR up to 90 days but not to 180. In particular, the point estimate for the 180 day event window in table 8.1 is very close to zero. Keeping in mind the qualifications about long event windows mentioned in the methodology section, there does not seem to be any evidence in the results of this study that the stock market effects of data breaches persist even six months after.

Finally, Table 8.5 shows the results of regressing the five day CAR from both models on some variables using simple OLS. The zero coefficient in specifications (1) and (4) indicates that there is no trend for the amount of stock market punishment to increase or decrease over the sample period. This contradicts the conclusion of Gordon et al. (2011). The nearly zero coefficients in specifications (2) and (5) indicates that the amount of stock market punishment does not depend strongly on the amount of records leaked[14]. Specifications (3) and (6) regress CAR on the types of data lost. There is only one significant coefficient, which is for credit card data, and this is only significant for one model, so it is most likely spurious. What is interesting to note here is the fact that CAR does not seem to react strongly to consumer data being leaked, even though as it has been already discussed most firm specific variables do depend strongly on this. This observation is evidence for the mechanism whereby stock returns fall because investors internalise information about the firm's infrastructure and management rather than changing their expectations of future earnings. In some sense, this finding is consistent with the results of the *firm study*, which showed that all of the negative effects of data breaches appear to be temporary, and therefore, any lost earnings would have only a nominal impact on net present value of future earnings.

6 Conclusion

This paper evaluates a number of mechanisms through which firms may incur losses as a result of suffering a data breach. Using a database of data breaches in the Unites States spanning the years 2005-2017, effects on revenues, expenses, profits and stock returns were tested. There is some evidence that consumers will punish firms for leaking their private data, especially those firms that do so repeatedly, and smaller firms. However, these results are noisy and seem to be driven by some large outliers, because the average effects do not seem to be significant. On the other hand, there is no evidence that firms systematically increase advertising, security, or other expenditures as a result of a data breach. Likewise, there is no

[14] However note that it is possible that the scale of a data breach may not always be understood five days after the announcement.

evidence that data breaches result in large non-operating expenses such as legal expenses. The fact that instead, operating and particularly advertising expenses may decrease after a data breach seems to be indicative of temporary downsizing or cutbacks. Consistent with the observation that the negative effect on revenues is larger than that on operating expenses, there is evidence that net income falls, but only temporarily in the wake of a data breach.

In contrast to the lukewarm response of variables in the *firm study*, the *stock market study* demonstrated a strong negative response in firms' stock returns to data breaches. The strength of this response does not seem to be changing over time, nor does it appear to depend on the severity of the data breach. Furthermore, the stock market response seems to be driven by the information about the firm that the data breach reveals, rather than any future losses that the firm may face. Taken together the negative effects of a data breach seem sufficient to justify the fall in stock market value, so this does not seem to represent an arbitrage opportunity.

Considering these conclusions in a broader context, the results of this paper imply that firms do — to a limited degree — face corrective action from market forces for leaking data. However, these effects do not seem to be especially strong, and there is no evidence that firms actually respond by taking preventative or remedial action. This does not necessarily mean that there is market failure as forecast by Roberds and Schreft (2009). It is possible that firms choose not to invest in increased security, advertising, or other measures because such measures are in fact to be ineffectual, or because firms already allocated an efficient amount of spending to these areas before the data breach. Finally, it is possible that firms do indeed increase investment in these measures, but that the increase in spending is proportionally so small that it does not stand out against noise in this study.

This paper illuminates a number of avenues for future research. Firstly, there is a need for research into factors which predict the risk for data breaches. While this paper attempted to skirt around the issue of endogeneity of data breaches, if it is possible to identify an exogenous source of variation in the risk for data breaches it would allow studies like this one to more strongly identify their results. Secondly, in order to evaluate the possibility of market failure, further research will be required to understand the preventative and remediation strategies available to firms, and their cost effectiveness[15]. Finally, in order to provide strong evidence that governments should step in to ensure privacy protections, future studies should attempt to estimate not only the direct cost but also the broader social costs of data breaches including all stakeholders such as firms, consumers, and public institutions, as this is necessary to calculate the socially optimal amount of security investment.

[15] I put this second because conducting this research may first require a strong understanding of the risk factors.

References

Acquisti, A., Friedman, A., and Telang, R. (2006). Is there a cost to privacy breaches? an event study. *ICIS 2006 Proceedings*, page 94.

Berger, J., Sorensen, A. T., and Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29(5):815–827.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business (2019). Daily stock file (2005-2017). data retreived from https://wrds-web.wharton.upenn.edu/wrds.

Davis, G., Garcia, A., and Zhang, W. (2009). Empirical analysis of the effects of cyber security incidents. *Risk Analysis: An International Journal*, 29(9):1304–1316.

Edwards, B., Hofmeyr, S., and Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14.

Facebook (2018). Security update [press release].

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22.

French, K. (2018). Fama-french 5 research factors (2005-2017). data retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Goel, S. and Shawky, H. A. (2009). Estimating the market impact of security breach announcements on firm values. *Information & Management*, 46(7):404–410.

Google (2018). Google trends data (2005-2017). data retrieved from https://trends.google.com/trends/.

Gordon, L. A., Loeb, M. P., and Zhou, L. (2011). The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security*, 19(1):33–56.

Livingston, J. A. (2005). How valuable is a good reputation? a sample selection model of internet auctions. *Review of Economics and Statistics*, 87(3):453–465.

Morse, E. A., Raval, V., and Wingender Jr, J. R. (2011). Market price effects of data security breaches. *Information Security Journal: A Global Perspective*, 20(6):263–273.

NCSL (2018). Security breach notification laws [press release].

Ponemon-Institute (2018). Cost of a data breach study.

Privacy-Rights-Clearinghouse (2018). Data breaches database (2005-2017). data retrieved from https://www.privacyrights.org/.

Roberds, W. and Schreft, S. L. (2009). Data breaches and identity theft. *Journal of Monetary Economics*, 56(7):918–929.

Romanosky, S., Telang, R., and Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2).

Rosati, P., Deeney, P., Cummins, M., Van der Werff, L., and Lynn, T. (2019). Social media and stock price reaction to data breach announcements: Evidence from us listed companies. *Research in International Business and Finance*, 47:458–469.

Standard & Poor's Compustat Services (2019). Quarterly data (2005-2017). data retreived from https://wrds-web.wharton.upenn.edu/wrds.

Tables and Figures

1 Summary Statistics

Table 1.1: Summary statistics for financial variables

Statistic	N	Mean	St. Dev.
Net income (Millions USD)	20,736	392.4	1,690.8
Revenue (Millions USD)	17,855	4,991.9	11,374.3
Operation Expenses (Millions USD)	20,663	3,848.5	9,512.7
Non-Operating Income (Net NO Expenses) (Millions USD)	20,497	−47.7	620.9
Google Trends Index (Company Name)	19,596	24.8	27.7
Google Trends Index (Company Ticker)	20,359	31.1	28.5

Table 1.2: Summary of types of data loss

Statistic	Mean
Customer	320
Credit Card	148
Social Security Number	353
Name	423
Address	268
Total Breaches	759

Table 1.3: Summary of magnitude of data loss

Statistic	N	Mean	Min	Max	St. Dev.
Records leaked per breach	20,796	1,889,737	0	167,000,000	13,959,847

Figure 1.1: Histogram of number of records leaked per breach

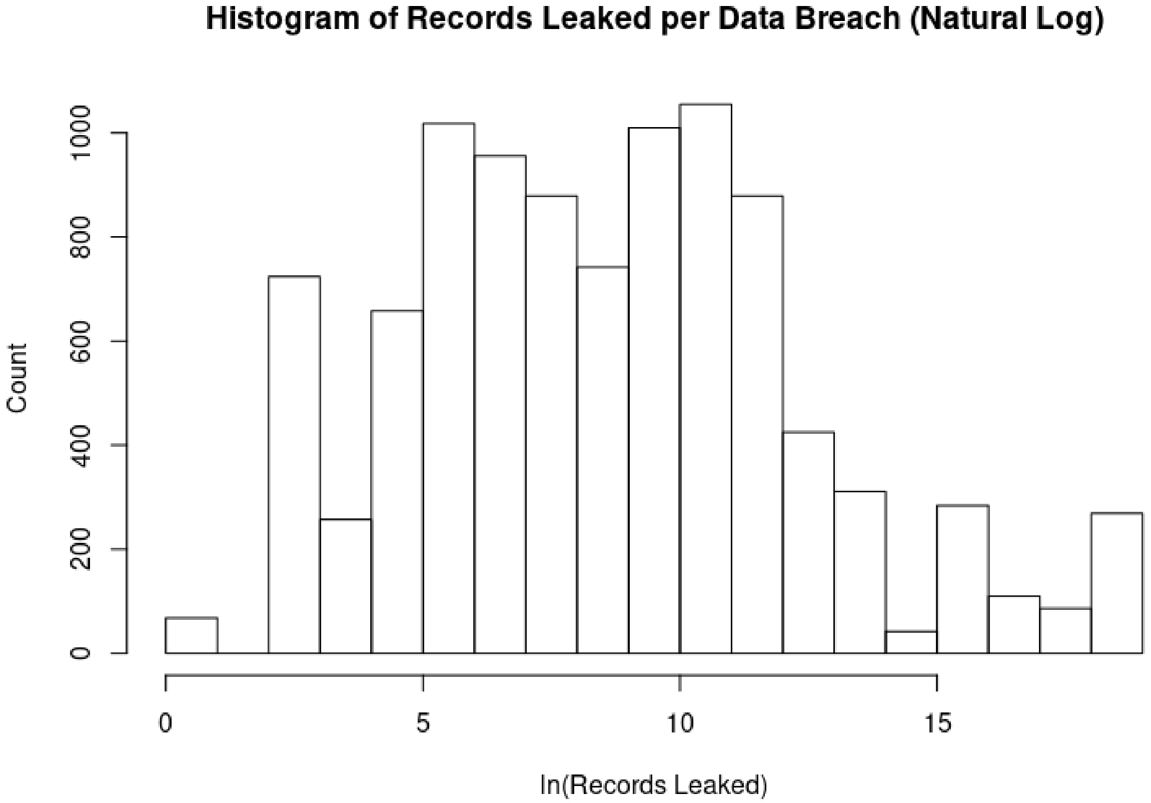


Table 1.4: Summary of stock market data

Statistic	N	Mean	St. Dev.
Daily Firm Return	936,151	0.0005	0.03
Value Weighted Market Return	936,404	0.0003	0.01
Risk Free Market Return	936,404	0.04	1.19
SMB factor	936,404	0.004	0.57
HML factor	936,404	0.001	0.65

2 Revenue

Table 2.1: Revenue: main specification and controls

	Dependent variable:				
	Revenue				
	(1)	(2)	(3)	(4)	(5)
After Breach	90.126	−1,075.740*	−1,590.337**	−25.749	−167.876
	(276.882)	(584.361)	(702.747)	(821.692)	(550.014)
After Breach x Quarter		44.540**	52.681**	47.527**	44.146**
		(18.645)	(20.761)	(19.217)	(18.957)
Records Leaked (log) x After Breach			42.130*		
			(24.940)		
Google Search Index			15.579		
			(9.899)		
Google Search Index x After Breach			11.765		
			(10.769)		
After x Revenue Quartile 1				−2,045.009**	
				(795.993)	
After x Revenue Quartile 2				−1,660.845**	
				(804.454)	
After x Revenue Quartile 3				−1,313.745*	
				(773.529)	
Customer Data Leaked x After breach					−1,169.773***
					(423.595)
Credit Card Leaked x After breach					−18.776
					(316.451)
SSN Leaked x After breach					−852.194*
					(478.507)
Name Leaked x After breach					−344.489
					(384.326)
Address Leaked x After breach					607.888**
					(293.087)
Dependent Mean	5149.94	5149.94	5149.94	5149.94	5149.94
Dependent SD	11548.84	11548.84	11548.84	11548.84	11548.84
Observations	10,574	10,574	10,023	10,574	10,574
R^2	0.943	0.943	0.942	0.944	0.944
Adjusted R^2	0.940	0.941	0.940	0.941	0.941

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 2.2: Revenue: various event period lengths (robustness check)

	Dependent variable:			
	Revenue (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
After Breach	−915.479	−1,075.740*	−1,070.575*	−1,000.952*
	(561.245)	(584.361)	(595.469)	(591.975)
After Breach x Quarter	44.145**	44.540**	38.889**	34.038*
	(18.661)	(18.645)	(18.954)	(18.679)
Dependent Mean	4969.33	5149.94	5285.29	5340.45
Dependent SD	11224.58	11548.84	11839.15	12066.49
Observations	9,623	10,574	12,114	13,369
R^2	0.944	0.943	0.940	0.938
Adjusted R^2	0.941	0.941	0.937	0.936

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

Table 2.3: Revenue: first data breaches for each firm (robustness check)

	Dependent variable:			
	Revenue (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
After Breach	−179.939	−220.091	−235.265	−171.523
	(293.063)	(306.158)	(282.210)	(278.508)
After Breach x Quarter	10.889	13.349	14.222*	12.265
	(8.186)	(8.871)	(8.380)	(9.125)
Dependent Mean	3179.51	3227.27	3280.89	3261.11
Dependent SD	7995.85	8143.4	8456.55	8621.45
Observations	7,432	8,035	9,077	9,941
R^2	0.955	0.954	0.955	0.952
Adjusted R^2	0.953	0.952	0.952	0.949

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 2.4: Revenue: breakdown by number of data breaches a firm has

	Dependent variable:			
	Revenue (Number of Breaches for Firm)			
	(1)	(2)	(3)	(4+)
After Breach	−125.512	−479.309	−321.079	−624.802
	(157.839)	(916.204)	(438.449)	(1,884.327)
After Breach x Quarter	9.659	5.170	18.815	82.551
	(6.522)	(23.129)	(12.656)	(52.310)
Number of Data Breaches	231	139	85	176
Dependent Mean	1933.71	7302.19	5438.41	15062.09
Dependent SD	5942.93	14370.67	5799.58	18875.12
Observations	5,826	2,348	1,050	1,350
R^2	0.955	0.940	0.926	0.929
Adjusted R^2	0.952	0.937	0.920	0.925

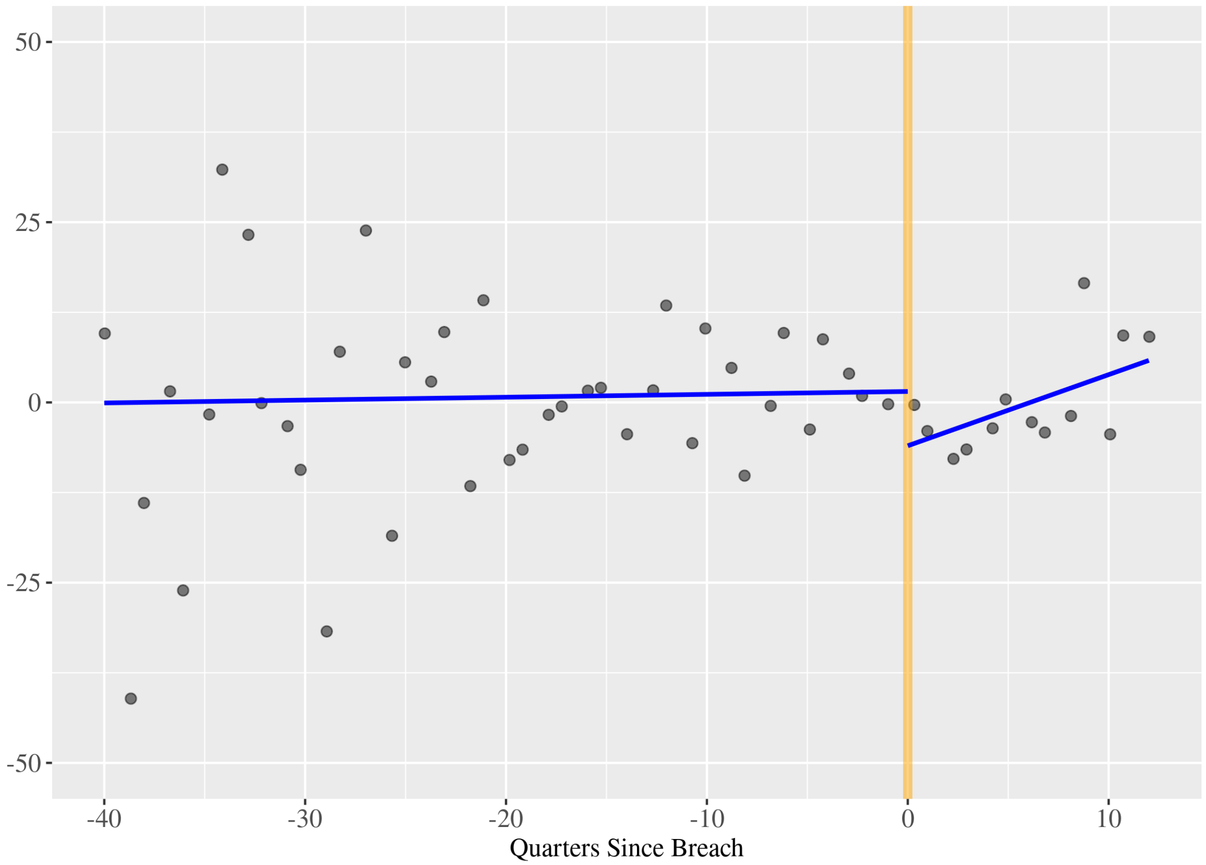
Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 2.5: Revenue: offset event dates (robustness check)

	Dependent variable:				
	Revenue (Offset of Event Date)				
	(0)	(-2 Years)	(-3 Years)	(+2 Years)	(+3 Years)
After Breach	$-1{,}075.740^{*}$	-729.132^{*}	-751.626	-708.919	$-1{,}151.524^{**}$
	(584.361)	(424.132)	(752.797)	(578.295)	(543.060)
After Breach x Quarter	44.540^{**}	27.944^{*}	30.224	15.326	19.878
	(18.645)	(16.653)	(32.716)	(19.200)	(17.307)
Dependent Mean	5149.94	5149.94	5149.94	5149.94	5149.94
Dependent SD	11548.84	11548.84	11548.84	11548.84	11548.84
Observations	10,574	6,153	4,519	13,369	14,437
R^2	0.943	0.957	0.957	0.938	0.937
Adjusted R^2	0.941	0.954	0.954	0.936	0.935

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Figure 2.1: Mean residual revenue (fixed effects removed)



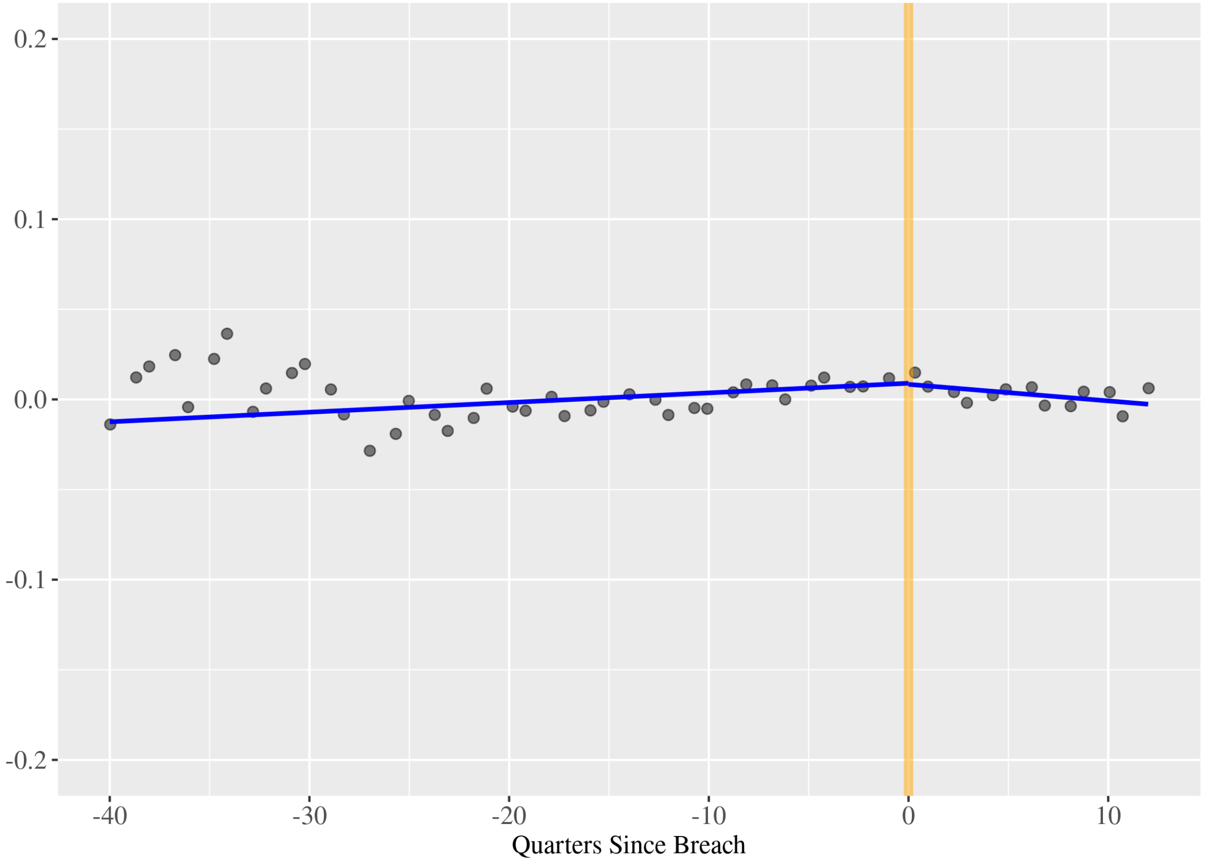
3 Log Revenue

Table 3.1: Log Revenue: main specification and controls

	Dependent variable:			
	Log Revenue (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
	(1)	(2)	(3)	(4)
After Breach	0.006	−0.006	−0.004	0.019
	(0.040)	(0.043)	(0.047)	(0.046)
After Breach x Quarter	0.001	0.001	0.001	0.00002
	(0.001)	(0.002)	(0.002)	(0.002)
Dependent Mean	6.92	6.96	6.99	6.99
Dependent SD	2.3	2.28	2.25	2.25
Observations	9,623	10,574	12,114	13,369
R^2	0.966	0.964	0.961	0.953
Adjusted R^2	0.965	0.962	0.959	0.951

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Figure 3.1: Mean residual log revenue (fixed effects removed)



4 Non-Operating Income/Expenses

Table 4.1: Non-operating income: main specification and controls

	Dependent variable:				
	Non-Operating Income — Net of NO Expenses				
	(1)	(2)	(3)	(4)	(5)
After Breach	13.472	101.301*	112.690*	110.329*	88.530*
	(11.547)	(59.933)	(66.504)	(58.722)	(52.293)
After Breach x Quarter		−3.381*	−3.708*	−3.342*	−3.231*
		(1.930)	(2.157)	(1.947)	(1.844)
Records Leaked (log) x After Breach			−1.635		
			(1.402)		
Google Search Index			−0.834		
			(0.736)		
Google Search Index x After Breach			−0.178		
			(0.313)		
After x Revenue Quartile 1				−25.918	
				(28.945)	
After x Revenue Quartile 2				−8.147	
				(20.196)	
After x Revenue Quartile 3				−2.787	
				(20.822)	
Customer Data Leaked x After breach					−50.464*
					(26.211)
Credit Card Leaked x After breach					42.168
					(27.045)
SSN Leaked x After breach					0.377
					(13.176)
Name Leaked x After breach					7.775
					(16.626)
Address Leaked x After breach					53.018**
					(25.807)
Dependent Mean	-65.91	-65.91	-65.91	-65.91	-65.91
Dependent SD	717.84	717.84	717.84	717.84	717.84
Observations	11,833	11,833	11,241	11,833	11,833
R^2	0.904	0.905	0.911	0.905	0.905
Adjusted R^2	0.900	0.901	0.907	0.901	0.901

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 4.2: Non-operating income: Various event period lengths (robustness check)

	Dependent variable:			
	Non-Operating Income — Net of NO Expenses (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
	(1)	(2)	(3)	(4)
After Breach	115.521^{*}	101.301^{*}	73.271	66.795
	(66.091)	(59.933)	(52.060)	(47.986)
After Breach x Quarter	-3.889^{*}	-3.381^{*}	-2.463	-2.192
	(2.235)	(1.930)	(1.623)	(1.476)
Dependent Mean	-63.22	-65.91	-64.89	-61.75
Dependent SD	706.92	717.84	715.08	701.14
Observations	10,757	11,833	13,603	15,089
R^2	0.902	0.905	0.893	0.892
Adjusted R^2	0.898	0.901	0.889	0.888

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 4.3: Non-operating income: first data breaches for each firm (robustness check)

	Dependent variable:			
	Non-Operating Income — Net of NO Expenses (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
	(1)	(2)	(3)	(4)
After Breach	−4.001	−1.470	−6.028	−2.050
	(16.809)	(15.982)	(16.026)	(15.697)
After Breach x Quarter	−0.288	−0.134	−0.062	−0.176
	(0.489)	(0.449)	(0.443)	(0.452)
Dependent Mean	3.37	3.08	3.1	2.85
Dependent SD	314.88	324.88	329.88	326.7
Observations	8,303	8,974	10,157	11,173
R^2	0.870	0.879	0.802	0.787
Adjusted R^2	0.862	0.873	0.792	0.778

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 4.4: Non-operating income: breakdown by number of data breaches a firm has

	Dependent variable:			
	Non-Operating Income — Net of NO Expenses (Number of Breaches for Firm)			
	(1)	(2)	(3)	(4+)
After Breach	−12.411	−9.012	20.905	278.056
	(8.119)	(25.436)	(31.826)	(173.812)
After Breach x Quarter	0.380	−0.298	−0.976	−8.247
	(0.238)	(0.452)	(1.108)	(5.477)
Number of Data Breaches	255	158	85	210
Dependent Mean	6.62	41.86	43.42	-615.51
Dependent SD	160.95	382.08	121.59	1765.81
Observations	6,518	2,670	1,052	1,593
R^2	0.725	0.820	0.455	0.917
Adjusted R^2	0.711	0.810	0.409	0.912

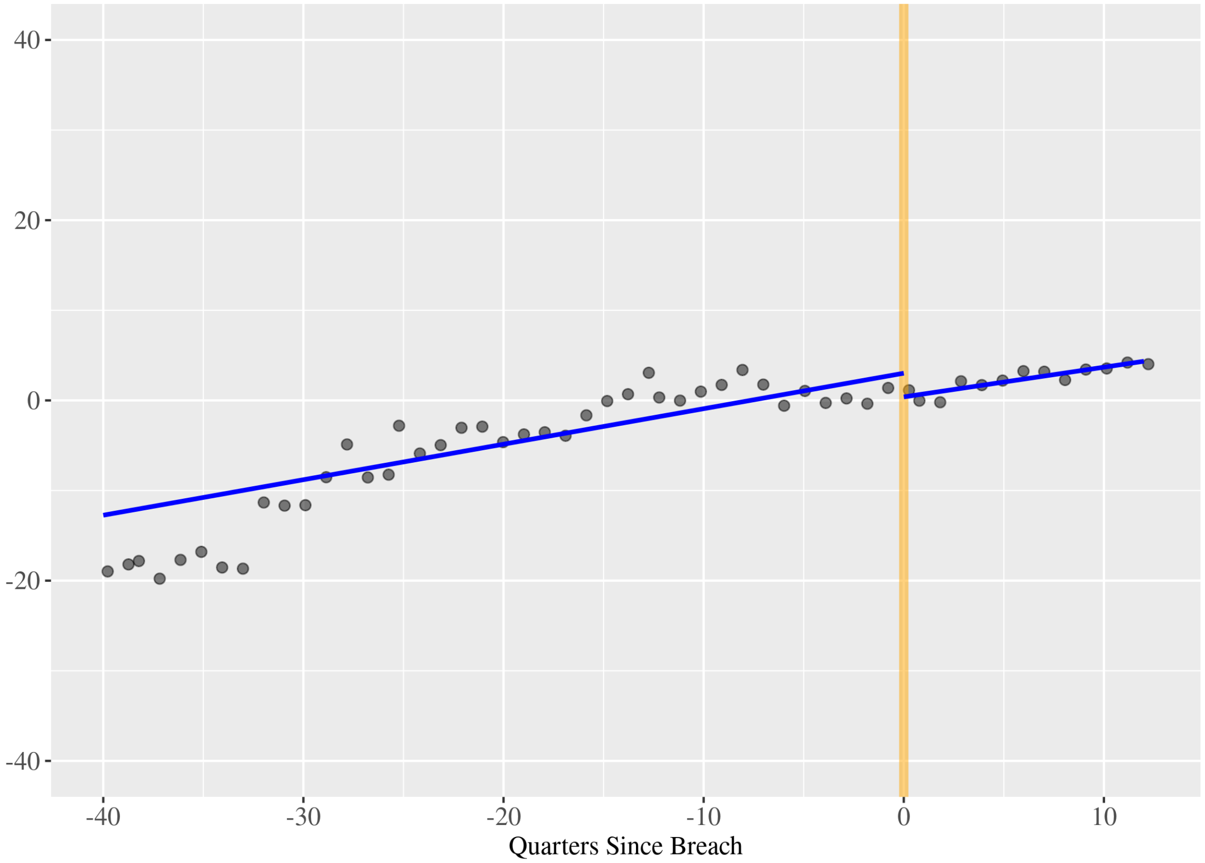
Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 4.5: Non-operating income: offset event dates (robustness check)

	Dependent variable:				
	Non-Operating Income — Net of NO Expenses (offset of event date)				
	(0)	(-2 Years)	(-3 Years)	(+2 Years)	(+3 Years)
After Breach	101.301^{*}	10.626	-41.256	8.116	17.646
	(59.933)	(15.892)	(36.421)	(35.636)	(33.804)
After Breach x Quarter	-3.381^{*}	-0.368	1.650	0.274	-0.120
	(1.930)	(0.462)	(1.526)	(0.952)	(0.938)
Dependent Mean	-65.91	-4.02	4.06	-61.75	-57.65
Dependent SD	717.84	389.3	302.53	701.14	684.37
Observations	11,833	6,796	4,991	15,089	16,357
R^2	0.905	0.907	0.872	0.892	0.885
Adjusted R^2	0.901	0.900	0.862	0.888	0.882

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Figure 4.1: Mean residual non-operating income — net of NO expenses (fixed effects removed)



5 Operating Expenses

Table 5.1: Operating expenses: main specification and controls

	Dependent variable:				
	Operating Expenses				
	(1)	(2)	(3)	(4)	(5)
After Breach	57.473 (197.373)	−709.532 (463.608)	−1,044.119* (553.269)	290.304 (642.460)	−37.672 (392.606)
After Breach x Quarter		29.528* (15.240)	34.846** (17.650)	30.933** (15.483)	29.259* (15.491)
Records Leaked (log) x After Breach			31.075 (19.970)		
Google Search Index			12.429* (6.904)		
Google Search Index x After Breach			7.622 (6.637)		
After x Revenue Quartile 1				−1,700.488*** (562.244)	
After x Revenue Quartile 2				−1,330.164** (566.521)	
After x Revenue Quartile 3				−1,161.428** (548.598)	
Customer Data Leaked x After breach					−797.924*** (306.535)
Credit Card Leaked x After breach					−78.503 (242.769)
SSN Leaked x After breach					−485.310 (306.365)
Name Leaked x After breach					−482.602 (316.239)
Address Leaked x After breach					618.402** (246.422)
Dependent Mean	4214.72	4214.72	4214.72	4214.72	4214.72
Dependent SD	10047.91	10047.91	10047.91	10047.91	10047.91
Observations	11,899	11,899	11,309	11,899	11,899
R^2	0.945	0.946	0.946	0.946	0.946
Adjusted R^2	0.943	0.943	0.943	0.944	0.944

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 5.2: Operating expenses: various event period lengths (robustness check)

	Dependent variable:			
	Operating Expenses (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
After Breach	−641.238	−709.532	−684.455	−611.873
	(429.111)	(463.608)	(483.710)	(486.512)
After Breach x Quarter	30.817**	29.528*	25.035	21.037
	(14.763)	(15.240)	(15.687)	(15.651)
Dependent Mean	4214.72	4214.72	4214.72	4214.72
Dependent SD	10047.91	10047.91	10047.91	10047.91
Observations	10,820	11,899	13,686	15,190
R^2	0.946	0.946	0.943	0.939
Adjusted R^2	0.944	0.943	0.941	0.937

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 5.3: Operating expenses: first data breaches for each firm (robustness check)

	Dependent variable:			
	Operating Expenses (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
After Breach	−395.290	−409.098	−419.899*	−319.190
	(269.070)	(264.033)	(242.161)	(238.595)
After Breach x Quarter	15.983**	17.202**	17.924**	14.280*
	(8.095)	(8.141)	(7.647)	(8.051)
Dependent Mean	2443.43	2480.56	2525.9	2506.17
Dependent SD	6840.51	6967.15	7280.56	7399.64
Observations	8,350	9,019	10,207	11,229
R^2	0.951	0.950	0.951	0.942
Adjusted R^2	0.948	0.948	0.949	0.939

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

Table 5.4: Operating expenses: breakdown by number of data breaches a firm has

	Dependent variable:			
	Operating Expenses (Number of Breaches for Firm)			
	(1)	(2)	(3)	(4+)
After Breach	−115.905	13.270	−298.018	−925.356
	(135.296)	(719.891)	(342.430)	(1,439.271)
After Breach x Quarter	8.364	−10.026	18.026*	70.763
	(5.631)	(20.137)	(9.306)	(45.528)
Number of Data Breaches	255	158	85	210
Dependent Mean	1487.61	5293.86	4243.98	12907.84
Dependent SD	5034.45	11605.58	4796.78	16230.26
Observations	6,533	2,726	1,050	1,590
R^2	0.951	0.943	0.951	0.930
Adjusted R^2	0.948	0.940	0.947	0.925

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors clustered at the company level

Company and quarter fixed effects in all specifications

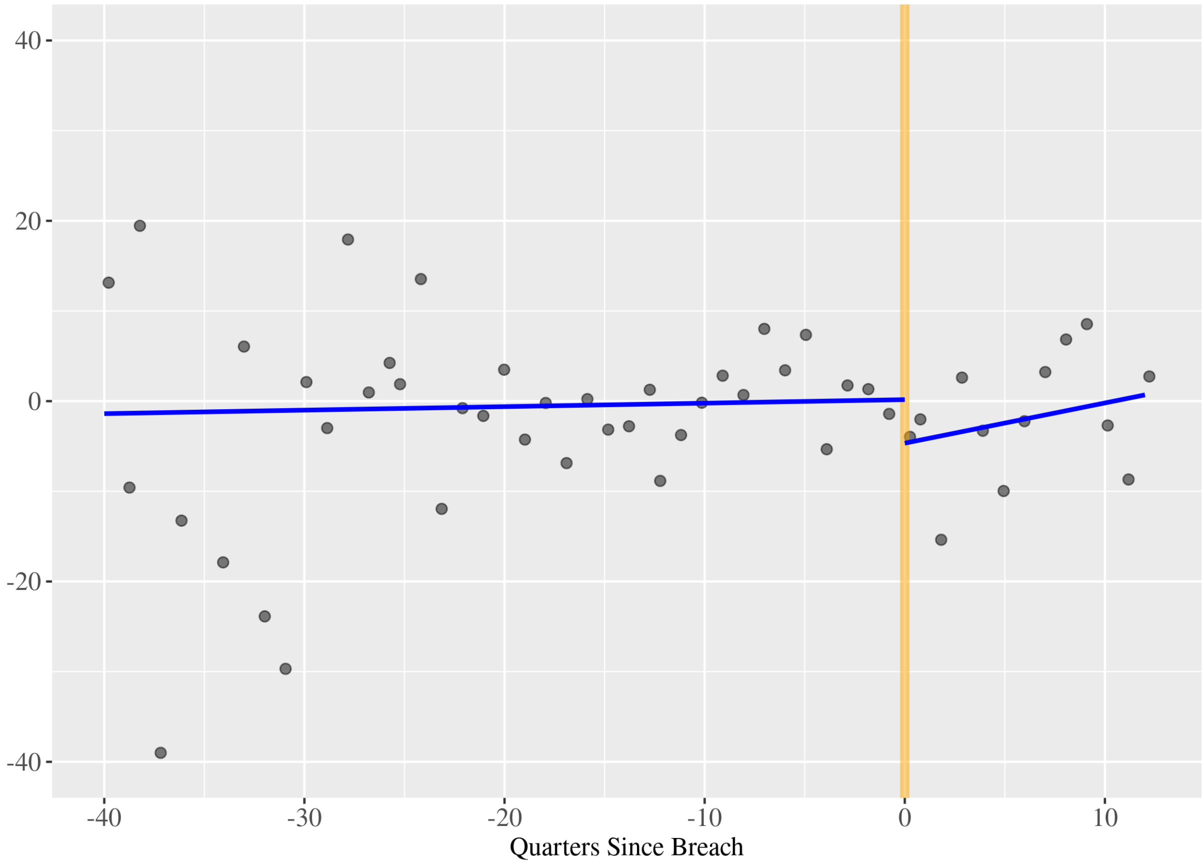
Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 5.5: Operating Expenses: Offset event dates (robustness check)

	Dependent variable:				
	Operating Expenses (Offset of Event Date)				
	(0)	(-2 Years)	(-3 Years)	(+2 Years)	(+3 Years)
After Breach	−684.455	−407.196	−424.522	−687.913	−964.997**
	(483.710)	(327.361)	(648.659)	(482.451)	(414.988)
After Breach x Quarter	25.035	16.132	20.579	15.328	19.619
	(15.687)	(12.365)	(27.597)	(16.063)	(13.684)
Dependent Mean	4214.72	4214.72	4214.72	4214.72	4214.72
Dependent SD	10047.91	10047.91	10047.91	10047.91	10047.91
Observations	13,686	6,836	5,016	15,190	16,471
R^2	0.943	0.956	0.955	0.939	0.939
Adjusted R^2	0.941	0.953	0.951	0.937	0.937

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Figure 5.1: Mean residual operating expenses (fixed effects removed)



6 Net Income

Table 6.1: Net income: main specification and controls

	Dependent variable:				
	Net Income				
	(1)	(2)	(3)	(4)	(5)
After Breach	−86.377 (66.703)	−390.372** (171.958)	−453.561** (206.152)	−383.633 (290.527)	−94.755 (146.636)
After Breach x Quarter		11.697** (4.871)	12.728** (5.103)	11.893** (4.850)	11.997** (4.842)
Records Leaked (log) x After Breach			2.999 (5.099)		
Google Search Index			−1.874 (1.954)		
Google Search Index x After Breach			1.279 (2.623)		
After x Revenue Quartile 1				−83.509 (226.990)	
After x Revenue Quartile 2				16.797 (217.332)	
After x Revenue Quartile 3				49.027 (214.735)	
Customer Data Leaked x After breach					−362.903*** (100.321)
Credit Card Leaked x After breach					75.753 (99.050)
SSN Leaked x After breach					−247.812** (115.114)
Name Leaked x After breach					10.909 (83.450)
Address Leaked x After breach					−91.312 (75.771)
Dependent Mean	427.34	427.34	427.34	427.34	427.34
Dependent SD	1816.55	1816.55	1816.55	1816.55	1816.55
Observations	11,952	11,952	11,359	11,952	11,952
R^2	0.290	0.291	0.288	0.291	0.294
Adjusted R^2	0.262	0.263	0.259	0.262	0.265

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 6.2: Net income: various event period lengths (robustness check)

	Dependent variable:			
	Net Income (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
After Breach	-371.171^{*}	-390.372^{**}	-410.271^{**}	-422.288^{**}
	(192.567)	(171.958)	(169.343)	(165.827)
After Breach x Quarter	12.304^{**}	11.697^{**}	11.839^{**}	11.604^{***}
	(5.570)	(4.871)	(4.649)	(4.446)
Dependent Mean	419.67	427.34	432.49	427.88
Dependent SD	1818.62	1816.55	1790.5	1839.08
Observations	10,862	11,952	13,743	15,247
R^2	0.282	0.291	0.307	0.309
Adjusted R^2	0.250	0.263	0.283	0.287

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 6.3: Net income: first data breaches for each firm (robustness check)

	Dependent variable:			
	Net Income (Event Period)			
	(6 Months)	(1 Year)	(2 Years)	(3 Years)
	(1)	(2)	(3)	(4)
After Breach	−53.339	−41.920	−31.425	−3.151
	(106.911)	(94.414)	(87.813)	(89.715)
After Breach x Quarter	1.726	1.490	1.333	0.157
	(2.702)	(2.448)	(2.298)	(2.533)
Dependent Mean	229.02	231.74	226.05	210.84
Dependent SD	914.37	925.65	946.25	1123.7
Observations	8,371	9,046	10,234	11,256
R^2	0.579	0.585	0.540	0.400
Adjusted R^2	0.555	0.562	0.517	0.374

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications

Table 6.4: Net income: breakdown by number of data breaches a firm has

	Dependent variable:			
	Net Income (Number of Breaches for Firm)			
	(1)	(2)	(3)	(4+)
After Breach	−74.962	−258.433	20.965	−712.170
	(122.444)	(217.311)	(233.661)	(611.408)
After Breach x Quarter	2.261	6.755	−0.261	22.401
	(3.132)	(5.614)	(7.770)	(16.178)
Number of Data Breaches	255	158	85	210
Dependent Mean	109.83	443.13	544.93	1609.38
Dependent SD	680.93	1200.23	1157.95	4170.13
Observations	6,549	2,733	1,052	1,618
R^2	0.347	0.657	0.400	0.196
Adjusted R^2	0.314	0.640	0.348	0.150

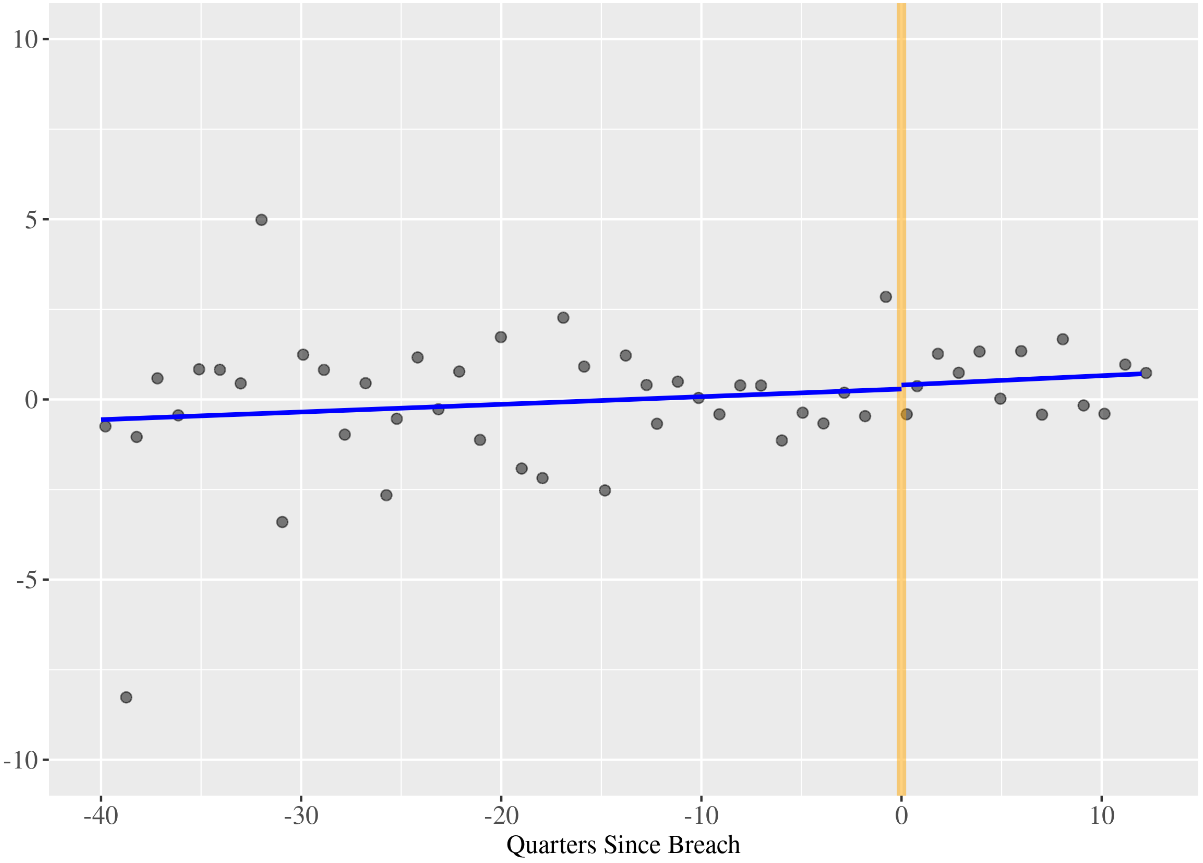
Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Table 6.5: Net income: offset event dates (robustness check)

	Dependent variable:				
	Net Income (Offset of Event Date)				
	(0)	(-2 Years)	(-3 Years)	(+2 Years)	(+3 Years)
After Breach	-390.372^{**}	-136.829^{*}	-103.339^{*}	-574.147^{*}	-336.764^{*}
	(171.958)	(76.415)	(53.155)	(294.070)	(200.623)
After Breach x Quarter	11.697^{**}	4.395	4.208^{*}	15.567^{**}	8.819
	(4.871)	(2.784)	(2.327)	(7.746)	(6.164)
Dependent Mean	427.34	427.34	427.34	427.34	427.34
Dependent SD	1816.55	1816.55	1816.55	1816.55	1816.55
Observations	11,952	6,852	5,029	15,247	16,533
R^2	0.291	0.732	0.743	0.308	0.311
Adjusted R^2	0.263	0.714	0.723	0.286	0.291

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the company level
Company and quarter fixed effects in all specifications
Prediction period is up to 10 years before breach, and event period up to 1 year after

Figure 6.1: Mean residual profit (fixed effects removed)



7 Other Outcomes

Table 7.1: Other outcomes: main specification

	Dependent variable:					
	Sales, General and Other Expenses	Total Shareholders' Equity	Earnings per Share (Basic)	Number of Employees	Google Searches (Company Name)	Google Searches (Stock Ticker)
	(1)	(2)	(3)	(4)	(5)	(6)
After Breach	−426.048***	−6,758.839**	−6.142	−2.253	−2.136*	−1.610
	(157.999)	(2,672.161)	(5.979)	(6.006)	(1.171)	(1.162)
After Breach x Quarters Since Breach	14.497***	223.119***	0.121	0.153	0.052	0.029
	(5.098)	(81.262)	(0.115)	(0.188)	(0.039)	(0.040)
Dependent Mean	1045.61	14123.25	3.03	63.31	26.12	30.44
Dependent SD	2255.5	34818.35	72.91	154.54	27.84	28.34
Observations	9,430	11,914	11,771	6,357	11,401	11,866
R^2	0.946	0.927	0.240	0.990	0.865	0.858
Adjusted R^2	0.944	0.924	0.209	0.989	0.859	0.853

Note: *p<0.1; **p<0.05; ***p<0.01

Standard Errors clustered at the quarter level

Company and quarter fixed effects in all specifications

Prediction period is up to 10 years before breach, and event period 1 year after

8 Stock Market

Figure 8.1: Capital Asset Pricing Model stock market event study

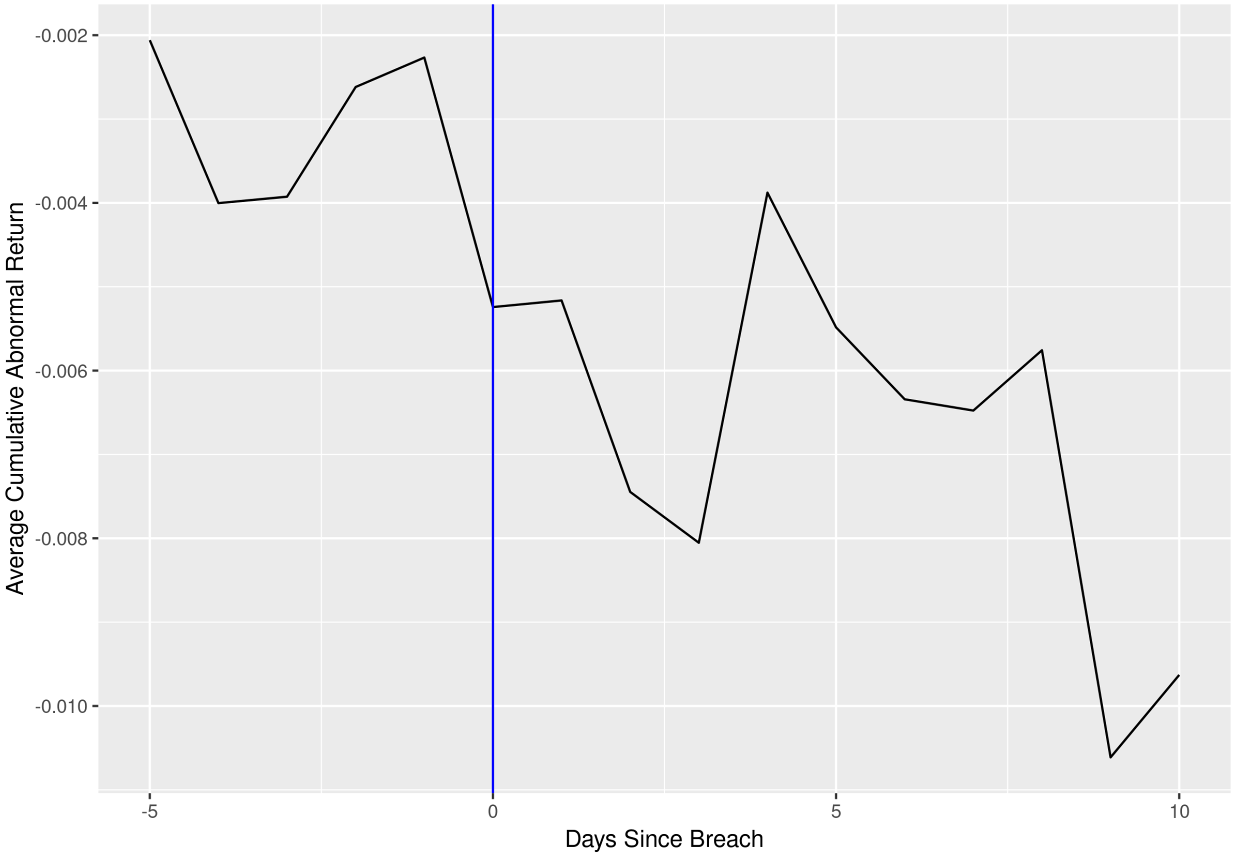


Figure 8.2: Fama-French Model stock market event study

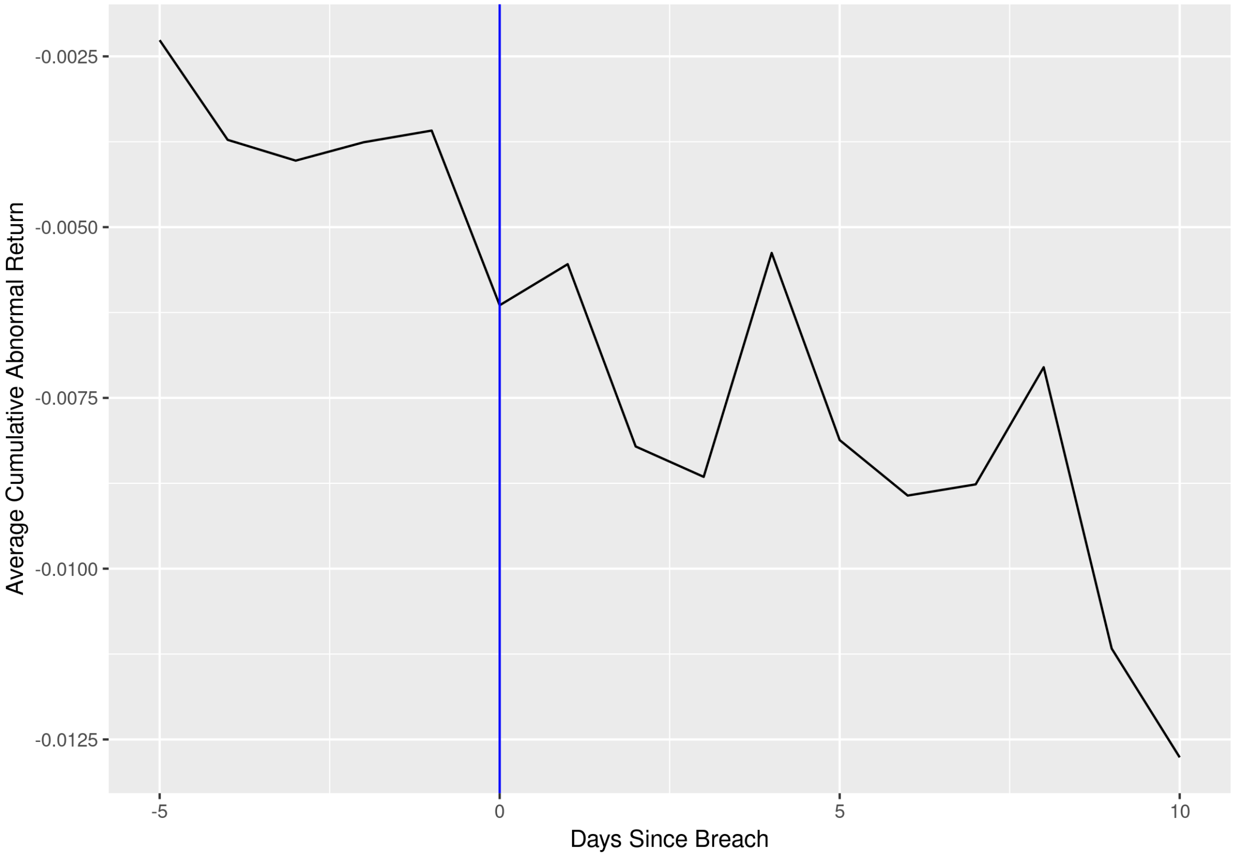


Table 8.1: CAR t-test (CAPM Model)

Event Window	2 days	5 days	10 days	90 days	180 days
N	473	473	469	439	409
Mean	-0.44	-0.54	-0.45	-2.03	0.05
SD	0.23	0.26	0.31	0.98	2.00
t-stat	-1.95	-2.08	-1.49	-2.07	0.02
p	0.03**	0.02**	0.07*	0.02**	0.51

Note: *p<0.1; **p<0.05; ***p<0.01
Returns given as percentages

Table 8.2: CAR t-test (Fama -French Model)

Event Window	2 days	5 days	10 days	90 days	180 days
N	473	473	469	439	409
Mean	-0.57	-0.71	-0.64	-2.54	-1.30
SD	0.22	0.25	0.30	0.93	1.97
t-stat	-3.00	-3.00	-2.00	-3.00	-1.00
p	0.01**	0.00***	0.02**	0.00***	0.26

Note: *p<0.1; **p<0.05; ***p<0.01
Returns given as percentages

Table 8.3: CAR t-test (CAPM) (with two days restriction)

Event Window	2 days	5 days	10 days	90 days	180 days
N	438	372	275	51	24
Mean	-0.51	-0.67	-0.59	-3.87	0.33
SD	0.22	0.28	0.39	3.25	6.82
t-stat	-2.26	-2.39	-1.52	-1.19	0.05
p	0.01**	0.01**	0.06*	0.12	0.52

Note: *p<0.1; **p<0.05; ***p<0.01
Returns given as percentages

Table 8.4: CAR t-test (Fama-French) (with two days restriction)

Event Window	2 days	5 days	10 days	90 days	180 days
N	438	372	275	51	24
Mean	-0.65	-0.88	-0.84	-3.71	-1.38
SD	0.22	0.27	0.37	3.23	6.11
t-stat	-3.00	-3.00	-2.00	-1.00	0.00
p	0.00***	0.00***	0.01**	0.13	0.41

Note: *p<0.1; **p<0.05; ***p<0.01
Returns given as percentages

Table 8.5: Regress CAR on various explanatory factors

	Dependent variable:					
	Mkt Model CAR			FF Model CAR		
	(1)	(2)	(3)	(4)	(5)	(6)
Days Between Jan 1, 2005 and Breach	−0.001			−0.002		
	(0.002)			(0.002)		
Records Leaked (log)		−0.017			−0.016	
		(0.030)			(0.030)	
Customer Data Leaked			−0.907			−0.828
			(0.729)			(0.716)
Employee Data Leaked			−0.926			−0.718
			(0.795)			(0.781)
Credit Card Leaked			-1.482^{**}			−1.159
			(0.718)			(0.705)
SSN Leaked			0.963			0.853
			(0.666)			(0.654)
Name Leaked			−0.835			−0.629
			(0.681)			(0.669)
Address Leaked			−0.447			−0.536
			(0.645)			(0.634)
Observations	372	372	372	372	372	372
R^2	0.050	0.049	0.078	0.074	0.072	0.094
Adjusted R^2	0.015	0.015	0.031	0.041	0.038	0.047

Note: *p<0.1; **p<0.05; ***p<0.01
Year fixed effects in all specifications
CAR from 5 day event period