Informed Trading and Cybersecurity Breaches*

Joshua Mitts† Eric Talley‡

December 2, 2017

Abstract

Cybersecurity has become a significant concern in corporate and commercial settings, and it has significant implications for firm value. Consequently, the potential or realized threat of a cybersecurity breach can have material effects on firm value. This paper interrogates whether advance knowledge of a cybersecurity breach and/or vulnerability represent arbitrage opportunities for informed traders from a theoretical and empirical perspective. We first demonstrate the plausibility of such opportunities in in a theoretical securities market setting. We than make use of a novel data set to study trading activity preceeding the public announcement of a breach. Using a matched sample control group, we find significant trading abnormalities in the put options market for hacked firms, both in open interest and volume. Our results are robust to a variety of matching techniques and to both cross-sectional and logitudinal analysis. Our results are consistent with the proposition that arbitrageurs have early notice of impending disclosures of breaches. Our results help frame important policy questions about whether / how to regulate informed trading about cybersecurity risks.

*PRELIMINARY AND INCOMPLETE DISCUSSION DRAFT. DO NOT QUOTE OR CITE WITHOUT PERMISSION. We thank [___] for helpful comments and discussions and Hanna K. Song for excellent research assistance. All errors are ours.

†Columbia Law School. Email: joshua.mitts@law.columbia.edu

‡Columbia Law School. Email: etalley@law.columbia.edu.

1 Introduction

The ascendancy and impact of the information economy during the last two decades is as stunning as it is unprecedented. In the mid-1990s, fully one fifth of the preeminent Dow Jones Industrial index consisted of Eastman Kodak, Bethlehem Steel, F.W. Woolworth, International Paper, Sears Roebuck and Union Carbide. Amazon and Google were little-known startups. Apple was a fading icon of the 1980s. And Facebook and Bitcoin were still a decade away from existence. How times have ever changed. The digitization of the world's economy has hastened profound changes in commerce, record-keeping, law enforcement, personnel policy, banking, insurance, securities markets, and virtually all aspects of services and manufacturing sectors.

At the same time, the key pillars of the digitization economy – scale economies, standardization, and the ease of accessing/copying/distribution – frequently are also its Achilles Heel, in the form of cybersecurity risk. The massive and cataclysmic data breach of Equifax in September 2017, compromising highly confidential information of tens of millions of clients (including Social Security numbers), is hardly the first of its kind. For well over a decade, firms and organizations that store confidential data digitally have been potential (and often actual) targets of similar types of attacks often with analogously cataclysmic implications for victims.

Within securities-market settings, of course, one person's catastrophe is another's arbitrage. In the late summer of 2016, a well known short hedge fund, Muddy Waters Capital, opened a confidential line of communication with MedSec, a start-up cybersecurity firm that claimed to have discovered a serious security flaw in the pacemakers produced by St. Jude Medical, a then-public medical device company (in the process of being acquired by Abbot Laboratories). Only after taking a substantial short position in St. Jude did Muddy Waters publicly disclose the claimed vulnerability,[1] causing an immediate fall in St. Jude's stock price in excess of eight percent. (Goldstein et al. 2016). Similar episodes of material changes in value after disclosure of a cybersecurity event are now commonplace.[2]

[1] See http://d.muddywatersresearch.com/tou/?redirect=/content/uploads/2016/08/MW_STJ_08252016_2.pdf

[2] To take a current example, Uber's recent disclosure of a cybersecurity loss of client payment records caused an outside investor (Softbank) to reduce its valuation assessment of Uber by nearly

The anecdotal account of Muddy Waters' securities-market play around St. Jude is perhaps unsurprising—particularly when (a) cybersecurity breaches have material price effects; and (b) the underlying breach involves potentially confidential data. Trading in the securities of compromised issuers is far safer than trafficking directly in the compromised information itself. Indeed, fencing such protected data directly is almost always a criminal offence under state and federal law. In contrast, buying low and selling high (or selling high and buying low) in securities markets is a time-honored capitalist activity. At the same time, the St. Jude / Muddy Waters saga raises intriguing questions about how widespread such cybersecurity-related trading is, whether significant arbitrage rents are appreciable, and who tends to earn them. And, to the extent that appreciable arbitrage rents exist, might they directly or indirectly subsidize cyber-hacking—effectively catalyzing destructive activity solely for the purpose of treading on the basis of the harms and risks it creates? Is it possible to detect such activities by observing the footprint of trading patterns? Should such coordinated behavior be more heavily regulated by authorities?

In this paper, we consider cybersecurity vulnerabilities and/or breaches more generally, and how they interact with trading in sophisticated securities markets. Specifically, we consider whether advanced knowledge of a cybersecurity vulnerability or breach constitutes a material arbitrage opportunity for an informed trader. Using a theoretical securities trading model, we demonstrate that such arbitrage opportunities are eminently plausible equilibrium phenomena, and arbitrage opportunities exist whenever there is sufficient independent trading (e.g,. by liquidity or noise traders) to provide cover for the informed arbitrageur. Our model predicts that the informed trader will take short positions against the hacked firm that should be reflected in both price and volume of the underlying securities. We then endeavor to test these predictions empirically, making use of a novel data set corporate data breaches involving publicly traded companies (from the Identity Theft Resource Center). Using a variety of means to match these firms against comparator firms with no announced vulnerabilities, we find significant trading abnormalities in the put-option market for hacked firms, measured both through open interest and trading

a third. See Financial Times, "SoftBank share purchase discounts Uber by 30%" (Nov. 27, 2017).

volume. These results appear robust to a variety of matching techniques as well as to cross-sectional and time-series analysis. We view these results as consistent with the proposition that arbitrageurs tend to have early notice of impending cybersecurity breach disclosures.

Although our focus here is predominantly positive in nature—presenting evidence of informed cybersecurity trading—our results help frame and inform larger normative / prescriptive debates about whether such trading practices warrant additional legal proscription. Under current law, it is almost certainly illegal for third parties to coordinate an outright cybersecurity attack on a firm; it would similarly be unlawful for a party to spread *false* information about a cybersecurity risk in order to manipulate stock prices. That said, if third parties were simply to use publicly available investigatory tools to expose *bona fide* cybersecurity vulnerabilities (as Muddy Waters and MedSec are said to have done), they would face few if any legal impediments in arbitraging that information. They would not run afoul of insider trading laws, which generally require the breach of a confidential relationship.[3] And they would not violate market manipulation proscriptions, which require the injection of inaccurate information into the market.[4]

Whether such activities *should* elicit regulatory scrutiny is a sligthly different (and more difficult) question. On the one hand, outside arbitrageurs are routinely permitted (if not encouraged) to profit off of other forms of information they independently uncover about undisclosed risks at the market, industry or firm level. Such profit opportunities aid in price discovery and the rapid dissemination of relevant information to market participants. On the other hand, the availability of significant arbitrage rents from advance knowledge of cybersecurity risks can distort investment

[3]United States v. O'Hagan, 521 U.S. 642 (1997). That said, some recent case law suggests that hacking into a confidential serve and then trading on the information accessed might constitute a "deceptive practice" under Rule 10b-5. See, e.g., S.E.C. v. Dorozhko, 574 F.3d 42, 51 (2nd Cir. 2009) ("misrepresenting one's identity in order to gain access to information that is otherwise off limits, and then stealing that information is plainly "deceptive" within the ordinary meaning of the word. It is unclear, however, that exploiting a weakness in an electronic code to gain unauthorized access is "deceptive," rather than being mere theft. Accordingly, depending on how the hacker gained access, it seems to us entirely possible that computer hacking could be, by definition, a "deceptive device or contrivance" that is prohibited by Section 10(b) and Rule 10b–5.")

[4]See, e.g., SEC v. Masri, 523 F. Supp. 2d 361, 373 (S.D.N.Y. 2007).

decisions, both by cybersecurity firms attempting to find (and expose) weaknesses that would not otherwise come to light, and by the issuers themselves attempting to frustrate such zealous scrutiny. Such expenditures represent real economic costs, which plausibly might justify regulatory oversight of informed cybersecurity trading. Although we offer a few conjectures as to how our results might inform this policy debate, we save the broader normative analysis for another day.

Our analysis contributes to a growing literature on cyber-security threats, assimilating to a larger literature on informed trading in securities markets. Gordon & Loeb (2002) develop a framework for analyzing self-protection decisions among firms that are potential cybersecurity risks, arguing that firms may optimally concentrate on information resources of "mid-range" value when high vulnerability resources would be prohibitively expensive to protect. LeLarge (2012) extends this framework to a network of interconnected agents, showing that self-protection levels are often sub-optimal due to network effects and collective action problems. Kunruther & Heal (2003) study a similar framework based on a terrorism scenario. Lakdawalla & Talley (2006) analyze strategic target "hardening" by potential victims when the attacker makes strategic choices about where to attack, arguing that it may make sense to allow attacked parties to lodge a cause of action against non-attacked entities. Bohme & Moore (2016) study a similar model and relate it to whether targets should invest pro-actively or reactively. They find that when reactive investment is possible to mitigate an existing attack, it is optimal to under-invest in proactive technology. Although we are unaware of any existing literature on cybersecurity trading, the seminal framework of Kyle (1985) provides a baseline for our analysis, which predicts equilibrium information trading (contra Milgrom & Stokey 1982) in the presence of liquidity / noise traders who create volume that can mask the informed trader's activities. Empirically, Zhang et al (2014) lay out key network misconfiguration flags that tend to predict hacker vulnerability. Liu (2015) uses machine learning approaches to forecast cybersecurity vulnerabilities, exposing periods of latency of between one and twelve months before disclosure.

Finally, a literature in computer science studies how stock prices react to the disclosure of cybersecurity breaches. Spanos & Angelis (2016) present a meta-

analyzis of 37 papers containing 45 empirical studies of the effect of information-security breaches on public-company stock prices from 2003 to 2015. They find that 75.6% of the studies measure statistically significant stock-price reactions to the disclosure of cybersecurity breaches. 20 out of 25 studies find negative and significant stock-price reactions for victim firms, and none of these find significant positive reactions for victim firms. Three studies find positive and significant stock-price reactions for information security firms, plausibly reflecting the additional demand for their services in the wake of security breaches. Consistent with our findings, several studies also find evidence of "leakage" of cybersecurity vulnerabilities prior to the official announcement. Arcuri, Brogi & Gandolfi (2014) find that the mean cumulative abnormal return to 128 cybersecurity disclosures is -.029 in the (-20,+20) window, but shrinks to -0.003 in the (-1,1) window. That said, we are unaware of any prior study measuring trading patterns in the months preceding the disclosure, as we explore here.

Our analysis proceeds as follows. Section 2 develops a theoretical framework under which a collection of firms face hacking threats, whose realization can affect market pricing. We demonstrate that under fairly broad conditions, hacking threats can persist in equilibrium even in the presence of available protective measures by affected firms, and that an informed trader can make arbitrage profits in equilibrium from advanced information about such breaches. Section 3 turns to our empirical analysis,

2 A Model of Cyber-Hacking, Preventative Measures, and Trading

In this section we develop and analyzes a theoretical framework incorporating cybersecurity threats, security decisions made by potential targets, and market traders who may become informed of cyber-security breaches in advance of their disclosure. We develop the model in advance of the empirical portion of the paper both to fix our ideas concretely and to deliver predictions about equilibrium behavior that lends

itself to informed empirical testing. Like many economic models, the equilibria may turn on the informational environments of the players, and we thus take some care to make sure that our predictions are robust across multiple such environments.

2.1 Basic Framework:

Consider an economic setting involving N identical firms (indexed by $i \in \{1, ..., N\}$), each with widely-dispersed ownership of shares. Firms' respective operations are independent, in that they do not coordinate with one another on cyber-security defenses.[5] Each firm is entirely equity financed, with a diversified, risk-neutral shareholder base. For current purposes, we assume away any manager-owner agency costs (which are not central to our arguments); and thus we assume that firms are faithfully managed to maximize expected shareholder value. Each firm's business plan consists of a Bernoulli-like project, achieving either "success" or "failure" with respective probabilities of q and $(1-q)$, where $q \in (0,1)$. Without loss of generality, assume that "success" yields normalized payoff at 1, while "failure" yields a normalized payoff of 0. Consequently, and in the absence of cybersecurity breaches and preventative measures (both discussed below), the expected gross valuation of each firm is equal to q.

A risk-neutral cybersecurity hacker (denoted as player H) receives a payoff from successfully breaching a firm's security protocols. When the hacker succeeds in breaching a firm's security protocols, he obtains hedonic benefit of B. (In the baseline case, we assume that B is constant). The hacker is assumed to select his target from among the N firms, with mixed strategies permitted. Consequently, the hacker's strategy profile consists of choosing an attack probability (or 'hacking intensity') for each firm, denoted $\gamma \equiv \{\gamma_1, \gamma_2, ..., \gamma_N\}$, where $\gamma_i \in [0,1]$ and $\Sigma_i \gamma_i = 1$. (Because of the adding up condition, $\overrightarrow{\gamma}$ is in actuality an $N-1$ dimensional vector.)

Each firm i is permitted to take preventative measures to bolster its cybersecurity protocols. Those measures are given by the vector $C \equiv \{c_1, ..., c_N\}$. Pre-

[5]The firms' production plans may be interdependent, however, such as in the case of oligopolistic competition.

ventative measures are helpful to the firm, in that conditional on being hacked, the firms protection level can defeat the hacker's efforts with probability c_i. Thus, given a hacking intensity γ_i, firm i's expected gross valuation is:

$$(1-\gamma_i)\,q + \gamma_i\left[(1-c_i)\cdot 0 + c_i\cdot q\right] = (1-\gamma_i)\,q + \gamma_i c_i q \tag{1}$$

Intuitively, the expected marginal benefit of preventative measures is the salvaging of $q\gamma_i$ of firm value. Preventative measures are costly to install, however. We assume that in order to install c_i, firm i must incur costs of $\frac{c^2}{2}$. We will consider two canonical information structures below: In the first, the components of C can be observed by the hacker before selecting γ; in the second, the components of C are not observable—a case that is isomorphic to when γ and C are chosen simultaneously.[6]

Finally, we assume that there exists an outside trader who may find out (with probability π) which firm the hacker has targeted. The informed trader may then trade in the attacked firm's securities. (To simplify the analysis, we constrain the informed trader to transact only in the affected firm, though the model is easily generalized to allow trading in the entire portfolio of N issuers.)

We solve the game backwards, starting with the hacker's choice, followed by the firms' protection choice, followed by the trading behavior of informed insiders.

2.2 Case a: Hacker Does Not Observe Firm Protections

The simplest case is when the hacker has no information about C, the vector of firm-level protections. In this case, the hacker's decision will turn critically on his equilibrium *conjectures* about firms' expected preventative measures—which we define as $\hat{C} = \{\hat{c}_1, \hat{c}_2, ...\hat{c}_N\}$. Consider the set of firms we denote as $\hat{k} \equiv \{i \in N | \hat{c}_i = \hat{c}_{\min}\}$, where $\hat{c}_{\min} \equiv \min\left\{\hat{C}\right\}$. The set $\hat{k}$ represents those firms whose expected prevention measures under the hacker's equilibrium conjectures are minimal—a group one might use predatory jargon to describe as the "slowest animals in the pack." Depending

[6]It is also possible in addition to study a hybrid case where the hacker observes preventative measures with some probability ρ; a generalization that adds additional notational complexity, but yields no significant beyond these two polar cases.

on the hacker's equilibrium conjectures, $\hat{k}$ may have as few as 1 and as many as N members. Regardless, each member of $\hat{k}$ is equally vulnerable to hacking efforts, the hacker expects an identical marginal benefit of $(1-\hat{c}_{\min})\cdot B$ from targeting such firms. Given the equality of marginal benefit across them, the hacker would be willing to mix in any proportion among the members of $\hat{k}$. In contrast, consider the remaining $\left(N-\hat{k}\right)$ firms for whom $\hat{c}_i > \hat{c}_{\min}$ under the conjectured equilibrium. For these firms, the hacker's marginal benefit is strictly lower than $(1-\hat{c}_{\min})\cdot B$, and he would not rationally target any of them with positive probability.

Now consider the firms' incentives, and suppose that they conjecture the hacker will select strategy $\hat{\gamma} = \{\hat{\gamma}_1, ..., \hat{\gamma}_N\}$. Each will choose a level of protection to maximize his own expected net payoff, solving:

$$\max_{c_i \geq 0} (1-\hat{\gamma}_i)\, q + \hat{\gamma}_i c_i q - \frac{c_i^2}{2},$$

which yields an optimal protection level for each firm of

$$c_i\left(\hat{\gamma}_i\right) = \hat{\gamma}_i q.$$

Note that the firm will never expend any resources on protection if it conjectures that it is not being targeted ($\hat{\gamma}_i = 0$). Moreover, among those firms being targeted with positive probability ($\hat{\gamma}_i > 0$), the level of their protection will be directly proportional to their conjecture $\hat{\gamma}_i$. From this reasoning the following proposition immediately follows (all proofs are in the appendix):

Proposition 1: *When the hacker cannot observe firm protection levels, there is a unique equilibrium of the hacking/protection game, in which the hacker targets each firm with equal probability $\gamma_a = \frac{1}{N}$ and each firm installs preventative measures of $c_a = \frac{q}{N}$. Firms' equilibrium payoffs (and market values) are given by:*

$$V_a = \frac{q\left(N^2 - N + \frac{q}{2}\right)}{N^2}$$

Note that the firms' payoff is increasing in their baseline success probability

(q), but also increasing in the number of targets, reflecting that an increasing population of targets requires the hacker to divide his attention across a larger population. In the limit, in fact, $V(\gamma_a, c_a|q, N)$ approaches the baseline expected gross value of the firm in the absence of hacking, or q.

Before moving onto the case where the hacker can observe protection levels, it is worth taking note of what fair market value the firms would have if one knew the realization of γ_a—that is, which firm had been actually selected under the hacker's mixed strategy for attack. In this case, the targeted firm (assuming it remained unaware) would be playing equilibrium strategy $c_i = c_a = \frac{q}{N}$, but would face attack with (ex post) probability one, yielding an expected payoff of:

$$V(\gamma_a, c_a|Breach, q, N) = \frac{q^2}{N^2}\left(N - \frac{1}{2}\right)$$

And thus the difference between the ex ante and ex post value of the breached firm is:

$$\Delta_a \equiv V(\gamma_a, c_a|q, N) - V(\gamma_a, c_a|Breach, q, N) = \frac{q(N-q)(N-1)}{N^2} > 0$$

And thus, knowing that a firm had been selected to be hacked in this equilibrium would give rise to mispricing of the firm, and that selling short an $x-$fraction shares of the firm would yield profits of $\Delta_a \cdot x > 0$. It is easily confirmed that Δ_a is strictly increasing in N, and thus the benefit of inside information increases as the market increases in size.

Finally, note that when $N > 2$, the value of the information is strictly larger for the breached firm than for the non-breached firms, so that $|\Delta_a(Breach)| > |\Delta_a(NoBreach)|$. In general, then, it is more profitable to trade on information about cybersecurity breach victims than non-victims.

2.3 Case b: Hacker Does Observe Firm Protections

Now consider the alternative information structure where the hacker can observe C before selecting his own strategy γ. Here the hacker's ex ante conjectures about firm

activity levels are no longer relevant, since he can observe their actual realization. After observing $C = \{c_1 ... c_N\}$, the hacker will select the components of γ to solve:

$$\max_{\gamma_i \in [0,1]} \left\{ B \cdot q \cdot \sum_{i=1}^{N} \gamma_i (1 - c_i) \right\} \text{ s.t. } \Sigma_i \gamma_i = 1.$$

Here it is straightforward to show that the hacker will devote all his time to the firm(s) with the lowest protection level(s) c_i. Similar to above, denote $c_{\min} = \min \{c_{1,} c_{2,} ... c_N\}$, and define $K \equiv \{i \in N | c_i = c_{\min}\}$. Similar to the analysis above, K has cardinality $k \in \{1, ..., N\}$, and the hacker will spread his efforts among those k firms comprising K, according no efforts to breach towards the remaining $N - k$ firms. Following the reasoning from Proposition 1a above, when $k \geq 2$, the equilibrium strategy of the hacker entails setting:

$$\gamma_i = \begin{cases} \frac{1}{k} & \text{if } i \in K \\ 0 & \text{else} \end{cases}$$

Now consider the perspective of firm i, which conjectures that of the other $(N-1)$ firms will pursue strategies $\hat{C}_{-i} = \{c_1 ... c_N \backslash c_i\}$. Denote $c_{\min}^{-i} = \min \left\{ \hat{C}_{-i} \right\}$ designating the lowest level of protection conjectured by firm i among the *other* firms. Thus, firm i's payoff function (incorporating the hacker's equilibrium behavior, shown above) is:

$$\pi (c_i | c_{-i}) = \begin{cases} q - \frac{c_i^2}{2} & \text{if } c_i > c_{\min}^{-i} \\ q \left(\frac{k-1}{k} + \frac{c_i}{k} \right) - \frac{c_i^2}{2} & \text{if } c_i = c_{\min}^{-i} \\ q \cdot c_i - \frac{c_i^2}{2} & \text{if } c_i < c_{\min}^{-i} \end{cases}$$

Note that firm i's payoff has several discontinuities, which in turn permit one to narrow down the firm's plausible equilibrium actions. First, the firm would never want merely to match $c_{\min}^{-i}$ when $c_{\min}^{-i} < 1$. Indeed, the firm could could always do better setting $c_i = c_{\min}^{-i} + \varepsilon$ for arbitrarily small ε, gaining $q \left(\frac{1}{2} - \frac{c_{-i}}{2} \right)$ at marginal cost ε. Thus, except in the extreme case where $c_{\min}^{i} = 1$, it never pays for a firm to pool with the least-well protected firms. In addition, note that if there were only one firm in the industry (i.e., $N = 1$), it effectively is the case that $c_{\min}^{i} = \infty$, so that

the optimal solution to the protection problem is to set $c_i = q$, generating payoff $\frac{q^2}{2}$. Similarly, if the firm were attempting to optimize under the assumption that $c_{\min}^{-i}$ were prohibitively large (so that firm i would have to be the "laggard" in preventative measures), it would also choose to set the same interior solution of setting $c_i = q$, with payoff $\frac{q^2}{2}$.

When $c_{\min}^{-i}$ is not prohibitively large, in contrast, the firm has a choice between (a) "just beating out" $c_{\min}^{-i}$ by setting $c_i = c_{\min}^{-i} + \varepsilon$, in which case it procures a payoff of (approximately) $q - \frac{\left(c_{\min}^{-i}\right)^2}{2}$, or (b) accepting its fate as a laggard in preventative measures once again setting $c_i = q$ and earning payoff $\frac{q^2}{2}$. Comparing these payoffs, it is easily confirmed that player i would only be content to remain a laggard if:

$$\begin{aligned} \frac{q^2}{2} &\geq q - \frac{\left(c_{\min}^{i}\right)^2}{2} \\ c_{\min}^{i} &> c^* = \sqrt{q\left(2-q\right)} \in (q, 1) \end{aligned}$$

The discontinuity in $\pi\left(c_i | c_{-i}\right)$ as well as the fact that firms would never be willing to pool in pure strategies suggests that the unique equilibrium in this case is a mixed strategy equilibrium, in which each firm i selects its self-protection strategy from a distribution $F\left(c|N\right)$. The implication of this logic is summarized in Proposition 2:

Proposition 2: *When the hacker can observe firm protection levels, there is a unique equilibrium of the hacking/protection game in which (a) the hacker targets each firm with equal probability $\gamma_b = \frac{1}{N}$, and (b) each firm installs preventative measures c_i according to a mixed strategy c_b,* defined over the interval $c_b \in \left[q, \sqrt{q\left(2-q\right)}\right]$ with *cumulative distribution function:*

$$F\left(c_b | q, N\right) = 1 - \left(1 - \frac{\left(c-q\right)^2}{2q\left(1-c\right)}\right)^{\frac{1}{N-1}}$$

Firms' equilibrium payoffs (and market values) are given by:

$$V_b = \frac{q^2}{2}$$

Note that the observability of the players' protective measures induces a type of "arms race" in this equilibrium, in which the realized value of c_b exceeds (with probability 1) the level of protection that a single firm would take (shown above to be $c_i = q$).

As in the previous case, consider the fair market value of the firm if one were able to observe on a proprietary basis the hacker's information and actions–which in this case means observing both the realization of γ_b and the minimal value of firm-level protections ($c_{\min}$). By hypothesis, the breached firm's realized level of protection is $c_{\min}$, and thus its expected payoff of conditional on being attacked is:

$$V\left(Breach|c_{\min}, q, N\right) = c_{\min}q - \frac{c_{\min}^2}{2}$$

And thus the expected difference between the ex ante and ex post value of the breached firm is:

$$\begin{aligned}
\Delta_b &\equiv E\left(V\left(\gamma_e, c_e|q, N\right) - V\left(\gamma_e, c_{\min}|Breach, q, N\right)\right) \\
&= \frac{1}{2}E\left\{\left(q - c_{\min}\right)^2\right\} \\
&= \int_q^{\sqrt{q(2-q)}} \frac{1}{2}\left(q - c_{\min}\right)^2 dF_{\min}\left(c_{\min}|N\right) > 0
\end{aligned}$$

where

$$F_{\min}\left(c_{\min}|N\right) = 1 - \left(1 - F\left(c_{\min}|N\right)\right)^N = 1 - \left(1 - \frac{\left(c_{\min} - q\right)^2}{2q\left(1 - c_{\min}\right)}\right)^{\frac{N}{N-1}}$$

It is easily confirmed that for all $N' > N \geq 2$, $F_{\min}\left(c_{\min}|N\right) < F_{\min}\left(c_{\min}|N'\right)$, and thus Δ_b is both positive and strictly increasing in N and the expected gain one can make from selling the hacked company's stock at market rates and realizing intrinsic value (i.e., potential arbitrage profits) is increasing in N.

2.4 Trading Stage (Incomplete)

From the previous subsection, it is clear that (a) knowing which firm is subject to hacking creates an appreciable gap between uninformed market price market value informed by knowledge of the hack; and that (b) this wedge persists regardless of whether the hacker is can observe firm-level protections (Δ_b) or cannot (Δ_a). We now turn to the trading stage, of the model, in which an informed trader may gain information about a cybersecurity attack at a firm, and then may trade on the basis of that information and in advance of its public disclosure. Our aim is to develop a general trading framework that is amenable to variations on differing information structures of the hacking stage discussed above.

Thus, suppose that in the absence of specific information, the expected value of firm i is given by $V > 0$, but if the firm is targeted by a cyber-hacker, its expected value is decreased to $V - \Delta$, where $\Delta \in (0, V)$. (The previous subsection confirms that this assumption is satisfied in both informational permutations of the model $\{(V_a, \Delta_a); (V_b, \Delta_b)\}$). For each firm i, suppose there is a volume x_i of sell orders from liquidity traders distributed according to a cumulative distribution function $G(x_i)$, distributed $N(0, \sigma^2)$, and associated probability distribution function $G'(x_i) = g(x_i) > 0 \; \forall \; x_i$.[7]

Suppose there is a potentially informed trader who (with probability π) learns which of the N firms has been targeted by the hacker. Other market participants, however, remain uninformed (at least directly). We suppose that the informed trader becomes can try to profit from the information by putting in a sell order on each firm in the amount ψ_i. In such a situation, we refer to the informed trader as being *active*. (For the moment, and without significant loss of generality, we confine the informed trader to transacting only in the affected firm). Viewed ex ante, then, each firm i anticipates that the informed trader will learn that firm i was targeted with probability $\pi_i = \frac{\pi}{N}$.

In pricing the security, uninformed market participants observe aggregate trading volume in the market, $z_i = (x_i + \psi_i)$, but they are unable to discern whether

[7](In later drafts we will explore relaxing the normality assumption.)

the volume is due to liquidity trading alone or insider orders as well. Armed with this observation and the ex ante probability an informed insider is present (π_i), market participants form a conjecture (denoted $\hat{\psi}_i$) about the informed trader's selling strategy if she is active. (This conjecture must be true in equilibrium, an issue we turn to later.)

From these inputs, capital market investors form a posterior probability estimate that an insider is active, denoted $\eta\left(z_i, \hat{\psi}_i\right) = \Pr\left\{Insider | z_i, \hat{\psi}_i\right\}$. Using Bayes rule, the market's posterior assessment of an active insider are given by:

$$\begin{aligned}
\eta\left(z_i, \hat{\psi}_i\right) &= \frac{g\left(z_j - \hat{\psi}_i\right) \cdot \pi_i}{g\left(z_j - \hat{\psi}_i\right) \cdot \pi_i + g\left(z_j\right) \cdot (1 - \pi_i)} \\
&= \frac{1}{1 + \frac{(1-\pi_i)}{\pi_i} \cdot \frac{g(z_i)}{g\left(z_i - \hat{\psi}_i\right)}}
\end{aligned}$$

Note that since $z_i \equiv (x_i + \psi_i)$, the expression for $\eta(.)$ can equivalently be re-written as:

$$\eta\left(x_i, \psi_i, \hat{\psi}_i\right) = \frac{1}{1 + \frac{(1-\pi_i)}{\pi_i} \frac{g(x_i + \psi_i)}{g\left(x_i + \psi_i - \hat{\psi}_i\right)}}$$

Given $\eta(.)$ and Δ, the resulting market value for the firm is:

$$\begin{aligned}
P_i &= \eta\left(x_i, \psi_i, \hat{\psi}_i\right) \cdot (V - \Delta) + \left(1 - \eta\left(x_i, \psi_i, \hat{\psi}_i\right)\right) \cdot (V) \\
&= V - \eta\left(x_i, \psi_i, \hat{\psi}_i\right) \cdot \Delta
\end{aligned}$$

This alternative expression is the most relevant for the informed trader, who strategically selects ψ_i holding conjectures $\hat{\psi}_i$ constant.

The informed trader chooses her selling strategy ψ_i holding constant the market's conjectures $\hat{\psi}_i$ as well as the pricing dynamics above. At the same time, the informed trader cannot perfectly predict the total realized volume of liquidity trades x_i, and thus cannot predict the realized total volume z_i of trades (beyond the trades she is responsible for, captured by ψ_i). Accounting for these elements, the informed

trader's payoff as a function of ψ_i consists of the Expected price she gets for each sale, $E_x\left(V-\eta\Delta\right)$, less the true value of the sale $V-\Delta$, resulting in a payoff associated with strategy ψ_i of:

$$\begin{aligned}\Sigma\left(\psi_i|\hat{\psi}_i\right) &= \psi_i\cdot E_x\left\{V-\eta\left(x_i,\psi_i,\hat{\psi}_i\right)\cdot\Delta-\left(V-\Delta\right)\right\}\\ &= \Delta\cdot\psi_i\cdot E_x\left\{1-\eta\left(x_i,\psi_i,\hat{\psi}_i\right)\right\}\end{aligned}$$

The first order conditions associated with an optimal size of sell order (denoted ψ_i^*) are:

$$\begin{aligned}\frac{d\Sigma\left(\psi_i|\hat{\psi}_i\right)}{d\psi_i} &= 0 \qquad\qquad \text{(OptimalOrder1)}\\ &\Leftrightarrow\\ \Delta\cdot E_x\left\{1-\eta\left(x_i,\psi_i^*,\hat{\psi}_i\right)\right\} &= \Delta\cdot\psi_i^*\cdot E_x\left\{\frac{\partial\eta\left(x_i,\psi_i,\hat{\psi}_i\right)}{\partial\psi_i}\right\}\Bigg|_{\psi_i=\psi_i^*}\end{aligned}$$

The interpretation of this condition is similar to a monopoly pricing problem: the informed trader's optimal sell order (ψ_i^*) is the one that balances the marginal arbitrage-pricing benefit of a small increase in her sell order (i.e., $\Delta\cdot E_x\left\{1-\eta\left(x_i,\psi_i^*,\hat{\psi}_i\right)\right\}$) against the marginal cost that her added activity has in cannibalizing the market price of her infra-marginal trades (i.e., $\Delta\cdot\psi_i\cdot E_x\left\{\frac{\partial\eta\left(x_i,\psi_i^*,\hat{\psi}_i\right)}{\partial\psi_i}\right\}$).

Analysis of the trader's optimization condition yields the following proposition:

Proposition 3: *All Perfect Bayesian Equilibria of the trading game entail the informed insider selecting $\psi_i^*>0$ for the attacked firm i.*

The content of Proposition 1 is intuitive; it states that the informed shareholder will always attempt to trade by placing a net sell order in the attacked firm. Although the informed trader's activity will drive down equilibrium price, she will nonetheless choose to take a net short position up to the point where the gains from an additional trade on the margin cannibalizes her payoff on infra-marginal trades.

Two additional corollaries follow directly from Proposition 1, which we take up further in the next empirical section.

Corollary 1: *The expected total volume of short positions / short sales increases when there is an informed trader.*

Corollary 2: *The expected market price of a targeted firm declines relative to other firms when there is an informed trader.*

Although Proposition 3 is not directly testable without observations informed traders and their transactons, Corollaries 1 and 2 are more directly testable with market data. These predictions in hand, we turn in the next section to our empirical tests.

3 Data

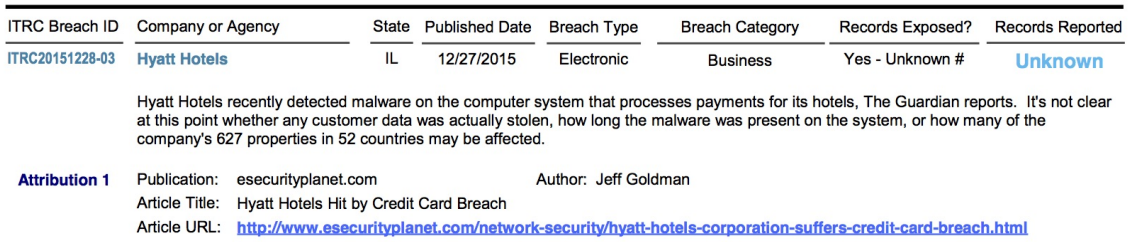
3.1 Corporate Data Breaches

We begin by collecting a list of corporate data breaches from the Identity Theft Resource Center (ITRC). Since 2005, the ITRC has collected and published an annual list of data breaches "confirmed by various media sources and/or notification lists from state governmental agencies." The ITRC's data breach report includes both exposure of personally identifying information — i.e., any incident "in which an individual name plus a Social Security number, driver's license number, medical record or financial record (credit/debit cards included) is potentially put at risk because of exposure" — as well as exposure of username and passwords that are not necessarily tied to an identifiable individual.

An example of an ITRC data breach report is given in Figure 1:

The categories of information included in the report are: (1) internal ITRC identifier of the breach, (2) the company which was attacked, (3) the state in which that company is located, (4) the date the breach was published, (5) the type of the breach, (6) the category of the breach, (7) whether personal records were exposed,

Figure 1: Example of ITRC Data Breach Report



ITRC Breach ID	Company or Agency	State	Published Date	Breach Type	Breach Category	Records Exposed?	Records Reported
ITRC20151228-03	Hyatt Hotels	IL	12/27/2015	Electronic	Business	Yes - Unknown #	Unknown

Hyatt Hotels recently detected malware on the computer system that processes payments for its hotels, The Guardian reports. It's not clear at this point whether any customer data was actually stolen, how long the malware was present on the system, or how many of the company's 627 properties in 52 countries may be affected.

Attribution 1
Publication: esecurityplanet.com Author: Jeff Goldman
Article Title: Hyatt Hotels Hit by Credit Card Breach
Article URL: http://www.esecurityplanet.com/network-security/hyatt-hotels-corporation-suffers-credit-card-breach.html

(8) how many records were exposed, and (9) a textual description of the breach. In addition, the ITRC provides details on the source of information about the breach, e.g., a news media report or disclosure by a governmental agency.[8]

The ITRC identified 4,580 data breaches from 2010 to 2016. While the vast majority of these involve private companies, nonprofits and governmental actors, out of this group, we were able to match 145 breaches to publicly traded companies.[9] To give a sense for the nature of the information contained in the textual descriptions of these 145 events, Figure 2 presents a "word cloud," which draws the most frequent words in these descriptions with a size proportional to the terms frequency — i.e., larger words appear more frequently in the textual descriptions. As Figure 2 shows, the most popular terms in these descriptions reflect the sort of information that would typically be the subject of a data breach, i.e., personal information, email address, credit cards, addresses, social security numbers, etc.

In order to conclude that transactions involving these victims of data breaches are not due to random chance alone, it is necessary to compare these data breaches to some sort of baseline (i.e., a "control" group). Even if there is no trading on corporate data breaches — i.e., even if we were to simply draw public companies and

[8]state privacy laws often require companies to notify individuals whose personal information may have been compromised (see, e.g., N.H. Rev. Stat. 359-C:19). Moreover, specific federal laws sometimes requires disclosure, e.g., when health concerns are implicated (HIPPA), or if the breach is sufficiently material to require disclosure by a publicly traded company under the securities laws. To be sure, there is no general duty to disclose material information under the securities laws, but cybersecurity vulnerabilities may fall into one of the enumerated categories of material event disclosure on Form 8-K.

[9]For reasons detailed below, we end up using a smaller sample to ensure adequate comparability between firms and industries.

Figure 2: Word Cloud of ITRC Data Breach Descriptions

This figure shows the relative frequency of two-word phrases that appear in the ITRC description of the data breach: larger phrases are more frequent.

calendar dates out of a hat by random chance — some firms will experience more or less trading activity for unrelated reasons. Thus, it is necessary to establish a baseline group that can serve as a counterfactual, i.e., such that "but-for" the hacker trading, our victim firms and the baseline group are similar in all other relevant ways, at least on average. Only if but-for causation holds can we conclude that observed differences may be attributable to hacker trading or tipping.

3.2 Put Options

We examine two primary sources of data in order to measure possible hacker trading and tipping. First, we look at relatively at-the-money (ATM) put options written on the common stock of victim firms. A put option is a downside bet on a firms stock: it gives its holder the right, but not the obligation, to sell the firm's stock at a particular price, known as the strike price until a specific date known as the "maturity" date of the put option. Taking advantage of this right to sell the firm's stock at the strike price is known as "exercising" the put option.

If we denote the strike price of a put option as K, its maturity date as T, and the firms stock price at date T as S_T, the holder of a put option will receive the greater of KS_T or zero. In other words, she receives the difference between the strike price and the stock price at maturity if the former exceeds the latter. If the stock price at maturity is higher than the strike price, she will rationally not exercise the put option because that would cost her money; she is better off doing nothing.

To understand how put option payoffs work, consider a brief numerical example. Suppose that the stock price at maturity is \$5 and the strike price is \$10. Further suppose that the holder of the put option already owns the underlying stock, which, again, is worth \$5 at maturity. Then she can profit by exercising the put option, selling the stock at \$10 instead of its current value of \$5, and receive \$10 (the sale price) $-$\$5 (the value of the stock sold) = \$5. What if the holder of the put option does not currently the underlying stock? She can simply buy the shares at \$5 and immediately sell them at \$10. Thus, her profit from the option is \$10 - \$5 = \$5, regardless of whether she owns the underlying shares of stock or not.

Put options are a downside bet on the firm's stock because the value of a put option increases as the firm's stock price at maturity decreases. Put simply, the lower the stock price, the more the put option is worth: put options are thus directionally negative bets on the value of the firm. Because the directional implications of a data breach are unambiguously negative for a victim firm — that is, one would be hard-pressed to find an example of a successful data breach that should lead to an increase in the stock price of the victim firm — put options are likely to become

more valuable upon revelation of a successful data breach. This implies that market demand for put options may reflect at hackers or their "tippees" may seek to exploit information, known only to them, about a successful data breach.

We restrict our analysis to put options that are close to "at the money" — that is, they have a "delta" between 0.4 and 0.6.[10] Put simply, that means that the strike price is likely to be relatively close to the current price of the firm's stock. We do so because a put option that is "out of the money" is likely to be less responsive to changes in the underlying price of the firm's stock.

We measure market demand for put options in two ways. The first is open interest, which refers simply to the number of outstanding put-option contracts on the stock of a particular underlying firm. The second is volume, which refers to the quantity of put-option contracts that change hands between buyers and sellers over a particular window of time. Both measure the extent to which traders in the market are seeking to place downside bets on the prospects of victim firms.

In order to facilitate meaningful comparisons that are straightforward to interpret, we aggregate our dataset to the firm-event level. That is, the unit of analysis in our study is an average measure of trading in a given firm's put options over a time window relative to a data breach event. For example, we refer below to average open interest of put options for a particular firm over the two months prior to disclosure of the data breach. If, hypothetically, there were two events and two firms for each events, there would be four observations, each reflecting the average open interest for each firm in the two months prior to each event. In the following Section 4, we describe how we design our empirical study to maximize the reliability of inferences as to the link between corporate data breaches and the demand for put options.

4 Empirical Design

Our goal is to evaluate empirically whether there is heightened trading in put options prior to the announcement of corporate data breaches. To do so, we rely on the

[10]The "delta" of an option refers to its sensitivity to changes in the underlying stock price.

well-developed literature on causal inference in empirical economics. To be sure, our hypothesis is inherently "descriptive" in nature—we do not suppose that data breaches *causally* increase put option trading, but rather that individuals who are aware of data breaches prior to the rest of the market may be directly trading or tipping others as to the presence of these vulnerabilities prior to disclosure. Formally speaking, this thesis requires only a correlation between the execution of corporate data breaches and market demand for put options.

Nonetheless, we are aware that an analysis of this sort is vulnerable to spurious correlations. The problem of forming a valid *counterfactual* — what level of put option trading would have emerged even in the absence of a data breach — is a vexing challenge that applies to our study just as much as with a classical causal inference project. For this reason, we employ methods in the literature to estimate the "average treatment effect" of data breaches, keeping in mind the importance of forming a valid counterfactual to evaluate whether observed put option demand can actually be attributed to data breaches.

We thus estimate two basic kinds of empirical designs, each of which relies on a different dataset. The first is a cross-sectional estimation, which simply asks: is there a heightened level of open interest and trading volume in the put options of data breach targets, *prior to* revelation of the data breach by the victim firm? To minimize the likelihood that this simple comparison between firms for each event is contaminated by other events that may give rise to put option trading, this estimation focuses on the two months immediately preceding announcement of the data breach. In this specification, we ask whether the average level of open interest and trading volume during this two-month is higher for firms who are the victims of data breaches. As described below, we employ propensity-score matching (Abadie and Imbens, 2006) to ensure that treatment and control firms are as similar as possible.

This cross-sectional specification, however, is vulnerable to the critique that firms may differ for unobserved reasons that can lead to greater overall demand for put options. To address this concern, we consider an alternative "difference-in-differences" design which allows each firm-event in our dataset to have a baseline level of open interest and trading volume of put options. In this difference-in-differences

specification, we compare the *change* in open interest and volume of put options from a baseline period — eight to sixteen months prior to announcement of the data breach — to the period of interest — eight months prior to the day of announcement.

In our difference-in-differences design, we use this eight-month cutoff for two reasons. First, this corresponds roughly to the average period of time during which a hacker is aware of a successful data breach.[11] Moreover, a visual inspection of the data shows that this is also approximately the time when time trends begin to diverge between treatment and control firms—prior to this point, they are roughly parallel, as we show below.

We aggregate pre-post differences to the firm-event level and compare these differences between treatment and control firms. As with the cross-sectional design, we employ propensity score matching on observable covariates to ensure that similar firms are compared to each other. This heightens the plausibility of the counterfactual inference that treatment and control firms would have similar counterfactual outcomes. Along with showing that the parallel trends assumption is satisfied, this evidence suggests that observed differences in put option trading are likely to be linked to corporate data breaches and not spuriously arising as a result of other differences between firms.

As noted previously, both of our specifications employ propensity-score matching (Abadie and Imbens, 2006), which matches each treatment observation to one or more control observations which are similar along several covariates. We generate a propensity score and thus matching observations by estimating a logistic regression on the following covariates: (1) 4-digit SIC industry code (i.e., an indicator for each), (2) log of market capitalization, (3) log of total assets, (4) log of net income, and (5) log of total liabilities. In our view, it is essential to compare within industry because firms in different industries are very different from each other.

For these reasons, we are forced to drop those firms in industries which are too small to allow for obtaining a meaningful matched control group. Indeed, while many of these smaller industries contain several firms, many small-cap firms are too illiquid

[11]Research by Symantec has shown that hackers tend to exploit security vulnerabilities for an average of ten months prior to discovery by the affected firm (Bilge and Dumitras, 2012).

to have frequent options trading. Limiting the sample to those firms for which we have sufficient information over the relevant time periods yields 46 treatment firm-event pairs and 3,319 control firm-event pairs in the difference-in-differences dataset and 51 treatment firm-event pairs and 3,425 control firm-event pairs in the difference-in-differences dataset.[12] The following Tables 1 and 2 present summary statistics on these datasets.

[Table 1]

[Table 2]

4.1 Balance Tests

The validity of our propensity-score matching method to estimate causal effects turns on the extent to which the treatment and control groups are "balanced," that is, likely to exhibit the same counterfactual outcomes even in the absence of treatment. Of course, there are a relatively small number of public companies with liquid options in each 4-digit SIC code industry, so any matching procedure will fall short of achieving perfect balance. Nonetheless, we perform a series of tests to verify balance in the distribution of treatment and control firms.

We begin by visually comparing the distribution of the propensity score for both the cross-sectional and difference-in-difference datasets when estimated using the full set of covariates. Figures 3 and 4 show this distribution before and after matching for the cross-sectional and difference-in-difference datasets, respectively.[13] The similarity in the density of the two propensity scores suggests that the two groups are balanced on the propensity score.

[12]The latter contains more firms than than the former because it covers a longer time period.

[13]In these figures, the propensity score is estimated on the subsample which contains nonzero open interest, but the results are virtually identical when estimating on the subsample that contains nonzero trading volume.

Figure 3: Propensity Score Balance Test: Cross-Sectional Dataset

This figure plots the distribution of the propensity score in the cross-sectional dataset. The figure on the left shows the distribution in the original sample, while the figure on the right shows the distribution in the matched sample. As the figure shows, the matching produces a sample that is virtually identical with respect to the observed propensity score.

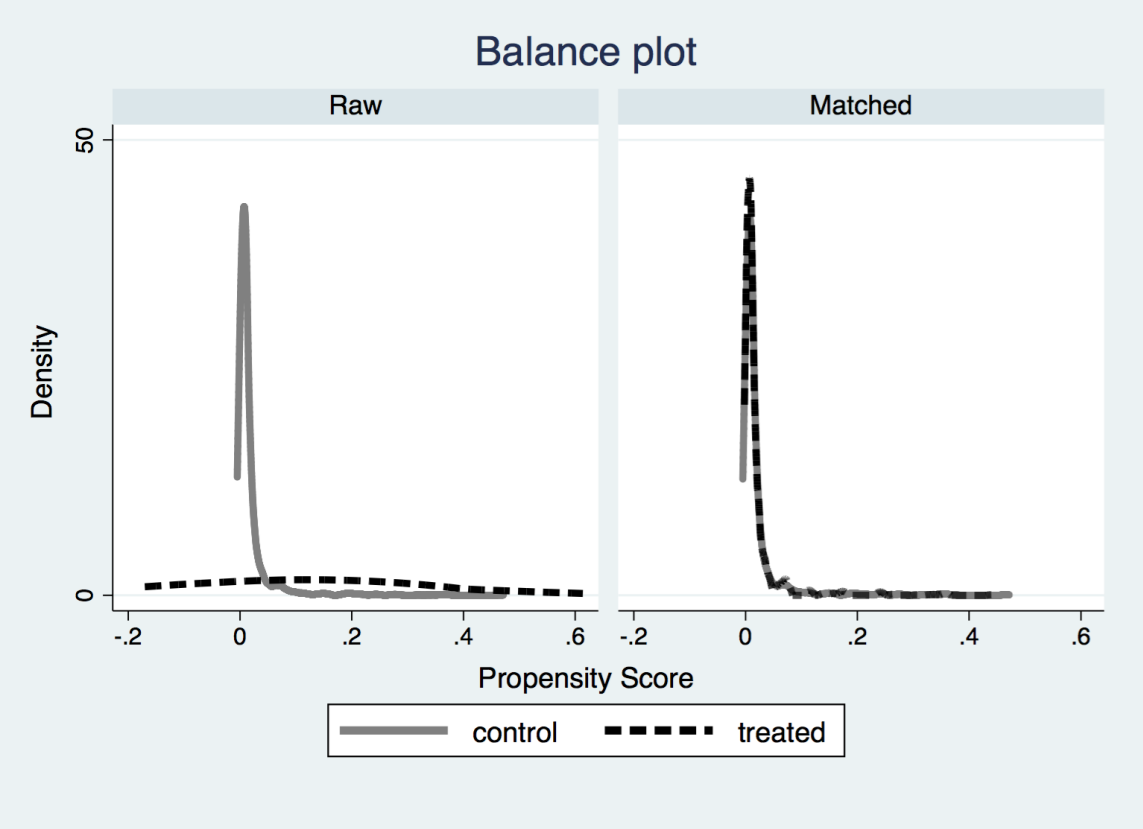
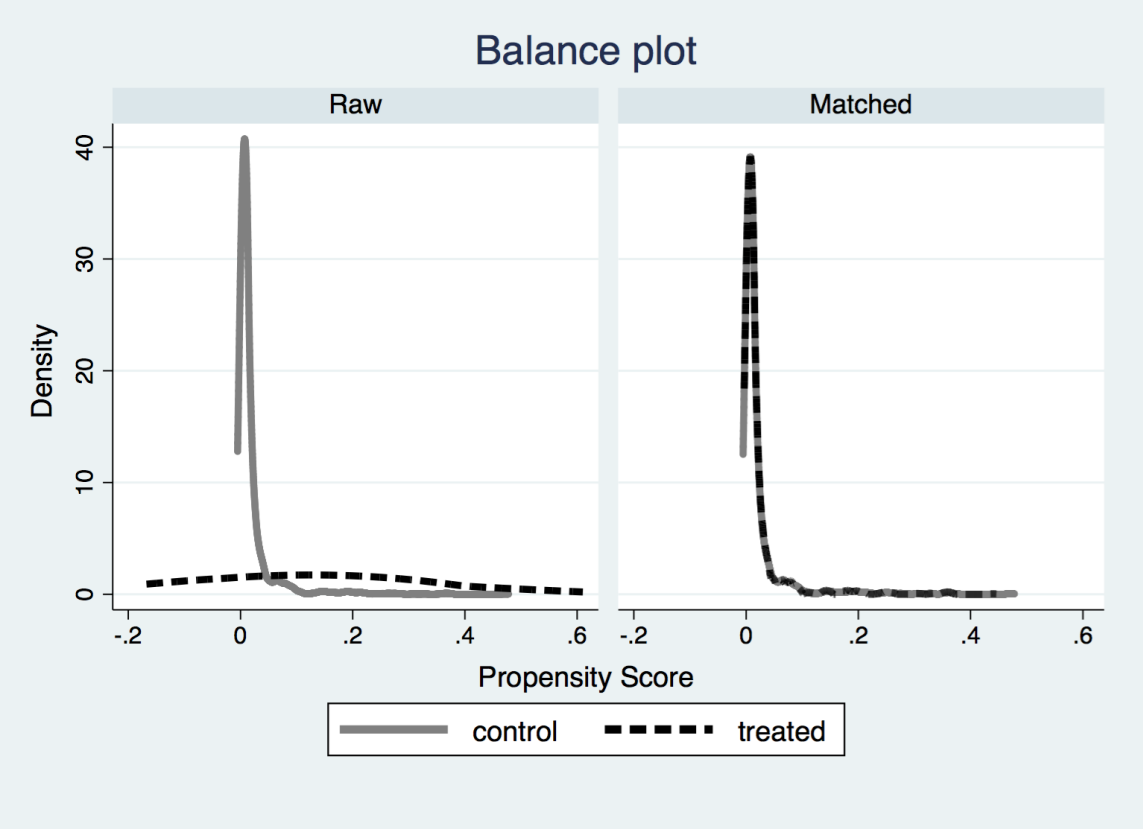


Figure 4: Propensity Score Balance Test: Difference-in-Differences Dataset

This figure plots the distribution of the propensity score in the difference-in-differences dataset. The figure on the left shows the distribution in the original sample, while the figure on the right shows the distribution in the matched sample. As the figure shows, the matching produces a sample that is virtually identical with respect to the observed propensity score.



What about individual covariates? Due to the relatively small number of public firms with liquid equity options in each SIC code, achieving greater balance on one covariate inevitably involves a loss of balance on another (to some extent). For this reason, in Section 5, we present results using propensity-score matching on individual covariates, as well as all of the covariates together, to illustrate that the results do not depend on which covariates are included.

Tables 3 and 4 compare covariate means in the cross-sectional and difference-in-differences dataset between the raw and matched samples. While the matching is

unable to achieve perfect balance across all of the covariates simultaneously, this table shows that each specification leads to near-perfect balance on a different covariate. As shown in Section 5 below, The consistency of the coefficient estimates across these different specifications in significance and magnitude strongly suggests that the results are not driven by spurious variation in covariate balance.

5 Primary Results

5.1 Cross-Sectional Estimation

We begin by estimating the average treatment effect (ATE) for the treatment group by propensity score matching (Abadie and Imbens, 2006) over the interval $[-60, 0]$, i.e., two months prior to disclosure of the data breach.[14] Here, identification of the treatment effect is based on the assumption that this interval is likely to be unknown to anyone other than the hacker and corporate officers aware of the data breach. First, we estimate the difference in log open interest on outstanding put options between treatment and control firms. The results are shown in Table 5.

[Table 5]

As Table 5 shows, there is an average increase of between .55 and .70 log points in the open interest of the put options written on target firms, and the result is consistent and statistically significant across specifications.

Next, we estimate differences in log trading volume of outstanding put options between treatment and control firms. The results are shown in Table 6.

[Table 6]

As Table 6 shows, there is an average increase of between .50 and 1.27 log points in trading volume of put options written on target firms. The result is significant and increases in magnitude as additional covariates are included in the propensity score matching, indicating that initial statistical insignificance may simply reflect

[14] We show below that the results are not driven by the choice of this interval.

estimation noise driven by over-weighting of firms that are more different from each other.

To evaluate how sensitive these results are to the propensity score method (Abadie and Imbens, 2006), we re-estimate the ATE with all covariates across three other matching schemes for identifying treatment effects: inverse-probability weighting (Imbens, 2000), inverse-probability weighting with regression adjustment (Wooldridge, 2007), and regression adjustment (Lane and Nelder, 1982). The results are shown in Table 7 and Table 8.

[Table 7]

[Table 8]

Table 7 and Table 8 show that the effect is consistently positive and significant regardless of the matching method used.

To get a sense for the magnitude of the effect, recall that the unconditional log average open interest is approximately 4.60. Thus, the point estimate of 0.70 in the full model corresponds to roughly $0.70/4.60 \approx 15\%$ additional open interest. Similarly, the unconditional log average volume is approximately 1.62. The point estimate of 1.28 in the full model corresponds to roughly $1.28/1.62 \approx 79\%$ additional trading volume of put options in the targets of corporate data breaches.

To verify that the results are not driven by the choice of a two-month interval, we re-estimate the models matching on the full set of covariates using different time windows. The results for open interest and volume are shown in Tables 9 and 10. While some subsamples yield higher t-statistics than others, the point estimates are consistent in sign and magnitude regardless of the time window.

[Table 9]

[Table 10]

5.2 Difference-in-Difference Estimation

One concern with the results in the prior Section is that they may be driven by baseline differences between our treatment and control firms. To address this concern, we estimate a difference-in-differences specification which estimates a baseline level of open interest and volume on outstanding put options of target firms over the interval $[t-480, t-240)$, i.e., sixteen months to eight months prior to disclosure of the data breach.[15] The difference-in-difference design compares the difference between this baseline period and the interval $[t-240, t]$, i.e., eight months prior to disclosure of the data breach, between treatment and control firms.

As explained previously, we aggregate the change in the log average open interest and log volume of put options between the two periods by firm-event, so there is one observation per firm-event. We then employ propensity score matching with robust standard errors (Abadie and Imbens, 2006) to ensure that treatment and control firms are as balanced as possible on observable covariates and proceed to estimate the ATE on this outcome (i.e., the difference in log open interest and log volume).

The key identifying assumption of a difference-in-differences analysis is that treatment and control firms follow parallel trends in the matched sample. We plot these parallel trends on log open interest in Figure 5 and Figure 6:

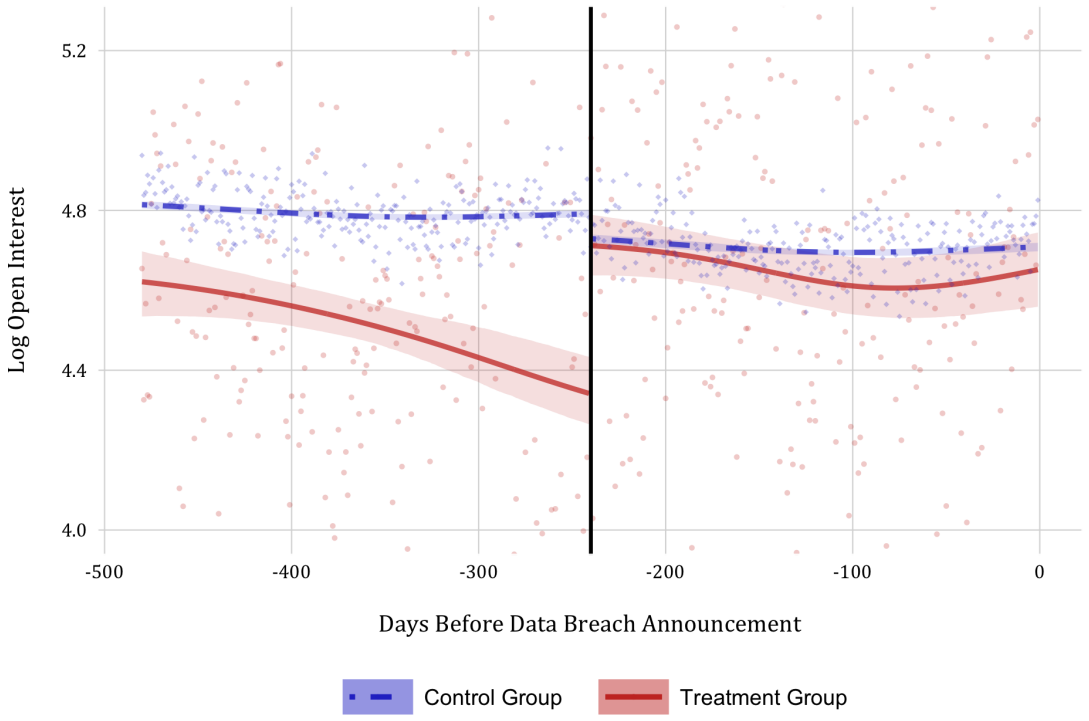
These parallel trends figures suggest that, indeed, the two groups are following parallel trends prior to divergence during this eight-month period preceding disclosure of the data breach. This strengthens the causal interpretation of differences during this eight-month period. The parallel trend graph for open interest clearly shows the increase in the number of outstanding put options in the treatment group. Differences in volume, on the other hand, seem to be driven by a decrease in the control group. Both are valid estimations of the treatment effect in a difference-in-differences design.

Proceeding to the statistical analysis, as before we first estimate the difference in pre-post differences of log open interest between treatment and control firms. The

[15] We show below that the results are not driven by the choice of this specific interval.

Figure 5: Parallel Trends on Log Open Interest

This figure plots time trends for log average open interest on put options between treatment and control firms in the matched sample. The pre-treatment period is the interval $[t-480, t-240)$, i.e., sixteen to eight months prior to disclosure of the data breach, and the post-treatment period is the interval $[-240, t)$, i.e., the eight months prior to disclosure of the data breach.



results are shown in Table 11.

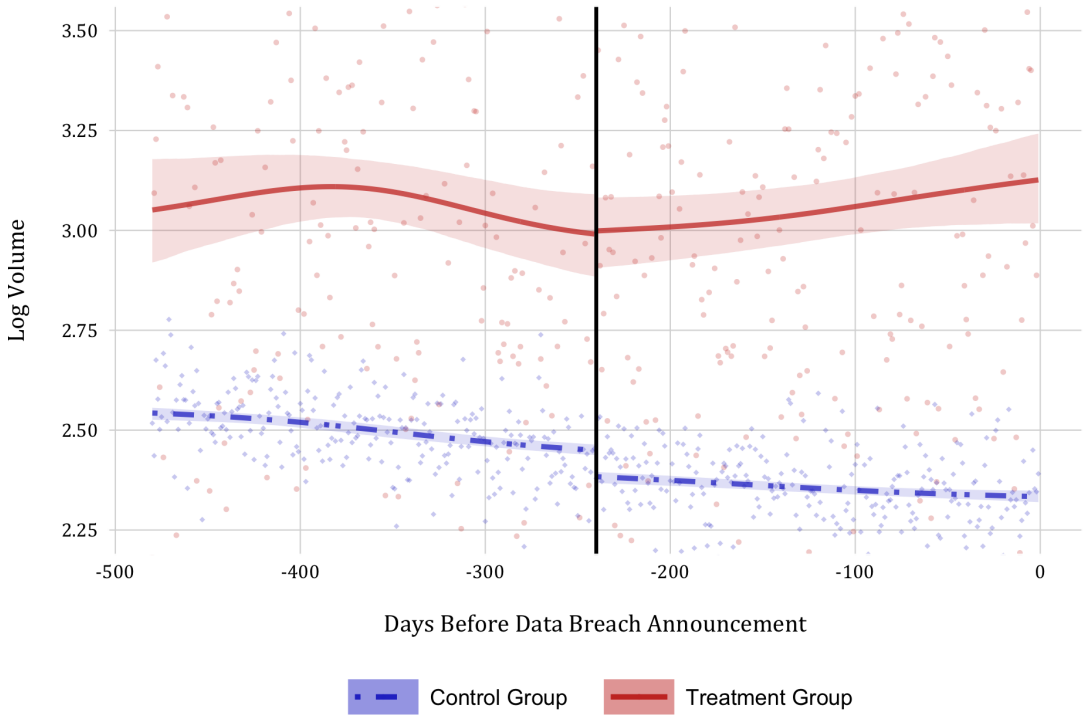
[Table 11]

Table 11 shows an average increase of between .26 and .32 log points in the pre-post difference in open interest of put options written on target firms, and the result is consistent and statistically significant across nearly every specification. The only insignificant specification has the fewest covariates included, but the point estimate is similar and thus the insignificance is likely to be driven by noise in the data.

Next, we estimate the difference in pre-post differences in log trading volume

Figure 6: Parallel Trends on Log Volume

This figure plots time trends for log average trading volume on put options between treatment and control firms in the matched sample. The pre-treatment period is the interval $[t-480, t-240)$, i.e., sixteen to eight months prior to disclosure of the data breach, and the post-treatment period is the interval $[t-240, t]$, i.e., the eight months prior to disclosure of the data breach.



of outstanding put options between treatment and control firms. The results are shown in Table 12.

[Table 12]

As Table 12 shows, there is an average increase of between .23 and .36 log points in the pre-post difference in average trading volume of put options written on target firms. As with the cross-sectional estimation, the result is significant and increases in magnitude as additional covariates are included in the propensity score matching, indicating that initial statistical insignificance may simply reflect estimation noise

driven by over-weighting of firms that are more different from each other.

To evaluate how sensitive these results are to the propensity score method, we re-estimate the ATE with all covariates across three other methods of identifying treatment effects: inverse-probability weighting (Imbens, 2000), inverse-probability weighting with regression adjustment (Wooldridge, 2007), and regression adjustment (Lane and Nelder, 1982). The results are shown in Table 13 and Table 14.

[Table 13]

[Table 14]

Table 13 and Table 14 show that the effect is consistently positive and significant for virtually every matching method, and similar in magnitude to the propensity-score estimation.

To verify that the results are not driven by the choice of the interval $[t-480, t-240)$, we re-estimate the models matching on industry, market value and total assets covariates using different intervals.[16] The results for open interest and volume are shown in Tables 15 and 16. While some subsamples yield higher t-statistics than others, the point estimates are consistent in sign and magnitude regardless of the time window.

[Table 15]

[Table 16]

5.3 Textual Analysis: Which Events Drive the Results?

To get a better sense for the kinds of events that are driving the results, we perform a textual regression of log open interest for put options on firms in the Treatment $\times$ Post group on the words and bigrams in the ITRC event descriptions. Of course, there are far more words and bigrams than events. We thus employ lasso regularization in a logistic regression using the coordinate descent method pioneered by

[16] We omit the log net income and log total liabilities covariates in the matching due to the loss of observations arising from negative values, which leads to collinearity at certain windows.

Friedman et al. (2010). We report the top words and bigrams with non-zero coefficients that predict log open interest for put options on firms in the Treatment $\times$ Post group. The results are given in Table 17 and 18.

[Table 17]

[Table 18]

6 Conclusion

TBD

7 Appendix (Proofs)

TBD

8 References

1. Arcuri, Maria Cristina, Marina Brogi, and Gino Gandolfi. "The effect of information security breaches on stock returns: Is the cyber crime a threat to firms?." European Financial Management Meeting, Rome. 2014.

2. Böhme, Rainer, and Galina Schwartz. "Modeling Cyber-Insurance: Towards a Unifying Framework." WEIS. 2010.

3. Böhme, Rainer, and Tyler Moore. "The "iterated weakest link" model of adaptive security investment." Journal of Information Security 7.02 (2016): 81.

4. Goldstein, Matthew, Stevenson, Alexandra and Picker, Leslie, 2016. "Unusual Pairing Makes Public Bet vs. Pacemakers." New York Times (Sept. 8, 2016 at B1).

5. Gordon, Lawrence A., and Martin P. Loeb. "The economics of information security investment." ACM Transactions on Information and System Security (TISSEC) 5.4 (2002): 438-457.

6. Gordon, L.A., M.P. Loeb, and W. Lucyshyn, "Sharing Information on Computer Systems: An Economic Analysis," J. Accounting and Public Policy, vol. 22, no. 6, 2003, pp. 461-485.

7. Grossklags, Jens, and Benjamin Johnson. "Uncertainty in the weakest-link security game." Game Theory for Networks, 2009. GameNets' 09. International Conference on. IEEE, 2009.

8. Hui, Kai-Lung, Wendy Hui, and Wei T. Yue. "Information security outsourcing with system interdependency and mandatory security requirement." Journal of Management Information Systems 29.3 (2012): 117-156.

9. Kunreuther, Howard, and Geoffrey Heal. "Interdependent security." Journal of risk and uncertainty 26.2-3 (2003): 231-249.

10. Kyle, Albert S. "Continuous Auctions and Insider Trading." Econometrica Vol. 53, No. 6 (1985), pp. 1315-1335

11. Lakdawalla, Darius N. and Talley, Eric L., Optimal Liability for Terrorism (October 2006). NBER Working Paper No. w12578. Available at SSRN: https://ssrn.com/abstract=935571

12. Lelarge, Marc. "Coordination in network security games: a monotone comparative statics approach." IEEE Journal on Selected Areas in Communications 30.11 (2012): 2210-2219.

13. Liu, Yang, et al. "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents." USENIX Security Symposium. 2015.

14. Milgrom, Paul; Stokey, Nancy. "Information, trade and common knowledge". Journal of Economic Theory. 26 (1): 17–27 (1982).

15. Smeraldi, Fabrizio, and Pasquale Malacaria. "How to spend it: optimal investment for cyber security." Proceedings of the 1st International Workshop on Agents and CyberSecurity. ACM, 2014.

16. Spanos, Georgios, and Lefteris Angelis. "The impact of information security events to the stock market: A systematic literature review." Computers & Security 58 (2016): 216-229.Zhang, Jing, et al. "On the Mismanagement and Maliciousness of Networks." NDSS. 2014.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *econometrica*, 74(1):235–267.

Bilge, L. and Dumitras, T. (2012). Before we knew it: an empirical study of zero-day attacks in the real world. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 833–844. ACM.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Lane, P. W. and Nelder, J. A. (1982). Analysis of covariance and standardization as instances of prediction. *Biometrics*, pages 613–621.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2):1281–1301.

Tables

Table 1: Summary Statistics: Cross-Sectional Dataset

This table reports summary statistics for the continuous variables in the cross-sectional dataset used in the paper.

	N	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
Treatment (0/1)	3,365	0.01	0.12	0	0	0	0	1
Avg. Open Interest	3,365	496.08	2678.74	1	33.09	98.99	278.27	64,708
Log Avg. Open Interest	3,365	4.60	1.64	0	3.50	4.60	5.63	11.08
Avg. Volume	3,365	25.59	122.36	0	0.46	3.54	14.44	4212.92
Log Avg. Volume	2,853	1.62	1.94	-4.47	0.34	1.71	2.95	8.35
Market Value	3,363	10,913	37,421	3.51	445.09	1364.47	4,138	540,659
Log Market Value	3,363	7.33	1.87	1.26	6.10	7.22	8.33	13.20
Total Assets	3,365	42077	225,130	0.08	252.44	1,081	7,046	2,807,491
Log Total Assets	3,365	7.28	2.41	-2.55	5.53	6.99	8.86	14.85
Net Income	3,224	542.04	22,343	-3,347	-24.32	21.88	143.04	23,057
Log Net Income	2,021	4.70	2.02	-3.41	3.44	4.58	5.74	10.05
Total Liabilities	3,362	37,055	207,757	0.42	77.92	478.94	5,167	2,736,580
Log Total Liabilities	3,362	6.51	2.80	-0.87	4.36	6.17	8.55	14.82

Table 2: Summary Statistics: Difference-in-Differences Dataset

This table reports summary statistics for the continuous variables in the difference-in-differences dataset used in the paper.

	N	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
Treatment (0/1)	3,476	0.01	0.12	0.00	0.00	0.00	0.00	1.00
Pre-Open Interest	3,476	490.75	2,130	1.00	41.89	113.96	327.79	56,933
Log Pre-Open Interest	3,476	4.79	1.54	0.00	3.74	4.74	5.79	10.95
Post-Open Interest	3,476	501.77	2,522	1.00	44.89	123.13	318.62	78,531
Log Post-Open Interest	3,476	4.82	1.50	0.00	3.80	4.81	5.76	11.27
Log O.I. (Post-Pre)	3,476	0.03	0.98	-5.87	-0.50	0.01	0.57	5.77
Pre-Volume	3,476	28.94	124.06	0.00	1.21	5.52	18.66	3,310
Log Pre-Volume	3,302	1.71	1.90	-5.83	0.47	1.83	3.00	8.10
Post-Volume	3,476	26.78	122.93	0.00	1.16	5.20	16.64	3,977
Log Post-Volume	3,268	1.67	1.86	-4.84	0.54	1.81	2.88	8.29
Log Volume (Post-Pre)	3,160	-0.07	1.22	-5.80	-0.65	-0.10	0.51	6.30
Market Value	3,474	11,260	39,095	2.59	442.99	1,390	4,195	540,659
Log Market Value	3,474	7.33	1.91	0.95	6.09	7.24	8.34	13.20
Total Assets	3,476	41,137	221,642	0.09	253.09	1,096.66	7,197	2,807,490
Log Total Assets	3,476	7.29	2.39	-2.47	5.53	7.00	8.88	14.85
Net Income	3,333	554.63	2,265	-3,347	-23.55	22.94	144.85	23,057
Log Net Income	2,078	4.75	2.00	-3.41	3.48	4.61	5.77	10.05
Total Liabilities	3,472	36,099	204,531	0.07	83.29	496.45	5,282	2,736,580
Log Total Liabilities	3,472	6.53	2.77	-2.60	4.42	6.21	8.57	14.82

Table 3: Balance Test on Individual Covariates: Cross-Sectional Dataset

This table reports the difference in covariate means in the matched sample for the cross-sectional dataset with the specifications given in Tables 5 and 6. The raw mean in the largest possible subsample for each covariate is given in the first column. While the matching is unable to achieve perfect balance across all of the covariates simultaneously, this table shows that each specification leads to near-perfect balance on a different covariate. The consistency of the coefficient estimates in Tables 5 and 6 across these different specifications in significance and magnitude strongly suggests that the results are not driven by spurious variation in covariate balance.

	Raw Mean	(1)	(2)	(3)	(4)
Market Value	0.7782	-0.1331	-0.0094	0.2582	0.1825
Total Assets	0.5584		-0.2617	-0.1218	-0.1458
Net Income	0.6117			0.0774	0.0202
Total Liabilities	0.2611				-0.1457

Table 4: Balance Test on Individual Covariates: Difference-in-Differences Dataset

This table reports the difference in covariate means in the matched sample for the difference-in-differences dataset with the specifications given in Tables 11 and 12. The raw mean in the largest possible subsample for each covariate is given in the first column. While the matching is unable to achieve perfect balance across all of the covariates simultaneously, this table shows that each specification leads to near-perfect balance on a different covariate. The consistency of the coefficient estimates in Tables 11 and 12 across these different specifications in significance and magnitude strongly suggests that the results are not driven by spurious variation in covariate balance.

	Raw Mean	(1)	(2)	(3)	(4)
Market Value	0.8335	-0.0870	0.0126	0.1881	0.0426
Total Assets	0.6226		-0.2797	-0.1762	-0.2880
Net Income	0.6117			0.0613	-0.0440
Total Liabilities	0.2611				-0.2990

Table 5: Cross-Sectional Estimation: Log Open Interest

This table reports the average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms over the two months preceding the data breach, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. In this table, the dependent variable is identical across all models, but each column reports the ATE with additional covariates included in the propensity score matching. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.5528**	0.3677**	0.7539***	0.7006***
	(2.14)	(2.18)	(4.02)	(4.23)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets		Y	Y	Y
Net Income			Y	Y
Total Liabilities				Y
Observations	3,363	3,363	2,019	2,016

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Cross-Sectional Estimation: Log Volume

This table reports the average treatment effect of corporate data breaches on the log trading volume in put options for target firms over the two months preceding the data breach, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. In this table, the dependent variable is identical across all models, but each column reports the ATE with additional covariates included in the propensity score matching. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.7980	0.5331	1.0103***	1.2798***
	(1.05)	(0.93)	(3.60)	(2.83)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets		Y	Y	Y
Net Income			Y	Y
Total Liabilities				Y
Observations	2,851	2,851	1,727	1,724

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Cross-Sectional Estimation: Log Open Interest (Alternative Models)

This table reports the average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms over the two months preceding the data breach, matching on an indicator for the 4-digit SIC industry code for the firm, log market value as of year-end, log total assets, log net income, and log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a different weighting scheme: (1) propensity score matching, (2) inverse probability weighting, (3) inverse probability weighting with regression adjustment, and (4) regression adjustment. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.7006***	0.6347***	0.6347***	0.8998***
	(4.23)	(2.94)	(2.94)	(3.69)
Control Mean		4.5933***	4.5933***	4.5946***
		(121.56)	(121.56)	(121.70)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y
Net Income	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y
Observations	2,016	2,016	2,016	2,019

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Cross-Sectional Estimation: Log Volume (Alternative Models)

This table reports the average treatment effect of corporate data breaches on the log trading volume in put options for target firms over the two months preceding the data breach, matching on an indicator for the 4-digit SIC industry code for the firm, log market value as of year-end, log total assets, log net income, and log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a different matching scheme: (1) propensity score matching, (2) inverse probability weighting, (3) inverse probability weighting with regression adjustment, and (4) regression adjustment. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	1.2798***	0.9711***	0.9711***	0.7731**
	(2.83)	(4.08)	(4.08)	(2.52)
Control Mean		1.8110***	1.8110***	1.8132***
		(38.03)	(38.03)	(38.10)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y
Net Income	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y
Observations	1,724	1,724	1,724	1,727

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Cross-Sectional Estimation: Log Open Interest (Alternative Windows)

This table reports the average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms, matching on an indicator for the 4-digit SIC industry code for the firm, log market value as of year-end, log total assets, log net income, and log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a different period of sample inclusion, from one to six months prior to disclosure of the data breach. t-statistics are reported based on robust standard errors.

	1 mo.	2 mo.	3 mo.	4 mo.	5 mo.	6 mo.
ATE	0.5842	0.7006***	0.6176***	0.2816	0.0838	0.1719
	(1.50)	(4.23)	(3.62)	(1.24)	(0.35)	(0.83)
SIC Industry	Y	Y	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y	Y	Y
Net Income	Y	Y	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y	Y	Y
Observations	1,884	2,016	2,052	2,083	2,146	2,156

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Cross-Sectional Estimation: Log Volume (Alternative Windows)

This table reports the average treatment effect of corporate data breaches on the log trading volume in put options for target firms, matching on an indicator for the 4-digit SIC industry code for the firm, log market value as of year-end, log total assets, log net income, and log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a different period of sample inclusion, from one to six months prior to disclosure of the data breach. t-statistics are reported based on robust standard errors.

	1 mo.	2 mo.	3 mo.	4 mo.	5 mo.	6 mo.
ATE	1.1293***	1.2798***	0.7210*	0.0466	0.4916	0.5141*
	(5.79)	(2.83)	(1.82)	(0.10)	(1.22)	(1.78)
SIC Industry	Y	Y	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y	Y	Y
Net Income	Y	Y	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y	Y	Y
Observations	1,519	1,724	1,812	1,868	1,943	1,975

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Difference-in-Differences: Log Open Interest

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. The outcome is the difference in log open interest between the periods $[t-480, t-240]$ and $(t-240, t)$ where t is the date of disclosure of the data breach. In this table, the dependent variable is identical across all models, but each column reports the ATE with additional covariates included in the propensity score matching. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.3234	0.2679***	0.2793***	0.3146**
	(1.45)	(3.29)	(3.94)	(2.33)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets		Y	Y	Y
Net Income			Y	Y
Total Liabilities				Y
Observations	3,479	3,479	2,069	2,066

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Difference-in-Differences: Log Volume

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average trading volume of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. The outcome is the difference in log trading volume between the periods $[t-480, t-240]$ and $(t-240, t)$, where t is the date of disclosure of the data breach. In this table, the dependent variable is identical across all models, but each column reports the ATE with additional covariates included in the propensity score matching. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.3626*	0.3044***	0.2300***	0.3425***
	(1.72)	(2.78)	(3.30)	(3.58)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets		Y	Y	Y
Net Income			Y	Y
Total Liabilities				Y
Observations	3,150	3,150	1,907	1,904

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Difference-in-Differences: Log Open Interest (Alternative Models)

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. The outcome is the difference in log open interest between the periods $[t-480, t-240]$ and $(t-240, t)$, where t is the date of disclosure of the data breach. In this table, the dependent variable is identical across all models, but each column reports a different weighting scheme: (1) propensity score matching, (2) inverse probability weighting, (3) inverse probability weighting with regression adjustment, and (4) regression adjustment. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.3146**	0.2971***	0.2971***	0.2614**
	(2.33)	(2.75)	(2.75)	(2.18)
Control Mean		-0.0028	-0.0028	-0.0040
		(-0.13)	(-0.13)	(-0.19)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y
Net Income	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y
Observations	2,066	2,066	2,066	2,069

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Difference-in-Differences: Log Volume (Alternative Models)

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average trading volume of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. The outcome is the difference in log trading volume between the periods $[t-480, t-240]$ and $(t-240, t)$, where t is the date of disclosure of the data breach. In this table, the dependent variable is identical across all models, but each column reports a different matching scheme: (1) propensity score matching, (2) inverse probability weighting, (3) inverse probability weighting with regression adjustment, and (4) regression adjustment. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)
ATE	0.3425***	0.4060***	0.4060***	0.1551
	(3.58)	(3.21)	(3.21)	(1.57)
Control Mean		-0.0212	-0.0212	-0.0221
		(-0.80)	(-0.80)	(-0.83)
SIC Industry	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y
Net Income	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y
Observations	1,909	1,909	1,909	1,912

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Difference-in-Differences: Log Open Interest (Alternative Windows)

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average open interest of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a comparison between a different period of sample inclusion: (1) $[t-360, t-180]$ vs. $(t-180, t)$, (2) $[t-420, t-210]$ vs. $(t-210, t)$, (3) $[t-480, t-240]$ vs. $(t-240, t)$, (4) $[t-540, t-270]$ vs. $(t-270, t)$, (5) $[t-600, t-300]$ vs. $(t-300, t)$ and (6) $[t-660, t-330]$ vs. $(t-330, t)$. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)	(5)	(6)
ATE	0.3165***	0.2783***	0.2679***	0.2756***	0.2361***	0.1422*
	(3.32)	(3.14)	(3.29)	(3.72)	(2.74)	(1.72)
SIC Industry	Y	Y	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y	Y	Y
Net Income	Y	Y	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y	Y	Y
Observations	3,443	3,429	3,474	3,467	3,444	3,418

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Difference-in-Differences: Log Volume (Alternative Windows)

This table reports the difference-in-differences average treatment effect of corporate data breaches on the log of the average trading volume of outstanding put options for target firms, with propensity score matching over the following covariates: (1) an indicator for the 4-digit SIC industry code for the firm, (2) log market value as of year-end, (3) log total assets, (4) log net income, and (5) log total liabilities. In this table, the dependent variable is identical across all models, but each column reports a comparison between a different period of sample inclusion: (1) $[t-360, t-180]$ vs. $(t-180, t)$, (2) $[t-420, t-210]$ vs. $(t-210, t)$, (3) $[t-480, t-240]$ vs. $(t-240, t)$, (4) $[t-540, t-270]$ vs. $(t-270, t)$, (5) $[t-600, t-300]$ vs. $(t-300, t)$ and (6) $[t-660, t-330]$ vs. $(t-330, t)$. t-statistics are reported based on robust standard errors.

	(1)	(2)	(3)	(4)	(5)	(6)
ATE	0.1092	0.3010*	0.3044***	0.4489***	0.4618***	0.3563***
	(0.58)	(1.88)	(2.78)	(3.83)	(3.74)	(2.67)
SIC Industry	Y	Y	Y	Y	Y	Y
Mkt. Val.	Y	Y	Y	Y	Y	Y
Total Assets	Y	Y	Y	Y	Y	Y
Net Income	Y	Y	Y	Y	Y	Y
Total Liabilities	Y	Y	Y	Y	Y	Y
Observations	3,049	3,071	3,158	3,174	3,188	3,172

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Top Word Predictors of Open Interest for Treatment × Post

This table reports the top words in the ITRC event description that predict a high level of log open interest in put options of target firms after they have been successfully attacked.

1	parties	341.78	17	november	19.82	33	early	0.59
2	fargo	246.41	18	today	16.19	34	healthcare	0.57
3	usernames	208.68	19	franchisees	12.61	35	local	0.48
4	continuing	182.89	20	copart	10.80	36	acquisitions	0.48
5	allegations	93.21	21	steps	6.16	37	associated	0.40
6	involved	81.28	22	database	5.25	38	find	0.17
7	mortgage	76.73	23	informed	5.11	39	thats	0.16
8	protect	64.21	24	administration	2.80	40	contractor	0.12
9	least	42.70	25	fraud	2.54	41	claims	0.08
10	provider	37.67	26	conclusions	1.95	42	complex	0.08
11	customers	37.65	27	called	1.56	43	ssns	0.04
12	indicates	35.27	28	fact	1.31	44	targeted	0.02
13	wells	25.02	29	experts	1.21	45	desktop	0.02
14	provided	22.98	30	evidence	0.76	46	appeared	0.02
15	date	22.00	31	ernst	0.60	47	paid	0.01
16	precaution	20.24	32	dark	0.60	48	intrusions	0.01

Table 18: Top Bigram Predictors of Open Interest for Treatment × Post

This table reports the top bigrams in the ITRC event description that predict a high level of log open interest in put options of target firms after they have been successfully attacked.

1	including names	351.67	21	access computer	15.79
2	wells fargo	298.79	22	accessed unauthorized	5.82
3	address itrc	282.80	23	cisco systems	2.67
4	usernames passwords	253.18	24	details turn	2.59
5	allegations data	129.29	25	complex attacks	2.56
6	attacks network	88.80	26	exposed usernames	2.23
7	information security	81.54	27	administration information	2.00
8	recently learned	56.54	28	december unauthorized	1.96
9	basic precautions	47.21	29	security experts	1.72
10	postal addresses	46.05	30	comcast denied	1.57
11	according statement	43.37	31	including sent	1.27
12	outside company	40.85	32	business relationship	1.08
13	desktop computer	37.15	33	cards south	0.99
14	date birth	34.88	34	provider theft	0.87
15	including credit	29.61	35	compromised according	0.70
16	accused stealing	28.68	36	behind series	0.68
17	account details	25.06	37	customers went	0.62
18	email address	21.53	38	area moved	0.60
19	accidentally emailed	20.12	39	forum today	0.56
20	claims million	16.38	40	appeared come	0.53