Information Lost

Will the 'paradise' that information promises, to both consumer and firm,
be lost on account of data breaches and loss of information?
The epic is playing out.

Conference Draft: Do Not Cite without Permission

Igor Karagodsky
PhD Candidate, Boston College and Economist Compass Lexecon
and
Catherine L. Mann
Rosenberg Professor of Global Finance, Brandeis University
CLMann@Brandeis.edu
(corresponding author)

Pre-conference Version: June 25, 2012
NBER Conference:
Economics of Digitization: An Agenda

Abstract

As consumers search and purchase on-line, firms gather more and more personal information. On the one hand, this information can enhance market efficiency and consumer surplus as firms tailor products to buyers. On the other hand, there is increased risk of personal information loss, either by accident or through theft. Starting with California law in 2003, legislation requires firms to reveal data breaches. Using data for 52 publically traded US companies that announced a data breach during 2003-2007, we conclude that there is a statistically significant negative impact of a data breach announcement on the cumulative abnormal equity return of the announcing firm. In the two days around the breach announcement, the mean CAR is -0.7%. We observe different effects for announcements by retail, finance, technology, and health sectors; mean negative CAR for health firms announcing a data breach is largest. On the other hand, evaluating the loss of shareholder value for individual firms, the dollar magnitude of the loss is generally small, Thus the equity market punishment of firms that lose data would appear to be too small to promote superior data protection. We discuss alternative ways to incentivize firms to protect data, including fines and legal action.

I. Introduction

The expanding scope of Internet use yields a widening array of firms with very large access to databases of information on individuals' search and transactions. The upshot is that buyers receive suggestions on complementary purchases, targeted news and advertising, which increase customer value, but also raise the potential for compromised privacy and ID theft. Similarly, firms have unprecedented windows into customer behavior and preferences, which can improve products and profits, but also raise the risk of losing customer information.

A U.S. state law, first introduced in 2003 in California as Senate Bill 1386, mandates that organizations that maintain personal information about individuals must disclose if the security of the information has been compromised. Moreover, the legislation stipulates that if there has been a security breach of a database containing personal data, the responsible organization must notify each individual for whom it maintained information. The law forced every firm doing business in California to comply. By 2007, most of the U.S. states had adopted similar versions of a security breach law. Disclosure of data breaches was a watershed piece of legislation; without disclosure, incidents of data loss could be ignored by the firm, even as the individual whose data are lost suffers the consequences, such as of unauthorized use of financial information.

When a firm discloses the fact that customer information has been lost, there are several possible channels through which corporate valuation could be affected. If the company is customer-facing, such as a retail firm, sales might drop as customers buy from competitors. If the company is a financial intermediary, such as a payment processor, it may be shunned or fined by other parts of the payment chain. If the company is a technology firm, corporate governance of its own activities may be questioned. If the company is in the health-care sector, its reputation may suffer. In all these cases, investors may shy away from or sell shares of the company, putting downward pressure on its equity price. If the loss in shareholder value is sufficiently large or sustained, firms may see an incentive to better protect customer data, or collect and retain different types of data.

Whether or not a firm will take action to reduce the probability or type of incidents of data loss depends not only on whether the equity market punishes the firm, but also on how and who bears the burdens of lost customer data. For example, the costs of notification and of ameliorating a data breach (for example, issuing new credit cards), as required by the California law, could exceed any discipline by the equity markets. Similarly, fines imposed within the self-regulatory hierarchy (for example between merchants, card issuers and payment processors) offers a disciplining device, as do fines levied by a regulatory agency such as the Federal Trade Commission. Finally, legal suits brought by those suffering the data loss could be sufficiently threatening, or actually costly enough, to encourage firms to enhance their data security or design their data systems differently.

We start, in the next section, by considering to what extent information and information security exemplifies the classic type of market imperfection—the deviation between social and private costs and/or benefits—which therefore, may elicit a regulatory response.

Then, in Section III, we focus on one specific kind of market response – whether the equity market punishes companies that lose customer data. We use stock market valuation around the time that a breach is announced to explore how equity markets react to a breach announcement. We use the metric of cumulative abnormal return, CAR, of a given company's stock price relative to the stock valuation of a mutual fund portfolio of its sector, as well as its stock price relative to broad measures of market performance (NYSE and NASDAQ composites), to evaluate whether there is an equity-market response to a breach announcement, and whether it is large and sustained. In this section, we also evaluate whether the CAR of a firm announcing a data breach is associated with the characteristics of the breach (number of records lost, type of data, how breach happened) and characteristics of the firm (size, sector).

Section IV reviews complementary strategies of market discipline, based on existing disclosure legislation and FTC examination, to change incentives toward protecting data from unwarranted intrusion. This section also considers the challenge of cross-border data flows and potential for forum shopping.

Section V concludes. We find that there is a statistically significant negative impact of a data breach announcement on the cumulative abnormal equity return of the announcing firm. Consistent with the theory section, we observe different effects for announcements for the retail, finance, technology, and health sectors. The statistical effect is most prominent when Social Security data are lost, and in the health and banking sectors. Yet, the absolute magnitude of the loss to stockholder equity is small and short-lived. Thus the equity market punishment of firms that lose data would appear to be too small to promote superior data protection. As a complimentary market discipline, we cite data that show that for the financial sector at least, the remediation costs associated with disclosed data breaches exceed (as large as….) the stock market discipline, thus suggesting that data disclosure laws and pressures to reallocate the burdens of data loss among data players may be an effective strategy.

II. On the market for information and information loss

Are there imperfections in the market for information such that social and private costs/benefits diverge? Do these imperfections warrant intervention by self-regulatory or governmental regulatory bodies or other adjudicating parties, such as the legal profession? Or is the market for information sufficiently robust that market discipline can generate the socially optimal outcome?

Numerous authors have taken up this question—usually in the context of the privacy of personal information. The several papers reviewed below offer a clear and concise discussion of the issues, although they do not reach a final assessment on the existence of market failure or what to do about it. Moreover, because the whole notion of whether there is a ‘right’ to private information differs between countries, there is not even a common starting point to the discussion, so obviously not to any policy strategy either.

Perfect markets, but not when aggregation yields externalities

In a perfect markets framework, full information and an instrument to protect it from unauthorized use, somewhat akin to the Arrow-Debreu world, yields the optimal outcome. As in the A-D world, keeping some information private limits the creation of individual-specific information-based products, and therefore creates inefficiencies. But, also, an incomplete set of instruments to protect information (or agreements on authorized use of information?) also puts us in the second-best world.

A key difference with the information market is that information can flow to third parties. Can an agreement between a consumer (data source) and the initial data receiver (firm) be made binding? That is, is the mapping between the protection instrument and the state of nature complete? Probably not since third parties aggregate their information with that aligned between the first two parties. Third party flows could generate either positive or negative externalities to the data source and data recipient.

This narrow point is indicative of the major difference with the information marketplace even in the complete markets framework. Because the value of information in aggregate (database) is greater than the sum of the parts (individual behaviors and transactions), the independence of the state of nature and the perfect mapping between state of nature and instrument does not hold. Both positive and potentially negative externalities can result from data aggregation even in the perfect markets framework.

Environmental model for externalities

Hirsch (2006) presumes that collecting personal information generates negative externalities. ‘There is a growing sense that the digital age is causing unprecedented damage to privacy…. digital economy businesses often do not bear the cost of the harms that they inflict’. Just as pollution as an externality is an outcome of production, so too is harm to privacy an externality of the information ‘production’ activity itself. There need not be any information loss to generate harm.

Hirsh continues with the environmental analogy and reviews the evolution of policy strategy from ‘command and control’ compliance to ‘second-generation’ or ‘outcome oriented’ policy whereby the regulated entities find their own cost-effective strategy to achieve the legislated goal.

While environmental economics offers some analogs for the information marketplace it is stretched because …. More evidence of positive externalities…

Tang, Hu, and Smith, Michael D. (2007) nicely model this kind of regulation, and find that mandatory regulation raises consumer prices and reduces firm profits, just as would be expected.

Trade-offs and limits to rationality model including externalities

Acquisti (2010) argues that the information marketplace is all about trade-offs. “In choosing the balance between sharing or hiding one's personal information (and in choosing the balance between exploiting or protecting individuals' data), both individuals and organizations face complex, sometimes intangibles, and often ambiguous trade-offs. … But trade-offs are the natural realm of economics.” What opens the door to incomplete markets, or regulation of some sort, is the limit to consumer rationality and transactions costs, which might affect the distribution of benefits and costs. If consumers don’t know the value of their information, they cannot do the trade-offs to achieve social optimum.

Romanosky and Acquisti (2009) use a systems control strategy to map three legislative approaches to reducing harm to privacy, which in their case happens when data are lost, not just (as in Hirsch’s case) when data are collected. Two of the three approaches draw from accident legislation: ex ante safety regulation in the context of information would include promulgation and adherence to, say, Payment Card Industry standards. But, they argue that ex ante standards focus on inputs (encryption) rather than outcomes (harm); so are not efficient. Ex poste liability law would include negligence in treatment of personal data. But, effectiveness of ex-post liability is reduced because courts have been unwilling to award damages based on the probability of some future harm coming as a consequence of a data breach. A third mechanism is information disclosure, such as the California data breach disclosure law. Information disclosure offers the greatest promise to close the gap between private and social outcomes, but consumer cognitive bias (misperception of risk) and transactions costs mean that the gap cannot be completely closed.

Romanosky and Acqusiti use their framework to outline an example of where cognitive bias and transactions costs problems have become less apparent and where disclosure has been key to that happening: The relationship between credit-card issuing institutions and firms that hold (and lose) credit-card data. They argue that information disclosure has enabled the internalization of the costs of remediation by the data holders (and losers). Why? First, a sufficient number of data breaches have occurred such that these costs have been quantified (to be discussed in Section IV below). Second, the number of affected intermediaries (card issuers) is sufficiently small that they have power, whereas individuals are too numerous so do not. Third, the chain of causation between data loss and required remediation is known because of disclosure. Therefore, the transfer the remediation costs from the card issuers to the loser of the data can be affected, and at least some of the externality internalized. (This does not address the costs to the individual card holders, however, so that not all the costs of data loss have been internalized.)

Externalities of aggregation: Asymmetries and multi-player frameworks

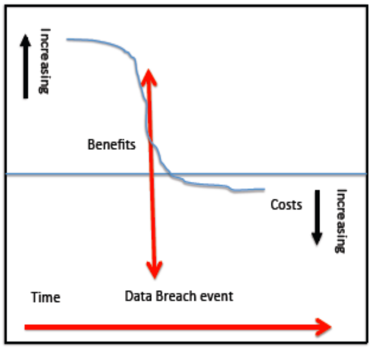
Most of the literature discussing externalities in the information marketplace uses a two-player framework– so-called data subjects (such as customers that ‘provide’ the information) and so-called data holders (such as payment processor that aggregate the information). Often there is a third player in the information marketplace, including data aggregators. The interactions between these three are where important market imperfections exist. In addition, there can be asymmetric externalities (costs vs benefits) depending on the type of information gathered, aggregated, and potentially lost. Individuals may get disproportionate benefits in some examples of information aggregation (free mobile phone apps), but bear disproportionate costs in cases of certain kinds of data loss (such as financial or medical information).

To start, presume that individual pieces of information have relatively small value, but databases (aggregations of information, over individuals or for an individual over time) are quite valuable. Databases enable tailoring of products and services that would not be possible with only individual pieces of data. There are plenty of examples: One of the first, Amazon aggregated individual buying decisions and linked it to a specific individual's order: If you buy book A, you might like to buy book B because other people did so. A more complex example of database value: OpenTable, the mobile phone app for restaurant recommendations and reservation requires the individual's location, a real-time database of restaurant reservations, and also a link to a database of comments and recommendations about the restaurants.

Without a doubt, benefits of tailored products and services accrue to the individual who enjoys both book A and book B, and finds an open table at a nearby Thai restaurant that is highly recommended. Are there privacy costs of providing the personal information in order to get these benefits? Perhaps. Benefits also accrue to the end-producers: the author of book B and the Thai restaurant that sells the additional meal. There is a third (and perhaps fourth) beneficiary, the owner of the database and the creator of the software: Amazon gets a cut and OpenTable gets a cut. So the allocation of the benefits to the various parties is something that the market is working out. If anything, the externalities of databases appear to go in favor of the individual whose own data at a point in time is not very valuable, but whose data is quite valuable when aggregated over time and/or over other individuals.

It is a fact that data are lost. Who bears the costs of data loss or misuse? Are the costs of a data breach disproportionately borne by individuals? In the previous example, suppose the database of customer buying preferences is breached or the database of which restaurants have open tables, or the location of the person. None of these losses of data appears to be inordinately offensive to either the individual or the associated firms. Perhaps the individual's information could be used inappropriately, such as e-mail or phone-text spam, and perhaps exposing the reading or eating preferences or geo-location could be viewed as an invasion of privacy. (And the example of Path Social's surreptitious downloading of the users' mobile phone directory means you never know what personal information you are revealing).[1] If these databases were breached, would there be evidence, would any individual even know if their taste in books or food were known? Therefore, if we draw a cost-benefit function associated with the existence of database, it would look something like this: A lot of benefit to both sides and not much loss to anyone in the case of a data breach.

[1] Anger for Path Social Network After Privacy Breach - NYTime...



Do data differ?

Neither of the two examples incorporate the central concern that individuals have about data loss, which is ID theft. ID theft (associated with loss of social-security numbers and financial information, and perhaps medical information) has been the leading concern as measured by complaints to the Federal Trade Commission for XX years running.[2] Are costs and benefits of information aggregation for financial or medical information allocated differently by the market, as compared to the dynamic discussed in the previous example for books and restaurants?

Consider financial transactions. Individuals and merchants are the two end points of the financial transaction. A credit-card issuing institution (Bank of America) provides the tool to make the transaction. A fourth player is the intermediary, who processes payments (Heartland), and there may be yet another player who aggregates the financial transactional information into a database along with other information (about books and restaurants) (Choicepoint).

What are the benefits of financial databases? For individuals, there is a convenience factor associated with credit-card information retained by on-line firms that the individual frequents. As in the cases above, aggregated information exceeds the value of any individual's information. Databases of financial transactions enable financial firms to differentiate products and services. For example, they can develop customer loyalty by sending summaries of buying behavior to individuals using their transactions history, which is also useful to the individual. They can reduce their own costs by using transactions history of an individual to prevent fraudulent use of that card. The aggregation of transactional history of many individuals produces a real-time picture of spending in the economy, by income, region, story-type, etc, which represents a new product (VISA Insights).

The relative magnitude of these benefits across individuals, merchants, card issuers, payment processors, and information aggregators no doubt varies based on relative market power. But, as for non-financial databases, the allocation of benefits appears to be something the market can work out—externalities in the classic sense appear to be small.

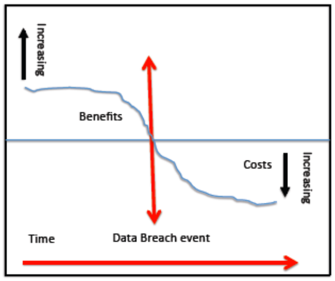
[2] FTC reference. Also, use Google Insights to track US vs. Europe.

What about the cost of data loss? For the individual, if credit card data are compromised, the individual may be inconvenienced, but there is a maximum $50 exposure (because of previous legislation). Given this individual limit, the cost of fraudulent use is borne by the card issuing institution. Although an important caveat is that if the financial information is a window into other key data, including social security numbers, the losses for the individual mount. This is ID theft.

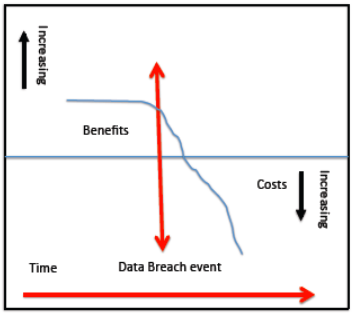
Therefore, the externality that may be evident in the market for financial information is that data loss can take place at either individual (lost card), the merchant (BJ Wholesaler, TJX), at the payment processor (Heartland) location, or aggregator (Choicepoint). In terms of the relative magnitude of loss, an individual lost card has a far smaller externality than a merchant (where many transactions are made, although size of the merchant matters) or payment processor and aggregator (who hold the databases of financial and other information).

The merchant has an incentive to avoid fraudulent card use by an individual because of charge-backs. But, it has less incentive to prevent data loss because of a smaller probability that they would be on the receiving end of subsequent fraudulent transactions undertaken using stolen credit card numbers. The payment processor or consumer information aggregator has no incentive to protect data. If they lose data (in the absence of a disclosure law), many fraudulent transactions may ensue, to which the card-issuing institutions need to respond, but linking the transactions back to the breached database would be very difficult. Free-rider/moral hazard problems because all players foist protection off to some one else.

Therefore, for financial transactions, the benefit-cost function may look like the previous one—everyone get benefits, but no one bears costs, except for ID theft where the individual gets both big benefits and big costs, as below.



Finally, for the card-issuing institutions, the benefit-cost function may like this—small benefits, but large costs in the case of data breach.



Considering these various shapes of the benefit-cost functions, we might expect that the market reaction to the data breach could differ depending on the sector of the firm announcing the breach. This is exactly what we want to investigate in the empirical analysis.

III. Equity-market assessments of data loss

"TJX disclosed the higher costs in its second-quarter earnings report, released yesterday. For that quarter alone, costs related to the data theft lowered TJX's profit by $118 million, or 25 cents a share, after accounting for taxes. … After the disclosure yesterday, shares fell … 8 percent below their level the day before TJX disclosed the security breach in January." Cost of data breach at TJX soars to $256m, By Ross Kerber, Boston Globe | August 15, 2007

Literature review

Five important analyses studying data breach incidents precede ours (Table 1). These papers all use the same methodology, cumulative abnormal returns, which is one of the approaches we use (more on this method below). These papers differ somewhat in the time horizon over which they calculate the 'normal' return as well as the window over which they calculate the CAR. They differ somewhat in the measure of the market against which to assess the abnormal returns: Kannan et. al (2007) considers the 'market' to be an SIC code or the S&P aggregate; Gatzlaff and McCullough (2010) use the value-weighted CRSP; the others use the NYSE or the NASDQ.

The previous studies differ most in terms of the time period of the analysis and the number of events. Campbell, et al. (2003) and Cavusoglu et. al. (2002) considered 43 and 78 breach incidents, respectively, during the period approximately 1997/7-2001/2. Acquisti et. al. (2007) consider 79 breach announcements over 2000-2006. Gatzlaff and McCullough (2010) examine 77 events between 2004 and 2006.

The predominant conclusion is that there is a negative, short term, statistically significant effect of a breach announcement on the financial performance of the announcing firm. When the type of data lost are considered, the conclusion appears only when classified customer information was lost. Campbell sums up the findings: "we do not find a significant market reaction when we examine security breaches that are not related to confidentiality. In contrast, we find a highly significant negative reaction for those breaches that relate to violations of confidentiality."

Firm characteristics may play a role, although the conclusions are mixed. Gatzlaff and McCullough who find strong and persistent effects up to 35 days after the event do not find that the type of data lost matters, although firm characteristics, such as higher market-to-book ratio exacerbate CAR whereas larger firms mitigate the negative impact. Cavusoglu et. al. find similar results that stock valuation of larger firms appear to be less affected. In contrast, Acquisti et. al. who also consider firm characteristics suggest that large firms might be more significantly affected by negative reports about their privacy practices as a result of irreversible damage to their reputation.

How the data were lost (e.g. mistake vs. hacker) plausibly may make a difference for the stock market's attitude, but apparently not, once other characteristics of the firm and size of data breach are considered.

Table 1: Summary of literature review of equity market effect of data breach

Author	Days to calculate market model	Market index	Interval for CAR calculation	# events in the dataset	Time period covered	Mean CAR % (check)by window (reported if significant)
Campbell, et. al.	121	NYSE AMEX NASDAQ	-1 to +1	43	1997-2000	−0.02
Acquisti, et. al	92	NYSE NASDAQ	0 to +1 0 to +2 0 to +5 0 to +10	79	2000-2006	-0.58 -0.46 0.21 1.3
Cavusoglu, et. al	160	NASDAQ	2 days Day 0 Day +1	78	1996-2001	Not signif –0.0086 –0.0123 (check magnitudes
Kannan, et. al	50	SIC codes control group S&P 500 index	-1 to +2 -1 to +7 -1 to +29	72	1997-2003	-0.65 -1.4 2.22
Gatzlaff and McCullough	245	Value-weighted S&P500 index	Day 0 0 to 1 0 to x in one day increments to 0 to +35	77	2004-2006	-0.57 -0.84 avg: -0.74

Building on this literature, our paper makes the following contributions:

- Sector benchmarks: A key contribution is to consider the firm's CAR relative to its sector-specific market benchmarks, as well as to the two broad-market benchmarks (NYSE, NASDAQ).
- Characteristics of data breach and CAR: We consider characteristics of the data breach, including firm size, amount and type of data lost, and how the data breach occurred (stolen vs. lost, insider vs. outsider), as well as sector of the data breach.

Our Data

This key ingredient to this analysis is the date and nature of the data breach, which was available from the DLDOS -open security foundation public database. Further information about the database can be found at the following URL: http://attrition.org/dataloss/. (No longer available for public download; access has been requested.)

52 data breaches were reported during the January 2003 to October 2007 time period. (Bank of America reported four different incidents, so the total number of breach announcing companies we analyze in this study is 48.) The database includes detailed information about the firms that experienced the data breach, including: country, business type and the sector where the firm is operating; as well as a description of the breach, including number of records lost, breach type such as lost, fraud, or stolen data, and whether the breach was by an insider or an outsider to the firm.

Other data include information on company equity returns, risk-free rate, and market value-weighted market indicators, such as NYSE and NASDAQ, all from the Wharton Research Data Services. In addition, sector aggregates for equity portfolio benchmarks by sector come from Ken French's website. More details on these data follow.

Statistical Design

There are two objectives of the analytical section of the paper. First is to estimate whether a breach announcement event is associated with an abnormal return to the share value of the announcing firm. Second is to analyze the relationship between the cumulative abnormal return of a firm's share value and various characteristics of the breach announcement.

The first step constructs the firm's abnormal return as the difference between its expected (e.g. predicted through estimation) return and its actual return in the market for a particular day. The cumulative abnormal return is the accumulation, over some time window around the data breach announcement, in the abnormal daily returns.

We start with the general formula for calculating the abnormal return. Then we turn to three different calculations of the expected return, which is the key component of the abnormal return.

Calculating Abnormal Return

$$\text{Abnormal Return}_{j,k,t} = \text{Share Return}_{j,t} - \text{Calculated Expected Return}_{j,k,t}$$

$$\text{Share Return}_{j,t} = (\text{Share Value}_{j,t} - \text{Share Value}_{j,t-1}) / \text{Share Value}_{j,t-1}$$

The *calculated expected return* $_{j,k,t}$ represents the predicted returns to a stock *j* based on historical relationships between the specific stock *j* and broader market indicators *k* using data prior to the breach event. The key ingredient to the calculated expected return $_{j,k,t}$ are the *estimated parameters* $a_{j,k}$ and $b_{j,k}$.

$$\text{Calculated Expected Return}_{j,k,t} = a_{j,k} + b_{j,k} * \text{market indicator}_{k,t}$$

Estimating market parameters $a_{j,k}$ and $b_{j,k}$: Standard CAPM

Estimated parameters $a_{j,k}$ and $b_{j,k}$ are derived from a regression of stock j's return against market indictors *k* for an historical time window well before the breach event. The market indicators that we use are NYSE composite, NASDAQ composite, sector mutual funds, and Ken French portfolios that match our stock *j*'s. The historical time window over which we estimate the relationship between the stock and the market is -200 to -30 days counting from day 0, which is the day that stock *j* announced a data breach.

We use historical returns (firms' returns, market returns, and risk free rates) from the Wharton Research Data Services. Since most of the firms in the data base are public, when calculating their share returns, we account for the dividend announcements by using the holding period return option on the CRSP (Center for Research in Security Prices) website. The risk free rate represents a monthly Treasury bill rate distributed on a daily basis. In addition to share growth rates, the CRSP website reports each firm's total number of outstanding shares and the growth of market indicators. These indicators reflect the Value-Weighted Return (including distributions) of commonly used market composites such as the NYSE and NASDAQ. The time frame used to calculate the slope and intercept coefficient stops 30 days prior to the event to avoid the "gossip effect." Such a phenomenon occurs when investors change their behavior a short period prior to the event date due of rumors regarding the upcoming stock changes.

Therefore, $a_{j,k}$ and $b_{j,k}$ are the coefficients $\alpha_{j,k}$ and $\beta_{j,k}$ from the following regression:

$$\text{Expected Excess Return}_{j,k,t} = \alpha_{j,k} + \beta_{j,k} * (\text{Market Indicator}_{k,t} - \text{RFR}_t) + e_{j,k,t};$$
$$t = -200 \text{ to} -30$$

Using this method, there is an estimated $a_{j,k}$ and $b_{j,k}$, unique for each of firm. These $a_{j,k}$ and $b_{j,k}$, will be used to calculate the expected return for the stock, and the abnormal return for the stock.

Table 2 reports the estimated $\alpha_{j,k}$ and $\beta_{j,\ k}$ for each stock j using market indictor NYSE.

Company name	Beta	Alpha	Segnificant at 95% level
Bank of America	0.809515	0.044155	Yes
Ameritrade	1.597428	-0.06001	Yes
Bank of America / Wachovia	0.909902	-0.00561	Yes
Citigroup	0.951342	-0.02444	Yes
Bank of America	0.845712	-0.00578	Yes
M & T Bank	0.992993	-0.01153	Yes
American International Group	0.854579	-0.00029	Yes
Equifax Inc.	0.464974	0.052918	Yes
Chase Card Services	1.205912	0.088662	Yes
BMO Bank of Montreal	0.601862	0.075928	Yes
American Family Insurance	0.554189	0.049684	Yes
KeyCorp	0.736539	-0.03327	Yes
MoneyGram	1.709547	-0.0401	Yes
Talvest Mutual Funds	0.378127	0.024528	Yes
Halifax	0.476763	-0.01818	No
JPMorgan Chase	1.1557	-0.00223	Yes
Merrill Lynch	1.63055	-0.09474	Yes
Ameritrade	1.597428	-0.06001	Yes
Hartford Group	1.171217	-0.04346	Yes
Dollar Tree	0.662583	0.077613	Yes
Williams-Sonoma	1.206029	-0.11175	Yes
Gymboree	1.478702	0.24056	Yes
Starbucks	1.378458	0.045895	Yes
TJX Companies Inc.	1.002428	0.012311	Yes
CVS Corp.	0.856392	-0.01758	Yes
Gap Inc.	0.807073	-0.10318	Yes
Home Depot	0.908155	-0.06388	Yes
Blockbuster	1.043819	-0.10824	Yes
IBM	1.24202	0.130181	Yes
America Online	1.106194	-0.05053	Yes
Intuit	1.14083	-0.20429	Yes
MCI	0.077484	0.099323	No
McAfee	0.645002	0.039091	Yes
Hewlett Packard	0.725709	0.142677	Yes
Impac	1.924228	-0.48563	Yes
Cablevision	0.673902	0.009784	Yes
Verizon Wireless	0.865567	0.025792	Yes
Electronic Data Systems (EDS)	0.696615	0.018928	Yes
Western Union	1.10503	-0.02203	Yes
Ameriprise Financial	1.317326	0.367269	Yes
Aetna Inc.	1.215014	-0.00245	Yes
LabCorp	0.324451	0.001354	Yes
Pfizer	0.737753	-0.00046	Yes
Humana Medicare	1.166329	-0.00014	Yes

Estimating market parameters a_k and b_k : Ken French Sector portfolio benchmarks

It is most common to use the broad measure of the market, such as NYSE or, as did some of the authors reviewed earlier, the S&P500 and calculate a unique $\alpha_{j,k}$ and $\beta_{j,\ k}$ for each stock. However it may be of interest to address the breach announcing firms' stock behavior compared to other firms in its sector. This method provides an opportunity to identify a negative effect of the data breach announcement on firms' financial performances relative to other firms in the same sector.

For example, let us consider a case where there is a crisis in the financial markets. Suppose that this financial instability did not importantly affect the Retail, Tech, or Health sectors. In such case, if we observe a breach announcing company such as Bank of America and compare our results against market indicator we might not have a clear picture of the full consequent effect of the data breach announcement on Bank of America's performance. It is possible that we observe a strong negative effect of the breach announcement on the stock value of Bank of America relative to the NYSE or the NASDAQ composites. However, a closer consideration reveals that during that time period, the effect of the crisis in the financial market is much larger relative to the negative effect of the announcement. Therefore even though at first it may have seemed as the announcement effect was very substantial for the abnormal returns of BofA, a more careful review implies that, relative to the generally poor performance of finance sector, Bank of America was not doing as bad as appears.

On his website, Ken French publishes data for various market portfolios. The 48 industry mutual fund portfolios best matches our sample of firms: it differentiates between insurance and finance companies in SIC48. SIC49 also distinguishes between the hardware and software technology companies, but this is not a point of interest for our analysis. We use SIC30 for the healthcare mutual fund performance.

The following Table maps the firms to the Ken French portfolios. The first column of the Table describes the symbol of the company whose name is labeled in the following column. The number in the third column refers to the specific industry in the SIC disaggregation. The fourth column is associated with the number and the industry aggregation mutual fund within the SIC48. The fifth column describes the type of the industry where the firm is operating. The last column provides a more detailed information about the type of the industry as indicated in the forth and the fifth column.

Company Symbol	Company name	Ken French Sctor	Sector	Description
BAC	Bank of America	45 Banks	Banking	National commercial banks
AMTD	Ameritrade	47 Fin	Trading	Security and commodity brokers
BAC	Wachovia	44 Banks	Banking	National commercial banks
C	Citigroup	44 Banks	Banking	National commercial banks
BAC	Bank of America	44 Banks	Banking	National commercial banks
PBCT	People's Bank	44 Banks	Banking	Savings institutions
AMP	Ameriprise Financial	47 Fin	Trading	Security and commodity brokers
BAC	Bank of America	44 Banks	Banking	National commercial banks
WFC	Wells Fargo	44 Banks	Banking	Commercial banks
MTB	M & T Bank	44 Banks	Banking	National commercial banks
AIG	American International Group	45 Insur	Insurance	Fire, marine, property-casualty ins
ING	ING U.S. Financial Services	47 Fin	Trading	Security and commodity brokers
EFX	Equifax Inc.	34 BusSv	Business	Services - credit reporting agencies, collection services
ALL	Allstate	45 Insur	Insurance	Fire, marine, property-casualty ins
BK	Aflac	45 Insur	Insurance	Accident and health insurance
JPM	Chase Card Services	44 Banks	Banking	National commercial banks
BMO	BMO Bank of Montreal	44 Banks	Banking	Banks
TMK	American Family Insurance	45 Insur	Insurance	Insurance agents
KEY	KeyCorp	44 Banks	Banking	National commercial banks
MGI	MoneyGram	44 Banks	Banking	Functions related to deposit banking
CM	Talvest Mutual Funds	47 Fin	Trading	Investment offices
PJC	Piper Jaffray	47 Fin	Trading	Security and commodity brokers
HX	Halifax	35 Comps	Computer	Services - computer programming and data processing
JPM	JPMorgan Chase	44 Banks	Banking	National commercial banks
WU	Western Union	34 BusSv	Business	Services - misc business services
MER	Merrill Lynch	47 Fin	Trading	Security and commodity brokers
AMTD	Ameritrade	47 Fin	Trading	Security and commodity brokers
HIG	Hartford	45 Insur	Insurance	Insurance agents
DLTR	Dollar Tree	42 Rtail	Retail	Retail - variety stores
WSM	Williams-Sonoma	42 Rtail	Retail	Retail - home furnishings stores
GYMB	Gymboree	42 Rtail	Retail	Retail - gasoline service stations
SBUX	Starbucks	43 Meals	Restaraun	Retail - eating places
TJX	TJX Companies Inc.	42 Rtail	Retail	Retail - apparel & acces
CVS	CVS Corp.	42 Rtail	Retail	Retail - drug & proprietary stores
GPS	Gap Inc.	42 Rtail	Retail	Retail - apparel & acces
HD	Home Depot	42 Rtail	Retail	Retail - lumber & other building mat
BBI	Blockbuster	7 Fun	Entertain	Services - video rental
IBM	IBM	35 Comps	Computer	Office computers
TWX	America Online	35 Comps	Computer	Services - information retrieval services
INTU	Intuit	35 Comps	Computer	Services - computer programming and data processing
MCI	MCI	32 Telcm	Communication	Telephone communications
MFE	McAfee	34 BusSv	Business	Services - misc business services
HPQ	Hewlett Packard	35 Comps	Computer	Office computers
IMH	Impac	47 Fin	Trading	REIT
CVC	Cablevision	32 Telcm	Communication	Cable and other pay TV services
VZ	Verizon Wireless	32 Telcm	Communication	Telephone communications
SAI	SAIC	34 BusSv	Business	Services - research, development, testing labs
EDS	Electronic Data Systems (EDS)	34 BusSv	Business	Services - computer processing, data prep
AET	Aetna Inc.	8 Hlth	Healthcare	Healthcare, Medical Equipment, Pharmaceutical Products
LH	LabCorp	8 Hlth	Healthcare	Healthcare, Medical Equipment, Pharmaceutical Products
PFE	Pfizer	8 Hlth	Healthcare	Healthcare, Medical Equipment, Pharmaceutical Products
HUM	Humana Medicare	8 Hlth	Healthcare	Healthcare, Medical Equipment, Pharmaceutical Products

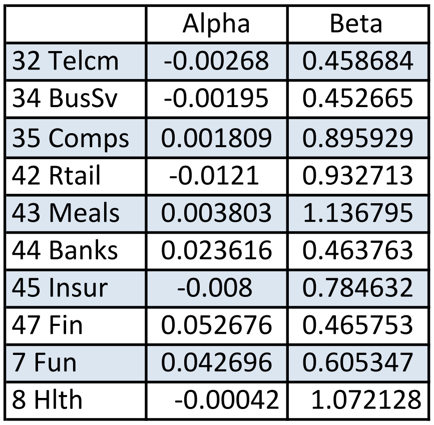
Estimating α_k and β_k for market indictor Ken French sector k follows this strategy: For each sector k, there are a set of breach announcing firms j. Each of these stocks j have an historical window of days -200 to -30 and a matching market return and risk free rates for those days, jt. Since the breach announcement is on different days for different firms, the right hand side variables will be matched for the -200 to -30 for the breach announcing firm and the estimated α_k , β_k will be for a sector k.

$$\text{Expected Excess Return}_{j,k,tj} = \alpha_k + \beta_k * (\text{Market Indicator}_{k,\ tj\ -}\ \text{RFR}_{tj}) + e_{j,k,\ tj;}$$

tj= -200 to-30 for each j; e.g. each tj is unique to the breach announcing firm j

This method derives an a_k and b_k for each sector rather than each firm as in the first method. These estimated a_k and b_k will be used to estimate expected returns for each of the firms j in sector k as an input to calculated the abnormal return for that firm.

Table (was Tbl 6) : Summary of alphas and betas for the Ken French disaggregation



	Alpha	Beta
32 Telcm	-0.00268	0.458684
34 BusSv	-0.00195	0.452665
35 Comps	0.001809	0.895929
42 Rtail	-0.0121	0.932713
43 Meals	0.003803	1.136795
44 Banks	0.023616	0.463763
45 Insur	-0.008	0.784632
47 Fin	0.052676	0.465753
7 Fun	0.042696	0.605347
8 Hlth	-0.00042	1.072128

The data analysis using Ken French's using the SIC48 disaggregation provides a closer look at the announcing company share's value relative to its related mutual fund. When we aggregate up all the breach effects using the Ken French portfolio, we expect to get results that are similar to the broad-market. So, this approach is a robustness check that confirms our results for the standard CAPM method.

Normal distribution:

Besides standard approaches as described above, I also use a slightly different method to calculate the expected return. Let us assume that a share return of any given firm might follow a normal random distribution around the market return. In such a case, the expected return would simply be calculated as the market return. Under the assumption that alpha=1 and beta=0, I used the NYSE and the NASDAQ composites as the market indicators for the expected return forecast.

In this section, I use the NYSE and NASDAQ composites as market indicators to approximate the expected return of the breach-announcing firms. Table (formerly Table 9) describes the slope and intercept coefficient that the regression analysis yields when studying the relationship of the NYSE and NASDAQ composites to the share returns of the our firms.

Table (was tbl 9): Summary of alphas and betas for different market indicators

includes health	Intercept Coefficient	Slope Coefficient	Slope Coefficient Significant at 99% Confidence Level?
NYSE	-0.00023	0.8721519	Yes
NASDAQ	7.2592E-06	0.7248956	Yes
Ken French	-9.465E-05	0.8506427	Yes

Original: does not include health firms…. Big difference….	Alpha	Beta
NYSE	0.000879	0.954661
NASDAQ	0.0001072	0.8175344

Analysis of Results

This paper addresses the way financial markets respond to a data breach announcements. To examine this effect, I calculated the abnormal return using the several methods to approximate the expected excess return as a robustness check. The following graph describes the abnormal return of the breach-announcing firms relative to the two broad market indicators (NYSE and NASDQ) and the Ken French sectors.

Chart 1: Abnormal Returns: Aggregates

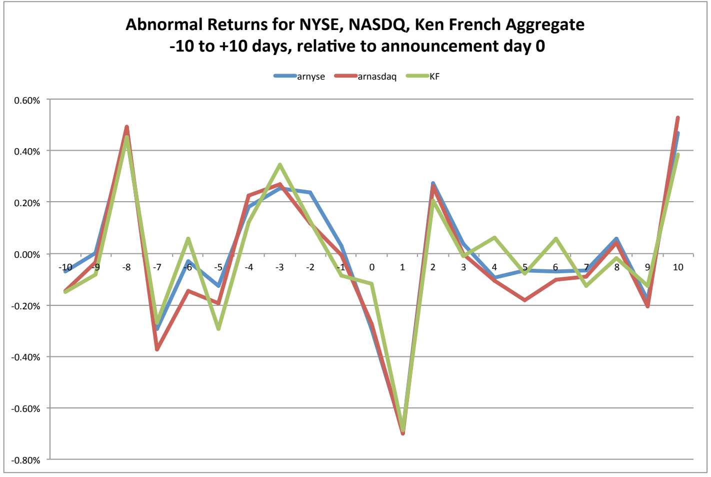
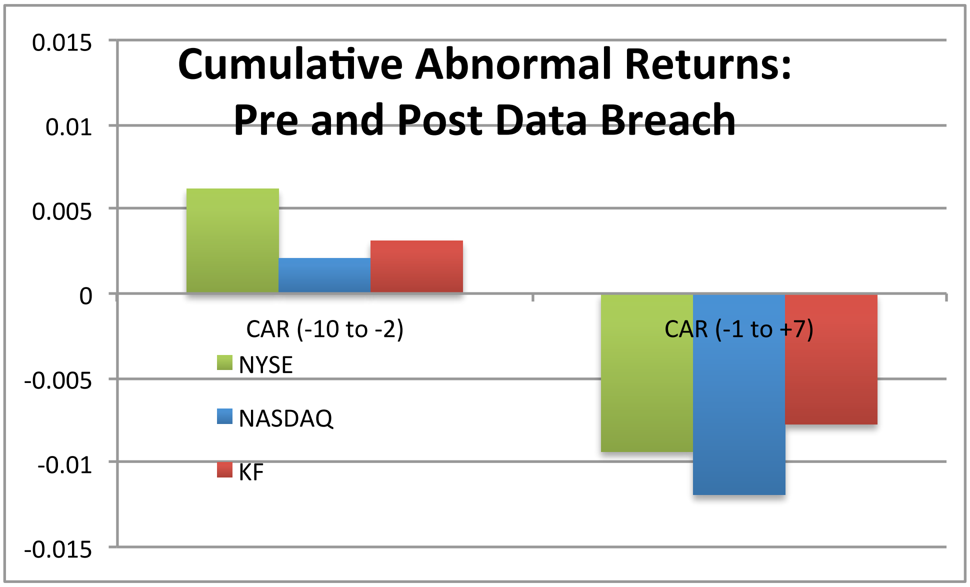


Chart 1 shows the abnormal return for 10 days prior and following the announcement event. Every data point on the graph indicates an average of the AR values for all firms in the dataset for a given day. For example, the blue line data point at day 0 represents the average of the AR values calculated using NYSE composite as the market indicator.

By all three measures, the AR starts to decline several days before the announcement day 0; but this decline is within the realm of recent daily experience. However, the decline in day +1 is larger than any abnormal return over the -10,+10 window. The abnormal return on day +1, after the announcement represents a share loss on that day of about 0.7%., which is a bit larger, but in the same ball park, as previous studies. The abnormal return on day +2 represents a big rebound (may be a consequence of 'buy on the dips' program trading). But otherwise, until the last day of the 10-day period, the abnormal daily return tends toward the negative terrain, suggesting some persistence in punishment for the data breach.

Suppose we compare the cumulative abnormal return over the (-10 to -2) day period (9 days cumulated) compared to the cumulative abnormal return in the (-1 to +7) day period (also 9 days cumulated to match the pre-announcement period and including day -1 in the data breach period because of the gossip effect). Chart 2 shows these before and after cumulative abnormal returns. It appears that the data breach effect persists for at least some time period. The cumulative loss in shareholder value is about 1 to 1.3 percent, somewhat larger but in the same ball park as the effects observed in previous analyses.

Chart 2:

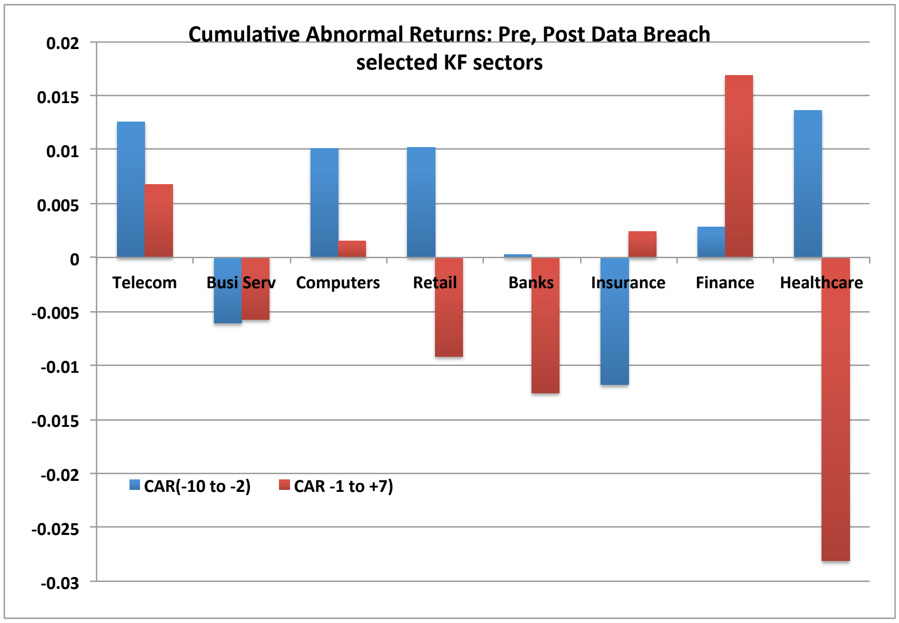


It is notable that the three measures behave rather similarly—a bit of a robustness check that the signs are the same pre and post data breach! . (some more discussion here on the difference, and whether is statistically significant or not) This is not so surprising for the two broad market indicators, but is perhaps less expected for the Ken French (KF) aggregate because the method of estimating the alphas/betas differs. This is a robustness check on the KF methodology—when aggregated across sectors the method and the sectors approximate the market.

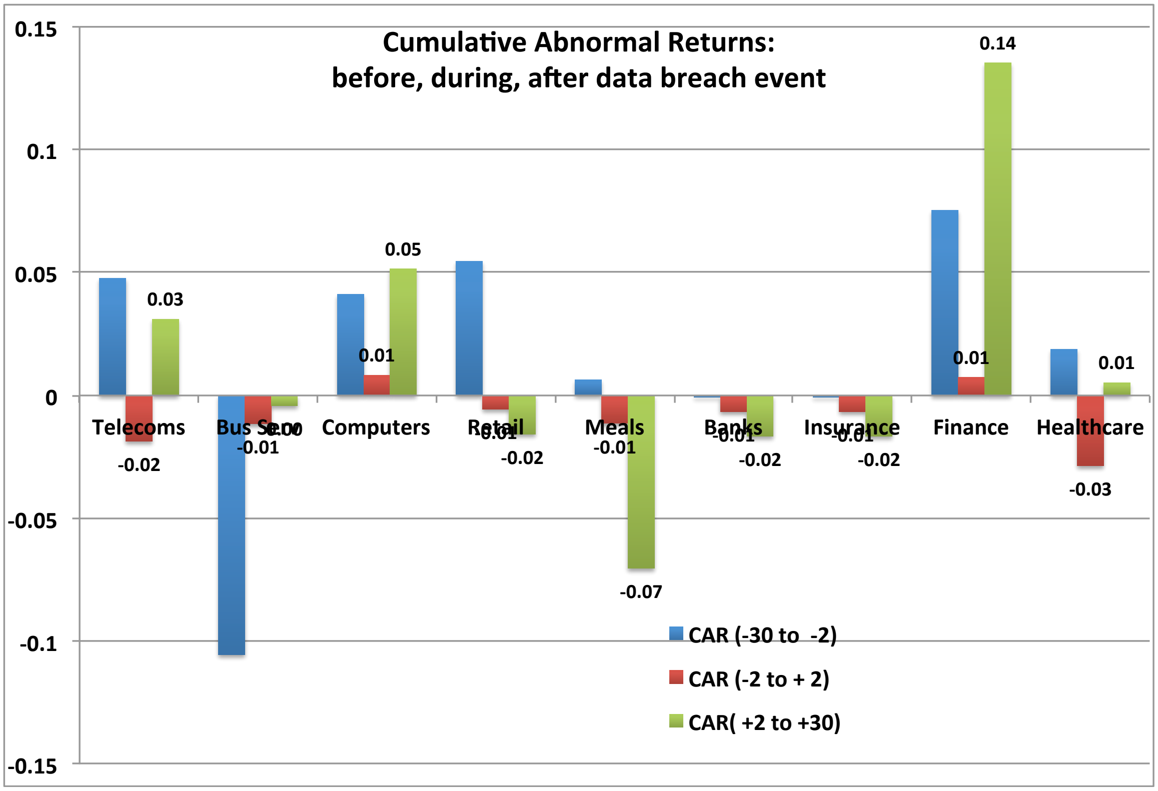
The purpose of the KF method was to examine behavior of disaggregated sectors. The theory section suggested that benefits and costs of data breaches might differ by sector. Chart 3 shows the cumulative abnormal returns for each sector for the pre and post data breach periods over the same period as above for most of the KF sectors (not shown are Meals and Entertainment).

First, most of the sectors do have the expected CAR behavior of negative CAR in the post data breach period. The negative CAR is particularly apparent for banks (including BofA, Citigroup, Wachovia, Wellsfargo) with a mean CAR loss of about 1.2% and for health care (including Humana, Labcorp) with a mean CAR loss of about 2.5%. These are two of the sectors where the personal information revealed by the breach might be of higher value than say in the telecom (Verizon) or computer (IBM, HP) sectors where the CAR post-breach remained positive, although much smaller than the CAR in these sectors prior to breach. For the banks, the cost of remediation might be high if credit cards are involved. Similarly, the data revealed by retail breaches (TJX, CVS), with mean CAR loss post breach of some 1% might be representing data on financial or even medical transactions and the cost of remediation rather than the clothes or personal care product buying patterns of the individual.

On the other hand, insurance (AIG, Allstate) and especially financial firms (Ameritrade, Ameriprise, Merrill Lynch) do not have the expected pattern. These show higher returns after the data breach event. (hum… because they foist the costs onto someone else or?)



Looking over a longer time horizon, around 30 days (G and M found significant negative CAR through day 34) before and after the data breach event and highlighting the CAR around the time of the event (day -2 to day +2) reveals impact of event (see telecons, computers, finance, health) but then a rebound. Negative CAR during event window, but then persistence and worsening of negative CAR for retail, meals, banks, insurance.



In sum, the Ken French sectoral analysis does reveal differences in the extent to which the stock market punishes firms. This suggests that sectoral analysis has traction and is worth exploring in more detail. In comparison to the findings of other researchers, we find that the relatively larger CAR could be coming from abnormal returns in bank and to the healthcare firms.

Dollar magnitude of stock market punishment

Are these results economically large? So far, we have assessed the cumulative abnormal return in percentage terms per share value. To get a more comprehensive understanding of the impact on the firm following the announcement in money terms, let's look at four representative firms from each sector (Bank, Retail, Computers, Health) and calculate the cumulative decreases in the firms' value 30 days following the breach announcement event. This is done by multiplying the CAR by the share value and number of shares outstanding.

Case 1: J.P Morgan Chase (Bank sector)
CAR (30 days following the event) = -0.1022032
Number of shares outstanding =3,461,700
Share value for 3/19/2007 = 47.58
Lose of value per share = 47.58* (-0.1022032) = -0.0486282826
Total value loss = 47.58* (-0.1022032)* 3461700 = -$168,336.53

Case 2: Gap Inc. (Retail sector)
CAR (30 days following the event) = -1.263334106
Number of shares outstanding = 815,925,000
Share value for 8/16/2007 = 16.27
Lose of value per share = 16.27*(-1.263334106)= -0.205544459
Total value loss = 16.27*(-1.263334106)* 815925=-$167,708, 870

Case 3: IBM (Computer sector)
CAR (30 days following the event) = -0.79170208
Number of shares outstanding = 1,690,088
Share value for 12/16/2002= 81.62
Lose of value per share = 81.62*(-0.79170208)= -0.646187238
Total value loss = 81.62*(-0.79170208)* 1690088= -$1,092,113,000

Case 4: Pfizer (Health Sector)
CAR (30 days following the event) = -4.01808%
Number of shares outstanding = 7,490,000,000
Share value for 7/23/2007= 25.03
Lose of value per share =25.03*(-4.01808%)= -0.0100572542
Total value loss = 25.03*(-4.01808%)* 7.49B = -$7.53B

As we noticed from the calculations above, even though we established that there exists a negative impact on the breach announcing firms' performance within a short time window around the announcing event, there is generally only a small percentage decrease in associated share returns. This small negative in percentage terms generates rather small dollar impact, at least in Cases 1-3. For Pfizer, both the percentage damage after the breach and the large number of shares outstanding leads to a huge financial impact.

Cases 1-3 suggest that firms have relatively little incentive to implement better data protection, at least to the extent that implementing data protection costs more than what the firm has lost in terms of shareholder value.

Characteristics of the Data Breach and the CAR

In this section, we explore the relationship between the characteristics of the data breach and the evolution of the CAR. In the regression analysis, the evolution of the cumulative abnormal return for each firm over a -2 to +10 or a -2 to +30 window is the dependent variable. Among the factors that might affect the CAR include the amount and type of information exposed, the size of the firm, and sector of the firm. The calculation of CAR (using the NYSE method or the KF sector method) might also be relevant.

- The CAR (-2 to +30) window yields results with more significance on characteristics. Is this because the sample includes more days, such that persistent effects can be observed?

- NYSE method vs. KF method of calculating key inputs to the CAR does not appear to be that important in that the coefficient estimates are not very different across the two regressions.

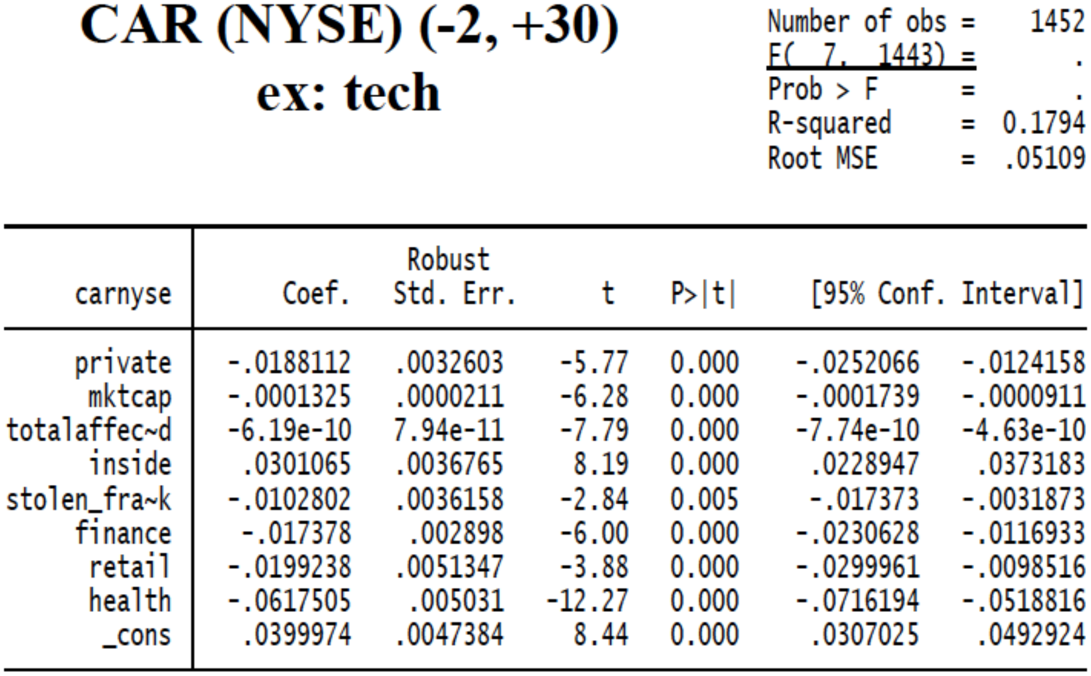
- By type and amount of data lost:
  - CAR more negative (~2%) when data lost include SSN data
  - CAR more negative (~very small) when number of record lost is greater.
    - Consistent with Poneman survey: costs rise with records lost.

- By nature of the breach:
  - CAR less negative (~3%) if breach caused by insider (vs by outside the firm. (e.g. computer goof is less bad than an intrusion.
    - This variable is significant and about the same magnitude both windows
  - CAR more negative (~1%) if breach due to stolen, fraud, hack (vs. just lost)
    - Results are consistent with Poneman survey. (elaborate)
- By firm characteristics:
  - CAR more negative for larger firms, as measured by market cap
  - CAR for health, finance, retail firms more negative compared to tech.
    - Health is significantly more negative than all other categories in shorter window, other categories not statistically different from each other.
  - Results are consistent with Poneman survey, which has indirect costs of churn etc very low for retail, and quite low for tech. Indirect costs are highest for finance and healthcare firms that lose data.

Table: Summary of right hand side variables

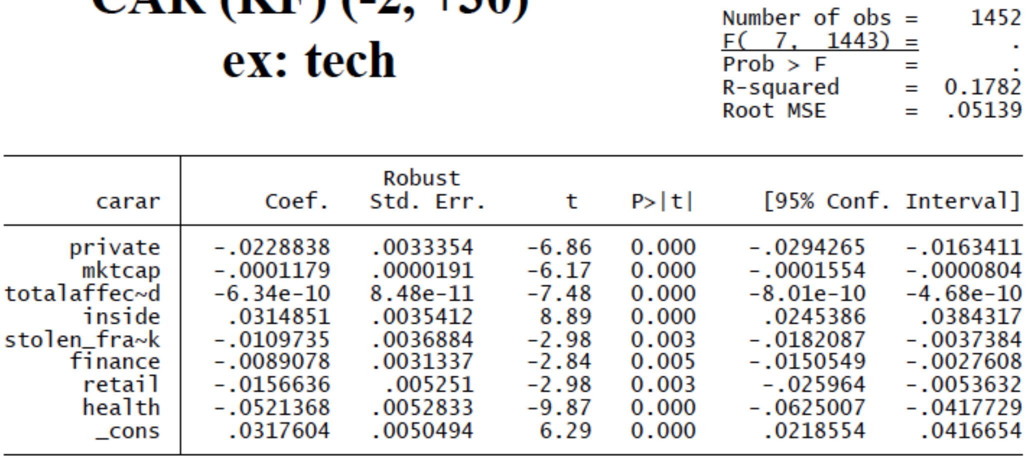
Independent variable	Definition	Type
Private	1 if stolen information includes social security, 0 if otherwise	Binary
Total Affected	The Number of effected customers	Continuous
Mktcap	market capitalization	Continuous
Inside	1 in case of a breach due to insider to firm, 0 otherwise	Binary
Stolen, fraud, hack	1 in case of breach due to these methods, 0 otherwise (e.g. lost)	Binary
Finance	1 in case of a finance (and bank??) company , 0 otherwise	Binary
Tech	1 in case of a tech company , 0 otherwise	Binary
Health	1 in case of a healthcare company, 0 otherwise	Binary

CAR (NYSE) (-2, +30)
ex: tech

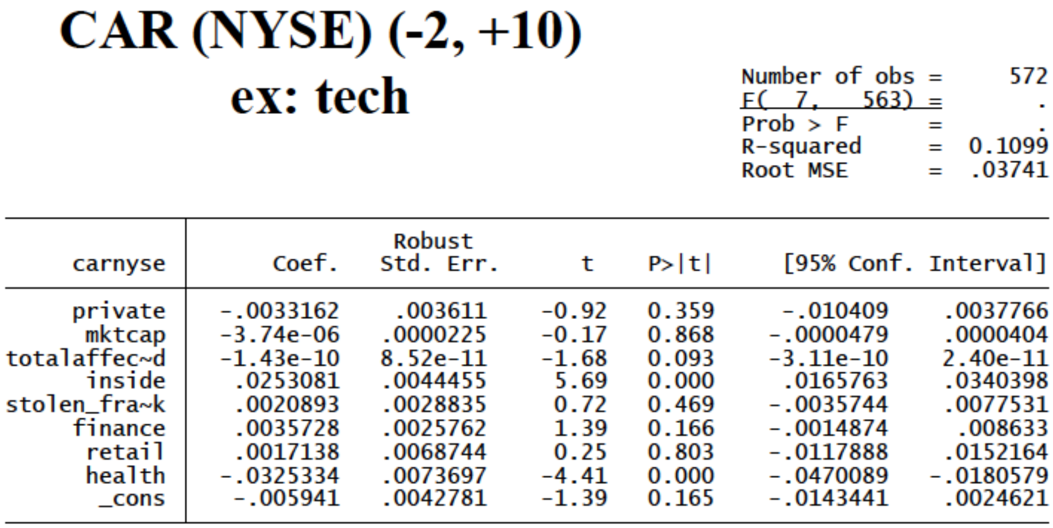


```
                                                  Number of obs =     1452
                                                  F(  7,  1443) =        .
                                                  Prob > F      =        .
                                                  R-squared     =   0.1794
                                                  Root MSE      =   .05109

------------------------------------------------------------------------------
             |               Robust
     carnyse |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |  -.0188112   .0032603    -5.77   0.000    -.0252066   -.0124158
      mktcap |  -.0001325   .0000211    -6.28   0.000    -.0001739   -.0000911
totalaffec~d |  -6.19e-10   7.94e-11    -7.79   0.000    -7.74e-10   -4.63e-10
      inside |   .0301065   .0036765     8.19   0.000     .0228947    .0373183
stolen_fra~k |  -.0102802   .0036158    -2.84   0.005     -.017373   -.0031873
     finance |   -.017378    .002898    -6.00   0.000    -.0230628   -.0116933
      retail |  -.0199238   .0051347    -3.88   0.000    -.0299961   -.0098516
      health |  -.0617505    .005031   -12.27   0.000    -.0716194   -.0518816
       _cons |   .0399974   .0047384     8.44   0.000     .0307025    .0492924
------------------------------------------------------------------------------
```

CAR (KF) (-2, +30)
ex: tech

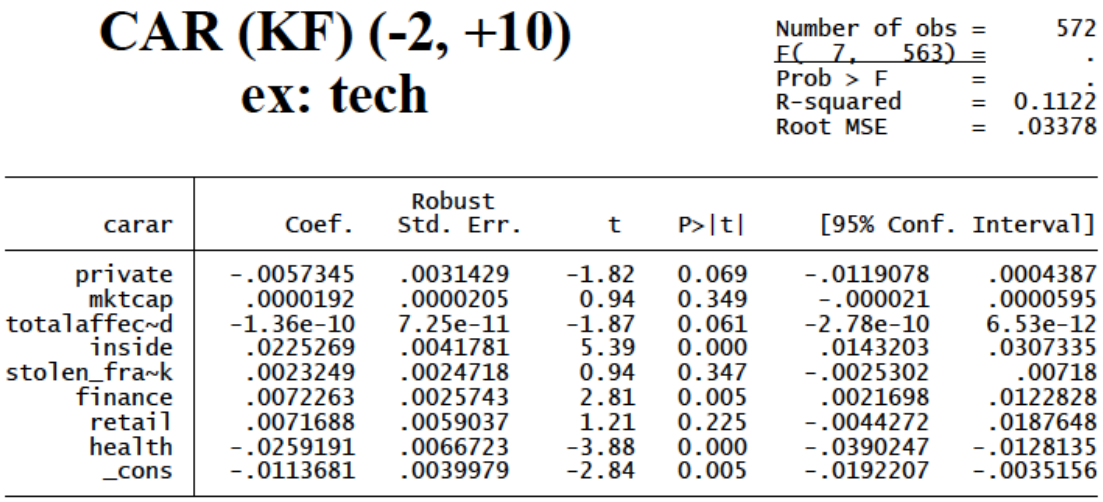


```
                                                  Number of obs =     1452
                                                  F(  7,  1443) =        .
                                                  Prob > F      =        .
                                                  R-squared     =   0.1782
                                                  Root MSE      =   .05139

------------------------------------------------------------------------------
             |               Robust
       carar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |  -.0228838   .0033354    -6.86   0.000    -.0294265   -.0163411
      mktcap |  -.0001179   .0000191    -6.17   0.000    -.0001554   -.0000804
totalaffec~d |  -6.34e-10   8.48e-11    -7.48   0.000    -8.01e-10   -4.68e-10
      inside |   .0314851   .0035412     8.89   0.000     .0245386    .0384317
stolen_fra~k |  -.0109735   .0036884    -2.98   0.003    -.0182087   -.0037384
     finance |  -.0089078   .0031337    -2.84   0.005    -.0150549   -.0027608
      retail |  -.0156636    .005251    -2.98   0.003     -.025964   -.0053632
      health |  -.0521368   .0052833    -9.87   0.000    -.0625007   -.0417729
       _cons |   .0317604   .0050494     6.29   0.000     .0218554    .0416654
------------------------------------------------------------------------------
```



CAR (NYSE) (-2, +10)
ex: tech

```
                                                       Number of obs =     572
                                                       F(  7,   563) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.1099
                                                       Root MSE      =  .03741

------------------------------------------------------------------------------
             |               Robust
     carnyse |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |  -.0033162    .003611    -0.92   0.359     -.010409    .0037766
      mktcap |  -3.74e-06   .0000225    -0.17   0.868    -.0000479    .0000404
totalaffec~d |  -1.43e-10   8.52e-11    -1.68   0.093    -3.11e-10    2.40e-11
      inside |   .0253081   .0044455     5.69   0.000     .0165763    .0340398
stolen_fra~k |   .0020893   .0028835     0.72   0.469    -.0035744    .0077531
     finance |   .0035728   .0025762     1.39   0.166    -.0014874     .008633
      retail |   .0017138   .0068744     0.25   0.803    -.0117888    .0152164
      health |  -.0325334   .0073697    -4.41   0.000    -.0470089   -.0180579
       _cons |   -.005941   .0042781    -1.39   0.165    -.0143441    .0024621
------------------------------------------------------------------------------
```



CAR (KF) (-2, +10)
ex: tech

```
                                                       Number of obs =     572
                                                       F(  7,   563) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.1122
                                                       Root MSE      =  .03378

------------------------------------------------------------------------------
             |               Robust
       carar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |  -.0057345   .0031429    -1.82   0.069    -.0119078    .0004387
      mktcap |   .0000192   .0000205     0.94   0.349     -.000021    .0000595
totalaffec~d |  -1.36e-10   7.25e-11    -1.87   0.061    -2.78e-10    6.53e-12
      inside |   .0225269   .0041781     5.39   0.000     .0143203    .0307335
stolen_fra~k |   .0023249   .0024718     0.94   0.347    -.0025302      .00718
     finance |   .0072263   .0025743     2.81   0.005     .0021698    .0122828
      retail |   .0071688   .0059037     1.21   0.225    -.0044272    .0187648
      health |  -.0259191   .0066723    -3.88   0.000    -.0390247   -.0128135
       _cons |  -.0113681   .0039979    -2.84   0.005    -.0192207   -.0035156
------------------------------------------------------------------------------
```

IV: Regulatory strategies and the international dimension

Regulatory strategy:

- Command and control: standardize privacy policies because people do not read them (Oussayef, 2008) but generally this is not direction authors are going.
- So-called second generation (environmental regulation model): Shifts to risk assessment and incorporating privacy into the product like 'green' products. Bamberger-Mulligan CISO survey says firms are doing this… "privacy by design".

- Notification/disclosure is key. Is ‘patchwork’ states ok, or need federal level? Bamberger-Mulligan CISO survey reports that patchwork is not a problem… ambiguity increases the incentive to protect to stay out of the newspaper.

- FTC’s role:
  - Fines can be large: $800,000 fine Spokeo under Fair Credit Report Law [3]
  - Mandated audit: Many years, big price-tag. Should change the balance between which is more costly: to protect vs. probability * cost of lose.
    - o But the advocacy of FTC has political lifespan

Legal recourse.

- No standing: Courts have not found for plaintiffs – hard to measure costs of losing information. Potential for future cost is not legal grounds. And, no pain and suffering award for the potential for future loss based on data breach. Have not determined that credit checks, etc are a cost. Yields problem of moral hazard.
  - Unless firm experiencing data breach did not employ ‘industry standard’ -- more likely courts find against, indeed especially if data used inappropriately. (YouRock)

- Class action suits: ‘Ambulance chasing’ is increasingly important. (Gibson Dunn) Poneman indicates that legal defense costs have risen steadily, from accounting for 6% of costs (2006) to 15% of costs in 2011. Increased legal costs and threats of legal costs increase incentives for firms to take evasive action/or protecting data to avoid becoming embroiled, even if the case won’t go against them.

- Contract law: terms of service based on privacy. This is where FTC is going.

- Property law: Not really in scope in the US since notion of personal data being private property not so much.

Market response;

- Indirect cost of loss is much bigger than direct remediation (Poneman)
  - o Direct cost per record $59 (2011) down from high $73
  - o Indirect costs (consumer churn, lost sales etc) much more important (avg $135 in 2011, down from 2009, but up from $78 in 2005.
  - o Churn varies by market sector with finance, health, telecoms highest and retail lowest.

- Role for reputation – Bamberger-Mulligan survey says is important

[3] Edward Wyatt, F.T.C. Levies First Fine Over Internet Data NYTimes.com, June 12, 2012,

- Banks do the remediation, and bear the cost (new credit cards, etc) if cannot pass back to who lost the data – TJX vs Heartland
    - Individual does not pay, but issue of ID theft.

The international dimension

EU approach to privacy and new law[4]

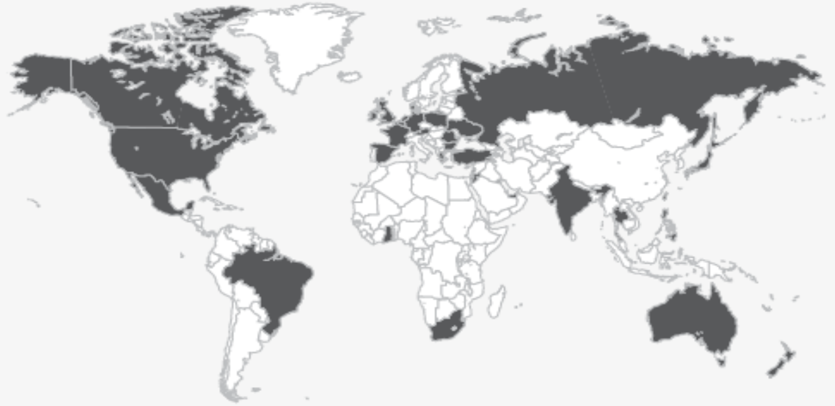
Do not track working group

- Value of personal data: 65% reduction in effectiveness of advertising comparing EU websites with personal w/o personal data.
-

Nationality of data loss

Verizon, DBIR 2012, http://www.verizonbusiness.com/about/events/2012dbir/index.xml

Figure 7. Countries represented in combined caseload



Countries in which a breach was confirmed

Australia
Austria
Bahamas
Belgium
Brazil
Bulgaria
Canada
Denmark
France
Germany
Ghana
Greece
India
Ireland
Israel
Japan
Jordan
Kuwait
Lebanon
Luxembourg
Mexico
Netherlands
New Zealand
Philippines
Poland
Romania
Russian Federation
South Africa
Spain
Taiwan
Thailand
Turkey
United Arab Emirates
Ukraine
United Kingdom
United States

We set a high mark in 2010 with 22 countries represented, but smashed that record in 2011 with a whopping 36 countries hosting organizations that fell victim to a data compromise.

[4] SOMINI SENGUPTA, Europe Weighs a Tough Law on Online Privacy and User Data ...
http://www.nytimes.com/2012/01/24/technology/europe-weighs-...
1 of

Nationality of perpetrator

Verizon, DBIR 2012, http://www.verizonbusiness.com/about/events/2012dbir/index.xml

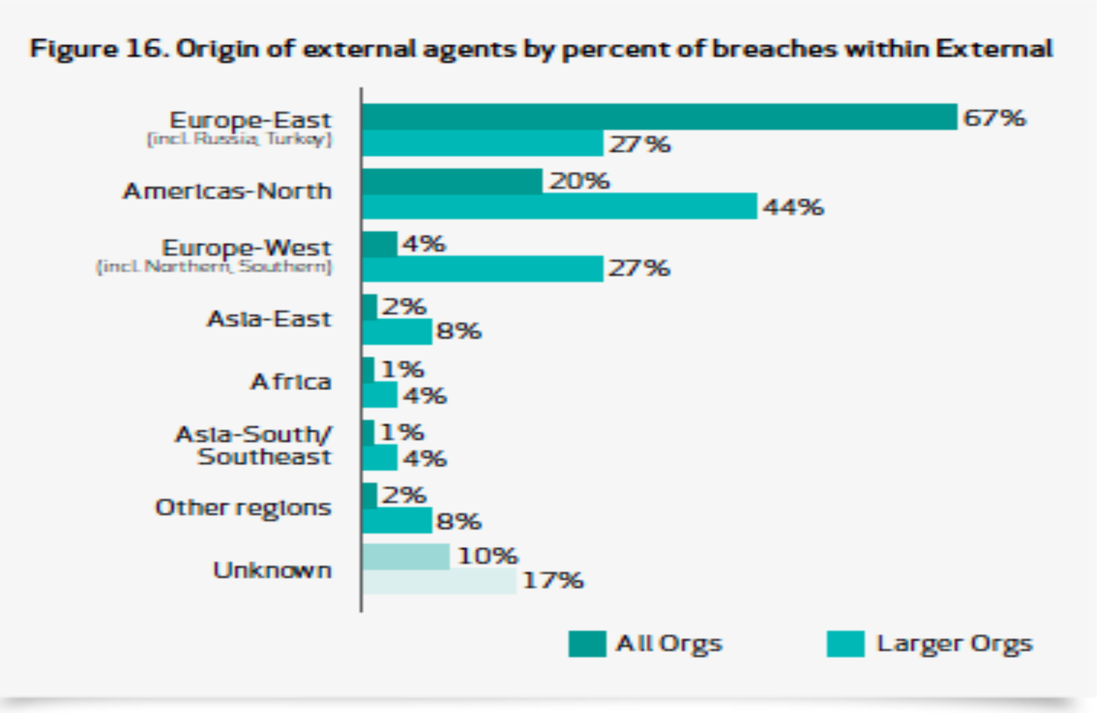


Figure 16. Origin of external agents by percent of breaches within External

V: Conclusion

TBA

Appendix: Share Values Follow Normal Distribution

For robustness check, an alternative method is to calculate expected return by assuming that a share return of stock j in our breach announcement dataset follows a normal distribution around the market return. In such a case, the expected return would simply be calculated as the market return. Therefore in the calculation of the a $_{\mathbf{j,k}}$ and b $_{\mathbf{j,\ k}}$, are assumed to be 0 and 1. It is worthwhile to investigate these assumptions.

In the methodology section, we discussed various approaches for expected return calculation. Regression analysis helps us understand if there exists statistically significant difference between expected return calculations using unrestricted firm specific intercept and slope coefficients, and assuming normally distributed expected return.

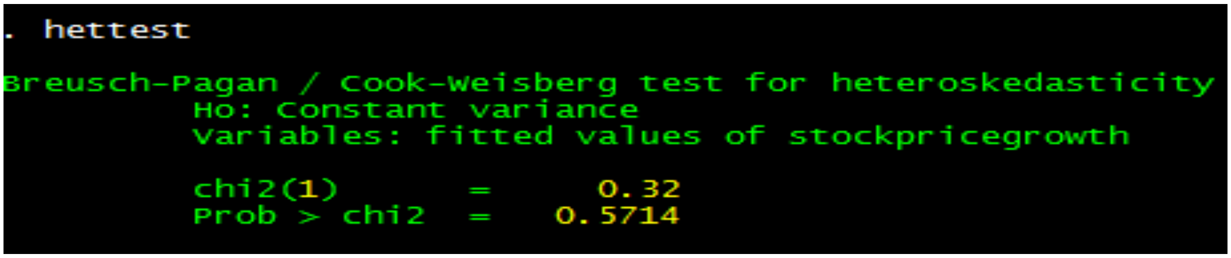
In order to test statistical significance between the two approaches outlined, we use the Housman test. The test evaluates the significance of a restricted estimator compared with an unrestricted estimator. We consider the restricted variable to be abnormal return calculated using the assumption that alpha=0 and beta=1. We assume that the market expected return follows a normal distribution around the market indicator return. NYSE composite serves as the market indicator in this model .The unrestricted variable in this case will be abnormal return with the calculated alpha and beta based on regression analysis of 170 days prior to the event. We use the following formula to calculate the test statistic which follows a chi-squared distribution.

$$F = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1-R^2_{unrestricted})/(n-k_{unrestricted}-1)}$$

In this formula the R squared – restricted indicates the R^2 for the restricted regression. The R^2 unrestricted represents the R^2 for the unrestricted regression.

The letter q indicates the number of restrictions under the null hypothesis that there is no significant difference between the estimators. K represents the number of regressors in the unrestricted and restricted regressions.

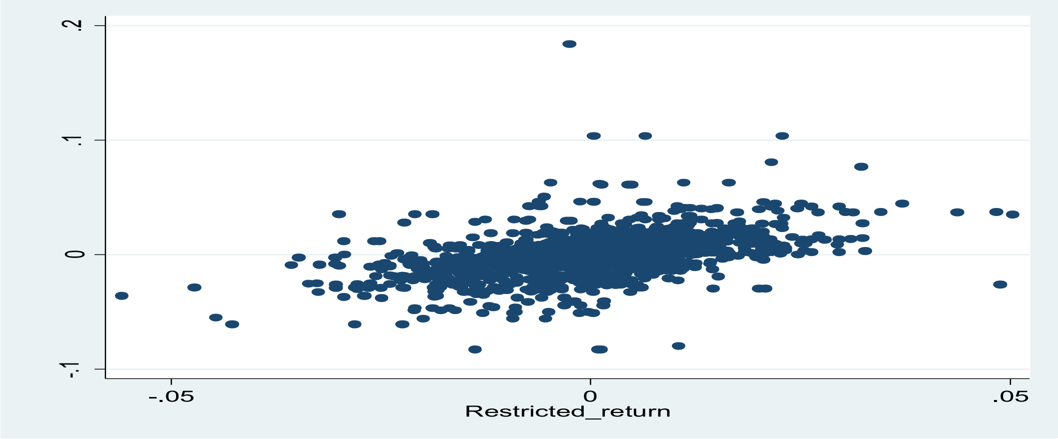
In order to apply the chi squared distribution for the results of the F statistic, there is a need to show homoskedastic errors. The h test tests the null hypothesis that for constant error variance. The following output describes the result of the test.



```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of stockpricegrowth

         chi2(1)      =     0.32
         Prob > chi2  =   0.5714
```

In this case the probability that there are homoskedastic errors exceeds 57 percent, therefore it's reasonable to apply the chi squared distribution for the Housman test result.

The following graph emphasizes the relationship of the stock return to the restricted return. In this case the restricted return is the return of the market indicator. When replacing values 0 and 1 for the intercept and the slope coefficients of the expected return, we have a normal distribution around the market indicator return.



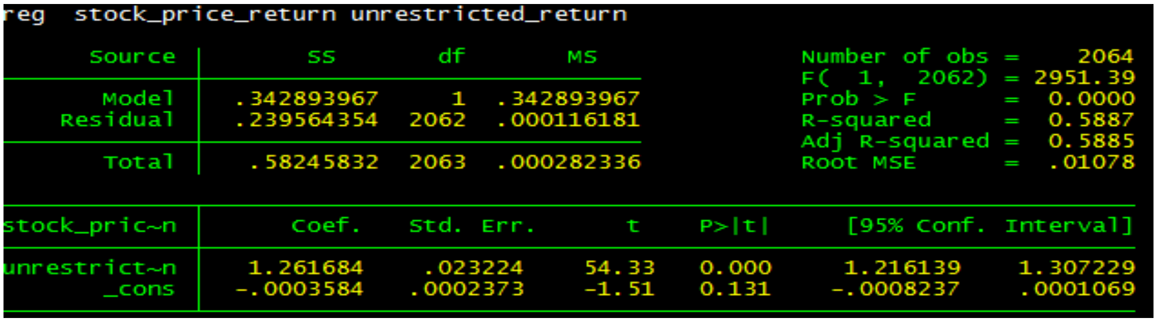
The relationship of the share and the restricted expected return

Finally, In order to achieve homoskedasticity, the error terms of the y variable should be constant across the X term. This graph emphasizes fairly constant relationship of the share return across the expected restricted.

To derive an F statistic result for the Housman test, we used previously calculated values for the R^2 in the restricted and unrestricted case. In this case q=2 as there are only two restrictions of α=0 and β=1. The unrestricted K value will be equal to one since there is only one regressor in the unrestricted regression.

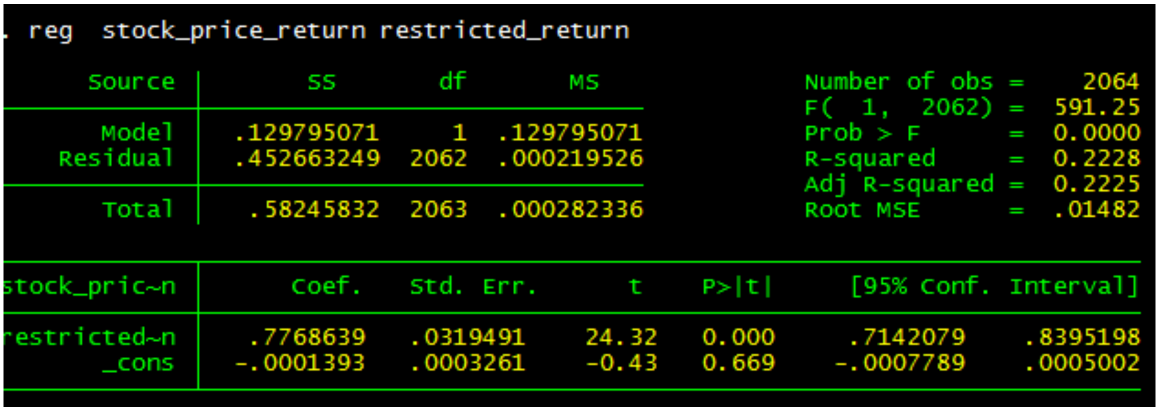
The unrestricted and restricted R squared is an output of the following two regressions of share price return on the restricted and unrestricted returns

.



```
reg  stock_price_return unrestricted_return

      Source |       SS       df       MS              Number of obs =    2064
-------------+------------------------------           F(  1,  2062) = 2951.39
       Model |  .342893967     1  .342893967           Prob > F      =  0.0000
    Residual |  .239564354  2062  .000116181           R-squared     =  0.5887
-------------+------------------------------           Adj R-squared =  0.5885
       Total |   .58245832  2063  .000282336           Root MSE      =  .01078

------------------------------------------------------------------------------
stock_pric~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unrestrict~n |   1.261684    .023224    54.33   0.000     1.216139    1.307229
       _cons |  -.0003584   .0002373    -1.51   0.131    -.0008237    .0001069
------------------------------------------------------------------------------
```

Non robust regression, unrestricted return as the regressor



```
. reg  stock_price_return restricted_return

      Source |       SS       df       MS              Number of obs =    2064
-------------+------------------------------           F(  1,  2062) =  591.25
       Model |  .129795071     1  .129795071           Prob > F      =  0.0000
    Residual |  .452663249  2062  .000219526           R-squared     =  0.2228
-------------+------------------------------           Adj R-squared =  0.2225
       Total |   .58245832  2063  .000282336           Root MSE      =  .01482

------------------------------------------------------------------------------
stock_pric~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
restricted~n |   .7768639   .0319491    24.32   0.000     .7142079    .8395198
       _cons |  -.0001393   .0003261    -0.43   0.669    -.0007789    .0005002
------------------------------------------------------------------------------
```

Non robust regression, restricted return as the regressor

Assuming homoskedastic errors and use the result for R squared from the restricted and unrestricted non robust regressions for the following calculation.

[(0.5887-0.2228)/2] / [(1-0.5887) * (2064 -1 -1)] = 0.000215717334

This result follows the chi squared over two distributions. Multiplying the result by two yields a result that follows the chi squared distribution with two restrictions.

2*{[(0.5887-0.2228)/2] / [(1-0.5887) * (2064 -1 -1)]} =

= 2*0.000215717334 = 0.000431434667

This result implies that we fail to reject the null hypothesis that there is no statistically significant difference between the restricted and unrestricted indicators. Therefore, it is reasonable to assume that the stock price returns will follow a normal distribution around its market indicator (i.e NYSE composite).

Appendix: PERMNO

Appendix table: Names of firms and their identifying PERMNO numbers.

Firm	PER MNO	Firm	PERMNO
Bank of America	59408	Halifax	67942
Ameritrade	84597	JPMorgan Chase	47896
Bank of America / Wachovia	59408	Western Union	91461
Citigroup	70519	Merrill Lynch	52919
Bank of America	59408	Ameritrade	84597
People's Bank	57510	Hartford Financial Services Group	82775
Ameriprise Financial	90880	Dollar Tree	81481
Bank of America	59408	Williams-Sonoma	83011
Wells Fargo	50024	Gymboree	78972
M & T Bank	35554	Starbucks	77702
American International Group	66800	TJX Companies Inc.	40539
Equifax Inc.	52476	CVS Corp.	17005
Chase Card Services	47896	Gap Inc.	59010
BMO Bank of Montreal	81284	Home Depot	66181
American Family Insurance	62308	Blockbuster	90337
KeyCorp	64995	IBM	12490
MoneyGram	90213	America Online	77418
Talvest Mutual Funds	85636	Intuit	78975
Piper Jaffray	10120	MCI	56565
McAfee	77976	Cablevision	68857
Hewlett Packard	27828	Verizon Wireless	65875

References

Articles

Acquisti, Alessandro (2010) “The Economics of Personal Data and the Economics of Privacy” draft dated November 21, 2010.

Acquisti, Alessandro, Allan Friedman, and Rahul Telang. (2006) Understanding the Impact of Privacy Breaches *35th Research Conference on Communication, Information and Internet Policy (TPRC).*

Bamberger, Kenneth A. and Deirdre K. Mulligan (2011) “ PRIVACY ON THE BOOKS AND ON THE GROUND,” STANFORD LAW REVIEW, Vol. 63:247, 247-315.

Campbell, K., L.A. Gordon, M. P. Loeb and L. Zhou, (2003) The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market, *Journal of Computer Security, Vol. 11, No. 3.*

Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. (2004) The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers, *International Journal of Electronic Commerce / Fall, Vol. 9, No. 1, pp. 69–104*

Gatzlaff, Kevin M. and Kathleen A. McCullough. (2010) The Effect of Data Breaches on Shareholder Wealth, *Risk Management and Insurance Review, Vol. 13, No. 1, 61-83*

Gibson Dunn (2012) “2011 Year-End Data Privacy and Security Update” February 7.

Hirsh, Dennis D. (2006) “Protecting the Inner Environment: What Privacy Legislation Can Learn from Environmental Law” *Georgia Law Review*, vol 41 no. 1, p1-62.

Kannan, Karthik, Jackie Rees, and Sanjay Sridhar. (2007) Market Reactions to Information Security Breach Announcements: An Empirical Analysis, *International Journal of Electronic Commerce / Fall, Vol. 12, No. 1, pp. 69–91*

Oussayef, Karim Z. (2008) “Selective privacy: Facilitating Market-Based Solutions to Data Breaches by Standardizing Internet Privacy Policies,” Boston University Journal of Science and Technology Law, Vol. 14:1, 104-131.

Romanosky, Sasha and Alessandro Acquisti (2009) “Privacy Costs and Personal Data Protection; Economic and Legal Perspectives” Berkeley Technology Law Journal vol 24 no 3, pp 1061-1101.

Tang, Zhulei; Hu, Yu Jeffrey; and Smith, Michael D. (2007) "Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor," Heinz Research. Paper 49. http://repository.cmu.edu/heinzworks/49

Data sources:

Kenneth R. French. Data Library, U.S. Research Returns Data, 48 Industry Portfolios [Daily].

DLDOS, open security foundation public database. Further information about the database is available at http://attrition.org/dataloss/dldos.html. [this database is no longer, as of first quarter 2012, available for immediate download].

Ponemon Institute LLC (2011) “2011 Cost of Data Breach Study United States’, Sponsored by Symantec, March.