Regression Model for the Impact of a Data Breach for a Financial Institution

Thomas Lee\textsuperscript{2}, Jason Hegland\textsuperscript{1}, Spencer Graves\textsuperscript{2}

1. Stanford University Law School, Stanford Security Litigation Analytics, Palo Alto California
2. VivoSecurity Inc., Los Altos CA,
Introduction

• An impactful data breach is a tail event often addressed in scenario analysis using expert judgment.

• The questioner-studies, which are a fixed amount per record, are often used.

• We have performed a regression analysis to aid expert judgment.

• We find variables considered but eliminated, as well as variables retained provide insights to expert.
# Sources & Methods

## Summary

### Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advisen</td>
<td>• Principle source</td>
</tr>
<tr>
<td></td>
<td>• Data breach cost</td>
</tr>
<tr>
<td></td>
<td>• Lawsuit probability</td>
</tr>
<tr>
<td>EDGAR</td>
<td>• Research cost in 10-K SEC filings</td>
</tr>
<tr>
<td>California Attorney</td>
<td>• Comparison analysis</td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>Maryland Attorney</td>
<td>• Comparison analysis.</td>
</tr>
<tr>
<td>General</td>
<td>• Estimate total number of data breaches</td>
</tr>
<tr>
<td>HHS</td>
<td>• Comparison analysis.</td>
</tr>
<tr>
<td>Interviews</td>
<td>• Challenge assumptions</td>
</tr>
</tbody>
</table>

### Tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>• Filtering</td>
</tr>
<tr>
<td></td>
<td>• Aggregation</td>
</tr>
<tr>
<td>R-Studio</td>
<td>• Modeling</td>
</tr>
<tr>
<td>Python</td>
<td>• Additional variables</td>
</tr>
<tr>
<td>Excel</td>
<td>• Manual corrections</td>
</tr>
</tbody>
</table>

### Process

<table>
<thead>
<tr>
<th>Step</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Correction</td>
<td>• Created independent and meaningful variables</td>
</tr>
<tr>
<td>BMA</td>
<td>• Screen many variables</td>
</tr>
<tr>
<td>LM</td>
<td>• Final cost modeling</td>
</tr>
<tr>
<td></td>
<td>• Establish SE</td>
</tr>
<tr>
<td>GLM</td>
<td>• Lawsuit Probability</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>• Combine cost model with lawsuit probability model</td>
</tr>
</tbody>
</table>
Establish Scope

Total Cost of PII Data Breach

- Impact from breach of PII that triggers legal reporting requirements:
  - PHI, PII, CHD, PFI

- Does not include:
  - Impact from intellectual property loss
  - Denial of services
  - Ransomware
  - Data corruption or loss
  - Fraud
Total Cost of PII Data Breach

**Breach Company**
- Fines & settlements
- Credit monitoring & privacy insurance.
- Business loss; theft of money & goods
- Call center
- Notification
- Remediation
- Investigation

**Public & Other Businesses**
- Business Loss
- Damage to personal credit
- Theft of money & goods
- Credit card replacement costs

Transfer via suits
Mitigate
Suitability of Data

No Significant Size vs Frequency Bias

Characterization of Size vs Frequency, for Advisen Data

Characterization of Size vs Frequency, whole United States, based upon Maryland Attorney General

Regression Model for the Impact of a Data Breach
Suitability of Data

Significant drop in Incidents per Year

Could Affect Lawsuit Probability

Data breaches are increasing

Unexplained drop-off in incidents

Maryland Breach Notices

- Year: 2018 (132)
- Year: 2017 (1080)
- Year: 2016 (792)
- Year: 2015 (482)
- Year: 2014 (333)
- Year: (3)
Data Transformation

Affected Count

Best transformed as log

Regression Model for the Impact of a Data Breach
Data Transformation

Calculated Total

Best transformed as log

Regression Model for the Impact of a Data Breach
Data Transformation

Lawsuits

Best transformed as log
## Best Cost Model

### Full Formula

\[
Cost = e^{\left(12-3.6 \times BrchY - 1.4 \times (IncidentOther) + 0.7 \times \ln(1 + Lawsuits) + \frac{BrchY \times \ln(Affctd)}{2}\right)}
\]

| Variable          | Description                                                                 | \(Pr(>|t|)\) |
|-------------------|-----------------------------------------------------------------------------|--------------|
| BrchY             | Data was breach, leading to 1) notification, 2) call center, 3) privacy insurance, 4) possible lawsuits | 4.9e-4       |
| Lawsuits          | Number of lawsuits filed as a result of the data breach.                    | 5.1e-2       |
| Affctd            | Number of people affected by the data breach (number of people with exposed PII data) | 1.7e-9       |
| IncidentOther     | Data breach was caused by any the following: Malicious Insider, Lost or Stolen Device or Accident. Data breach was NOT caused by a Malicious Outsider. | 6.0e-3       |

\[
R-Sqrd = 0.69
\]
Forecast Accuracy

10 observations from training set, NAICS=52

Regression Model for the Impact of a Data Breach
Forecast Accuracy

- 19 Observations found after model development
- Red bars are median costs,
- error bars are 2x range
Interpretation

• Based upon Variables
• Based upon Confidence Interval
Interpretation

Condensing of Incident Type

<table>
<thead>
<tr>
<th>Final Model</th>
<th>Initially Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malicious Outsider</td>
<td>Malicious Outsider</td>
</tr>
<tr>
<td>Malicious Insider</td>
<td>Malicious Insider</td>
</tr>
<tr>
<td>Lost/Stolen</td>
<td>Lost/Stolen</td>
</tr>
<tr>
<td>Accident</td>
<td>Accident</td>
</tr>
</tbody>
</table>

Breach Company

- Fines & settlements
- Credit monitoring & privacy insurance.
- Business loss; theft of money & goods
- Call center
- Notification
- Remediation
- Investigation

Transfer via suits

Public & Other

- Business Loss
- Damage to credit
- Theft of money
- Card replacement

Mitigate

The only costs that would be different
Interpretation

Based upon Variables

Breached=Y vs Breached=N

Same cost relationship between MO and Other suggests investigation cost is an importance difference between MO and Other.
Overview

Breached=Y

- Two types of PII breach:
  - *Malicious Outsider*
  - *Other*
- *Malicious Outsider* is 4x costlier
- Cost increases by sqrt of people affected
- One lawsuit doubles the cost

\[
\text{Cost}_{MO} = 4,500 \times (1 + \text{Lawsuits})^{0.7} \times \sqrt{\text{Affected}}
\]

\[
\text{Cost}_{\text{Other}} = 1,100 \times (1 + \text{Lawsuits})^{0.7} \times \sqrt{\text{Affected}}
\]
Interpretation

Based upon Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Importance</th>
<th>Interpretation, Guide the Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Small</td>
<td>We can draw upon lessons learned from past data breaches.</td>
</tr>
<tr>
<td>NAICS=52</td>
<td>Small</td>
<td>We can draw upon lessons learned across industries. Look towards costs that are the same across industries: investigation, notification, credit monitoring, call center</td>
</tr>
<tr>
<td>Company Demographic</td>
<td>Small</td>
<td>Suggests reputation damage may be less important</td>
</tr>
<tr>
<td>Data.Breached</td>
<td>Large</td>
<td>Consistent with Investigation cost being significant when variable is considered with Incident Type,</td>
</tr>
<tr>
<td>Data Type</td>
<td>Small</td>
<td>Suggests damage to customers is small. Look towards costs that are independent of data exposed: investigation, notification, call center.</td>
</tr>
<tr>
<td>Incident Type</td>
<td>Large</td>
<td>Consistent with investigation costs being significant.</td>
</tr>
<tr>
<td>Affected Count</td>
<td>Large, (Sqrt)</td>
<td>Should not use a constant multiplier per person affected since there is efficiency with scale. Don’t use record count, use people affected. Consistent with Investigation and Notification costs being major costs.</td>
</tr>
<tr>
<td>Lawsuits</td>
<td>Large</td>
<td>Focus on reducing probability of lawsuits for large data breaches. Most costs will be experienced over several years.</td>
</tr>
</tbody>
</table>

Regression Model for the Impact of a Data Breach

- **Breach Company**
  - Fines & settlements
  - Credit monitoring & privacy insurance.
  - Business loss; theft of money & goods
  - Call center
  - Notification
  - Remediation
  - Investigation

- **Public & Other**
  - Transfer via suits
  - Mitigate
    - Business Loss
    - Damage to credit
    - Theft of money
    - Card replacement
Interpretation

Based upon Confidence Interval

Cost of a Data Breach Affecting 10M People, caused by Malicious Outsider

Regression Model for the Impact of a Data Breach
Based upon Confidence Interval

Likelihood of Cost

- Median Cost, $23,688,721
- 80% Confidence, $122,557,359
- 90% Confidence, $289,370,426

Since a data breach affecting 10M people is rare, the 80% Confidence interval is CRAZY rare

A data breach affecting 10M People is a rare event across all industry

Regression Model for the Impact of a Data Breach
**Interpretation**

**Based upon Confidence Interval**

Most costs are within the control of the company and managed as part of the incident response plan.

**Regression Model for the Impact of a Data Breach**

- **Breach Cost**
  - Median Cost, $23,688,721
  - 80% Confidence, $122,557,359
  - 90% Confidence, $289,370,426

**Likelihood of Cost**

Expert Judgment can select confidence interval based upon the company's incident response plan vis-à-vis industry best practice and industry peers.

**Breach Company**

- Fines & settlements
- Credit monitoring & privacy insurance
- Business loss; theft of money & goods
- Call center
- Notification
- Remediation
- Investigation

**Public & Company**

- Business loss
- Damage to reputation
- Theft of money & goods
- Card recovery

**Most Important**

Transfer via suits

Mitigate
Interpretation
Based upon Confidence Interval

*Incident Response* best practice:
• All access logs turned on
• Access logs saved in a read-only manner
• Access logs saved in a uniform format
• Supporting policies, procedures, training and records
• Tools to aid investigation: Carbon Black, end-point detection and response technology
• Experienced third party evaluation of readiness
Conclusion

• It is possible to develop a model that characterizes the cost of a PII data breach

• The forecasting accuracy of such a model is acceptable, over a large range of Affected Count and incident types

• It is important to characterizes incident type independent of methods used to cause the data breach:
  • Malicious Insider
  • Malicious Outsider
  • Accidents
  • Lost Stolen

• Both variables eliminated and variables kept can inform expert judgment

• Expert judgment can assess most reasonable confidence interval by evaluating incident response plan