To notify or not to notify? Do organizations comply with U.S. data breach notification laws? An empirical study.

Bernold Nieuwesteeg[1]

Abstract

Data Breach Notification Laws (DBNLs) oblige organizations to notify personal data breaches. In theory, DBNLs mitigate damage after a data breach and incentivize companies to invest in information security. The regulatory enforcement of the DBNL is based on deterrence, because penalties are imposed, varying from $1,000 to $750,000 between states. It is uncertain whether DBNLs are deterrent enough to prevent organizations from concealing data breaches, especially because organizations suffer reputational costs from a notification. This study empirically tests compliance, by relating the adoption and characteristics of different U.S. DBNLs to actual observed data breach notifications based on the privacy breach clearinghouse dataset (2005-2012). After the adoption of the law, a 50% increase of notifications is observed. But, the absolute number of notifications is low, merely 0.05% of the U.S. companies notified. This indicates low compliance, possibly caused by high costs of notifying and low costs of concealing a notification. Unexpectedly, higher sanctions did not have an effect, but limited commensurability of the different sanctioning regimes prohibits a permanent statement.

This paper recommends enhancing DBNLs by increasing both the benefits of notifying and deterrence. Benefits are increased by incorporating rewards for good behavior by assisting companies in mitigating damage and continuously reward companies that are compliant by sharing knowledge about threats. Deterrence is increased by higher penalties and more stringent enforcement.

Key words: Data Breach Notification Law - empirical legal analysis - compliance – effectiveness - data privacy.

Endorsed by prof. Michel van Eeten (professor in the governance of cyber security at Delft University of Technology) and prof. Louis Visscher (professor in law and economics at Erasmus University Rotterdam)

[1] Bernold Nieuwesteeg is a PhD candidate in the economic analysis of cyber security regulation at the European Doctorate in Law & Economics, Rotterdam School of Law, Rotterdam University. He holds degrees in European Law as well as Systems Engineering, Policy Analysis and Management. The statistical analysis of this paper is based on the master thesis 'The Legal Position and Societal Effects of Security Breach Notification Laws (July 2013)' by the same author.

Electronic copy available at: https://ssrn.com/abstract=2431174

TABLE OF CONTENTS

INTRODUCTION

The need for a data breach notification law

During the last decennium, the world has experienced an unprecedented growth of the digital economy. The Internet accounts for 21% GDP growth in developed economies between 2006 and 2011. (Pélissié du Rausas, 2011) But, the embedment of the Internet also has made societies more vulnerable for data breaches.[2] There is a wide belief that quick awareness and open information about cyber security breaches reduces the social costs of cyber insecurity. Disclosure allows for fast mitigation of damage for third parties. Moreover, breach announcements serve as 'sunlight as disinfectant'; it is believed that other organizations will improve security practices to avoid breaches and foster cooperation. (Romanosky, Hoffman, & Acquisti, 2011; Schwartz & Janger, 2007)

[2] See for instance the Diginotar affair, which caused major disruptions in the Dutch governmental security systems. Diginotar was a Dutch certificate authority company for main Dutch governmental websites. Certificates were used to establish a safe connection between a consumer and the owner of a websites. The certificates were hacked by Iranian hackers through a man in the middle attack, which made Dutch governmental website extremely vulnerable. Diginotar decided not to disclose the breach and aimed for an internal fix, but failed to do so. However, the breach was discovered by a third party that eventually led to the companies bankruptcy.

Electronic copy available at: https://ssrn.com/abstract=2431174

Unfortunately, while disclosure of data breaches has public benefits, organizations incur private costs while making a notification. Notifying breaches causes administrative burdens and (perceived) high reputational costs. (Cavusoglu, Mishra, & Raghunathan, 2004; Goel & Shawky, 2009) Because of these private costs, organizations have incentives to conceal breaches in situations where the (perceived) costs of a notification are higher than its benefits. Hence, in theory, the market tends to show inefficiencies. The benefits of notifications lie at society while the cost of a notification lies at the individual.

46 U.S. states have adopted a DBNL in 2013.[3] California was the first state to adopt legislation in 2002 and other states quickly followed. The company Choicepoint was the first company to disclose a data breach affecting 145,000 people in accordance with Californian data breach legislation.[4]

The costs and benefits of complying with DBNLs

How are DBNLs supposed to work? The regulatory enforcement of the DBNL is mainly based on deterrence, because penalties are imposed for non-compliance. (Winn, 2014:1144) The deterrence theory provides a framework for the behavior of organizations in complying with regulations. (Becker, 1968) The deterrence theory is based on the assumption that complying with a regulation is to a large extent a cost benefit analysis. Organizations will comply if the cost of compliance is higher than the cost of non-compliance.

In the case of data breaches: the cost of compliance consists of direct costs. These are costs of notifying to consumer and mitigation and recovery of the personal data. Apart from that, there are also (perceived) reputational costs. The cost of non-compliance consists of the expected liability: the probability of detection times the fine imposed. If a data breach is detected, an organization will also incur direct costs and (more severe) reputation costs.

Cost of compliance	Cost of non-compliance		
Reputation costs	Likelihood of detection	*	Sanction
Direct costs (notification, mitigation, recovery)			Reputation costs
			Direct costs

Table 1: costs of compliance and non compliance

[3] 'Overview Security Breaches' (*NCSL*, 2013) <http://www.ncsl.org/issues-research/telecom/overview-security-breaches.aspx> accessed 2 February 2013; The variety in the design of a data breach notification laws has resulted in initiatives to pass a federal data breach notification law (European Commission, 2009). Also the European Union recently proposed initiatives for DBNLs.

[4] An example of an American DBNL can be found in the Texas Business and commerce code, § 521.03: "A person who conducts business in this state and owns or licenses computerized data that includes sensitive **personal information** shall disclose any breach of system security, after discovering or receiving notification of the breach, to any individual whose sensitive personal information was, or is reasonably believed to have been, **acquired by an unauthorized person.**"

Electronic copy available at: https://ssrn.com/abstract=2431174

US data provides an opportunity to empirically assess compliance

It is unknown to what extent organizations comply with DBNLs. In other words, whether the adoption of a DBNL actually increases the number of data breach notifications. U.S. data provides an opportunity to empirically assess effects of DBNLs, in two ways.

First, there is empirical data about the number of notifications before and after the adoption of DBNLs. This allows for measuring the effect of the adoption of DBNLs as such.

Second, DBNLs have different characteristics, which allows for measurement of differences between the laws. (Veltsos, 2012:193)

- Sanctions between between states differ between $1,000 (Wyoming) to $750,000 (Maryland) are imposed for non-compliance. Some states do not give a prefixed penalty, but give a right to a consumer to start a civil liability action. Thus, in theory, the deterrence of U.S. DBNLs differs.
- Some laws have a broader scope of personal data than other laws, which could result in more notifications.[5]
- Most laws focus at informing consumers, but some also require notifying an official supervising institution, which could lead to better registration.[6]

Research objective: empirically exploring compliance with DBNLs

This paper aims to empirically explore compliance with DBNLs. This is done by relating the adoption of U.S. DBNLs to the number of data breach notifications in the privacy breach clearinghouse database. Also, characteristics of the law are constructed and related to the number of notifications in the database.

Structure of this paper

In this introduction, the origins of DBNLs have been discussed. Second, the dataset is constructed. Third, characteristics of these laws are quantified on a dichotomous scale. Fourth, based on deterrence theory, hypotheses are drawn regarding the effects the characteristics of the law. Fifth, the hypotheses are tested by means of an U.S. dataset containing data breach notifications. This is done by basic descriptive analysis and a comparison of means and medians to discover rough patterns in the data, and a more sophisticated fixed effects regression model to control more systematically for variables that differ over time and between states. Sixth, the limitations of this analysis are addressed. Finally, it is concluded that the adoption of the law as such has an effect on the number of notifications, but that most characteristics, such as sanctioning and scope, did not prove to have an effect. This results in recommendations for both enhancing the law and enhancing empirical measurement of effects.

---

[5] DBNLs with an extended scope include medical information or a taxpayer identification number apart from ordinary personal information such as a credit card number. See for precise elements of scope the construction of this variable at the paragraph 'quantifying data breach notification laws'

[6] See for example the Idaho Statutes §§ 28-51-104 to 28-51-107. The supervising authority is in this case the Attorney General.

Electronic copy available at: https://ssrn.com/abstract=2431174

QUANTIFYING DATA BREACH NOTIFICATION LAWS

This paper aims to test these theoretical considerations by empirically exploring compliance with DBNLs. This is done by relating the adoption of U.S. DBNLs to the number of data breach notifications in the privacy breach clearinghouse database.

	2005	2006	2007	2008	2009	2010	2011	2012
U.S. States with a DBNL	11	27	36	41	45	45	46	46
	Developing period				Mature period			

Table 2: number of DBNLs per year (out of 50 U.S. States)

Also, characteristics of the law are constructed and related to the number of notifications in the database.

The analysis of the effects of U.S. DBNLs requires a quantification of the law. For this purpose, the laws itself and different legal sources from U.S. government institutions and law firms are consulted in order to distinguish and map the different aspects. It would be most desirable to quantify those aspects on an interval/ratio scale, but this proved to be very difficult, as it is hard to distinguish equal differences between multiple values per aspect. Therefore, several aspects of the law have been classified on dichotomic scale. If the particular law has this aspect, it is classified '1', else '0' (dummy variables).

A summary of the independent variables is displayed below. Appendix A displays the construction of these variables in depth. In addition to the independent variables, four control variables have been constructed: the GDP, Internet penetration rate, the number of firms and the population per state.[7]

Description	Label	Number of laws (out of 50)
Adoption of the law	*Has_Law*	46 (from 2011)
Maximum sanctioning above $50k	*Sanctioning*	14
Private right of action	*Private_action*	14
Wider scope than the general definition	*Scope_law*	24
Obligation to notify Attorney General	*Not_ag*	17
Obligation to notify Customer Credit Reporting Agency	*Not_custcredit*	29
GDP per state	*GDPcap_per_state*	Control variable
Internet Penetration rate per state	*Internetpenrate_per_state*	Control variable
Number of firms per state	*Firms_per_state_0512*	Control variable

[7] 'US GDP' (US Government Revenue, 2013) <http://www.usgovernmentrevenue.com> accessed 13 June 2013; 'Internet penetration rate' (PEW Internet) <http://pewInternet.org/Reports/2012/Digital-differences/Main-Report/Internet-adoption-over-time.aspx> accessed 12 June 2013; 'Firms in U.S. states'; 'Population in U.S. states' (Internet World States, 2013) <http://www.Internetworldstats.com/unitedstates.htm> accessed 12 June 2013.

Electronic copy available at: https://ssrn.com/abstract=2431174

Population per state	*Pop_per_state*	Control variable

Table 3: summary of the independent variables

CONSTRUCTION OF THE DEPENDENT VARIABLE

With some exceptions, U.S. supervisory authorities do not record the number of official notifications flowing from the legal obligation to notify.[8] The intended quantitative analysis thus has to rely on secondary data of organizations that collect and register data breach notifications. The dataset of a Californian nonprofit organization, called the Privacy Rights Clearinghouse, is used for this purpose. [9]

The dataset has opportunities. The database solely registers U.S. data breach notifications and all U.S. states between 2005 and 2012 are covered. There are 3554 breach reports. It is remarkable that less than 0.05% of the U.S. organizations are represented, while it is likely that a multiple of that suffered a data breach between 2005 and 2012. To compare: A recent study in the United Kingdom published that 88% of the companies searched had experienced data theft in 2009. (Ponemon, 2010:5)

The raw database is restructured into a list of 400 cases containing the number of notifications per state for each year between 2005 and 2012. (50 states containing 8 years of data[10]) The database is constructed by means of underlying sources.

Some sources are biased because they focus solely on one state, period in time or sector. Therefore, a separate dataset of selected sources is constructed. This database contains 2506 breach notifications. The number of breach notifications is adjusted for the number of firms in a state. This results in two dependent variables:

- *Not_per_firm* (number of notifications per million firms)
- *Not_per_firm_sel.* (number of notifications per million firms, taking only selected sources into account)

A detailed description of the construction of the dataset is given in appendix B.

HYPOTHESES BASED ON COSTS AND BENEFITS

Deterrence theory shows that there are incentives to conceal data breach notifications and non-compliance with the DBNLs.

The expected liability of DBNLs is low because sanctions are relatively low. Most sanctions do not exceed $50,000. Civil liability actions happen very rarely. [11]

[8] 'The Privacy Rights Clearing House DataBase' (*PrivacyRights.org*, 2013) <https://www.privacyrights.org/data-breach> accessed 1 February 2013.
[9] The goals of this organization are amongst others to "document the nature of consumers complaints" and "answer questions about privacy in reports, testimony, and speeches and make them available to policy makers, industry representatives, consumer advocates and the media." A part of this work is the maintenance and construction of a dataset containing data breaches.
[10] For this research, the official fifty US states will be used for analysis. Thus, the District of Columbia, Puerto Rico and the Virgin Islands are omitted from the database.

Electronic copy available at: https://ssrn.com/abstract=2431174

(Romanosky et al., 2011) Moreover, the probability of detection is low. Currently, there is no evidence that supervisory authorities put much effort in enforcing DBNLs. On top of that, it is relatively easy to conceal breaches, because of information asymmetries between supervising authorities and regulatees. (Anderson, 2001)

The expected liability for not making a notification is low. On top of that, the cost of making a notification is high. The costs of informing consumers, mitigation, recovery can add up to tens of dollars per record.[12] Also, reputational costs are (perceived as) high. Reputation damage can be up to 1-2% of an organizations turnover. (Goel & Shawky, 2009)

In general, to some degree, there would be compliance with the law. However, given strong incentives to conceal breaches, this effect is not expected to be large. The independent variables based on the characteristics of the DBNL are clustered to construct hypotheses for the effects of adoption of the law as such and therein elements of sanctioning, scope and the notification authority. Hereafter the effects are summarized in a conceptual framework.

Effect of the law as such

The adoption of the law as such leads to more notifications, because of the deterrent effect. Moreover, notifications are better registered after the adoption of the law. We would not expect this effect to be large, due to the incentives to conceal a breach. Eventually, one would see fewer notifications because of enhanced internet security.

Hypothesis 1: the adoption of data breach notification laws has a positive effect on the number of notifications

Effect of sanctioning

High sanctions increase the cost of non-compliance. The independent variables that relate to sanctioning are *Sanctioning* and *Private_action*.

Hypothesis 2a: laws with a **maximum sanctioning above $50,000** have a **higher** number of breach notifications.
Hypothesis 2b: laws with a **private right of action** have a **higher** number of breach notifications.

This hypothesis can be criticized by opposing arguments. For example, the safety culture theory says that people and organizations have to be rewarded instead of punished in order to learn or admit mistakes (Hudson, 1999:3). The very successful leniency policy

[11] There is a distinction between the civil liability actions concerning the actual loss of data (I.e. plaintiffs that alleged an unauthorized disclosure of their personal information) and the civil liability action for not notifying. The former happens often and has been analyzed by Romanosky et al. The latter does happen very rarely to the authors knowledge.
[12] The Dutch government estimates that notifying a consumer costs about $ 10 (€8.30), see the explanatory memorandum of the proposed data breach notification law. See <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2013/06/21/memorie-van-toelichting-meldplicht-datalekken.html> Consulted 2 February 2014.

Electronic copy available at: https://ssrn.com/abstract=2431174

in competition law has shown that a mix of high fines and no punishments incentivizes organizations to notify a cartel.[13]

Effect of scope

A broad scope causes more notifications because more cases of breaches fall under the definition of a breach that should be notified.

Hypothesis 3: laws with a **wider scope than the general definition** have a **higher** number of notifications.

An opposing argument for this hypothesis is, the willingness to notify would be lower if the scope is very wide (a notification fatigue).

Effect of notifying to a notification authority

Some states require a notification to a notification authority. It is assumed that an obligation to notify the Attorney General or customer credit reporting agency will result in more notifications in the dataset because those agencies are a center for data collection. This does not necessarily imply an actual increase of compliance with the law.

Hypothesis 4a: laws with an **obligation to notify the Attorney General** have a higher number of notifications
Hypothesis 4b: laws with an **obligation to notify the Customer Credit Reporting Agency** have a higher number of notifications

Concluding: a conceptual framework for effects of the law

The following framework is constructed to summarize the hypotheses which are constructed. Apart from that, the effects of notification fatigue, and the impact of increased security because of the 'sunlight as disinfectant' effect of notifications laws is added.

[13] European Commission 'Guidelines on the method of setting fines imposed pursuant to Article 23(2)(a) of Regulation No 1/2003 (Leniency Policy)' [2006] OJ L210/02. Organizations that provide information about a cartel in which they participated might receive immunity or reduction from fines.

Electronic copy available at: https://ssrn.com/abstract=2431174

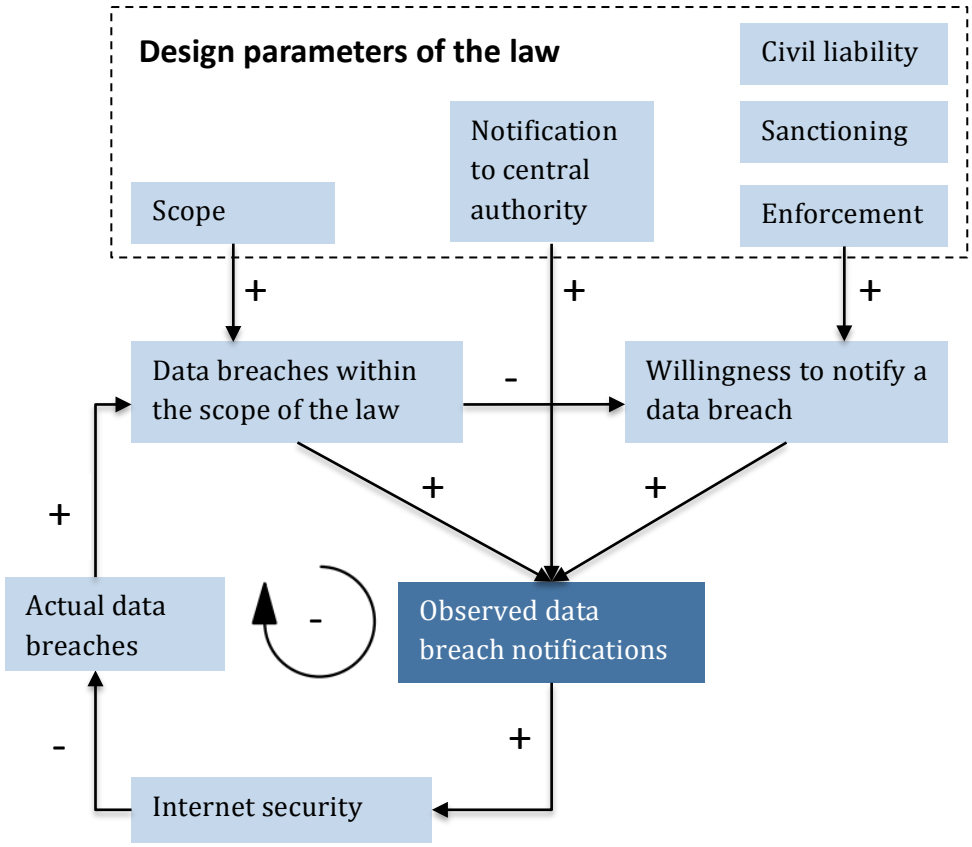

Figure 1: conceptual diagram of hypotheses

DESCRIPTIVES AND COMPARISON OF MEANS

Correlations are given for each variable, specifying whether there is a significant difference between the means and medians of the cases having '0' and '1'. Both a parametric test as well as a non-parametric Mann-Whitney test is performed for the whole (2005-2012), developing (2005-2008) and mature (2009-2012) period observed.[14]

- 'Yes' means that there are significant differences in any period and for every test.
- 'Mixed' means that only for some points in time a significant difference could be found.
- 'No' means that no significant difference could be observed. Descriptive statistics specify the direction of the difference.
- Positive means that '1' is higher than '0'.
- Negative means that '0'is higher than '1'.

Independent variable	Significant difference between '0' and '1'	Direction (from descriptive statistics)

[14] An extensive analysis of the correlations and descriptive statistics are displayed in appendix C.

Electronic copy available at: https://ssrn.com/abstract=2431174

	All sources	Selected sources	
Has_law	Yes	Yes	Positive
Sanctioning	No	No	-
Private action	Mixed	No	Positive
Scope	No	No	-
Not_ag	Mixed	Mixed	Positive
Not_custcredit	Mixed	Mixed	Positive

Table 4: descriptives and correlations

It is remarkable that only *has_law* shows a significant difference on all periods in time and at both tests. This means that the means and medians of average breach notifications per firm are significantly higher after the adoption of the law. Subsequently, there are no characteristics of laws that account for differences between notifications observed. This is especially interesting in the case of *sanctioning.* Based on the deterrence theory it is expected that higher sanctions lead to more notifications, and that thus the means and medians would differ.

FIXED EFFECTS REGRESSION MODEL

The comparison of means has some major drawbacks, which are partly solved by a fixed effects regression model. The comparison of means is vulnerable for differences in the number of breach notifications between states that are not caused by the differences between laws, but by other variables that are omitted from the analysis, apart from the number of firms in a state. The analysis is also vulnerable for effects over time. It is impossible to control for all variables that vary over time or between states, because it is impossible to identify all possible variables and to categorize them in a quantitative manner.

Moreover, the comparison of means indicates rough relations between the dichotomous independent variables and the dependent variable. A statistical model can improve the interpretation of the correlations. A possible solution for the problem of omitted variables is the fixed effects regression model, a standard econometric tool for longitudinal data.[15]

[15] The fixed effects regression functions as follows: one could imagine that the number of firms in a state determines the number of notifications in a state. A normal regression would indicate that large states have more effective laws, because large states generate more notifications. This is a false conclusion, because the variable *firms per state* is omitted from the analysis. The fixed effects regression measures the *changes* in the dependent variable from 2005 until 2012. A state with a large number of firms in 2012, such as Texas, already had a large number of firms in 2005. If there are no changes observed, than the adoption of the law in 2009 (probably) did not have an effect. In fact, the fixed effects regression assumes that omitted variables remain constant over time, because than any changes of the number of firms are not caused by the omitted variable. The same applies for variations over time, which are constant over states. The automatic control for these kinds of variables results in a more accurate model. However, the fixed effects regression does not control for variables that change over time and over states.

Electronic copy available at: https://ssrn.com/abstract=2431174

The model

Two models are constructed to capture the impact of the classifications of the law on the number of breach notifications per firm. The first model does not take the adoption of the law into account; it focuses entirely on the impact of the classifications of the law for the moment that states had adopted the law. This results in an unbalanced panel.[16] To acquire a somewhat more reliable balanced panel, the notifications that were collected without the existence of a law are entered into the model as well. For this purpose, the variable *Has_law* is used and interacted with the classifications of the law. The model is reconstructed for the independent variables that turned out to be significant.

Model 1: unbalanced panel (297 cases)
$Not\_per\_firm(\_sel) = \beta_0 + \beta_{Sanctioning} + \beta_{Private\_action} + \beta_{Scope\_law} + \beta_{Not\_ag} + \beta_{Not\_custcredit} + \gamma_s + \delta_t + \varepsilon_s$

Model 2: balanced panel with *Has_Law* (400 cases)
$Not\_per\_firm(\_sel) = \beta_0 + \beta_{Has\_law} + \beta_{Has\_law} * (\beta_{Sanctioning} + \beta_{Private\_action} + \beta_{Scope\_law} + \beta_{Not\_ag} \beta_{Not\_custcredit}) + \gamma_s + \delta_t + \varepsilon_s$

Alternative dependent variables
Not_per_firm_sel / Log_not / Log_not_sel

The independent variables are dummy variables. γ_s, δ_t and ε_s are respectively state fixed effects, time fixed effects and the error term. (Verbeek, 2012:345) As said, the fixed effects model is able to take correlations between several years of a law into account. This requires an assumption about the structure of the correlations between occasions. This is called the covariance structure. (Heck, 1999:164) Out of many types of covariance structures, the autoregressive and compound symmetry functions are the best candidates to fit the data.[17] The autoregressive structure is the most commonly used structure in these kinds of fixed effect regression and therefore used in the model. (Kincaid, 2005:30) The model below displays the coefficients and significance of the independent variables. The SPSS mixed models procedure does not produce an R-squared statistic, because "definitions for an R-square for become problematic in models with multiple error terms", caused by the multiple measurements.[18] Each model is repeated for the independent variables that were significant in the model. This iteration is displayed after the slash.

Independent variables (effect for = 0)	Not_per_firm [First run]/[Repeated]	Not_per_firm_sel [First run]/[Repeated]

[16] The panel is unbalanced because laws are adopted at different periods in time. Hence, a different number of laws are observed for each year in the dataset.
[17] Compound symmetry assumes equal variances and equal covariance across occasions that are constant over time. The autoregressive covariance structure assumes that the residual covariances between measurement occasions within subjects (=states) are correlated but decline exponentially with the distance. Hence, these structures are fairly similar.
[18] 'R-square statistics in SPSS Mixed Models' (*IBM support*, 9 July 2011) <http://www-01.ibm.com/support/docview.wss?uid=swg21481012> accessed 13 June 2013.

Electronic copy available at: https://ssrn.com/abstract=2431174

	Balanced	Unbalanced	Balanced	Unbalanced
Has_law	-41.7***/-79.6***	Not analyzed	-20.3/-37.4***	Not analyzed
Sanctioning	-	-	-	-
Private_action	-	-	-	-
Scope_law	-	-	-	-
Not_ag	-	-	-	-12.0**/-9.8*
Not_custcredit	-	-	-	-10.9**/-
Constant	95.3***/95.9***	96.6***/97.1***	68.9***/68.9***	69.5***/75.5***
Observations	400	297	400	297
Number of states	46	50	46	50
Significance: *** p<.01; **p<.05; p<.1; '-'p>.1				

Table 5: results of the fixed effects regression model

The results partly confirm the insights from the comparison of means but also display differences. The adoption of a law does have a significant effect on the number of breach notifications, as expected. This confirms the significant differences in means and medians. The absence of a law results in a reduction between on average 79.6/95.9=83% and 37.4/68.9=54%. Both *Sanctioning* and *Scope_law* do not have a significant impact in the model. This also corresponds with the findings in the comparison of means. But, *Private_action*, *Not_ag* and *Not_custcredit*, which partly had significant results concerning the means comparison, are not significant in this model except for *Not_ag* in the unbalanced selected sources dataset.[19] This can be caused by the fact that states with these laws already had a higher number of notifications at the start of the measurements, which cannot be attributed to this particular aspect of the law. Contrary to the comparison of means, the fixed effects regression controls for these kinds of errors. Furthermore, the balanced and unbalanced panels and the database with all sources and the database with selected sources perform quite similar.

Robustness

Robustness checks have already been made by constructing both a balanced panel and an unbalanced panel, and by running the model for both the dataset that includes all sources and the dataset that only includes selected sources. In addition to that, the dependent variable is transformed to the logarithmic transformation of the number of notifications, thus without dividing by the number of firms. The logarithmic transformation of the number of firms is used as a control variable in the model, because this variable varies between states and over time and cannot be filtered by the fixed effects regression. The outcome for the alternative model is displayed below:

Dependent variable (effect for = 0)	Log_not		Log_not_sel	
	Balanced	Unbalanced	Balanced	Unbalanced
Has_law	-/.86***	Not analyzed	-/-.58***	Not analyzed

[19] Except for the unbalanced model with selected sources, but this only results in *Not_ag* being significant on the 0.1 level, which cannot be regarded as a powerful result.

Electronic copy available at: https://ssrn.com/abstract=2431174

Sanctioning	-	-	-	-
Private_action	-	-	-	-
Scope_law	-	-	-	-
Not_ag	-	-	-.20**/.19**	-.19**/-.18**
Not_custcredit	-.16*/-	-.16*	-	-
Log_firms0512	.88***/.88***	.92***/.92***	.85***/.86***	.89***/.87***
Significance: *** p<.01; **p<.05; p<.1; '-'p>.1				

Table 6: robustness check fixed effects regression

Only the adoption of the law stands the robustness check. It is not surprisingly that the number of firms is highly significant. This corresponds with the high correlations found, when controlling for this variable. *Not_ag* produces some significant results at the .05% level for the dataset with the selected notifications. Hence, *Not_ag* is fairly robust on the unbalanced selected sources model.

LIMITATIONS

Before diving into conclusions, limitations regarding the statistical analysis are addressed, which are partly tackled by the approach followed.

Regarding the dataset

The representativeness of the dataset is questionable since only 0.05% of the U.S. organizations are represented within an eight year period. This could also be caused by non-compliance. In addition to that, the dataset is constructed out of multiple sources that are not mutually exclusive and representative. In order to tackle the latter problem, the dataset has been run separately for two representative selected sources and duplications have been filtered out.

Regarding the model

There are multiple inherently omitted variables that can change over time and between states. In the means comparison, this limitation is mitigated by separating a developing and a mature period and by controlling for the number of firms in a state. This issue is treated more thoroughly in the fixed effect regression. This method controls for state and time differences. However, the fixed effects regression does not control for variables that differ between states and over time simultaneously. For example, Internet security can vary between states but also over time. The results thus should be treated with care.

Regarding the validity of the analysis

Pesonal data breaches in the database sometimes contain records of inhabitants of different states while in the breach is only registered in one state. For instance, a security hack of the server of Virginia Commonwealth University concerned 176,567 records that contained personal information of former and current employees, students, staff and affiliates, also spread out over other states. If a data security breach concerns compromised records spread out over multiple states, multiple jurisdictions apply to the organization that has to notify. Since breach notifications in the database are only registered in one state (mostly the state where of the residency of the organization), one could not determine which jurisdiction had impact on the decision. Hence in these situations, it is an issue to tie the laws with the effects.

Electronic copy available at: https://ssrn.com/abstract=2431174

CONCLUSIONS

The aim of this paper was to systematically explore whether characteristics of the DBNL affect the number of breach notifications. The exploratory analysis suggests serious compliance harms.

First, as expected, the adoption of DBNLs leads to an increase of breach notifications. A notification increase of 50% can be attributed to the law, by a fixed effects regression analyzing differences in breach notification before and after the introduction of the law. The database is constructed by underlying sources that partly only register officially notified data breaches, which explain this high relative increase. From an absolute perspective, the effect is minor: less than 0.05% of the companies notified a security breach in America in the eight-year period that was researched. To compare: a recent study in the United Kingdom published that 88% of the companies surveyed had experienced data theft only in 2009.[20] (Ponemon, 2010) Due to the low absolute number of breach notifications, it is unlikely that DBNLs have such a large effect on improved security that they cause a decrease of the amount of subsequent notifications.

The low absolute number of notifications could be explained by incompleteness of the dataset and non-compliance. According to the deterrence theory, non-compliance is caused by the fact that notifying a breach costs more than concealing a breach. The costs of notification consist of direct costs (administration, mitigation and recovery) and reputational costs. Moreover, there are no benefits for an organization to notify. On the contrary, expected liability for concealing a notification is low. This is caused by the fact that it is hard to enforce a DBNL and that sanctions are low.

Second, characteristics of the law have been researched. It is clear that laws with a sanction higher than $50,000 do not differ in the number of notifications compared with laws that have lower or not predefined sanctions. Hence, this is a rejection of the hypothesis the height of the sanctions of the DBNLs scrutinized have an effect on compliance. This leads to four possible explanations

1. The present sanctions could be too low to have an effect on compliance. Only in the case of very high sanctions, organizations would have an increased willingness to notify.
2. Sanctions are not a perfect proxy for expected liability. Enforcement efforts can differ between states and are unobserved. (Insofar as enforcement efforts remain constant over time, they are filtered out by the fixed effects regression.)
3. Technical aspects concerning the construction of the variable hamper the exact determination of sanctions. The variable *Sanctioning* distinguishes laws with predefined sanction higher than $50,000 from laws that did not impose such a sanction. However, some laws do not have a predefined sanction, but these laws potentially leave open the possibility for attorney generals to impose sanctions above $50,000. Moreover, some laws also allow for private right of action, which can increase the total 'sanction' that can be imposed. However, a private action

[20] This figure is more illustrative, we would not expect 88% of the American organizations to notify because this survey mostly concerned large companies, but it expected that with full compliance, ratios would by far exceed 0,05% over an 8 year period.

Electronic copy available at: https://ssrn.com/abstract=2431174

happens very rarely and the separate *Private_action* variable did not show any significant results.

4. Sanctions do not matter at all and other factors, such as morality, determine compliance. (rejection of the application of the deterrence theory with regard to SBNLs)

Third, there are other characteristics of the law that did not have an effect. Laws that allowed for a private right of action or had an obligation to notify the Attorney General or the Customer Credit Reporting Agency were to some extent positively associated with higher laws in the comparison of means, but could not stand the fixed regression test. Also a broader scope did not have an effect on the amount of notifications in the database.

RECOMMENDATIONS

Acknowledging the fact that actors are to some extent rational in balancing the costs and benefits of their decisions, compliance of DBNLs can be strengthened either by increasing the benefits of a notification or increasing the cost of concealing a notification.

The benefit of notifying increases by adding cooperative elements in DBNLs. For instance, organizations that notify properly can be rewarded by giving them a label or quality mark to show that they a trustworthy. It is also be beneficiary for organizations to be a member of a group that shares information about upcoming cyber security risks.

The cost of compliance of DBNLs increases by intensifying deterrence. This not only concerns higher sanctions, but also intensified enforcement. Enforcement can be intensified by, for instance, stimulating private parties (bounty hunters) to search for concealed security breaches.

Apart from increasing compliance with DBNLs, the availability of empirical data needs to be improved. It is important to measure effects after the introduction of the law to scrutinize the claims about expected effects made before the introduction of the law. The dataset that is used in this analysis is flawed. Therefore, it is recommended to centrally register information about data breaches that are notified under the obligation flowing from the DBNL. The central registration of these notifications would provide a representative longitudinal dataset. This dataset can be used to measure to which extent the DBNL is capable of incentivizing companies to notify data breaches.[21]

Appendix A: construction of the independent variables

This appendix describes the construction of the independent variable in detail.

[21] This dataset can do this better than the dataset used in chapter 5, because it contains exclusively and exhaustively data breaches from the DBNL.

Electronic copy available at: https://ssrn.com/abstract=2431174

Introduction date

The introduction date is an important characteristic since it aids in determine the effects of the law as such.[22] For this purpose, data from the Commercial Law League America (CLLA) is used.[23] This database is an authoritative synthesis of legal analysis, which covers among others the introduction date of U.S. DBNLs and key provisions of the law. The data is updated until December 2011. After this date, no additional adoptions of DBNLs have taken place.

	2005	2006	2007	2008	2009	2010	2011	2012
U.S. States with a DBNL	11	27	36	41	45	45	46	46
	Developing period				Mature period			

Table 7: number of DBNLs per year (out of 50 U.S. States)

Most states adopted a DBNL in the years between 2005 and 2008. This is called the 'developing' period. Between 2009 and 2012 the vast majority of states have a law in place: the 'mature' period. Those periods will be used in the upcoming quantitative analysis. It is remarkable that the dataset also contains notifications from years and states that do not have a law. Those are most likely voluntary notifications or notifications from consumers or third parties, but indicate that a significant part of the notifications could come from other reporters than the companies affected.

Sanctioning laid down in the law higher than 50000 dollar

The sanction for not complying with the law differs between states. Not complying means not notifying or not notifying in due time or according to the formal requirements demanded. 14 laws can impose a maximum sanction of 50000 dollar or higher. Some states do not predefine a sanction but consider it a task of the Attorney General to impose a sanction, which could possibly be higher than 50000 dollar. The classification *sanctioning* therefore means that there is a predefined penalty of 50000 dollar or higher. These laws are labeled 1. For this classification, a legal analysis that contains aspects of all U.S. DBNLs made by the law firm Mintz Levin has been used. Mintz Levin is a large U.S. law firm with approximately 400 attorneys specialized in privacy and security, which made this chart for information purposes.[24] The classification is compared with a similar chart of Baker Hostetler, a similar U.S. law firm with 800 attorneys.[25]

---

[22] Most laws have been amended in some form after their adoption, but most amendments concern an incremental alteration of the laws. Thus, therefore, the introduction date is used for the analysis.

[23] 'Data Breach Notification Laws by State' (CLLA, December 2012) <http://www.clla.org/documents/breach.xls> accessed 12 June 2013.

[24] 'State Data Security Breach Notification Laws' (Mintz Levin, 1 December 2012) <http://www.mintz.com/newsletter/2007/PrivSec-DataBreachLaws-02-07/state_data_breach_matrix.pdf> accessed 13 June 2013.

[25] 'State Data Breach Stature Form' (Baker Hostetler, 2013) <http://www.bakerlaw.com/files/Uploads/Documents/Data%20Breach%20documents/State_Data_Breach_Statute_Form.pdf> accessed 13 June 2013.

Electronic copy available at: https://ssrn.com/abstract=2431174

Individuals have a private right of action

A possible important distinction between the U.S. laws is whether individuals have a private right of action. This means that people injured by a violation of the breach notification may institute a civil action to recover damages. There must however be a relation between the failure to notify and the damages caused. States that have a private right of action are labeled 1. 14 laws allow for a private right of action of individuals. This classification is based on the chart of Baker Hostetler.[26]

Scope of personal information is broader than general definition.

According to Baker Law, the general definition of personal information is "An individual's first name or first initial and last name plus one or more of the following data elements: (i) Social Security Number, (ii) driver's license number or state issued ID card number, (iii), account number, credit card number or debit card number combined with any security code, access code, PIN or password needed to access an account and generally applies to computerized data that includes personal information".[27] Baker law labeled cases with a broader definition of personal data. For example, because medical information a password or a taxpayer identification number is included.[28] Those are labeled 1. 24 laws have a scope of personal information, which is broader than the general definition.

Obligation to notify the Attorney General

A remarkable difference between U.S. laws is the obligation to notify the Attorney General in addition to the person whose data is breached. The Attorney General is a supervising authority and responsible for settling the procedure and potential procedures for damages. 17 of the 50 states require such a notification.[29] These are labeled 1.

Obligation to notify the customer credit reporting agency

A customer credit reporting agency is "a company that collects information from various sources and provides consumer credit information on individual consumers for a variety of uses. It is an organization providing information on individuals' borrowing and bill-paying habits". (O'Sullivan & Sheffrin, 2003:512) A notification to this authority can be required because these agencies maintain and compile personal information files on consumers. 29 laws contain this obligation. In some states, a company only needs to notify a customer credit reporting agency if the number of records per breach is above 500 or 1000 residents. Laws that have the obligation to notify the customer credit agency, regardless of the threshold, are labeled 1.[30]

Appendix B: construction of the dependent variable

[26] Ibid.
[27] Ibid.
[28] Ibid.
[29] Security Breach Notification Chart (Perkins, 2013) <http://www.perkinscoie.com/files/upload/LIT_09_07_SecurityBreachExhibits2.pdf> accessed 4 July 2013.
[30] Ibid.

Electronic copy available at: https://ssrn.com/abstract=2431174

The 3554 breach notification reports in the database contain several characteristics of a data breach, such as a resume of the breach, the number of records breached and the sector, the state and year in which the breach occurred.[31] A challenge is that the dataset is an aggregation of multiple data breach databases (hereafter: sources). Those sources are not mutually exclusive as they can contain the same notifications. Therefore, duplications have been filtered out. Apart from that, some sources are added in a later stage of the data collection, which can give a false impression of an increased number of data breaches over time. Besides, several actors, such as companies, consumers and third parties can be the reporters of a breach, while the interest of this research solely lies at organizations that notified their own breaches in order to make statements about compliance with the law.[32] The Privacy Breach Clearinghouse, however, claims that most of the notifications come from the harmed companies by stating that "if a breached entity has failed to notify its customers or a government agency of a breach, then it is unlikely that the breach will be reported anywhere".[33]

Breach: an occasion of a data breach.
Record: the number of people affected by a breach.
Source: the underlying database that collected the breach.

Description of the data breach notifications in the database

A descriptive analysis of the characteristics of a data breach in the dataset for the year 2012 shows that the size and fashion of data breaches widely varies. In 2012, 675 breach notifications were registered. In most of the cases, a breach contains multiple records: the number of people affected by the breach. The size of the breach varies between a few records and millions of records. There are small breaches (less than 1000 records), medium size breaches (between 1000 and 10000 records breached) and large breaches (more than 10000 records per breach). The size of 246 breach notifications is unknown. Almost all states have small, medium and large breaches. California, for instance, contains 122 breaches. This varies from 15 affected people (A hospital employee that used credit card information of cancer patients) to a big LinkedIn data breach, which contained 6.4 million encrypted passwords that were posted online by a group of hackers. In Washington, 16 breaches were reported: the smallest breach reported consists of 16 records (a small credit card fraud). The largest breach affected 35 million people by hacked password information an online gaming platform. In Virginia, there were also small breaches, such as a breach with 30 records in November 2012 in the healthcare sector. The largest breach (176567 records) contained a security hack of the server of Virginia Commonwealth University that contained personal information of former and current employees, students, staff and affiliates. Some notifications only include records such as passwords from a particular website such as the LinkedIn data breach, while in other cases people are affected directly, for example through the use of stolen creditcard information for fraudulent activities. Therefore, the

[31] The breaches are classed by sector: businesses (retail, financial/insurance and other), educational institutions, government and military, healthcare and non-profit organizations.
[32] Firms notify a their own breaches, but customers and third parties notify suspicious information on the Internet.
[33] The Privacy Rights Clearing House DataBase

Electronic copy available at: https://ssrn.com/abstract=2431174

number of records breached is not a very accurate unit for the impact of the breach. Based on this analysis, a smaller breach generally has a higher impact per record than a larger breach, but larger breaches compensate this by a multiplicity of records.

Below, the distribution of all notifications is given for 2012 and the entire dataset. It shows that most notifications are not categorized and that the size of breaches is distributed fairly evenly. In the eight-year period observed, in total, 600 million records were breached. The three largest breaches contain more than 300 million records.

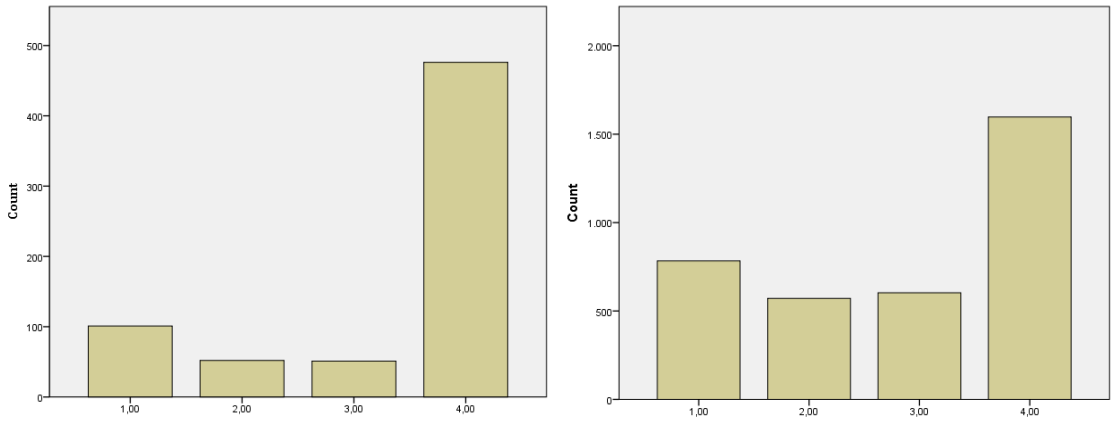

Figure 2: total records in 2012 (left) and 2005-2012 (right) (1=0-1000; 2=1001-10000; 3=10001+; 4=unknown)

Sources of the database

The database is an aggregation of several data breach notification sources. The number of notifications per source and the representation of the sectors are shown in the figures below:

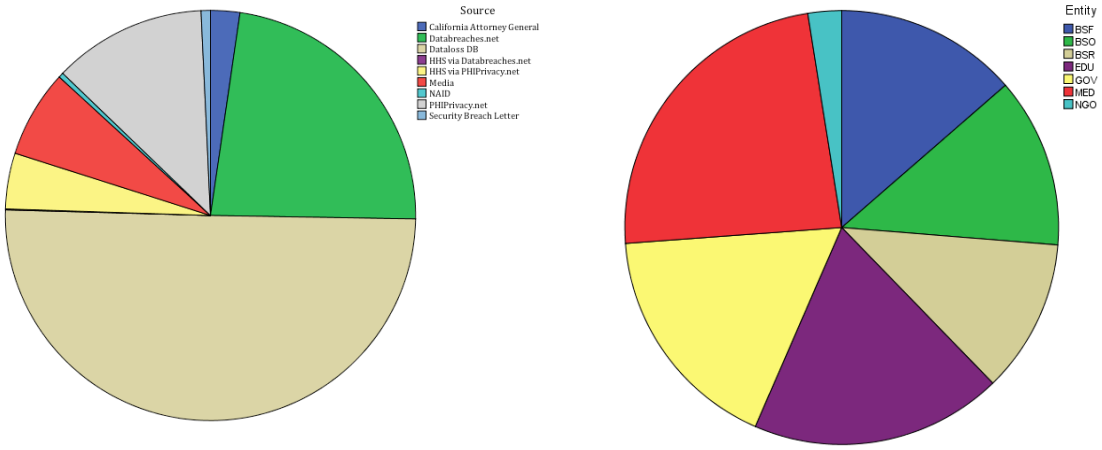

Figure 3: different sources in dataset from 2005 until 2012 and sectors.[34]

[34] The sectors are labeled as follows: BSO - Businesses – Other; BSF - Businesses - Financial and Insurance Services; BSR - Businesses - Retail/Merchant; EDU - Educational Institutions; GOV - Government and Military; MED - Healthcare - Medical Providers; NGO - Nonprofit Organizations.

Electronic copy available at: https://ssrn.com/abstract=2431174

Most of the notifications come from the Dataloss DB database. As of January 2010, the sources Databreaches.net, PHI Privacy and NAID are included. As of March 2012, the list of the California Attorney General is included.[35] As said, the sources are not mutually exclusive, which entails that one occasion of a data breach can both be found in Dataloss DB, the media, and by the Attorney General of California. PHIPrivacy, 'HHS via' and NAID mostly include medical breaches, which is also problematic from a representativeness point of view. Hereafter, two examples of sources that are not representative are displayed: the source 'California Attorney General' represents mostly breaches from California and PHIPrivacy.net contains mostly breaches in the medical sector.

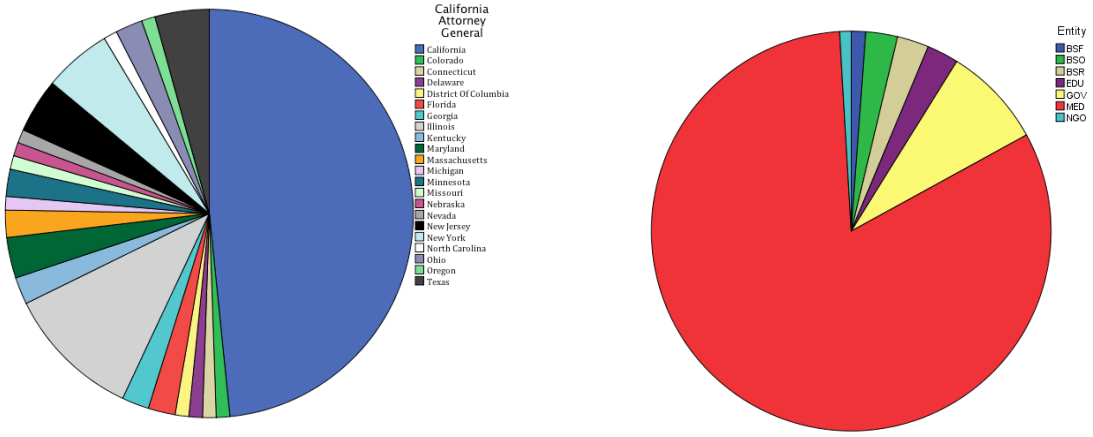

Figure 4: representativeness issues: notifications per state from the source 'California Attorney General' and sector of notifications of PHIprivacy (medical is red).

Sources added in a later stage that also have representative issues from a state or sector point of view should be filtered out. In fact, only the sources from Dataloss DB and databreaches.net are not added in a later stage and do not have representative issues. Therefore, within the database, these two sources are selected as the main database to perform the analysis.

However, the exclusion of sources that are not representative for the population or added in a later stage could be problematic. Although the database does not contain duplications, but notifications in excluded sources originally also could be represented in sources that are representative, before they were filtered out because of the duplication exclusion. Those notifications would be falsely excluded if the sources that are not representative would be excluded. Therefore, for comparison and as a robustness check, also analysis is run with the database containing all sources.

[35] The Privacy Rights Clearing House DataBase' under 'FAQ'.

Electronic copy available at: https://ssrn.com/abstract=2431174

Restructured data: notifications per year and per state.

In order to analyze notifications per year and per state, the raw database is restructured into a list of 400 cases containing the number of notifications per state for each year between 2005 and 2012. (50 states containing 8 years of data) [36] The following graph only displays year after year effects. The total number of notifications of 50 states is summarized per year.

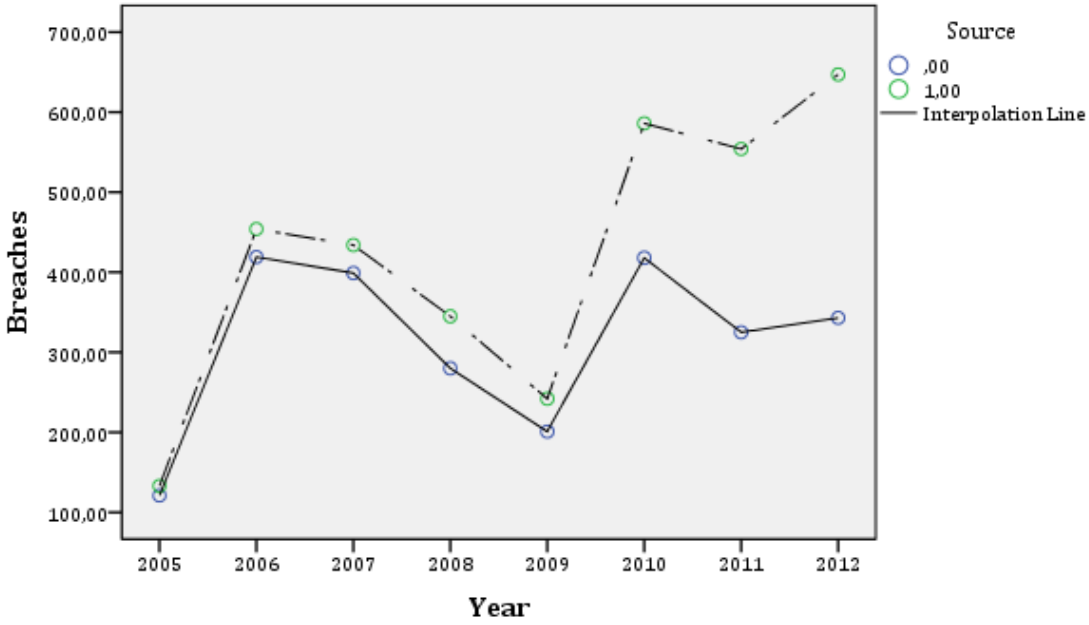

Figure 5: notifications per source (0=selected sources; 1=all sources)

An inconstant increase of the number of notifications can be observed. It must be noted that from 2010 on, the new sources that were added to the database can explain the increase of notifications. Apart from this, a remarkable decline is visible in 2008 and 2009. This could be related to the financial crisis, although this remains speculation. The following graph represents the number of records per source per year. It is clearly visible that the Dataloss DB and databreaches.net sources overlap. This overlap is one of the reasons that the two sources are selected together.

[36] The raw database contains of 3554 cases (all sources included) or 2506 cases (selected sources: Dataloss DB and Databreaches.net).

Electronic copy available at: https://ssrn.com/abstract=2431174

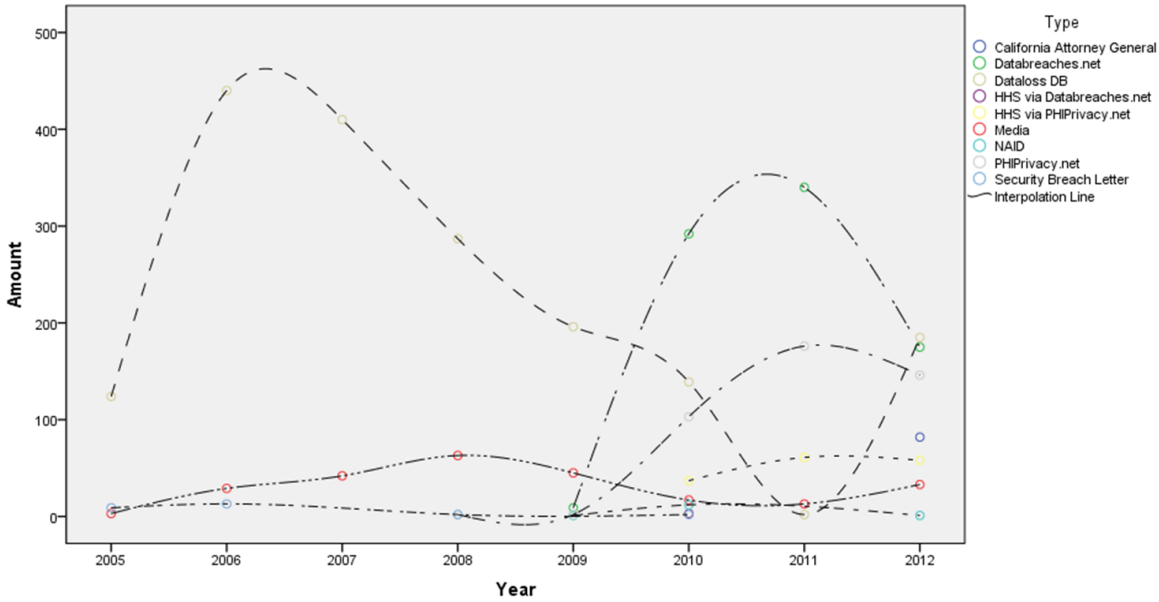

Figure 6: notifications per year per source

After the visualization of the number of notifications per year and per source the table of the restructured data and the distribution of the number of notifications per state and per year are displayed.

	2005	2006	2007	2008	2009	2010	2011	2012
Alabama	0	2	6	2	2	2	11	8
Alaska	0	1	1	0	1	2	2	3
Etc.	…	…	…	…	…	…	…	…
Total	133	454	434	345	242	586	554	647

Table 8: Visualization of restructured data: Notifications per state per year (all sources)

Electronic copy available at: https://ssrn.com/abstract=2431174

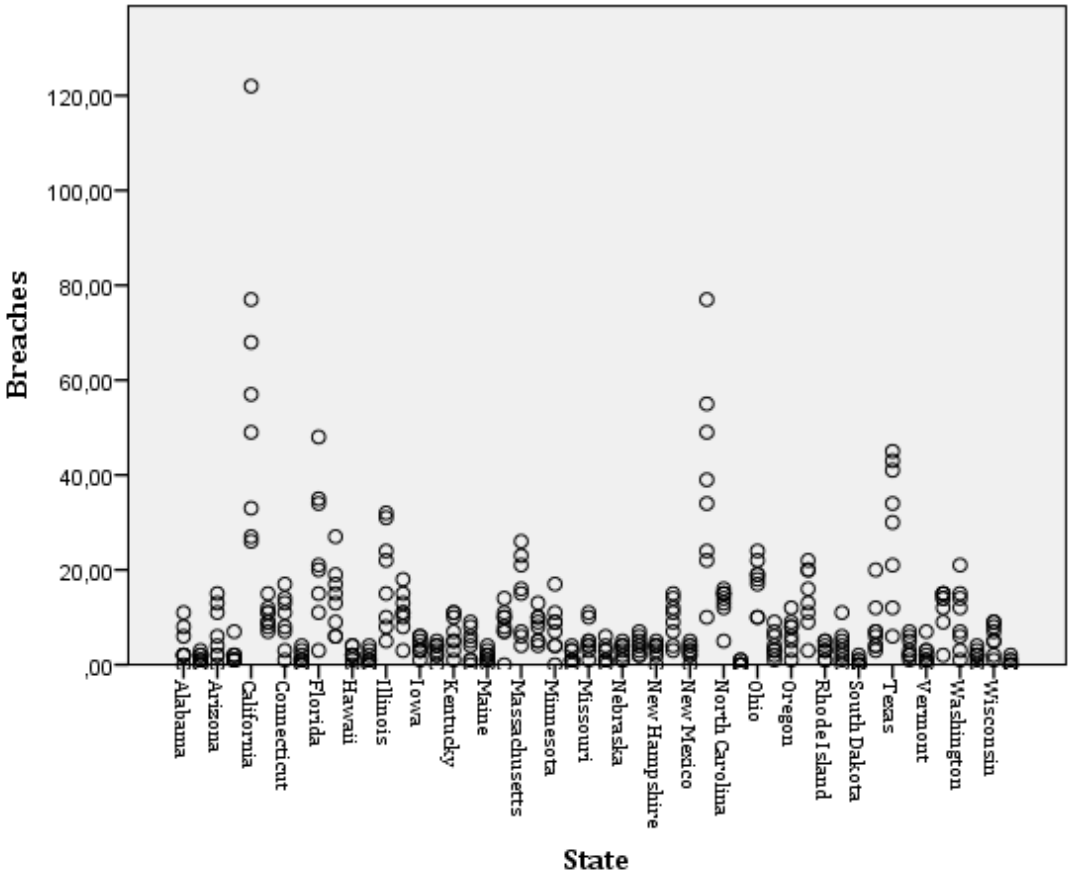

Figure 7: notifications (circle is number of notifications per year per state; all sources and 50 states included)

The low number of notifications per state in combination with the low number of firms represented gives the risk of representative errors, because measurement errors can change the picture of the data in the state. The distribution and the visual inspection of the table with the restructured data shows that most states had between 0 and 20 notifications per year and a few states have a lot more notifications. The average number of notifications per case is 9 (all sources included).

Apart from these representativeness errors, a few assumptions about the dataset need to be made. The first assumption is that the distribution of the number of records per breach is equal over all the states. With this assumption, there is no need for taking the wide variation of records into account. In the same way, an equal distribution of the origin of the notifications is assumed. Because it is impossible to exclude third parties and consumers of the database, it is assumed that they are distributed equally among all states. This means that the dataset is in some way representative for the number of breach notifications coming from firms that fall under a DBNL. Moreover, it is assumed that different sectors respond in a similar manner to data breaches. These assumptions need to be made because the subjects assumed can influence the outcomes of the analysis but cannot be filtered or controlled due to limitations in the model or time constraints. Therefore they will not be tested empirically, but are taken into account while making conclusions about the results of the quantitative analysis.

Electronic copy available at: https://ssrn.com/abstract=2431174

Variations between states: number of firms

There are many differences between states that can explain the differences between the number of notifications per state, for example, criminal activity, the cultural attitude to comply with laws, the GDP, population and Internet percentage rate. The difference between the numbers of notifications per state can possibly also partly be explained by the size of the state. Some states in America such as California and Texas are much larger than others. There are more data breaches reported in larger states because there are more firms that can be breached.[37] A possible controlling variable for this issue could be the number of firms, because firms make notifications. There are a few large states (for instance, California, Texas and New York), with a lot of firms and many small states with relatively few firms. The number of firms in a state for each year between 2005 and 2010 is used, which also embodies differences between the number of firms within a state size over time.[38] A scatter plot of the logarithm of notifications and firms per state shows a visual relationship.

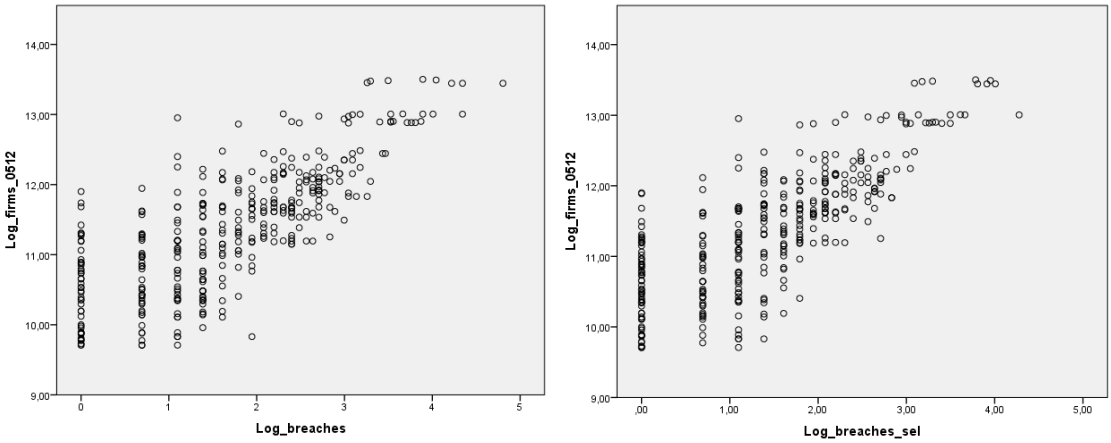

Figure 8: scatterplot of log_not and log_firms (Selected sources right)

A correlation analysis shows a significant correlation between the number of notifications and the number of firms.

Correlation *log_not* and *Firms per state*	Coefficient		Significance
	All sources	Selected sources	
Pearson	.805	.832	.000
Spearmans Rho	.771	.777	.000

When running a standard linear regression, it is shown that the number of firms in a state explains 65% (all sources) or 69% (selected sources) of the variance of the

[37] The concentration of vulnerable information technology services, such as in Silicon Valley in California can be another explanation for a higher number of breaches. This variation will not be discussed in this scientific paper.

[38] 'Firms in U.S. states' (*Census.gov*, 2013) <http://www.census.gov/econ/susb/> accessed 13 June 2013. The number of firms per state was available up to 2010. Therefore, 2011 and 2012 have 2010 values. The distribution of firm size per state is assumed to be equal.

Electronic copy available at: https://ssrn.com/abstract=2431174

number of notifications in a state. In a stepwise regression all other control variables (population, Internet percentage rate and gdp) are excluded. The visual information, the correlation and regression indicates that notifications can largely be explained by the number of firms in a state. Therefore, this study controls for this type of variance. In order to do this, the variable *Not_per_firm* is constructed: the number of notifications per state per million firms. The distribution of *Not_per_firm* is spread more evenly than *Notifications*. *Not_per_firm* will be used as a final dependent variable.

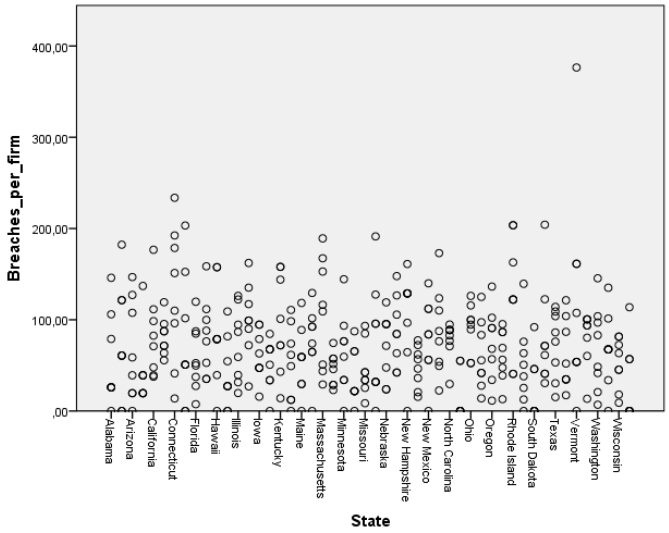

Figure 9: notifications per firm per state (all sources and 50 states included).

Appendix C: descriptive statistics and comparison of means

This appendix displays the descriptive statistics and comparison of means.

Effect of the law as such

Below the means and medians of *Not_per_firm* (and *Not_per_firm_sel)* for *Has_law* are displayed.[39] It is clearly visible that there are more notifications in the dataset when the law is adopted.

Database:	All sources				Selected sources			
Value:	0		1		0		1	
Mean/median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Has_law	55	50	77	72	37	32	45	37

Table 9: descriptive statistics of Has_law (highest marked green)

The mean and median also statistically differ on the .01 level as can be seen in the table below. The values of the means of the selected sources are lower because there are less

[39] Values are rounded.

Electronic copy available at: https://ssrn.com/abstract=2431174

notifications in the database with the selected sources compared with the database containing all sources.

Has_Law	All sources		Selected sources	
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.000	.000	.000	.000

Table 10: comparison of means and medians of the Has_law classification

Effect of higher sanctions

Below the means and medians of *Not_per_firm* for the classifications related to sanctioning are displayed. The means and medians of each classification *per year* is displayed in appendix C.

Database:	All sources				Selected sources			
Value:	0		1		0		1	
Mean/median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Sanctioning	78	72	75	75	56	47	54	53
Private_action	73	66	86	89	53	47	61	60

Table 11: descriptive statistics of sanctioning related classifications (highest marked green)

The descriptive analysis shows that the highest mean of *Sanctioning* lies at the states labeled 0, while the states that are labeled 1 have the highest median. Besides, the differences are very small. This suggests that the classification *Sanctioning* does not generate a lot of differences in the number of notifications in the database. *Private_action* has higher means and medians for states labeled 1. This direction corresponds with hypothesis 1b and 1c. The dataset that includes all sources does not show a different pattern than the dataset with selected sources.

The next analysis is a comparison of those means and medians using a parametric Independent Samples T-test and a non-parametric Mann-Whitney test. A comparison of all the cases covering all the years of data collection has been performed. As said, separate comparisons of means and medians are executed in order to analyze differences between the developing period and the mature period. The results are shown below.[40]

Sanctioning	All sources		Selected sources	
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.510	.916	.616	.582
2005-2008 (n=31)	.939	.910	.765	.965
2009-2012	.329	.774	.729	.464
Private_action				
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.039	.006	.115	.017

[40] A low number of cases (n>60) is mentioned. All other observation have a higher number of cases.

Electronic copy available at: https://ssrn.com/abstract=2431174

2005-2008 (n=30)	.618	.267	.289	.077
2009-2012	.049	.012	.232	.096

Table 12: comparison of means and medians of sanctioning related classifications (significant results marked green)

The results show that there are insignificant differences between means and medians for *sanctioning Private_action* shows significant differences for all sources, except for the developing period. The mature period contains more cases, thus is more likely to influence the total picture, which could explain the similarities in results. However, it does not produces significant results for selected sources except for the Mann-Whitney test. The fact that there are no differences in the developing phase could be attributed to the fact that most (aspects of the) laws need a certain period to become effective and that real impact can only be observed after a couple of years.

Effect of wider scope

Database:	All sources				Selected sources			
Value:	0		1		0		1	
Mean/median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Scope_law	77	71	77	74	54	48	56	52

Table 13: descriptive statistics of scope related classifications (highest marked green)

The descriptive analysis of the classifications related to hypothesis 2 shows that laws with a wider scope have slightly more notifications per firm for the selected sources. This is however not confirmed by the database with all sources. This database shows a mixed pattern, which indicates no significant difference.

Scope_law	All sources		Selected sources	
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.944	.910	.727	.588
2005-2008 (n=49)	.522	.786	.827	.777
2009-2012	.879	.945	.516	.670

Table 14: comparison of means and medians of scope (significant results marked green)

The comparison of means and medians shows no significant difference for scope. This was expected from the descriptive statistics.

Effect of notifying to a notification authority

Hypothesis 3 shows a major difference between the means and medians of the laws that do have the classifications *Not_ag* and *Not_custcredit* and the laws without this classification.

Database:	All sources				Selected sources			
Value:	0		1		0		1	
Mean/median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Not_ag	71	67	88	84	50	47	65	60
Not_custcredit	70	56	82	78	48	41	60	53

Table 15: descriptive statistics of notification authority related classifications (highest marked green)

Electronic copy available at: https://ssrn.com/abstract=2431174

The descriptive statistics of *Not_ag* and *Not_custcredit* show results that correspond with the hypotheses. There are, on average, more notifications if these obligations are present in the law.

Not_ag	All sources		Selected sources	
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.004	.009	.001	.001
2005-2008 (n=39)	.091	.063	.076	.041
2009-2012	.024	.066	.005	.014
Not_custcredit				
	Parametric	Mann-Whitney	Parametric	Mann-Whitney
2005-2012 (total)	.044	.016	.010	.002
2005-2008 (n=48)	.209	.032	.109	.011
2009-2012	.159	.199	.042	.054

Table 16: comparison of means and medians of notification authority related classifications (significant results marked green)

The parametric t-test and Mann-Whitney test show that both classifications render significant differences for the whole time span. The obligation to notify the Attorney General is even significant at the .01 level. The picture is less clear for the developing and mature period.

References

Anderson, R. (2001). Why information security is hard. Retrieved, 2013, from http://www.acsac.org/2001/papers/110.pdf

Becker, G. S. (1968). Crime and punishment: An economic approach. *The Journal of Political Economy, 76*(2), 169-217.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce, 9*(1), 69-104.

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

Heck, R. H. (1999). *Multilevel and longitudinal modeling with IBM SPSS* Routledge.

Kincaid, C. (2005). *Guidelines for selecting the covariance structure in mixed model analysis* . Unpublished manuscript.

Electronic copy available at: https://ssrn.com/abstract=2431174

O'Sullivan, A., & Sheffrin, S. M. (2003). *Economics, principles in action* Pearson Prentice Hall.

Pélissié du Rausas, M. (2011). *Internet matters: The net's sweeping impact on growth, jobs and prosperity.* (No. http://www.mckinsey.com/insights/high_tech_telecoms_Internet/Internet_matters).Mc Kinsey Global Institute.

Ponemon. (2010). *2010 annual study: U.S. cost of a data breach.* ().Ponemon.

Romanosky, S., Hoffman, D., & Acquisti, A. (2011). Empirical analysis of data breach litigation. Paper presented at the *, 3* 2242.

Schwartz, P. M., & Janger, E. J. (2007). Notification of data security breaches. *Michigan Law Review, 105*(5), 913-984.

Veltsos, J. R. (2012). An analysis of data breach notification as negative news. *Business Communication Quarterly, 75*(2), 192-207.

Verbeek, M. (2012). *A modern guide to econometrics* (Fourth Edition ed.) Wiley.

Winn, J. K. (2014). Are better security breach notification laws possible. *Berkely Technology Law Journal, 24*(3), 1133.

Electronic copy available at: https://ssrn.com/abstract=2431174