Empirical Analysis of Data Breach Litigation

WEIS

11th Annual
Workshop on the Economics of Information Security
WEIS 2012
Berlin, Germany, 25-26 June 2012

Sasha Romanosky
David Hoffman
Alessandro Acquisti

Carnegie Mellon
Heinz College

Problem: externalities caused by loss or theft of consumer information

- Modern IS, Web 2.0, and social media afford us many benefits.

- Many of these services are driven by the collection, analysis, and use of personal information (medical, financial, behavioral, etc.).

- However, use of personal information can impose externalities on consumers when their information is lost or stolen. E.g. identity theft, medical fraud, tax fraud, …

- For example…

Examples of data breaches

- Thief steals couple’s identity and files fraudulent tax refund.

- Pharmacy tosses medical files and employment applications in the public trash (In re Rite Aid Corp., FTC File No. 072-3121).

- Social Security Administration discloses the HIV results of a pilot to the FAA (Cooper v. FAA, 596 F. 3d 538).

- Heartland (credit payment processor) is hacked, compromising 130 million credit card numbers issued from over 650 banks. (In re Heartland Payment Systems, Inc. Securities Litigation).

Harm from breaches and idtheft

Consumer losses

- Tangible and intangible: e.g., psychological costs, but also lost opportunities, recovery efforts, increased cost of borrowing, etc.
- Reported no. of breaches since 2005: 2,725, ≈ 1/day.
- Est. no. of idtheft victims in 2011: 12 million.
- Est. cost of idtheft due to data breaches: $1 - $2.6 billion.

Firm losses

- Tangible and intangible: e.g., negative PR, stock market losses, but also consumer redress, recovery costs, legal fees, etc.
- Average cost of data breach: $5.5 million.
- Average per record cost of data breach: ≈ $200.

Sources: Privacy Rights Clearinghouse, Javelin Strategy and Research, Ponemon Research, Bureau of Justice Statistics.

IDENTITY THEFT
WHY YOU SHOULD BE WORRIED
Identity Theft is a form of fraud that steals someone else's personal information to access private resources, such as medical records or bank accounts, and often leads to expensive consequences for its victims.
64%
of Americans are very or extremely concerned about being a victim of identity theft.
THE NUMBER OF ADULT VICTIMS OF IDENTITY THEFT IN THE UNITED STATES FROM 2003 TO THE PRESENT:
2009 11.1 MILLION
2008 9.9 MILLION
2007 8.1 MILLION
2006 8.4 MILLION
2005 8.9 MILLION
2004 9.3 MILLION
2003 10.1 MILLION
1 IN EVERY 10 AMERICAN CONSUMER HAS BEEN A VICTIM OF IDENTITY THEFT.
1.6 MILLION HOUSEHOLDS have had their bank accounts and/or debit cards COMPROMISED.
THE POTENTIAL AMOUNT OF MONEY IN THE US AFFECTED BY 2009 FRAUDS AMOUNTS TO
$54 BILLION
(WORLDWIDE, BUSINESSES LOSE $221 BILLION YEARLY TO IDENTITY THEFT)
THE AVERAGE AMOUNT TAKEN FROM EACH IDENTITY THEFT VICTIM AMOUNTS TO
$4,841
(THAT'S ROUGHLY 3 MONTHS WORTH OF FULL-TIME WORK'S WAGES)
THE OUT-OF-POCKET COST FOR THE VICTIM TO RESOLVE IDENTITY THEFT DAMAGE:
$851 - $1,378
NEARLY 50% of victims learn of their identities being stolen within 3 months
ABOUT 15% of victims don't learn of their identities being stolen for FOUR OR MORE YEARS
70% Percentage of victims that have difficulty removing negative information stemming from the theft from their credit reports.
MOST COMMON METHODS OF STEALING IDENTITIES:
SKIMMING INFORMATION FROM TRANSACTIONS MADE IN A STORE
PAPER MAIL THEFT
HACKING COMPUTER NETWORKS AND DATABASES
REDUNDANT IT EQUIPMENT
"SHOULDER SURFING" WATCHING LOGIN CREDENTIALS OVER USERS' SHOULDERS
DUMPSTER DIVING
43% of victims knew the perpetrator who stole their identity.
330 HOURS
The average time it takes to repair the damage done by identity theft.
BUT IN SOME CASES, TO FULLY CORRECT THE DAMAGE DONE FROM THE THEFT CAN TAKE UP TO
5,840 HOURS
(THAT'S EQUIVALENT TO WORKING A FULL-TIME JOB FOR TWO YEARS)
25.9 million Americans now carry identity theft insurance.
HOW TO PROTECT YOURSELF AGAINST IDENTITY THEFT:
PREVENT CRIMINAL ACCESS BY PROTECTING YOUR DOCUMENTS.
When your SSN is requested, ask if you can provide alternate information
Shred any sensitive paper documents
Install anti-virus software and updated it regularly
Always secure digital information behind passwords
Turn off your phone's Bluetooth and Wifi when not in use
DETECT UNAUTHORIZED ACTIVITY IN ALL YOUR ACCOUNTS.
Monitor bank accounts at least weekly
Sign up for available mobile or email alerts by your institution
Monitor credit reports to spot unauthorized activity
RESOLVE YOUR IDENTITY FRAUD COMPLETELY AND THOROUGHLY.
Work through your bank and protection services provider to report problems immediately.
Use available services to restore the worthiness of your accounts and credit
Take full advantage of your financial provider's offers of loss
Sources:

How is US public policy addressing harms caused by data breaches?

- Both Congress and govt agencies are trying to find solutions: “Should a baseline data privacy legislation include a private right of action?” (Dept. of Commerce, 2010, 30).

- In the mean time, individuals are suing firms for alleged harms caused by data breaches.

- However, very little is known about the drivers, mechanisms, and outcomes of these suits.

- This makes it difficult to assess the effectiveness of litigation at balancing the tension between:
    - organizations’ use of personal information, and
    - individuals’ privacy rights.

- *Using a unique database of manually collected lawsuits, we analyze court dockets for over 230 federal data breach lawsuits from 2005 to 2010.*

Research questions

Q1: Which data breaches are being litigated at the federal level?

- Helps identify when firms are more likely to be sued, and what they can do to avoid litigation.

Q2: Which data breach federal lawsuits settle?

- Helps us understand how the legal system is addressing privacy harms.

Definitions

- Data breach: unauthorized disclosure of personal information.
- Disclosure: loss/theft hardware, cyberhack, or improper disposal.
- Personal information: SSN, CCN, medical, financial, email addresses, etc.

Related literature

- Legal scholarship of data breach lawsuits: Solove (2005), Citron (2007), Hutchins (2008), Lesemann (2009).
- Economics of data breaches: Campbell et al. (2003), Acquisti, Telang, Friedman (2006), Romanosky et al. (2010).
- Theoretical legal scholarship: Settlement rates (Priest and Klein, 1984); Legal disputes (Cooter and Rubinfeld, 1989).
- Empirical legal scholarship: Securities Class actions (Johnson et al.(2007), Choi (2007), Cox et al. (2008); Patents (Lerner, 2010); Docketology: Hoffman et al. (2007), Kim et al. (2009).

Theory of legal disputes (Cooter & Rubinfeld, 1989)

1. Accident
   - Injurer first balances expected cost of harm with expected cost of prevention.
2. Lawsuit
   - Victim (plaintiff) balances expected cost of litigation with expected damage award.
3. Settlement
   - Plaintiff and defendant each balance expected cost of further litigation with expected award at trial.

Data collection

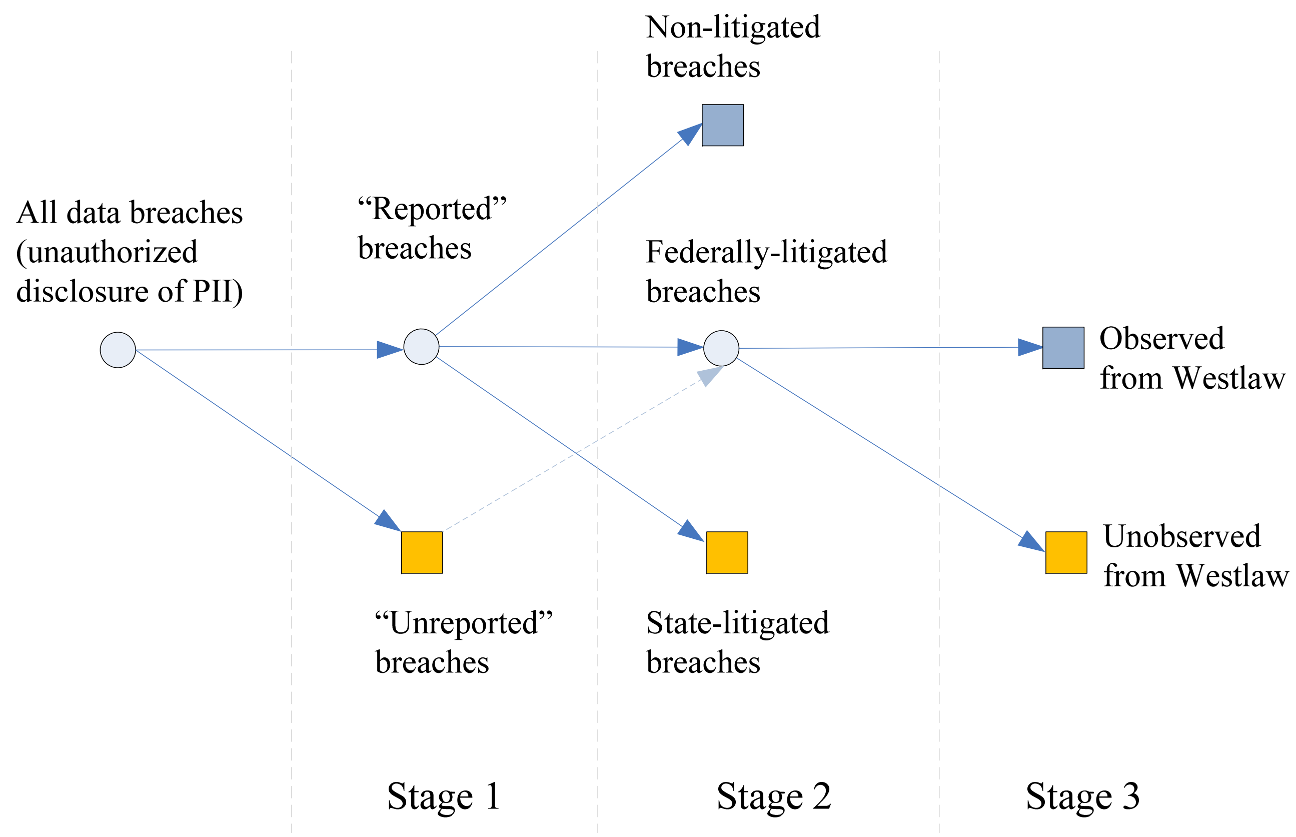
- Obtained list of all known data breaches (datalossdb.org).
- Used Westlaw to determine which breaches were federally litigated.
  - Systematically searched Westlaw for all suits matching key terms (e.g.: *"(data or security or privacy) breach," "personal information; identity theft"*)
- Purchased dockets, complaints, orders from PACER; manually coded dozens of variables.

- **≈ 1,772 data breaches in the 2005-2010 period, and 230 federal lawsuits,**

consisting of the following data:

  - Breach: types and number of records lost, firm industry, cause.
  - Case: outcome (settlement, dismissal), removal, jurisdiction, judge, class certification, law firms, number and types of causes of action.
  - Dates: date of breach, public notification, filing, disposition.
  - […]

Data generating process



- We focus on **federal suits** - a key to informing proposed legislation, and especially outcomes of most egregious cases.

What do suits typically look like?

- Usually private class actions (some public actions: FTC, SEC).

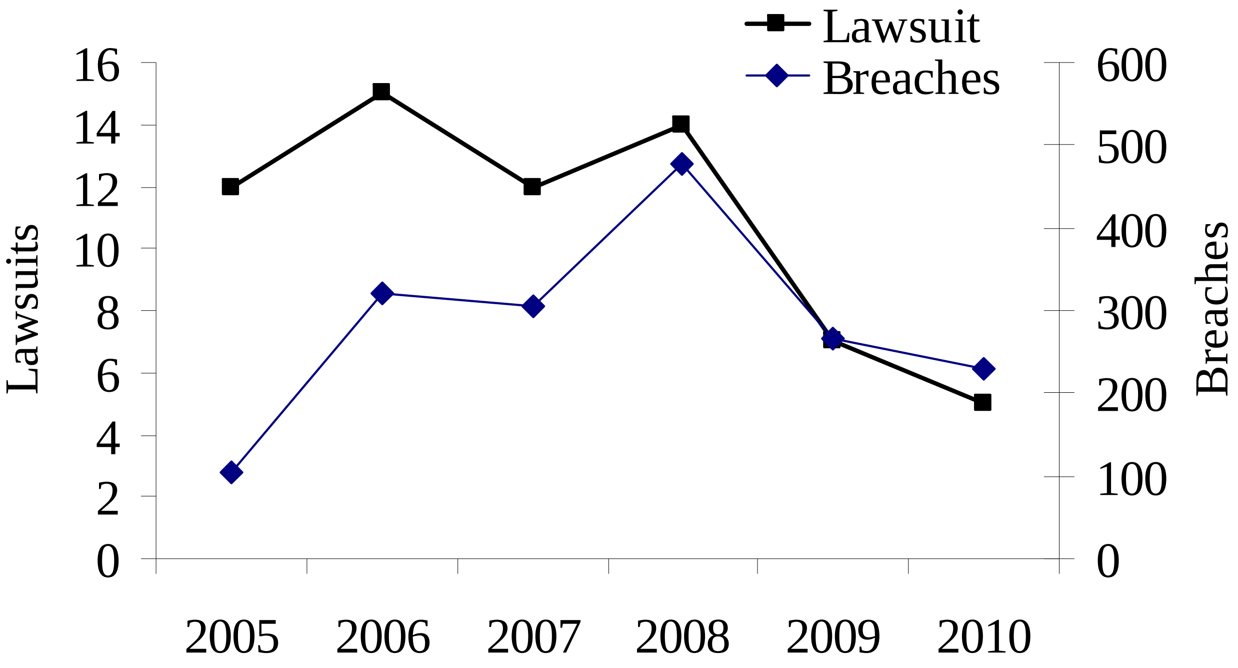
- Defendants are typically large firms (banks, retailers).

- Complaints allege both common law (tort, contract) and statutory causes of action (VPPA, DPPA). In fact, 87 unique COA for virtually the same event!

- Plaintiffs seek relief for: actual loss (identity theft), preventive costs (e.g. credit monitoring), potential future loss, emotional distress.

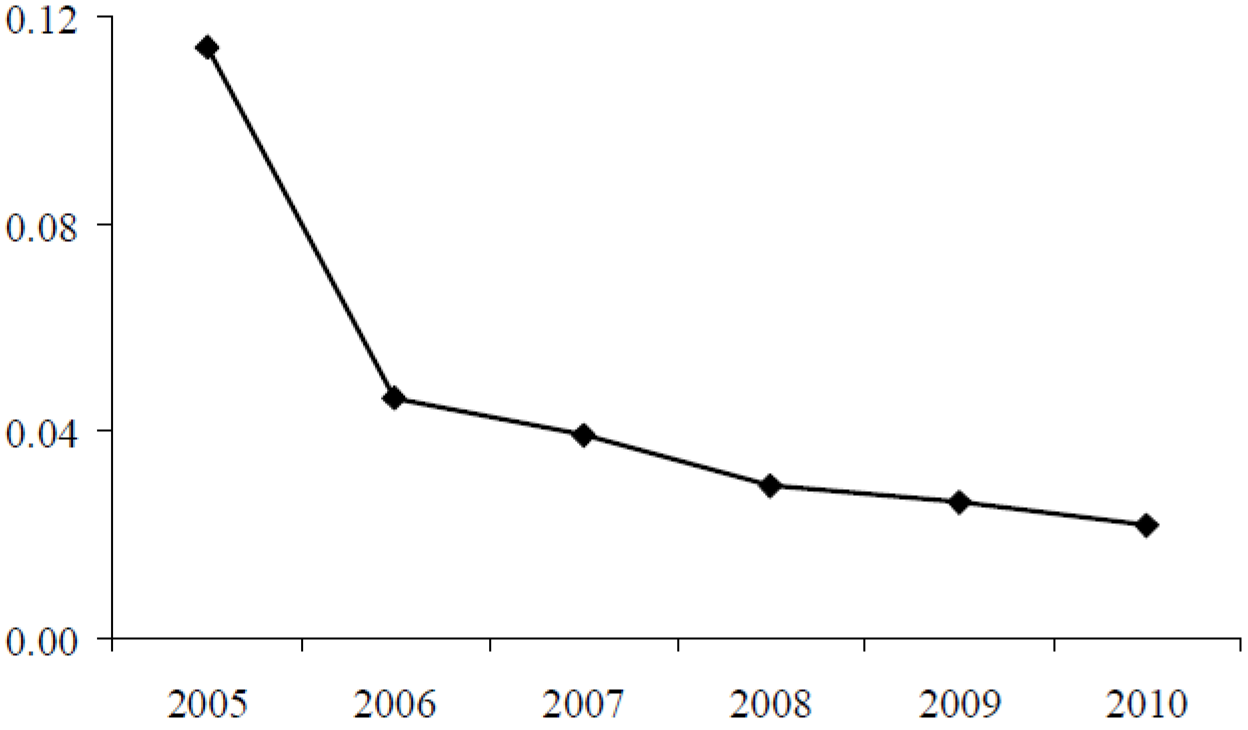
- Disposition: only 2 cases have reached trial, all others are either dismissed or settled.

Trends



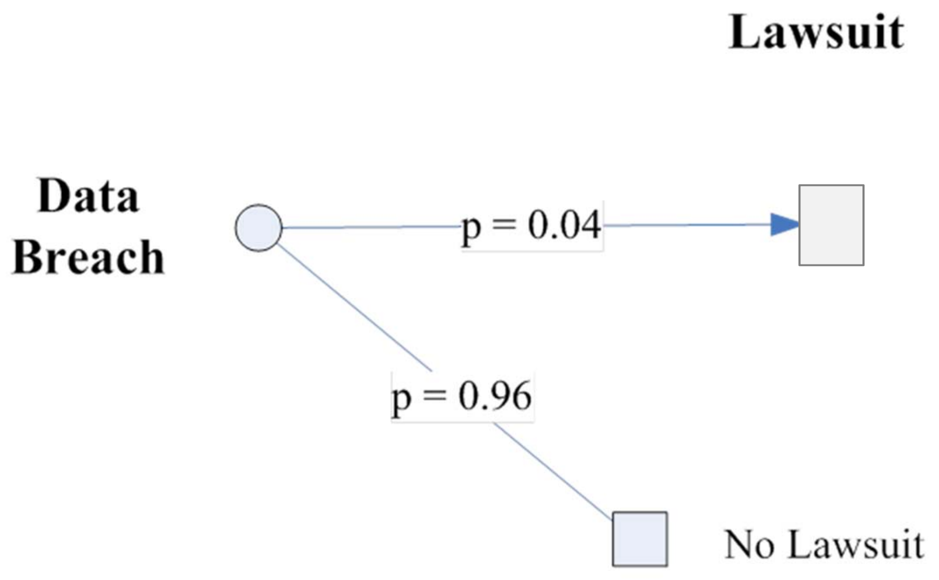
Both breaches and lawsuits decreasing since 2008.

Trends

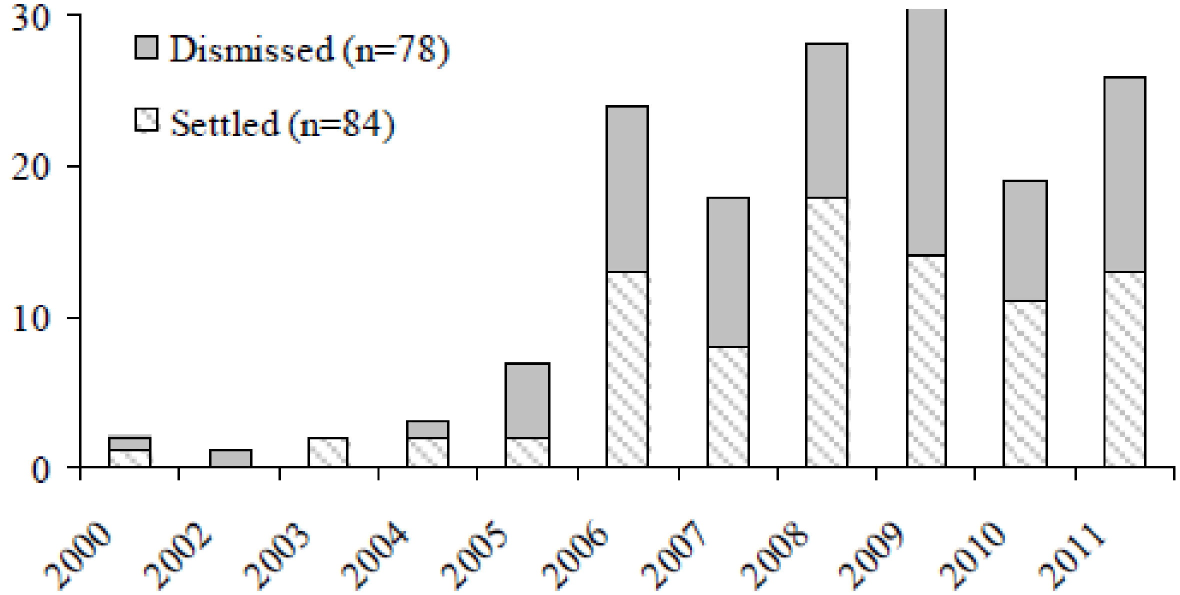


Ratio of lawsuits over breaches.

From data breaches to lawsuits



Trends



Dismissed vs. Settled lawsuits.

Q1: Which breaches are being litigated?

- Theory suggests: litigation increases with magnitude of award, probability of success.

- How does this apply to data breaches?

- Probability of lawsuit is <u>positively</u> correlated with breaches that:
  - *suffer greater number of records compromised,*
  - *show evidence of actual harm (financial loss),*
  - *required heightened level of protection of PII (CCN, medical, financial),*
  - *caused by improper disclosure of information, relative to the computer hack, or loss of hardware.*

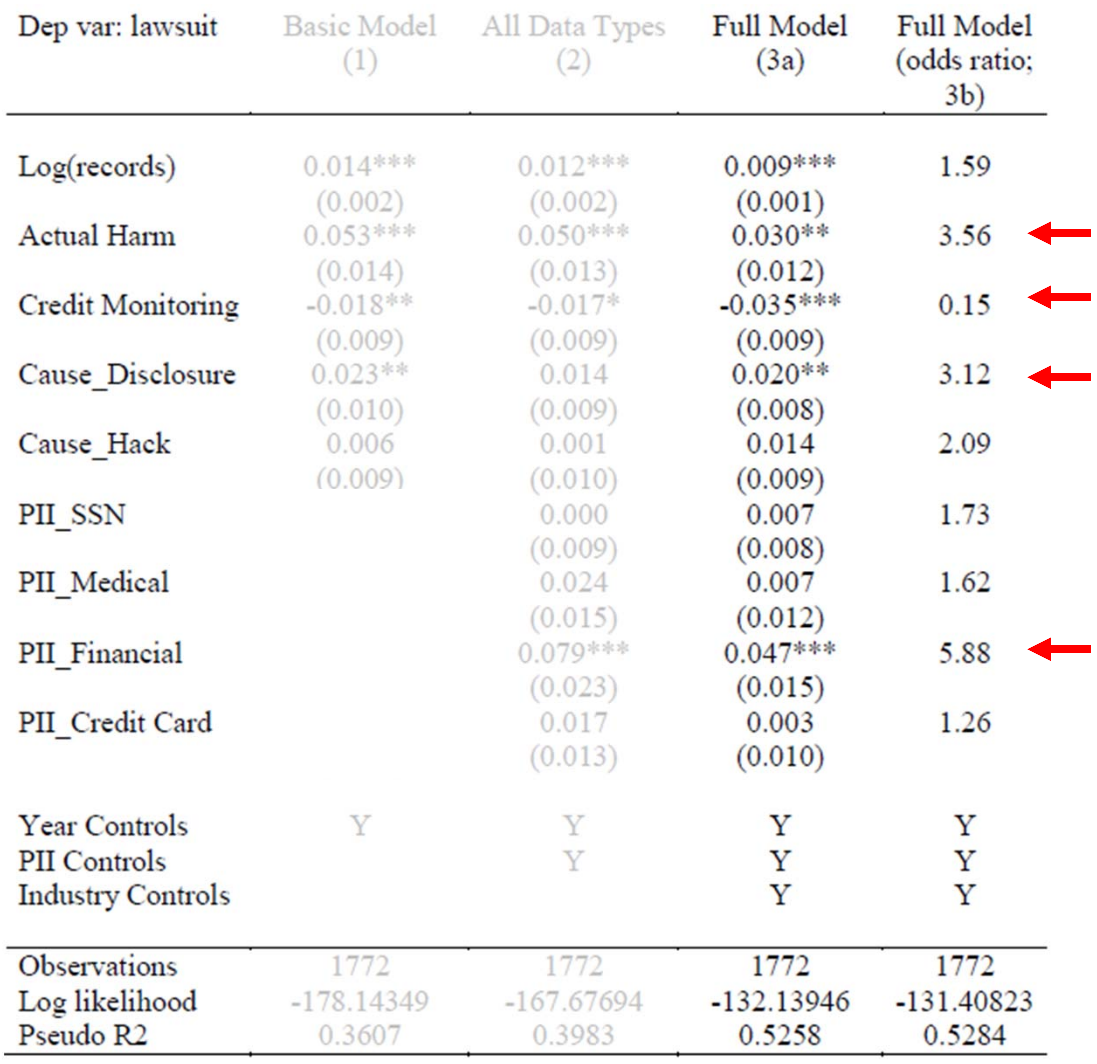
- <u>Negatively</u> correlated with instances of free credit monitoring.

Estimating model

- $Lawsuit_i = \alpha_0 + Size_i + ActualHarm_i + CreditMonitoring_i + Cause_i + PII_i + Controls_i + \varepsilon_i$

- *Lawsuit:* 1 if breach, i, was litigated.
- *Size:* log(number of records compromised).
- *ActualHarm:* 1 if evidence of financial loss from breach.
- CreditMonitoring: 1 if evidence of redress.
- *Cause:* categorical lost/stolen, improper disposal, cyberattack.
- *PII:* dummies for types of information compromised.
- *Controls:* firm industry, non-profit, publicly traded, year dummies.

Q1: Which breaches are being litigated?



Dep var: lawsuit	Basic Model (1)	All Data Types (2)	Full Model (3a)	Full Model (odds ratio; 3b)
Log(records)	0.014*** (0.002)	0.012*** (0.002)	0.009*** (0.001)	1.59
Actual Harm	0.053*** (0.014)	0.050*** (0.013)	0.030** (0.012)	3.56
Credit Monitoring	-0.018** (0.009)	-0.017* (0.009)	-0.035*** (0.009)	0.15
Cause_Disclosure	0.023** (0.010)	0.014 (0.009)	0.020** (0.008)	3.12
Cause_Hack	0.006 (0.009)	0.001 (0.010)	0.014 (0.009)	2.09
PII_SSN		0.000 (0.009)	0.007 (0.008)	1.73
PII_Medical		0.024 (0.015)	0.007 (0.012)	1.62
PII_Financial		0.079*** (0.023)	0.047*** (0.015)	5.88
PII_Credit Card		0.017 (0.013)	0.003 (0.010)	1.26
Year Controls	Y	Y	Y	Y
PII Controls		Y	Y	Y
Industry Controls			Y	Y
Observations	1772	1772	1772	1772
Log likelihood	-178.14349	-167.67694	-132.13946	-131.40823
Pseudo R2	0.3607	0.3983	0.5258	0.5284

Results show average marginal effects

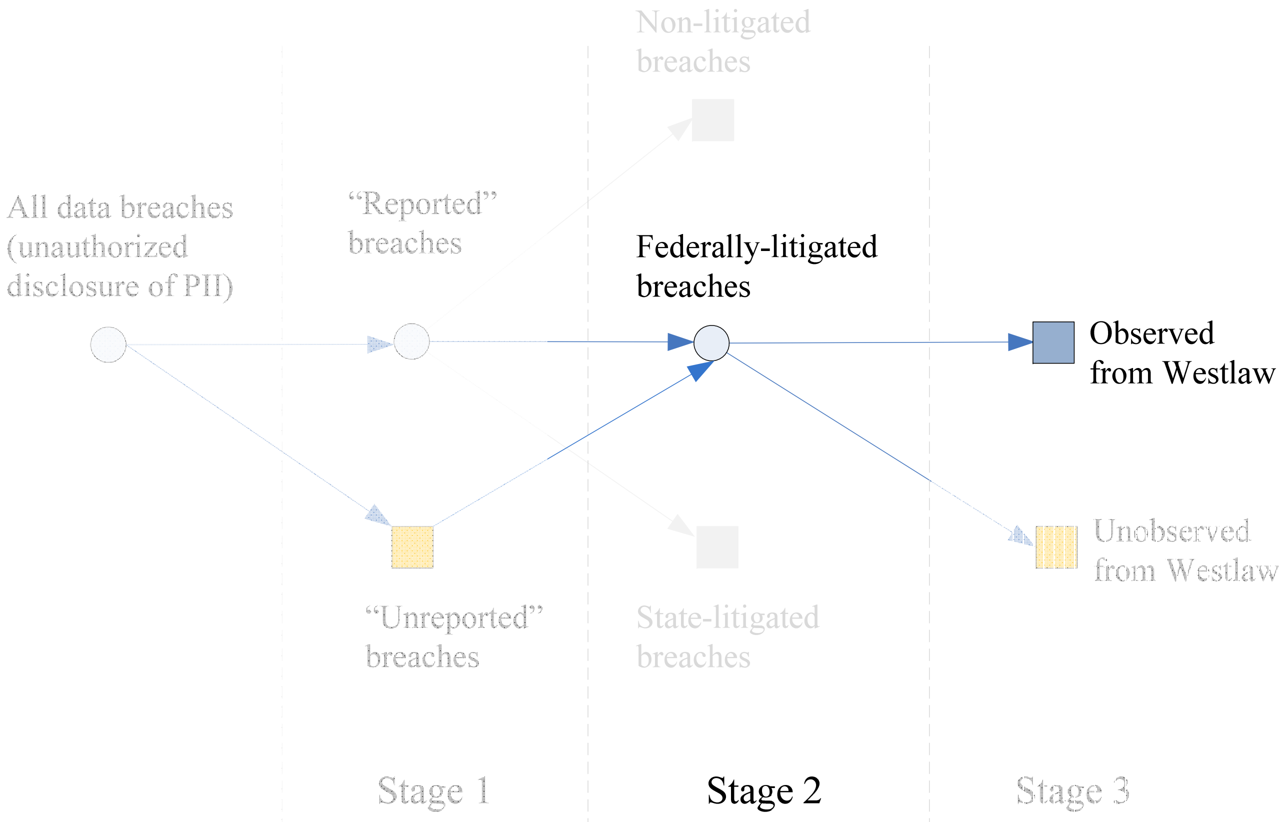
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

A possible causal interpretation for firms collecting PII, and how they should respond to a data breach

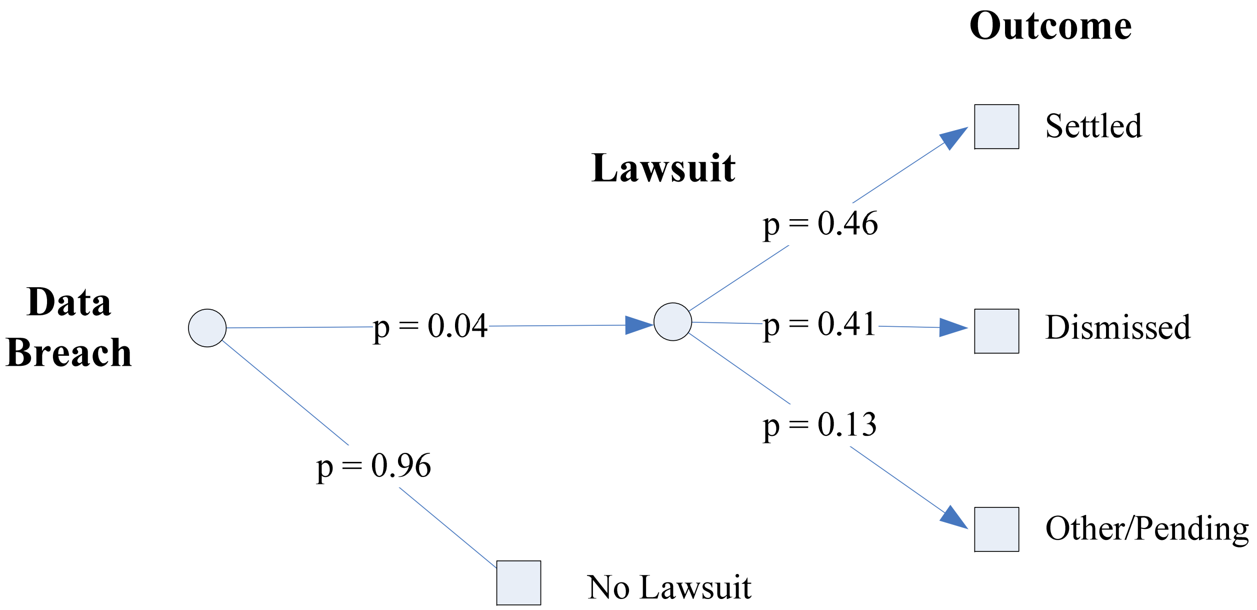
- While the overall probability of suit is small, the odds of a firm being sued is:
  - 3.5 times *greater* when actual loss occurs,
  - and almost 6 times *greater* when dealing with financial data,
  - but much *lower* when they provide free credit monitoring.

- Average marginal effects are small in magnitude, but statistically significant.

For Q2: All federal lawsuit observations



Descriptive data on lawsuit outcomes



- Settlement rate (46%) is lower than is ‘typical.’

Q2: Which data breach lawsuits settle?

- Theory suggests settlement increases with magnitude of award, probability of success.

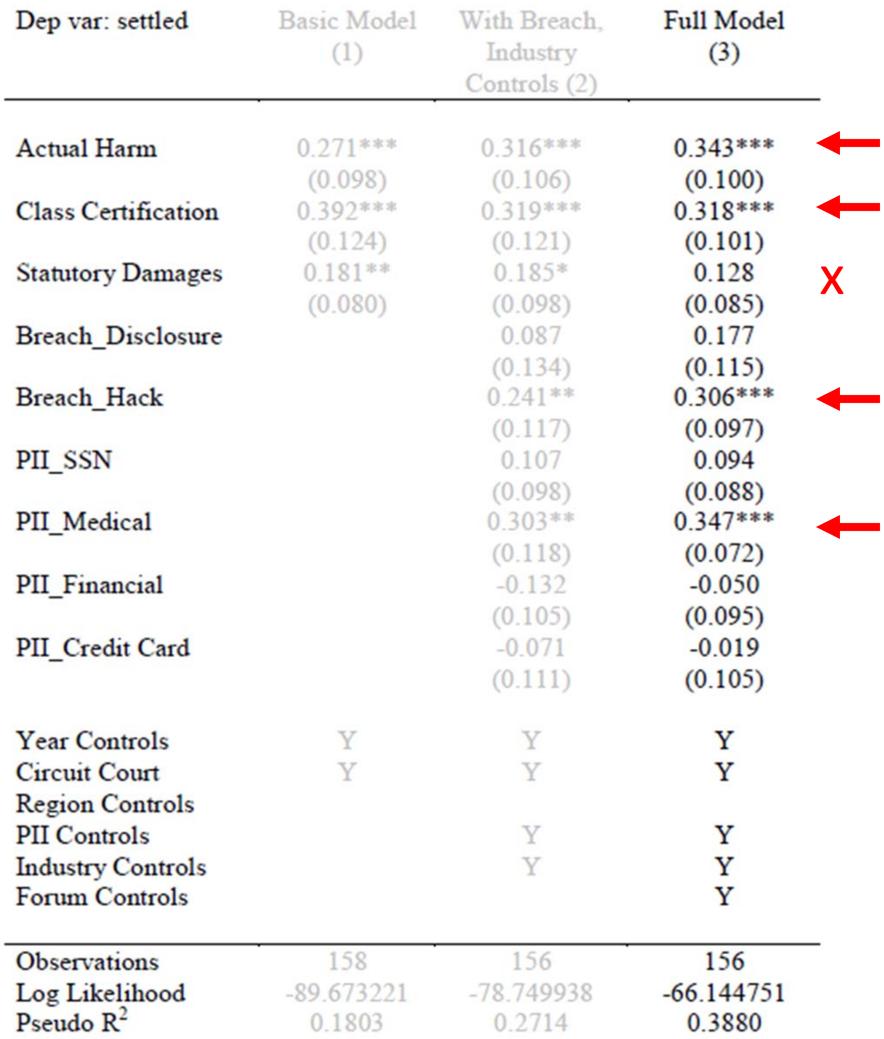
- The probability of settlement is positively correlated with lawsuits that:
  - *can demonstrate actual harm (measure of success),*
  - *achieve class certification (measure of magnitude),*
  - *seek statutory damages (measure of magnitude).*

$Settlement_i = \alpha_0 + ActualHarm_i + ClassCert_i + StatDam_i + Controls_i + \varepsilon_i$

- $ActualHarm_i$*:* financial loss asserted (not yet proven) in the complaint.

- $Controls_i$ *:* breach type, PII, forum shopping, year variables.

Q2: Which lawsuits settle?



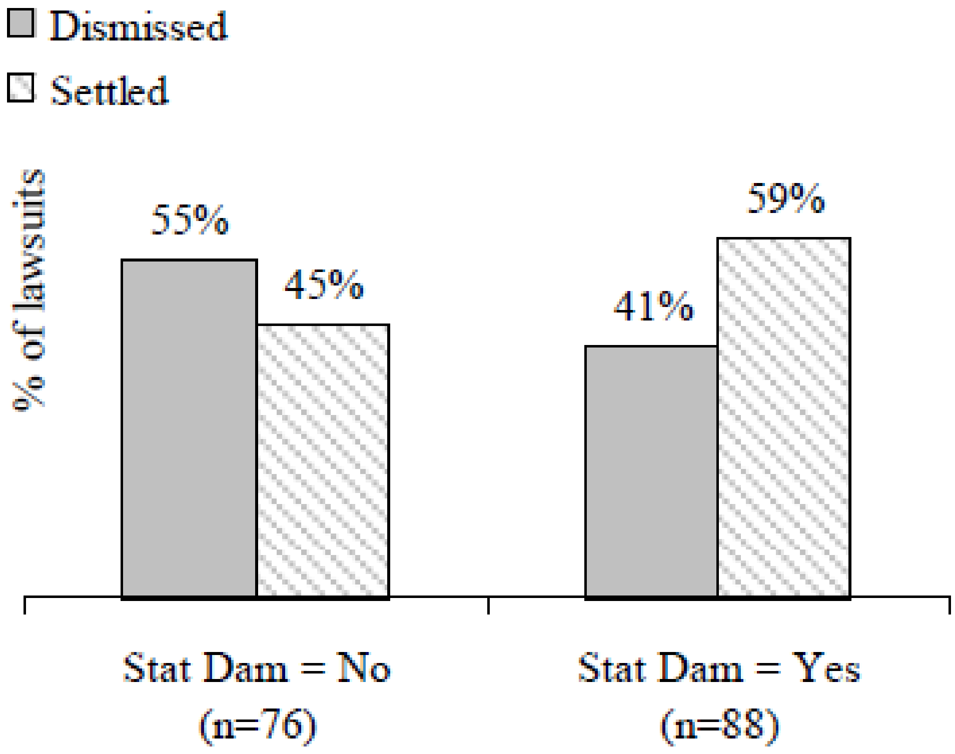
Dep var: settled	Basic Model (1)	With Breach, Industry Controls (2)	Full Model (3)	
Actual Harm	0.271***	0.316***	0.343***	←
	(0.098)	(0.106)	(0.100)	
Class Certification	0.392***	0.319***	0.318***	←
	(0.124)	(0.121)	(0.101)	
Statutory Damages	0.181**	0.185*	0.128	X
	(0.080)	(0.098)	(0.085)	
Breach_Disclosure		0.087	0.177	
		(0.134)	(0.115)	
Breach_Hack		0.241**	0.306***	←
		(0.117)	(0.097)	
PII_SSN		0.107	0.094	
		(0.098)	(0.088)	
PII_Medical		0.303**	0.347***	←
		(0.118)	(0.072)	
PII_Financial		-0.132	-0.050	
		(0.105)	(0.095)	
PII_Credit Card		-0.071	-0.019	
		(0.111)	(0.105)	
Year Controls	Y	Y	Y	
Circuit Court	Y	Y	Y	
Region Controls				
PII Controls		Y	Y	
Industry Controls		Y	Y	
Forum Controls			Y	
Observations	158	156	156	
Log Likelihood	-89.673221	-78.749938	-66.144751	
Pseudo R^2	0.1803	0.2714	0.3880	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Settlements

- Firms are about 30% more likely to settle when plaintiffs claim to suffer actual (financial) harm, and when class is certified (increase from 47% to about 60%).
- Surprisingly, statutory damages, were *not* found to drive settlement.
- Interestingly:
  - while loss of <u>financial</u> data and <u>careless handling</u> contributed to the probability of filing suit,
  - loss of <u>medical</u> data and <u>cyberattack</u> contributed to probability of settling a suit.

Pair-wise comparisons by settlement



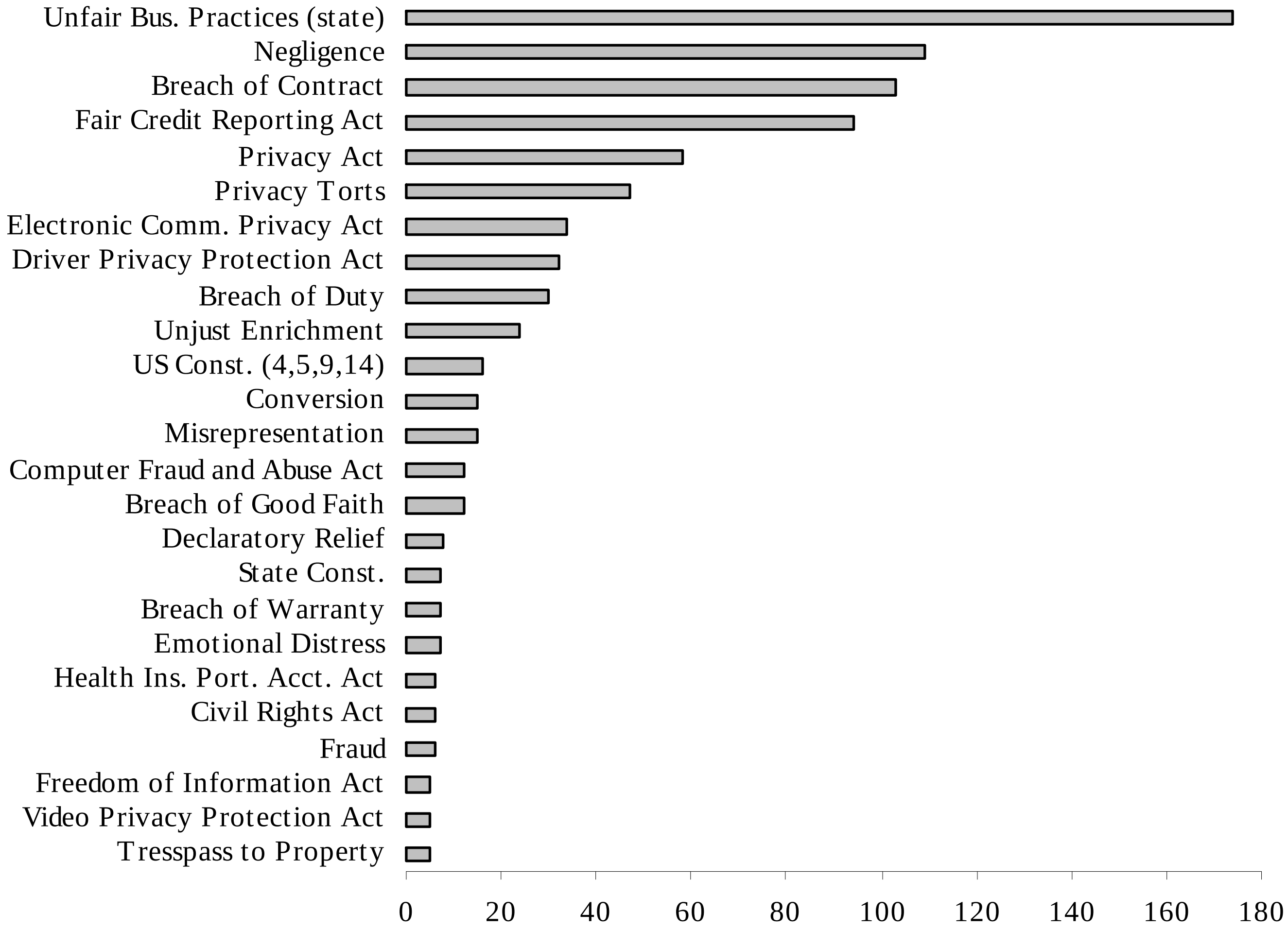
What do we know about settlement awards?

Known settlements:	28
Confidential settlements:	10
Unknown settlements:	48
Total settlements:	86

	Mean	Min	Max	N
Attorneys get:	$1.2m	$8k	$6.5m	15
Plaintiffs get:	$2.5k	$500	$15k	19

- Additional awards include redress for idtheft losses and expenses, cy pres awards to research, non-profits, charities.
- E.g. $50k, $2.8m, $5m, $6m, $8m, $9.5m.

What does variation suggest about effectiveness of current legal system?



What have we learned?

- Various potential policies can reduce the externalities caused by data breaches. Litigation is (a very contentious) one.

- Prescriptive guidance to firms:
  - Awareness of basic data handling practices appears to be the easiest way to avoid litigation.
  - Providing free credit monitoring is cheap way of avoiding costly lawsuit.
  - Financial and medical firms should pay particular attention.

- To policy makers:
  - If actual harm is appropriate measure of case merit, then litigation <u>does</u> appears to be resolving suits appropriately (both filing and outcome).

Limitations

- Not observing state suits is a limitation of this work. It prevents us from making inferences about *all* litigations.
    - However, Congressional activities and proposed legislation are key motivators for examining federal litigation.

- Discovery process is undocumented.
    - However, most firms will have discoverable liability insurance policies.

- We do not have a randomized experiment, and we are not testing a policy intervention.
    - However, if we believe our model, and the exogenous regressors, still possible to cautiously discuss about causality.

Thank you!

This research was supported by the National Science Foundation through CyLab grants DAAD19-02-1-0389 and W911NF-09-1-0273, from the Army Research Office, and by Temple Law School's Conwell Corps Program.