Do Banks Price Firms' Data Breaches?

Huang and Wang (2020)

Replication and Report by **Yunlin FENG**

April 22, 2024

1 Introduction

This report presents a replication of the paper "Do Banks Price Firms' Data Breaches?" by Huang and Wang, published in The Accounting Review in November 2020. The authors examine how reported data breaches affect firms' bank loan terms. They hypothesize that data breaches lead to higher default risk due to direct costs and reputation loss, as well as higher information risk, resulting in less favorable bank loan terms. The authors use a staggered difference-in-differences approach and find that breached firms experience significantly higher loan spreads, a higher likelihood of collateral requirements, and more covenants compared to control firms.

2 Replication

2.1 Sample Development

The replication follows the sample development steps outlined in the original paper (with few insignificant omissions). However, due to differences in data availability, the sample size differs slightly from the original study, as shown in Table 1. The replication starts with a larger number of data breaches (587) compared to the original study (551) from 2005 (2010) to 2014 (2020). After applying various filters, such as excluding firms with prior breach events, selecting the most significant events, and ensuring data availability for propensity score matching (PSM), the replication study obtains 228 event firms, compared to 213 in the original study. The replication then combines these event firms with matched control firms, resulting in a total of 456 firms, slightly more than the 426 firms in the original study. Finally, after merging with bank loan observations from 2003 (2008) to 2016 (2022) and applying additional filters, the replication study's final sample consists of 1,973 observations involving 149 data breach event firms, compared to 1,081 observations and 139 event firms in the original study.

Step	Authors'	Mine
# data breaches from 2005 (2010) to 2014 (2020)	551	587
Less:		
# firms with a prior breach event	(16)	(9)
# events that are not the most significant	(70)	(195)
# event firms lacking the data for (PSM)	(252)	(155)
# event firms after PSM	213	228
event firms + control firms	426	456
Bank loan observations from 2003 (2008) to 2016 (2022)	1,428	2,165
Less:		
financial services industries	(254)	(166)
bridge loans and non-fund-based facilities	(55)	(NA)
insufficient to calculate control variables	(37)	(26)
Final sample involving 139 (149) data breach event firms	1,081	1,973

Table 1: Sample Development

2.2 Probit Regression

The first step in the analysis is a probit regression to estimate the propensity scores for matching treatment and control firms. Table 2 presents the probit regression results, which are largely consistent with the original paper, although some variables are not available in the replication data. Table 2.2 presents the differences in variables between the treated and matched control group (replicated values in parenthesis). None of the differences are significant, indicating a succesesful match.

1

	Authors'	Mine
Firm Size$_{t-1}$	0.154***	0.182***
Leverage$_{t-1}$	-0.016	0.199*
ROA$_{t-1}$	0.089***	-0.164
Operational Risk$_{t-1}$	0.358*	0.761
Tangibility$_{t-1}$	-0.005	-0.331*
Z-score$_{t-1}$	-0.086	0.002
MB$_{t-1}$	-0.000**	0.010
IT Expertise$_{t-1}$	0.233***	-0.024
IT Reputation$_{t-1}$	0.222**	NA
Number of Segments$_{t-1}$	-0.009	NA
ICW$_{t-1}$	-0.134	NA
Intercept	-7.881***	-11.895
Industry/Year	Included	Included
Number of Observations	57,462	33,157
Pseudo R^2	0.166	0.243

Table 2: Probit Regression Results

Variable	Treated	Control	Diff.	p
Firm Size	8.308 (8.846)	8.190 (8.752)	0.118 (0.094)	0.591 (0.545)
Leverage	0.459 (0.653)	0.485 (0.679)	-0.026 (-0.026)	0.768 (0.397)
ROA	0.122 (0.134)	0.124 (0.130)	-0.002 (0.003)	0.859 (0.693)
Operational Risk	0.059 (0.039)	0.077 (0.036)	-0.018 (0.002)	0.137 (0.472)
Tangibility	0.431 (0.237)	0.449 (0.241)	-0.018 (-0.003)	0.643 (0.881)
Z-score	3.245 (4.965)	2.129 (4.770)	1.116 (0.196)	0.408 (0.809)
MB	2.330 (1.619)	2.455 (1.552)	-0.125 (0.067)	0.859 (0.683)
IT Expertise	0.364 (0.123)	0.341 (0.105)	0.023 (0.018)	0.616 (0.567)
IT Reputation	0.092	0.083	0.009	0.735
Number of Segments	2.055	1.853	0.203	0.252
ICW	0.028	0.009	0.018	0.154

Table 3: Difference in Variables for firms Matched by PSM

2.3 Descriptive Statistics

Table 4 presents the descriptive statistics of the key variables in the replication sample, which are largely consistent with the original paper (to facilate comparison, only mean is shown). Some differences worth pointing out:

- The loans in my sample are larger

- The loans in my sample are less likely to be secured

- The panel is not as balanced in my sample as post-breach loans are fewer than pre-breach loans

2.4 Difference-in-Differences Analysis

The main analysis uses a staggered difference-in-differences (DiD) approach to examine the effect of data breaches on bank loan terms. The regression specification is as follows:

$$
\begin{aligned}
\text{Loan Contract Terms} = {} & \beta_0 + \beta_1 \text{Data Breach} \times \text{Year}{-1} + \beta_1 \text{Data Breach} \times \text{Year0} \\
& + \beta_1 \text{Data Breach} \times \text{Year1} + \beta_1 \text{Data Breach} \times \text{Year2} \\
& + \beta \text{Controls} + \varepsilon
\end{aligned}
$$

Variable	Authors' Mean	My Mean
Bank Loan Characteristics		
Loan Spread	210.500	173.673
Loan Amount	0.954	1.781
Maturity	55.310	47.196
Performance Pricing	0.423	0.298
Secured	0.485	0.276
Total Covenants	3.096	5.104
Data Breach Variables		
Breach	0.543	0.568
Post	0.475	0.375
Firm-Level Variables		
Firm Size	8.779	9.422
Leverage	0.503	0.650
ROA	0.144	0.136
Operational Risk	0.043	0.032
Tangibility	0.568	0.285
Z-Score	2.883	3.966
MB	2.372	1.329
IT Expertise	0.391	0.114

Table 4: Descriptive Statistics

Table 5 presents the DiD regression results. Unlike the original paper, the replication does not find significant effects of data breaches on bank loan terms.

For comparison, Figures 1 presents the DiD regression results from the original paper.

3 Conclusion

The replication study does not find significant effects of data breaches on bank loan terms, in contrast to the original paper. Possible reasons for the discrepancy include:

- Different data breach databases. The paper's data breach sample is sourced from Privacy Rights Clearinghouse where as mine is based on WRDS. Since my sample are noticeably fewer than the authors', the database may be biased towards certain kinds of data breaches.

- Different sample period. My sample is 6 years later than the authors'. Perhaps as data breaches become more prevalent, banks no longer factor it as a risk when writing loans.

- Different control variable. Due to time constraint, the replication ignores several control variables such as firm characteristics and macroeconomic data. Including them may yield a different result.

	Loan Spread	Total Covenants	Secured
Breach:Year -1	0.009	−0.163	−0.033
	(0.063)	(0.330)	(0.041)
Breach:Year 0	−0.010	0.022	−0.028
	(0.064)	(0.318)	(0.041)
Breach:Year 1	−0.031	−0.202	−0.025
	(0.066)	(0.415)	(0.045)
Breach:Year 2+	−0.033	−0.177	0.008
	(0.045)	(0.251)	(0.030)
Log(Loan Amount)	−0.018	−0.090	−0.018*
	(0.016)	(0.103)	(0.010)
Log(Loan Maturity)	−0.010	−0.031	0.066***
	(0.022)	(0.124)	(0.014)
Performance Pricing	0.082***	0.172	0.043**
	(0.027)	(0.136)	(0.018)
Firm Size	−0.200***	−0.365*	−0.037
	(0.037)	(0.211)	(0.024)
Leverage	0.844***	1.012	0.203**
	(0.128)	(0.922)	(0.085)
ROA	−1.075***	−1.156	0.021
	(0.248)	(2.257)	(0.168)
Operational Risk	0.309	−3.580	0.170
	(0.638)	(3.449)	(0.418)
Tangibility	0.309	2.320	−0.398***
	(0.237)	(1.417)	(0.154)
Z-Score	0.029*	−0.103	−0.012
	(0.016)	(0.102)	(0.011)
MB	−0.079**	0.094	−0.039*
	(0.031)	(0.232)	(0.021)
IT Expertise	−0.161***	0.120	0.009
	(0.054)	(0.288)	(0.034)
Observations	1,608	850	1,889

Table 5: Difference-in-Differences Regression Results

Dependent Variable	(1) Ln(Loan Spread)	(2) Secured	(3) Number of Total Covenants
Data Breach * Year −1	−0.014	−0.008	−0.332
	(−0.17)	(−0.18)	(−0.96)
Data Breach * Year 0	0.052	0.067	−0.074
	(0.66)	(1.41)	(−0.21)
Data Breach * Year 1	0.221**	0.114**	0.783*
	(2.41)	(2.26)	(1.84)
Data Breach * Year 2+	0.037	0.097**	0.830**
	(0.49)	(2.14)	(2.13)
Ln(Loan Size)	−0.369***	−0.129***	0.196
	(−6.82)	(−3.95)	(0.82)
Ln(Loan Maturity)	0.197***	0.114***	−0.145
	(4.56)	(4.79)	(−0.91)
Performance Pricing	−0.193***	−0.029	2.135***
	(−5.25)	(−1.19)	(9.44)
Firm Size	−0.089	−0.024	−1.096*
	(−0.82)	(−0.43)	(−1.94)
Leverage	0.360*	−0.237*	−2.853**
	(1.93)	(−1.85)	(−2.46)
ROA	−3.625***	−0.848	−10.396**
	(−3.93)	(−1.57)	(−2.51)
Operational Risk	0.593	−0.070	6.207
	(0.56)	(−0.12)	(1.53)
Tangibility	−0.322*	0.089	−1.232
	(−1.66)	(0.69)	(−1.13)
Z-score	0.064	−0.038	−0.186
	(1.00)	(−1.30)	(−0.71)
MB	0.002	0.008	−0.029
	(0.20)	(1.51)	(−0.44)
IT Expertise	−0.044	−0.078*	0.065
	(−0.70)	(−1.85)	(0.20)
IT Reputation	0.089	−0.060	0.309
	(1.00)	(−1.10)	(0.86)
Number of Segments	−0.064***	−0.005	0.185*
	(−2.84)	(−0.33)	(1.68)
ICW	−0.162	0.045	0.608
	(−0.92)	(0.49)	(0.72)
Credit Spread	0.375***	0.130*	1.774***
	(3.49)	(1.87)	(3.60)
Term Spread	0.119**	0.092**	0.950***
	(2.08)	(2.51)	(3.33)
Intercept	5.339***	0.812	10.657*
	(5.12)	(1.32)	(1.80)
Firm/Year	Included	Included	Included
Number of Observations	1,081	1,081	1,081
R^2	0.756	0.672	0.609

Figure 1: Original Paper DiD Results