Sufi(2009), Sufi and Roberts(2009) Replication

Dexin Guo

February 2025

Sufi(2009)

By running 'BA952 Replication/Sufi/do_file/main.do', you are expected to get all empirical results in this write-up.
Definitions discussion

- `cash flow volatility`: Defined as the standard deviation of cash flows over a five-year rolling window in the main results and as the standard deviation of quarterly cash flows within a year in Appendix: Sufi(2009).

- `industry sale volatility`: Defined as the standard deviation of industry sales over a five-year rolling window in the main results and as the standard deviation of quarterly industry sales within a year in Appendix: Sufi(2009).[1]

- `firm age` and `Ln(firm age)`: Due to a high proportion of missing values in the firm-year observations of the Sufi (2009) dataset, missing `firm age` is replaced by the `mean firm age in industry` in the same year. Specifically, if `data year = IPO year`, we define `firm age==0`. Instead of using `Ln(firm age)`, we use `Ln(firm age + 1)` to ensure all values remain positive (1029 observations in our sample has firm age $\leq$ 1). And I think the author used another datasets or manually collect the missing `ipodate` for the public firm with missing `ipodate` in `compustat`.

- `Assets-cash`: to be aligned with the `winsorize` process in other variables report in `Table 1`, here we show it after winsorize `Assets-cash` at 5% level.

Table 1 Summary Statistics

	Full Sample			Random Sample		
	Mean	Median	St.Dev	Mean	Median	St.Dev
Line of Credit Variables						
Has line of credit 0,1	0.823	1.000	0.381	0.756	1.000	0.429
Total line of credit/assets				0.163	0.115	0.172
Unused line of credit/assets				0.104	0.069	0.127
Used line of credit/assets				0.059	0.000	0.099
Firm Characteristics						
Book debt/assets	0.201	0.175	0.179	0.203	0.178	0.183
EBITDA/(assets-cash)	0.048	0.122	0.294	0.057	0.122	0.292
Tangible assets/(assets-cash)	0.338	0.276	0.240	0.336	0.283	0.228
Net worth, cash adjusted	0.783	0.599	0.592	0.787	0.610	0.589
Assets-cash	813.203	116.757	1661.312	809.831	124.783	1658.685
Market-to-book, cash adjusted	4.414	2.495	7.102	4.339	2.467	6.668
Industry sales volatility	0.099	0.083	0.047	0.104	0.083	0.050
Cash-flow volatility	0.147	0.057	0.215	0.143	0.058	0.211
Not in an S&P index 0,1	0.223	0.000	0.416	0.235	0.000	0.424
Traded over the counter 0,1	0.430	0.000	0.495	0.420	0.000	0.494
Firm age (years since IPO)	6.199	6.000	3.490	6.003	6.000	3.536

---

[1]The reason is: since other control variables reflect the firm's performance, within-year-standard deviation could better reflect the performance in specific year, thus match the data level with other variables

- **Discussion of Regression Results**
  - The key regression results $(\frac{\text{EBITDA}}{\text{assets-cash}\,t\text{-}1})$are in align with Sufi (2009) in terms of the direction and significance of the coefficients. And the trend shown in Figure 1 is also consistent. However, several aspects deserve attention:
    * With `Ln(firm age + 1)`, the coefficient is **negatively significant** across all regressions, highlighting the importance of considering the sign of `Ln(firm age)`.
    * The coefficient of $\frac{\text{EBITDA}}{\text{assets-cash}\,t\text{-}1}$ becomes statistically insignificant in columns (3)–(6), possibly due to differences in variable definitions and datasets.
    * There are 1,676 fewer observations in the Full sample (column (1)) compared to Sufi (2009), all of which are **inactive firms**[2].
  - The trend in `Figure 1` is consistent with Sufi (2009) and remains robust when using alternative variable definitions (see Appendix: Sufi(2009)).
  - In summary, the slight change in definitions could have significant effects in regression that can't be ignored. The robustness of the results shown by Sufi(2009) need further discussion.
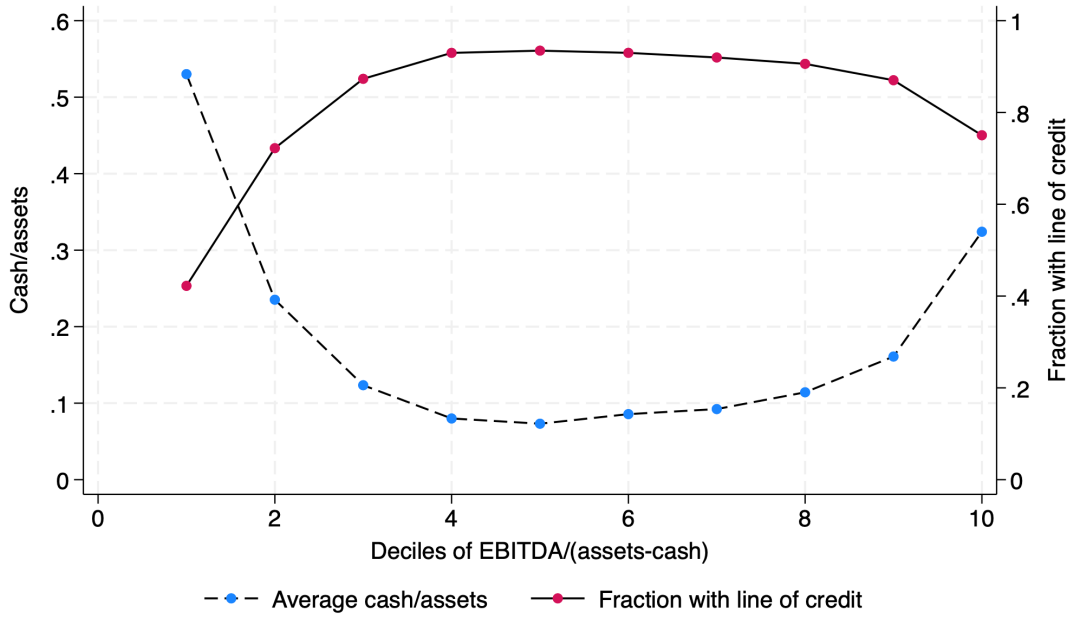

Figure 1 Use of line of credit versus cash holdings across cash-flow distribution

Discussion of the Reduced Form Regression

- This paper tries to answer the questions: **What factors determine the use of bank of credit**. And the two-stages are as following:

$$\text{First Stage:}\quad Pr(\text{get } \tilde{\text{use}} \text{ of credit}) = f(\text{possible factors})$$

$$\text{Second stage:}\quad Pr(\text{has line of credit})/\text{The amount/proportion of credit used} = g(Pr(\text{get } \tilde{\text{use}} \text{ of credit}))$$

However, we only have reduced form regression here in Table 3, which is not the whole story.

---

[2]See Appendix: Sufi(2009) for details on missing variables that led to the drop.

Table 3 Bank lines of credit and firm characteristics

	Has line of credit (0,1) Probit		Total line / (total line + cash) OLS		Unused line / (unused line + cash) OLS	
	Full	Random	Random	With Line of Credit	Random	With Line of Credit
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{EBITDA}{assets\text{-}cash}_{t\text{-}1}$	0.447***	0.580*	0.059	0.018	0.072	0.100
	(0.057)	(0.263)	(0.044)	(0.061)	(0.040)	(0.060)
$\frac{Tangible\ assets}{assets\text{-}cash}_{t\text{-}1}$	-0.005	0.149	0.000	-0.015	-0.011	-0.038
	(0.085)	(0.381)	(0.071)	(0.055)	(0.070)	(0.066)
Ln(assets-cash)$_{t\text{-}1}$	0.147***	0.134*	0.029**	0.002	0.040***	0.024*
	(0.013)	(0.059)	(0.011)	(0.008)	(0.010)	(0.009)
Net worth, cash adjusted$_{t\text{-}1}$	-0.613***	-0.913***	-0.284***	-0.352***	-0.220***	-0.243***
	(0.030)	(0.141)	(0.027)	(0.031)	(0.025)	(0.034)
Market-to-book, cash adjusted$_{t\text{-}1}$	-0.005**	-0.003	-0.002	-0.006***	-0.002	-0.005**
	(0.002)	(0.009)	(0.001)	(0.002)	(0.001)	(0.002)
Industry sales volatility$_{t\text{-}1}$	1.215*	-2.023	-0.306	0.123	-0.917*	-0.752
	(0.619)	(2.436)	(0.349)	(0.294)	(0.364)	(0.386)
Not in an S&P index 0,1	-0.034	0.171	0.035	0.041	0.046	0.048
	(0.051)	(0.171)	(0.034)	(0.026)	(0.034)	(0.029)
Traded over the counter 0,1	-0.006	0.055	-0.044	-0.045	-0.042	-0.037
	(0.040)	(0.153)	(0.034)	(0.030)	(0.033)	(0.032)
Firm age (years since IPO)$_{t\text{-}1}$	-0.256***	-0.422*	-0.109***	-0.107***	-0.090**	-0.088**
	(0.044)	(0.167)	(0.031)	(0.027)	(0.030)	(0.031)
Number of firms	3896	257	257	223	257	223
Number of observations	26108	1723	1723	1303	1723	1303

Standard errors in parentheses. Full = Full Sample, Random = Random Sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sufi and Roberts(2009)

By running 'BA952 Replication/Sufi and Roberts(2009)/do_file/main.do', you are expected to get all empirical results in this write-up.

- **Discussion of Variable Definitions**

    - Due to the long right tail of `Net Debt Issuance` and `Net Equity Issuance`(1% quantile=-1775,95% quantile=4253), we winsorize these variables using asymmetric quantile setting(1% and 85% for `Net Equity Issuance`). And the results are consistent with Sufi and Roberts(2009)

    - Instead of using `txditc` (yearly deferred taxes and investment tax credit) as in Sufi (2009), we use `txditcq` (quarterly level) for consistency with other variables. Book equity is defined as `book value equity = seq - pstkq + ceqq - txditcq`.

<div align="center">Table I Covenant Violations</div>

| | Capital Structure Variables | | |
	Mean	Median	SD
Capital Structure Variables			
Net Debt Issuance (basis point)	58.061	0.000	430.668
Net Equity Issuance (basis point)	40.080	0.560	267.713
$\frac{\text{Book Debt}_t}{\text{Assets}_t}$	0.273	0.232	0.243
Covenant Control Variables			
$\frac{\text{Net Worth}_t}{\text{Assets}_t}$	0.419	0.434	0.310
$\frac{\text{Net Working Capital}_t}{\text{Assets}_t}$	0.188	0.182	0.319
$\frac{\text{Cash}_t}{\text{Assets}_t}$	0.165	0.059	0.217
$\frac{\text{EBITDA}_t}{\text{Assets}_{t-1}}$	0.005	0.023	0.059
$\frac{\text{Cash Flow}_t}{\text{Assets}_{t-1}}$	-0.008	0.014	0.065
$\frac{\text{Net Income}_t}{\text{Assets}_{t-1}}$	-0.026	0.003	0.074
$\frac{\text{Interest Expense}_t}{\text{Assets}_{t-1}}$	0.006	0.005	0.007
Other Control Variables			
Market-to-book Ratio	1.317	0.890	1.631
$\frac{\text{Tangible}_t}{\text{Assets}_t}$	0.286	0.212	0.237
$\ln(\text{Assets}_t)$	4.511	4.486	2.292

Discussion of Regression Results

- **Our results are inconsistent with Sufi and Roberts(2009) in several aspects**, especially in first difference regression:

    - In Panel A:
        * the coefficient of `Covenant violation` in column (1) has different direction though statistically insignificant.
        * the coefficients of `lag Covenant violation` are consistent with Sufi and Roberts but more significant economically.

    - In Panel B:
        * the coefficients of `lag Covenant violation` are less economically significant compared with Sufi(2009) results.
        * the coefficients of `Covenant violation` are more economically significant though remain statistically insignificant for column(3)-(4)

- **Possible reasons**

    - **Use `txditcq` instead of `txditc`**, there are two effects brought by the change of definition:
        * First, the number of observations are changed due to the difference between `missing(txditcq)` and `missing(txditc)`
        * Second, `Market-to-book Ratio` **has economically and statically significant correlation with dependent variable**, for example, in all fixed effect regression, the coefficients of `Market to Book Ratio`$_t$ is significant at 1% and with value approximate 4.5.

- **Potential Problems in the order of Winsorize the lag control variables and generate lag variables**
  - In Sufi and Roberts (2009), all `lag control variables` are winsorized at 5%. In contrast, in my `'clean.do'`, I winsorize the variables before generating the lag terms to:
    * Prevent double-winsorization. For example, winsorizing `Assets (at)` automatically addresses outliers in `Lag Assets`.
  - Potential issue:
    * Lag terms are generated using observations outside the research period, leading to winsorization with extra observations and potentially introducing new information.[3]
  - The coefficients of key variables differ from Sufi and Roberts (2009) when using lagged one-quarter differences for control variables in the Panel B regression. Specifically, `Covenant violation` turns negatively significant, while `Lag Covenant violation` becomes insignificant (see Appendix Sufi and Roberts(2009)).

Table III Covenant Violations and Net Debt Issuance

Panel A: Fixed Effects				
Dependent Variable: Net debt issuance$_t$/assets$_{t-1}$ (Basis Points)				
	(1)	(2)	(3)	(4)
Covenant violation$_t$	-2.57	3.07	2.29	2.29
	(7.00)	(6.42)	(6.43)	(6.43)
Covenant violation$_{t-1}$	-79.31***	-61.97***	-62.58***	-62.58***
	(7.20)	(6.70)	(6.72)	(6.72)
Covenant control variables	None	Covenant control variables	Covenant control variables, covenant interaction control variables	Control variables, control variables squared, control variables to the third power, and quintile indicators for each control
Number of firm-quarters	105,938	105,938	105,938	105,938
Number of firms	6,782	6,782	6,782	6,782
R^2	0.001	0.084	0.092	0.092
Panel B: First Differences				
Dependent Variable: Change in Net debt issuance$_t$/assets$_{t-1}$ (Basis Points)				
	(1)	(2)	(3)	(4)
Covenant violation	11.23	12.40	14.51	17.14*
	(8.37)	(7.74)	(7.69)	(7.57)
Covenant violation$_{t-1}$	-34.45***	-16.75*	-15.67	-16.75*
	(8.66)	(8.06)	(8.01)	(7.91)
Covenant control variables	None	Covenant control variables	Covenant control variables, covenant interaction control variables	Control variables, control variables squared, control variables to the third power, and quintile indicators for each control
Number of firm-quarters	90,792	90,792	90,792	90,792
Number of firms	6,776	6,776	6,776	6,776
R^2	0.046	0.174	0.194	0.219

Notes: Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

---

[3]This step will be revised after submission due to time constraints.

Appendix: Sufi(2009)

The following results are generated using **within year standard deviation** definitions to calculate the `cash flow volatility` and `industry sale volatility`.

- Table 1A shows that under these definitions the variance and the scale of these variables are relatively small compared with Sufi(2009) and Table 1.

- Since it requires at least 2 quarters observations of cash flow/industry sale to generate within-year standard deviations, less observations are included in our datasets.

- Figure 1A shows the consistent result with Sufi(2009) and Figure 1.

Table IA Summary Statistics

|	*Full Sample*			*Random Sample*		
	Mean	Median	St.Dev	Mean	Median	St.Dev
Line of Credit Variables						
Has line of credit 0,1	0.826	1.000	0.379	0.761	1.000	0.427
Total line of credit/assets				0.163	0.115	0.169
Unused line of credit/assets				0.105	0.070	0.125
Used line of credit/assets				0.058	0.000	0.099
Firm Characteristics						
Book debt/assets	0.203	0.179	0.180	0.203	0.182	0.181
EBITDA/(assets-cash)	0.056	0.124	0.284	0.071	0.127	0.276
Tangible assets/(assets-cash)	0.349	0.287	0.245	0.350	0.304	0.230
Net worth, cash adjusted	0.776	0.593	0.586	0.779	0.604	0.580
Assets-cash	875.585	124.481	1796.391	899.323	144.541	1813.409
Market-to-book, cash adjusted	4.431	2.496	6.825	4.401	2.513	6.379
Industry sales volatility	0.001	0.001	0.001	0.001	0.001	0.001
Cash-flow volatility	0.035	0.013	0.052	0.033	0.013	0.049
Not in an S&P index 0,1	0.226	0.000	0.418	0.249	0.000	0.433
Traded over the counter 0,1	0.425	0.000	0.494	0.428	0.000	0.495
Firm age (years since IPO)	6.221	6.000	3.509	6.055	6.000	3.645
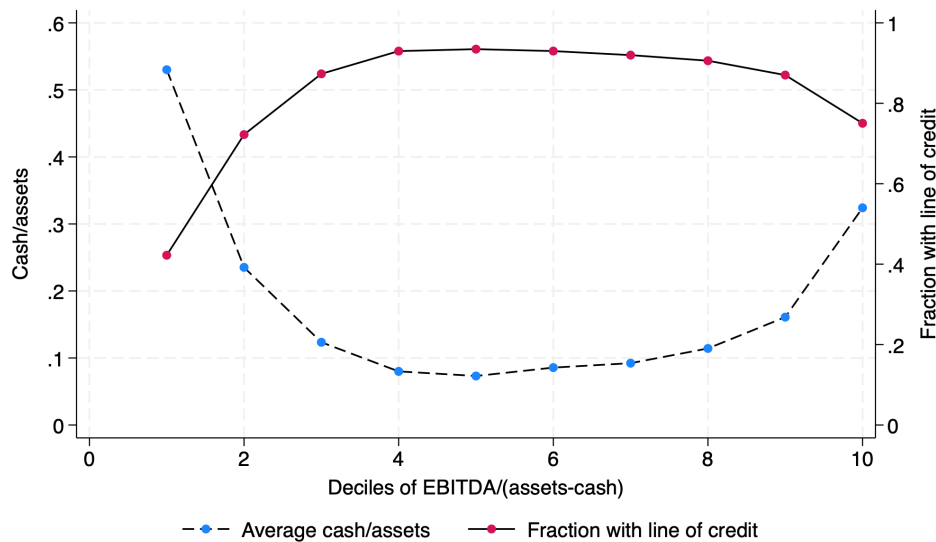
Figure 1A Use of line of credit versus cash holdings across cash-flow distribution

Table 3A Bank lines of credit and firm characteristics

	Has line of credit (0,1) Probit		Total line / (total line + cash) OLS		Unused line / (unused line + cash) OLS	
	Full (1)	Random (2)	Random (3)	With Line of Credit (4)	Random (5)	With Line of Credit (6)
$\frac{\text{EBITDA}}{\text{assets-cash}}_{t-1}$	0.457*** (0.064)	0.307 (0.310)	0.024 (0.048)	0.049 (0.067)	0.038 (0.044)	0.123 (0.068)
$\frac{\text{Tangible assets}}{\text{assets-cash}}_{t-1}$	0.002 (0.090)	0.172 (0.430)	-0.020 (0.078)	-0.052 (0.059)	-0.033 (0.075)	-0.089 (0.070)
$\text{Ln(assets-cash)}_{t-1}$	0.142*** (0.014)	0.137* (0.063)	0.031* (0.012)	0.006 (0.009)	0.042*** (0.011)	0.030** (0.011)
Net worth, cash adjusted$_{t-1}$	-0.633*** (0.033)	-1.083*** (0.163)	-0.288*** (0.029)	-0.318*** (0.037)	-0.223*** (0.028)	-0.209*** (0.041)
Market-to-book, cash adjusted$_{t-1}$	-0.007** (0.002)	0.002 (0.010)	-0.003* (0.002)	-0.008*** (0.002)	-0.003 (0.002)	-0.007** (0.002)
Industry sales volatility$_{t-1}$	-43.961 (33.269)	-150.755 (126.558)	0.857 (23.223)	22.473 (21.488)	15.052 (21.795)	38.421 (22.944)
Not in an S&P index 0,1	-0.052 (0.054)	0.247 (0.178)	0.049 (0.035)	0.043 (0.027)	0.062 (0.036)	0.055 (0.030)
Traded over the counter 0,1	-0.018 (0.044)	0.059 (0.168)	-0.043 (0.037)	-0.039 (0.032)	-0.041 (0.037)	-0.034 (0.035)
Firm age (years since IPO)$_{t-1}$	-0.286*** (0.049)	-0.442* (0.188)	-0.123*** (0.034)	-0.109*** (0.031)	-0.102** (0.034)	-0.092** (0.034)
Number of firms	3524	232	232	201	232	201
Number of observations	22100	1462	1462	1112	1462	1112

Standard errors in parentheses. Full = Full Sample, Random = Random Sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The following table shows factors that drive the drop of observations which are included in Sufi(2009) datasets:

Missing Variables Summary in Sufi(2009)

Variable	Obs=.	Obs .	Unique values	Min	Max
dcvt	8	2,328	199	0	4807.9
dlc	25	2,311	>500	0	6661
dltt	9	2,327	>500	0	20099
lt	1	2,335	>500	0	59805
oibdp	20	2,316	>500	-567.413	17001
ppent	5	2,331	>500	0	58530
pstkc	7	2,329	140	0	1570.2
txdb	776	1,560	357	0	1051

Appendix Sufi and Roberts(2009)

This table shows the Panel B for using lagged one quarter controls in the regression, detailed discussions are shown above.

Table III Appendix Covenant Violations and Net Debt Issuance

	Panel B: First Differences			
	Dependent Variable: Change in Net debt issuance$_t$/assets$_{t-1}$ (Basis Points)			
	(1)	(2)	(3)	(4)
d_violation	10.76	21.16**	21.68**	21.89***
	(8.30)	(6.70)	(6.69)	(6.58)
d_lag_violation	-34.72***	-6.59	-7.14	-6.92
	(8.55)	(6.88)	(6.85)	(6.76)
N	92973	92973	92973	92973

Notes: Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.