Advanced Corporate Finance - Assignment 1
Miguel A. Troppmair
Pedro Watuhã G. dos Santos

Abstract

This report explores firm capital structure, mainly the impact on their leverage and its explanation. We test different specifications to deal with endogeneity and other problems related to the data. Finally, we address causality concerns and propose a solution.

1 Introduction

This report aims to reinforce the findings in the capital structure literature by utilizing data from the Capital IQ database, covering the period from 1981 to 2014, specifically focusing on manufacturing firms. Capital structure, defined as the mix of debt and equity financing used by firms, is a crucial area of study in corporate finance, and this report seeks to provide additional empirical evidence to support existing theories.

To conduct this analysis, we generate several key variables from the dataset. Leverage is measured as the ratio of total debt to total assets, while CashFlow is estimated by the ratio of income and depreciation over the lagged net property. Firm Size is represented by the natural logarithm of total assets, and Q is proxied by the market-to-book ratio. Tangibility is calculated as the ratio of fixed assets to total assets.

Using these variables, we estimate the impact of each on leverage through various econometric methods and specifications. This approach allows us to address common issues in empirical finance research, such as endogeneity, inference, and outliers. By applying rigorous econometric techniques, we aim to provide robust and reliable findings that contribute to the understanding of capital structure determinants within the manufacturing sector.

In the subsequent sections, we will detail the data and methodology, present the results of our analysis, discuss the implications of our findings, and conclude with a discussion of how robust our findings might be.

2 Data

We use data from the Capital IQ database, covering the period from 1981 to 2014. The dataset specifically focuses on manufacturing firms, providing a overview of their capital structures during this period. We dropped observations which had no value for one of the variables, ending up with a total of 48,888 observations.

For all results presented we used $Q = \frac{TotalAssets - MarketCap - DefTaxAssetsCurr - BookValueShare}{TotalAssets}$. We find a large mean of 46. If we winsor the data, Q's mean goes to 2.4, which shows the great impact of outliers in the data.

Variable	Obs	Mean	Std. dev.	Min	Max
Leverage	48,880	1.132492	124.6743	0	27,300
CashFlow	48,880	-24.38178	1,064.727	-201,900	9,622.728
Tangibility	48,880	0.2405505	0.167345	0	1
Size	48,880	5.405497	2.580933	-6.907755	13.00952
Q	48,880	46.00271	8,891.949	-66.87971	1,965,801

Table 1: Descriptive Statistics

The impact of outliers can be seen by analysing the overall distribution of the variables at Figure 1. There is a concern for outliers and the skewness of the distributions. Observing the 3 variables in which Winsor was utilised, it is possible to realise the proportion of data in the most extreme quantiles and how greater (smaller) they are in magnitude when compared to the remaining data. On the other hand, the variables Size and Tangibility are less affected by outliers. Apart from that, there is also the fact that thoose outliers have a much greater magnitude than what is seem in the data, suggesting possible bias in the estimations.

Looking to the correlation, we can see that the Leverage and Q are (highly) positively correlated (0.991), and so is Tangibility and Size (0.186). The other variables are much less so. This very strong relationship between dependent and independent variables might cause a problem of co-linearity.

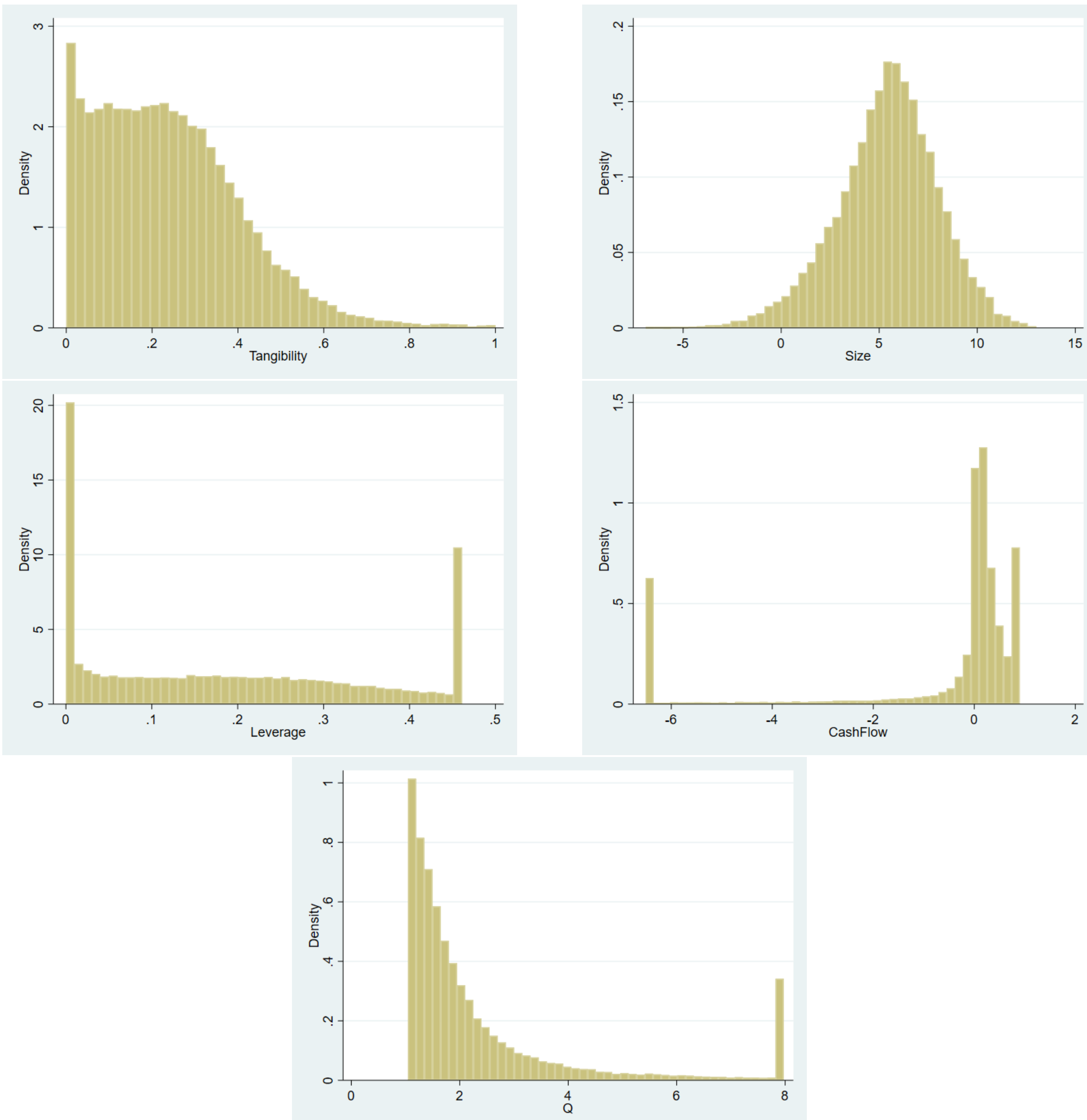

Figure 1: Data Distribution with Winsor for Data with Multiple Outliners

Size and Tangibility distributions are shown without any correction. For *Leverage*, *CashFlow* and *Q*, we plot them after winsorizing

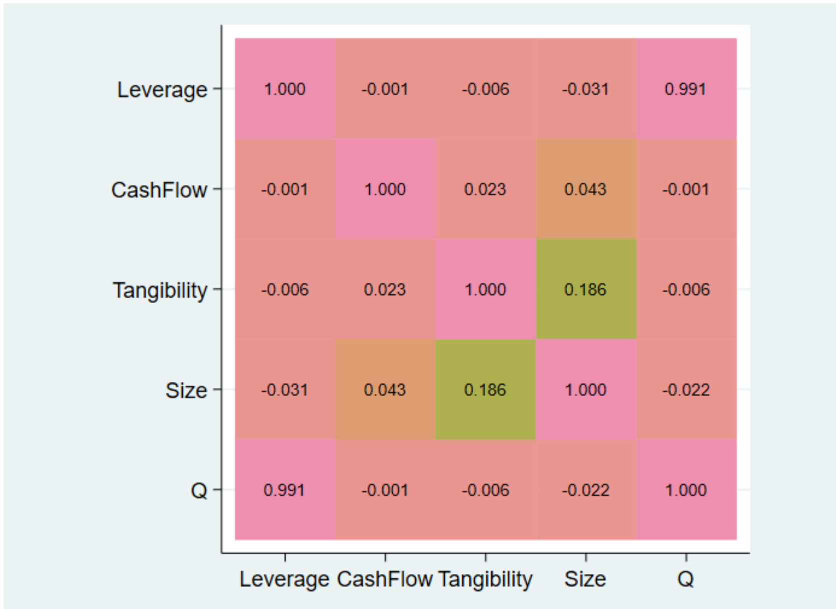

Figure 2: Heatmap for Correlations

3 Results

The baseline regression (1) is the following:

$$Leverage_{i,t} = \beta_0 + \beta_1 CashFlow_{i,t} + \beta_2 Tangibility_{i,t} + \beta_3 Size_{i,t} + \beta_4 Q_{i,t} + \varepsilon_{i,t}$$

This model however may suffer from endogeneity problems. As argued in Campello, Giambona (2013) - Real Assets and Capital Structure, firms choose the tangibility of their assets (the decision of which specific assets should be bought), making it an endogenous variable, i.e., the decision of tangibility is not necessarily independent on other firms' decisions such as optimal leverage.

This way, a firm could increase debt (and so leverage) to buy tangible assets creating a reverse causality relation. Another possibility is a firm with certain characteristics, such as solid fundamentals, simultaneously wanting more debt and more fixed assets, causing a omitted variable bias. This problem is difficult to be dealt with, and simple modifications may have little effect on controlling for endogeneity.

All variables except $CashFlow$ are highly significant, specially Q. This is no surprise, as seen on figure 2, Q and $Tangibility$ are extremely correlated.

The significance changes little when we add year, firm and industry fixed effects. The decrease in significance is due to part of the variability of $Leverage$ now being explained by the fixed effects. The point estimates are similar as well, with a particular increase in $Tangilibity$'s coefficient, from 1.8 to 2.9. This regression, although alleviates the effect of endogeneity, as the baseline, is likely affected due to the fact that the chosen fixed effect are not sufficient.

Table 2: Regression Results

	Standart	Fixed Effects	Winsorized	Trimmed	Lagged	Initial Leverage
	(1)	(2)	(3)	(4)	(5)	(6)
	Leverage	Leverage	Leverage_new	Leverage_trim	Leverage	Leverage
CashFlow	-0.0000567	-0.0000152	-0.00585***	-0.00591***	-0.000112	-0.0000512
	(0.0000707)	(0.0000747)	(0.000151)	(0.000409)	(0.00113)	(0.0000699)
Tangibility	1.828***	2.887**	0.293***	0.137***	9.925*	1.522***
	(0.457)	(1.006)	(0.00548)	(0.00671)	(3.853)	(0.453)
Size	-0.474***	-2.875***	0.00433***	0.00979***	-0.555*	-0.402***
	(0.0297)	(0.138)	(0.000389)	(0.000470)	(0.255)	(0.0294)
Q	0.0139***	0.0139***	-0.00495***	-0.0293***	0.0348**	0.0139***
	(0.00000846)	(0.00000850)	(0.000528)	(0.000859)	(0.0120)	(0.00000837)
InitialLeverage						0.526***
						(0.0160)
_cons	2.616***	16.46***	0.105***	0.164***	1.639	2.126***
	(0.192)	(0.877)	(0.00306)	(0.00380)	(1.662)	(0.191)
N	48880	48880	48880	30285	43989	48880

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We can see, however, considerable changes in the point estimates when we either trim or winsorize the data: all estimates decrease at least an order of magnitude, except for $CashFlow$, whose estimate increases sharply. Regarding significance, all variables are now significant, with an emphasis on $CashFlow$, now significant at 0.1%. This shows the large impact outliers have on the model. Although both estimates give similar interpretation, given the amount of observations we have left, trimming is prefered over winsoring due to the fact that winsoring severely modifies the data we are analysing while trimming works with a representative subsample of the distribution. Once again, this model probably has endogeneity problems.

A possible way out of the endogeneity problem is to use lagged variables, as shown bellow. The intuition for it is that variables in $t-1$ may not be correlated with the error in t, but still impact the dependent variable in t. This might help reduce simultaneity.

$$Leverage_{i,t} = \beta_0 + \beta_1 CashFlow_{i,t-1} + \beta_2 Tangibility_{i,t-1} + \beta_3 Size_{i,t-1} + \beta_4 Q_{i,t-1} + \varepsilon_{i,t}$$

As we can see in table 2, using lagged variables impact greatly the results: firstly, they are much less significant (only Q is significant at 0.01%). Secondly, the point estimated change sign and magnitude for many variables. It is not clear that using this model helps us to avoid endogeneity. One could still argue for the omitted variable problem here or that firms choose their capital structure in the long term, which would compromise the results.

The final specification we estimate is the baseline with the variable Initial Leverage. It is quite similar to the baseline both in term of point estimates and standard errors. Essentially, controlling for the leverage at the initial period is similar to adding a firm fixed effect in the sense that we are capturing a firm specific.

4 Discussion

Throughout the multiple specifications examined in our analysis, both the point estimates and the associated standard errors show significant variation. This fluctuation highlights the underlying issues with endogeneity that persist despite our the adjustments.

To mitigate the endogeneity problem, we used several strategies, including the use of lagged variables and the incorporation of initial leverage. These techniques are designed to reduce bias and provide more reliable estimates. However, our results indicate that these measures have not fully resolved the endogeneity issue.

Due to this problem, causal inference may not be possible. One approach to resolve this issue would be the introduction of additional instrumental variables in order to achieve a more robust solution.

Table 3: Regression Results

	Standart	Fixed Effects	Winsorized	Trimmed	Lagged	Initial Leverage
	(1)	(2)	(3)	(4)	(5)	(6)
	Leverage	Leverage	Leverage_new	Leverage_trim	Leverage	Leverage
CashFlow	-0.0000610	0.0000846	-0.00539***	-0.00360***	-0.0000648	-0.0000334
	(0.000174)	(0.000235)	(0.000166)	(0.000432)	(0.00142)	(0.000172)
Tangibility	1.827***	2.884**	0.284***	0.137***	9.919*	1.519***
	(0.458)	(1.006)	(0.00549)	(0.00671)	(3.853)	(0.453)
Size	-0.475***	-2.876***	0.00370***	0.00900***	-0.555*	-0.403***
	(0.0297)	(0.138)	(0.000393)	(0.000473)	(0.255)	(0.0295)
Q	0.0139***	0.0139***	-0.00375***	-0.0277***	0.0348**	0.0139***
	(0.00000846)	(0.00000850)	(0.000530)	(0.000866)	(0.0120)	(0.00000837)
InitialLeverage						0.526***
						(0.0160)
_cons	2.619***	16.46***	0.110***	0.166***	1.645	2.130***
	(0.193)	(0.877)	(0.00306)	(0.00378)	(1.662)	(0.191)
N	48880	48880	48880	30166	43989	48880

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$