Assignment Solutions

Elliot Beck elliot.beck@bf.uzh.ch
Ph.D. Class in Empirical Corporate Finance

March 31, 2023

Comments

The source code for all tables and regressions is written in python. All .tex files for the tables and regression outputs were generated automatically using the source code without any human intervention. For me there are two key take-aways after solving this Assignment:

1) It's the first time I used python with panel data. It was also the last time. While python is my preferred programming language overall, it doesn't offer the same ease-of-use as Stata or R for panel data.

2) Compustat is tricky and one needs a lot of practice to use the extracted data correctly.

Problem 1

Problem 1.c)

What is the CIK number? What is the gvkey? What is the CUSIP? What is the permno?
A Central Index Key (CIK) number is a unique number assigned to an individual, company, filing agent or foreign government by the United States Securities and Exchange Commission. The number is used to identify its filings in several online databases, including EDGAR. `https://en.wikipedia.org/wiki/Central_Index_Key`

The Global Company Key (GVKEY) is a unique six-digit number key assigned to each company (issue, currency, index) in the Capital IQ Compustat database. It is a company (issue, currency, index) identifier similar to a TICKER symbol. It represents the primary key for a company that is an index constituent. `https://www.alacra.com/alacra/outside/lei/info/gvkey.html`

A CUSIP is a nine-character numeric or alphanumeric code (e.g. 037833100 for Apple or 38259P508 for Google) that uniquely identifies a North American financial security for the purposes of facilitating clearing and settlement of trades. `https://en.wikipedia.org/wiki/CUSIP`

PERMNO is a unique permanent security identification number assigned by CRSP to each security. Unlike the CUSIP, Ticker Symbol, and Company Name, the PERMNO neither changes during an issue's trading history, nor is it reassigned after an issue ceases trading. `https://www.crsp.org/products/documentation/data-definitions-p`

Problem 1.d)

Find a way to select U.S. headquartered firms only.
I used the variable LOC – Current ISO Country Code - Headquarters (LOC) to identify US companies.

Why are there non-US firms in the Compustat North America files?
We have 42537 companies in total and 33749 in the US. Thus, roughly 20% are non-US. Compustat North America is a database of US and Canadian fundamental and market information on active and inactive publicly held companies. Companies from all over the world are listed on Canadian and US exchanges. Therefor it makes sense that we also have non-US companies in the files.

Problem 2

Problem 2.a)

I first tried to use the windorize function as provided by the python module scipy.stats.mstats. However, the results were not in line with simple, manual back-tests I performed. Therefor I switched to a manual approach using a for loop.

Problem 2.b)

How far off are means and medians?
For both Table 1 and Table 2 the mean and median diverge significantly for some variables such as Common Equity, Return on Equity and EBIT Interest Coverage. This implies skewed distributions. The outliers, which clearly exist in the data, have a strong impact on the mean. For other variables, such as the dividend variables, the differences are less pronounced.

What does this suggest for empirical work?
A researcher should potentially transform the data before doing further research. Potential transformations are windsorizing, manual outlier detection and log/root transformation of the data. If not excluded or adjusted, these outliers can have a large impact on the analysis.

Is there a difference between the overall sample and the large company sample?
There are a lot of very interesting, intuitive differences. Large caps invest less in R&D, they have more common equity and a larger share pays dividends. The dividend yield and the total payout ratio is higher on average, which makes sense too. For other variables such as book leverage variables and the capex ratio the differences are smaller.

Problem 2.c)

Draw a figure of the evolution of average asset tangibility, cash & short-term investment ratio, and the fraction of dividend-paying firms from 1970-2022. Compare your figures with the figures from Kahle and Stulz from lecture 1.

Table 1: 2ba) Summary statistics for winsorized data. (Full sample)

	count	mean	std	min	50%	max
Book Leverage 1	382716	0.33	0.67	0	0.21	13.98
Book Leverage 2	385736	0.8	1.82	0.02	0.61	42.96
Net Book Leverage 1	347416	0.19	0.84	-0.99	0.14	17.2
Common Equity	337197	1407.34	6177.17	0.11	88.45	126669.5
Market Leverage	288747	0.28	0.26	0	0.22	0.98
Asset Tangibility	378180	0.25	0.26	0	0.16	0.94
Cash & Short-term Ratio	348726	0.16	0.22	0	0.07	1
Return on Equity	109144	0.03	1.58	-16.01	0.07	13.3
Profit Margin	336151	-1.23	9.97	-240.3	0.03	1.28
CapEx Ratio	328321	0.06	0.08	0	0.04	0.59
R&D Ratio	163198	0.12	0.29	0	0.03	4.34
Dividend Yield	249258	0.02	0.03	0	0	0.25
Dividend Payer	323093	0.42	0.49	0	0	1
Total Payout Ratio	306938	0.32	1.15	-9.77	0	11.85
EBIT Interest Coverage	282162	-2.4	161.65	-4921.81	2.25	1585.02

Figure 1 is more or less in line with the figures as shown in the Kahle and Stulz paper. However, there are some minor differences. The ratio of dividend-paying firms is lower in the early 1970s, declines more in the early 2000s and increases more in the years after. Second, due to the longer time window, we can observe a sharp decline (incline) in cash and short-term investments (dividend payer) in 2022. This is probably due to the pandemic. Other than that, the pattern seems to be the same for asset tangibility and Cash and short-term investments.

Problem 2.d)

a.) Examine minimum and maximum. Are those numbers plausible? What do you suggest one should do in empirical research?

Initially, the negative minimum of net book leverage 1 caught my eye in Table 3. However, a company with only little or no debt and a lot of cash and short-term investments can have a net book leverage 1 ratio which is negative. Thus, this negative minimum value doesn't have to be problematic.

The maximum values for the book leverage variables well above one are interesting since this actually means that the companies have more debt than assets and therefor are technically insolvent. As shown in Problem 3, results of regression analysis change drastically when the maximum is clipped to one. As outliers have a lot of influence on regression estimations, I would suppose to clip, winsorize or remove these observations.

b. Compare book leverage 1 with net book leverage 1. What do you conclude from this comparison?

As commented above, for both ratios we observe maximum values well above one. Interestingly, the maximum value of net book leverage 1 is above the maximum for the book leverage 1 variable. Intuitively this doesn't make sense since cash positions usually should be positive.

Table 2: 2bb) Summary statistics for winsorized data. (Largest 25% of Companies by Assets)

	count	mean	std	min	50%	max
Book Leverage 1	94497	0.29	0.22	0	0.26	2.99
Book Leverage 2	96275	0.72	0.22	0.02	0.72	3.98
Net Book Leverage 1	76984	0.24	0.26	-0.93	0.23	2.87
Common Equity	65919	5471.22	12395.98	0.17	1235.03	126669.5
Market Leverage	64913	0.38	0.25	0	0.36	0.98
Asset Tangibility	92046	0.28	0.3	0	0.16	0.94
Cash & Short-term Ratio	77430	0.08	0.1	0	0.04	0.98
Return on Equity	27231	0.11	0.75	-16.01	0.1	13.3
Profit Margin	77818	0.03	1.34	-175.1	0.07	1.28
CapEx Ratio	67897	0.06	0.06	0	0.04	0.57
R&D Ratio	26250	0.03	0.04	0	0.01	0.68
Dividend Yield	54359	0.03	0.04	0	0.02	0.25
Dividend Payer	67105	0.78	0.41	0	1	1
Total Payout Ratio	63974	0.62	1.31	-9.77	0.45	11.85
EBIT Interest Coverage	64034	10.26	53.79	-1335.63	3.35	1585.02

The mean value of book leverage 1 is higher than the mean of net book leverage 1. This makes sense since we substract a (normally) positive quantity in the numerator.

The standard error of the net book leverage variable is larger which makes sense since the cash and short-term investments bring in additional variation into the ratio. The lower mean of the net book leverage is also explained by the substraction of cash and short-term investments.

Personally, I would stick to the book leverage 1 ratio since I am not sure why the maximum of net book leverage 1 ratio is larger.

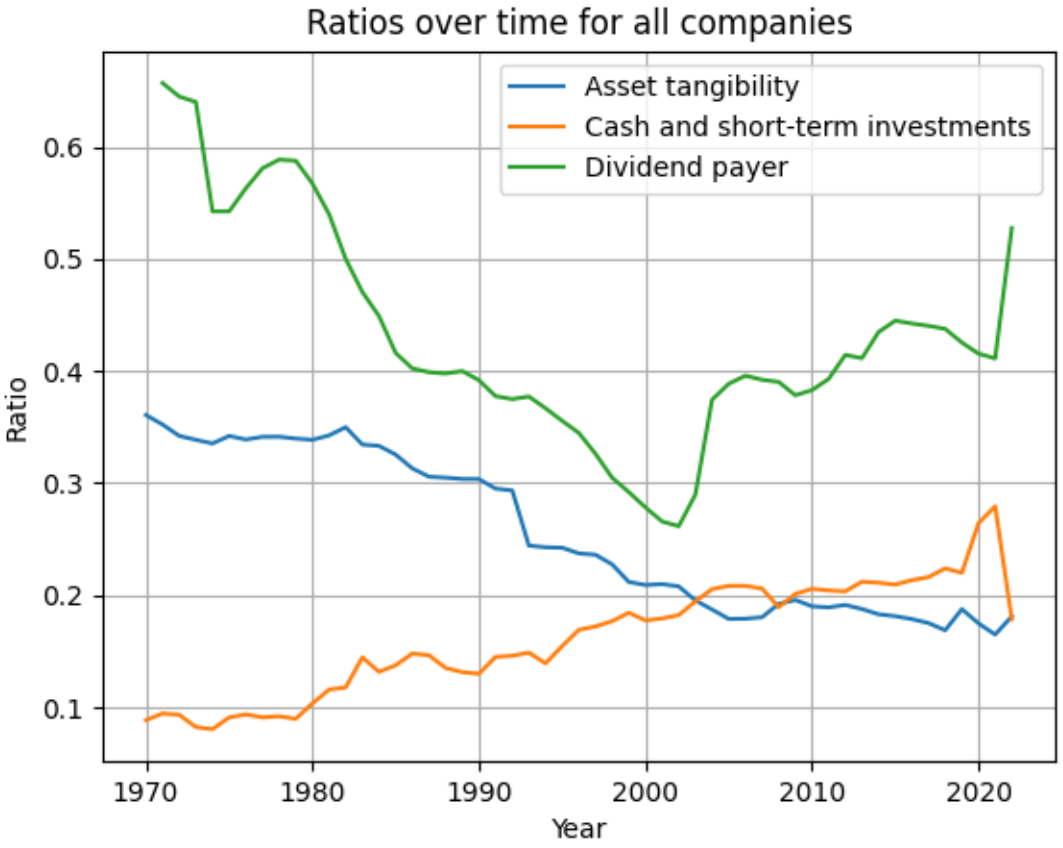
Table 3: 2d) Summary statistics of winsorized leverage variables. (Full sample)

	count	mean	std	min	50%	max
Book Leverage 1	382716	0.33	0.67	0	0.21	13.98
Book Leverage 2	385736	0.8	1.82	0.02	0.61	42.96
Net Book Leverage 1	347416	0.19	0.84	-0.99	0.14	17.2
Market Leverage	288747	0.28	0.26	0	0.22	0.98

Problem 3

The following regressions were estimated using PanelOLS and PooledOLS from the linearmodels.panel library in python. Rows containing NaNs were dropped.

Figure 1



Problem 3.a/b/c)

Estimate a regression of Book Leverage 1 on log (total assets), asset tangibility, R&D ratio, dividend payer, and profit margin using the unwinsorized data from 1970-2022. Estimate the same regression using the winsorized data. Compare the results.

The fits for a) and b) are shown in the first two columns of Table 4. Clearly, the intercept shrinks strongly towards 0 when the data is winsorized. This makes sense since we are aware about the right-skewed distribution with large maximum values (outliers) of the target variable book leverage 1. For the coefficients of the predictors the changes are mixed. For Asset Tangibility and the Profit Margin the coefficients stay the same or increase in magnitude. For the other predictors the coefficients decrease strongly. The T-Values, which are reported in brackets below the coefficients) increase for all predictors.

In addition to b), restrict book leverage to be between zero and one. Re-estimate the regression and compare. What do you conclude?

The fit for c) can be found in the third column of Table 4. The magnitude of all coefficients decreases. This makes sense since we apply an even harder restrictions on the dependent variable now. Large values of the dependent are completely replaced. The coefficients of the predictors remain highly significant. I conclude

that extreme values have a large impact on the magnitude of the coefficients and should be double-checked and potentially removed/adjusted before regression analysis is performed.

Table 4: 3a/b/c) Comparison of three different OLS fits

	(a)	(b)	(c)
Dep. Variable	Book Leverage 1	Book Leverage 1	Book Leverage 1
Estimator	PooledOLS	PooledOLS	PooledOLS
No. Observations	148052	148052	148052
Cov. Est.	Unadjusted	Unadjusted	Unadjusted
R-squared	0.0433	0.1054	0.0469
R-Squared (Within)	0.0435	0.1132	0.0402
R-Squared (Between)	-0.4020	-0.1614	0.0503
R-Squared (Overall)	0.0433	0.1054	0.0469
F-statistic	1338.9	3490.3	1455.8
P-value (F-stat)	0.0000	0.0000	0.0000
==========	==========	==========	==========
const	1.0993	0.2438	0.1971
	(13.074)	(67.357)	(120.77)
Total Assets	-0.2260	-0.0238	-0.0024
	(-14.158)	(-37.573)	(-8.2386)
Asset Tangibility	0.3583	0.3327	0.2740
	(1.8918)	(44.806)	(81.847)
R&D Ratio	1.3303	0.6560	0.0730
	(78.652)	(94.637)	(23.366)
Dividend Payer	0.3153	0.0352	-0.0089
	(3.6548)	(10.450)	(-5.8757)
Profit Margin	-0.0014	-0.0032	-0.0005
	(-7.8418)	(-26.914)	(-9.9848)

Problem 3.d)

Now include, in this order, in the regression: 1) year-fixed effects, 2) year fixed effects and industry-fixed effects, 3) year x industry-fixed effects. Compare the results and interpret differences.

The results are shown in Table 5. The coefficients are surprisingly stable for all three regressions. This implies that the effects of the predictors are roughly constant over time and across the different industries. The only difference I can find is the coefficient of dividend payer, which gets reduced by 50% when including entity fixed effects.

Problem 3.e)

Compare t-statistics to those obtained in part c). Why do they change so much?

We don't have independence between observations on the firm level and therefor it makes total sense to cluster the standard error on the firm level. The T-Values all decrease drastically in magnitude and some

Table 5: 3d) Comparison of three different OLS fits with fixed effects

	(1)	(2)	(3)
Dep. Variable	Book Leverage 1	Book Leverage 1	Book Leverage 1
Estimator	PanelOLS	PanelOLS	PanelOLS
No. Observations	148052	148052	148052
Cov. Est.	Unadjusted	Unadjusted	Unadjusted
R-squared	0.1112	0.1218	0.1234
R-Squared (Within)	0.1083	0.1110	0.1109
R-Squared (Between)	-0.1552	-0.2337	-0.2411
R-Squared (Overall)	0.1020	0.1017	0.1015
F-statistic	3704.6	4102.0	4080.1
P-value (F-stat)	0.0000	0.0000	0.0000
==========	==========	==========	==========
const	0.2825	0.2914	0.2915
	(75.541)	(73.850)	(72.960)
Total Assets	-0.0367	-0.0372	-0.0376
	(-50.286)	(-50.276)	(-50.366)
Asset Tangibility	0.3975	0.3712	0.3719
	(52.318)	(41.044)	(40.110)
R&D Ratio	0.6119	0.6862	0.6927
	(86.582)	(94.861)	(95.079)
Dividend Payer	0.0602	0.0342	0.0372
	(16.877)	(9.4270)	(10.139)
Profit Margin	-0.0025	-0.0026	-0.0027
	(-20.706)	(-21.832)	(-22.340)
==========	==========	==========	==========
Effects	Time	Entity	Other Effect (fyear_sic)
		Time	

coefficients are no longer significant. The results are shown in Table 6. The differences in T-Values are most pronounced for the predictors Asset Tangibility and Profit Margin. This makes sense as these predictors are very firm specific. To give an example: A large industrial company will often have consistently high ratios for these two predictors. Therefor we have correlation in the observations and we should cluster the standard errors on firm level.

Problem 4

a) What is the calendar date range over which the one-year change in the market value of equity has been calculated for each of these rows?

The results are shown in Table 7, I used winsorized data. Compustat has a very particular way how it calculates the the fiscal year for a company. The variable depends on the current fiscal year-end month. If the current fiscal year-end month lies between January and May, the Fiscal Year variable will be defined as

Table 6: Comparison of fit from exercise 3c) with fit from exercise 3e)

	(e)	(c)
Dep. Variable	Book Leverage 1	Book Leverage 1
Estimator	PanelOLS	PanelOLS
No. Observations	148052	148052
Cov. Est.	Clustered	Unadjusted
R-squared	0.0441	0.0469
R-Squared (Within)	0.0432	0.0402
R-Squared (Between)	-0.9411	0.0503
R-Squared (Overall)	0.0429	0.0469
F-statistic	1365.5	1455.8
P-value (F-stat)	0.0000	0.0000
==========	==========	==========
const	1.4418	0.1971
	(3.6420)	(120.77)
Total Assets	-0.3358	-0.0024
	(-3.5395)	(-8.2386)
Asset Tangibility	0.6997	0.2740
	(1.1534)	(81.847)
R&D Ratio	1.3222	0.0730
	(2.3976)	(23.366)
Dividend Payer	0.4885	-0.0089
	(3.8938)	(-5.8757)
Profit Margin	-0.0013	-0.0005
	(-0.9949)	(-9.9848)
==========	==========	==========
Effects	Entity	
	Time	

the current calendar year minus 1 year. If it falls between June and December it will be the current calendar year.

Therefor, calendar year = 2000 and fiscal year-end month $\in \{1, \ldots, 5\}$ means that the fiscal year would be 1999. If calendar year = 2000 and fiscal year-end month $\in \{6, \ldots, 12\}$ means that the fiscal year would be 2000. In other words, if a company's accounting year ends in Mai 2021, the corresponding fiscal year is 2020. if a company's accounting year ends in June 2021, the corresponding fiscal year is 2021.

This has implications on the data which is used to compute the ratios. The calendar date range for Fiscal Year 2000 and Month 5 is June 2000 to May 2001. The calendar date range for Fiscal Year 2020 and Month 6 is July 1999 to June 2000. Thus, finishing the accounting year one month later can lead to a different fiscal year.

b) In your own words, describe the two problems a researcher faces, explicitly linking them to the exercise you just did for fiscal year 2000/2001.

As the articles explains, the researcher has two main problems:

(1) **Market values and accounting values are (potentially) not aligned.**

The researcher has two options:

- Align the market and accounting variables such that they cover the same date range. If a researcher decides to align the market values to the fiscal year ends of the accounting variables, there will be some cross-sectional variation due to market variation over the year. This is especially problematic during times of financial turmoil.
- Ignore the issue. Fama et al. decide to ignore the issue.

(2) **Mixture of different fiscal yearends** This issues arises due to the mechanism which I explained in the first part of this exercise. The fiscal year (and as a consequence the reported data) changes throughout the year, depending on the choice of the company when to finish the accounting year. You can either match your market data to the correct fiscal year end or again ignore it. Fama et al. decide to ignore the different fiscal years and match the data on the calendar year.

Table 7: Percentage Change of Common Equity by Fiscal year and month

Fiscal Year	Month	Mean	SD
2000	1	0.05	0.74
2000	2	-0.02	0.8
2000	3	0.03	4.66
2000	4	-0.19	0.57
2000	5	0.01	0.73
2000	6	2.88	31.18
2000	7	0.46	2.15
2000	8	5.39	57.31
2000	9	0.59	2.9
2000	10	0.36	1.7
2000	11	-0.09	0.47
2000	12	0.34	13.93
2001	1	1.71	17.39
2001	2	3.25	16.61
2001	3	1.16	8.15
2001	4	0.93	3.67
2001	5	0.25	1.41
2001	6	0.14	1.99
2001	7	-0.07	0.64
2001	8	0.07	0.95
2001	9	-0.07	0.69
2001	10	1.08	16.72
2001	11	0.2	1.05
2001	12	0.62	9.12

Problem 5

Create an accurate list of book leverage and asset growth (i.e., (total assets year t / total assets year t- 1) − 1). Look at asset growth in the year of the merger and compare it to the other years. Under what circumstances could this become a problem?

The transaction happened in 1999. The companies involved were: MOBIL CORP with gvkey: 7476 and EXXON OIL CORP (gvkey: 4503). The last fiscal year for MOBIL CORP was 1998. In order to create an accurate list of the ratios, the two companies have to be combined before the merger took place. It is important to mention that one cannot just sum up the ratios of the two seperate companies and then calculate growth rates. Instead, one has to recalculate the ratios for the combined company and then calculate the growth rates.

The results are shown in Table 8. The growth rate of the total assets in the year of the merger was around 7%, which is a large, but not special for the combined company. If one does not correctly aggregate the companies, the growth rate for EXXON would be way too large since the assets of two companies are combined only in 1999. This is a very important finding since a "naive" approach to calculate the growth rates wouldn't consider this fact.

Table 8: Mobil Merger Book Leverage and Asset Growth in %

Date	Book Leverage 1	Book Leverage 2	Net Book Leverage 1	Total assets
1994-12-31	-0.046673	-0.003664	-0.038932	0.037473
1995-12-31	-0.202871	-0.027798	-0.223840	0.031143
1996-12-31	-0.012835	-0.000748	-0.102054	0.063709
1997-12-31	-0.041420	-0.030891	-0.138378	-0.016289
1998-12-31	0.022808	-0.005970	0.257190	-0.030360
1999-12-31	0.078496	0.019578	0.127158	0.067490
2000-12-31	-0.312832	-0.059572	-0.641577	0.030992
2001-12-31	-0.163637	-0.067868	-0.303751	-0.039101
2002-12-31	-0.066729	0.050912	-0.224282	0.066143
2003-12-31	-0.222169	-0.057793	-1.269057	0.141728
2004-12-31	-0.224514	-0.013014	11.254763	0.120371

Now pick the AOL – Time Warner merger and figure out how Compustat treated the respective firms in this case pre- and post-merger. Why is this so much more complicated here than for the ExxonMobile case?

This merger took place in 2000. Contrary to the easy case before, here we can't just add one company which died on Compustat to the other company which survived. We have three involved companies: Time Warner Inc-Old, Time Warner Inc and AOL. The two Time Warner companies have an overlap between 1992 and 1999. AOL entries in Compustat only start in 2005, which is due to the spin-off which happened later. Therefor, we can't just add Time Warner / Time Warner Old to AOL as we did in the easy case above.