FIN3210 Week 3 Assignment Report
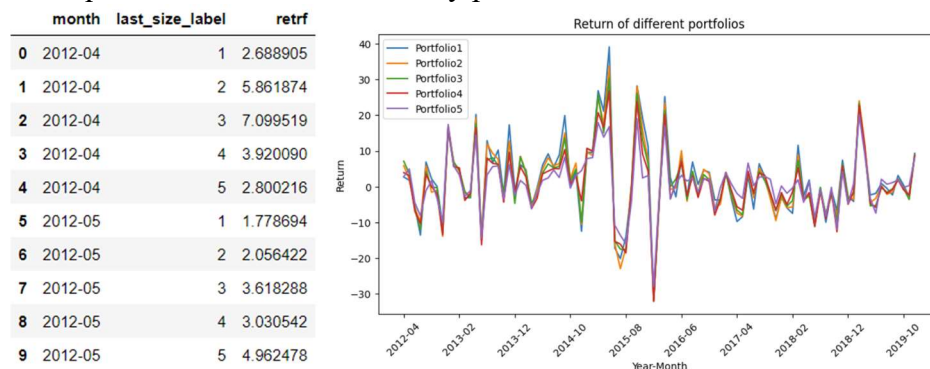Ma Kexuan 120090651

Abstract

This report provides monthly returns stats based on the size and institutional ownership of last quarter, and a panel regression result, as well as a figure containing the online sales and reported sales.

Data Preprocessing

The preprocessing procedures and some interpretations of the code are described in each code blocks in the appendix, please check.
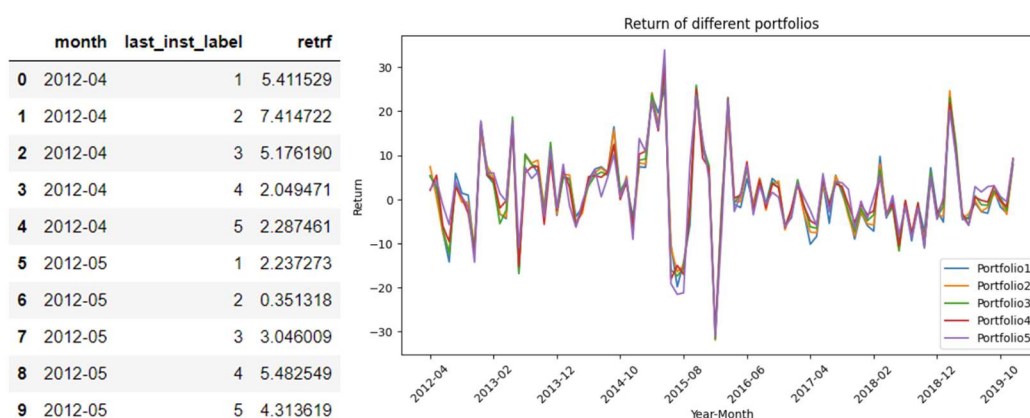
Questions

1) Using the data set of stock returns, sort stocks into quintiles by size every quarter, hold stocks over the quarter, and calculate monthly portfolio returns

	month	last_size_label	retrf
0	2012-04	1	2.688905
1	2012-04	2	5.861874
2	2012-04	3	7.099519
3	2012-04	4	3.920090
4	2012-04	5	2.800216
5	2012-05	1	1.778694
6	2012-05	2	2.056422
7	2012-05	3	3.618288
8	2012-05	4	3.030542
9	2012-05	5	4.962478

I divide the size of all the firms into 5 groups, and use the size in last quarter to invest, ultimately get the monthly return in the next quarter, the result is shown above. It has shown ambiguous relationship between the size and portfolio return, I think maybe performing a cumulative return result will be more convincing.

2) Using the data set of stock returns, sort stocks into quintiles by institutional ownership every quarter, hold stocks over the quarter, and calculate monthly portfolio returns

	month	last_inst_label	retrf
0	2012-04	1	5.411529
1	2012-04	2	7.414722
2	2012-04	3	5.176190
3	2012-04	4	2.049471
4	2012-04	5	2.287461
5	2012-05	1	2.237273
6	2012-05	2	0.351318
7	2012-05	3	3.046009
8	2012-05	4	5.482549
9	2012-05	5	4.313619

This time, I use the institutional ownership to divide the firms into 5 groups, the conclusion is the same as for Question 1, it's better to do a cumulative return to discover the relationship underneath.

3) Using the data set of stock returns, perform panel regression, and regress stock returns on firm characteristics such as size, book-to-market ratio, return12, roa, leverage, ppe, intang,

Dep. Variable:	retrf	R-squared:	0.0133
Estimator:	PanelOLS	R-squared (Between):	-1228.5
No. Observations:	85397	R-squared (Within):	0.0171
Date:	Thu, Oct 05 2023	R-squared (Overall):	-45.250
Time:	15:22:52	Log-likelihood	-3.2e+05
Cov. Estimator:	Clustered		
		F-statistic:	124.83
Entities:	2035	P-value	0.0000
Avg Obs:	41.964	Distribution:	F(9,83262)
Min Obs:	2.0000		
Max Obs:	92.000	F-statistic (robust):	14.436
		P-value	0.0000
Time periods:	92	Distribution:	F(9,83262)
Avg Obs:	928.23		
Min Obs:	377.00		
Max Obs:	1825.0		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
size	-4.0216	0.4473	-8.9917	0.0000	-4.8983	-3.1450
bm	0.2309	0.8461	0.2729	0.7849	-1.4274	1.8893
return12	-0.0023	0.0030	-0.7629	0.4455	-0.0081	0.0036
roa	4.6914	3.1983	1.4668	0.1424	-1.5772	10.960
lev	-1.3608	0.6284	-2.1655	0.0304	-2.5925	-0.1291
ppe	-0.0489	1.0345	-0.0473	0.9623	-2.0765	1.9788
intang	3.0578	2.5080	1.2192	0.2228	-1.8579	7.9736
numanalyst	0.0113	0.0163	0.6901	0.4901	-0.0207	0.0432
instown	0.0086	0.0160	0.5349	0.5927	-0.0228	0.0399

F-test for Poolability: 28.922
P-value: 0.0000
Distribution: F(2125,83262)

Included effects: Entity, Time

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
size	-0.0107	0.0508	-0.2100	0.8337	-0.1102	0.0888
bm	1.4906	0.8773	1.6991	0.0893	-0.2288	3.2101
return12	0.0036	0.0141	0.2555	0.7983	-0.0241	0.0313
roa	5.3103	7.6867	0.6908	0.4897	-9.7556	20.376
lev	0.5546	1.8904	0.2934	0.7692	-3.1504	4.2597
ppe	0.1788	1.2655	0.1413	0.8877	-2.3016	2.6592
intang	0.3477	1.2780	0.2721	0.7855	-2.1571	2.8526
numanalyst	-0.0084	0.0157	-0.5371	0.5912	-0.0392	0.0224
instown	0.0436	0.0165	2.6450	0.0082	0.0113	0.0759

For the controlled fixed effect and time group, we find that size and lev are at 95% significance level. There is a negative correlation between size and return, which may be caused by the reason that large companies are focusing on profiting, and the profit gained has been issued to investors, not using in further industrial developing, resulting in a lower growth, thus fewer people expect them to rise sharply, so lower return. The more leverage the firm has, may indicate that the firm has a larger debt to be paid, more bankrupt risk, so lower future return. For the not controlled group, only instown is at the same level of significance. The positive sign maybe indicate that more institutional investor likes this stock, so others follow, in order to get a higher return.

Finally I compare the two models, the R-square of the controlled one is 0.0133 and that of the uncontrolled one is 0.0062, which illustrates that both models are quite weak to dig out the concrete relationship, but the former one is better than the latter one. Overall, the introduction of fixed effects has illuminated crucial variations in the relationships between predictors and the dependent variable, emphasizing the importance of accounting for unobserved heterogeneity in panel data analyses.

4) Using the data set of Online sales, aggregate monthly online sales over quarters, download reported quarterly total sales from CSMAR, and plot figures including both online sales and reported quarterly sales.
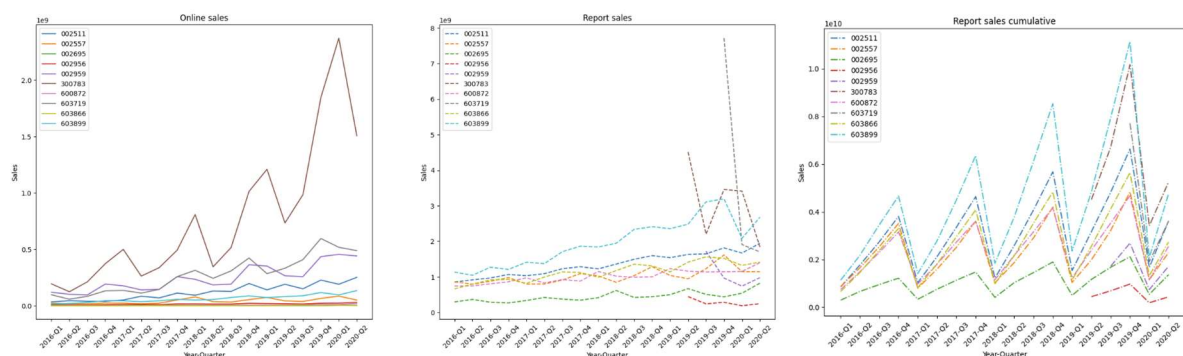
I have processed the data from CSMAR, removing the January data, since it's the same as the whole data of last year, then get the difference between each quarter, because the original data is cumulative. The result of both online sales and the reported sales from CSMAR are shown above. Also, I print out the cumulative one, that is the original data from CSMAR, you can check it as a reference. The 3 more limpid graphs are in the appendix below.