CFM 301 Winter 2024 – Financial Data Analytics
Dr. Alan Huang
Assignment 3
t54zheng (20939203)

Report/Summary of conclusions from the code

Task 1 – Estimating beta

Note that for beta, and for following variables, we will aggregate them in our summaries (instead of reporting them individually for each of our Nasdaq_100 stocks) since it appears that we are interested in their aggregate correlation in the assignment.

Furthermore, since each factor should be independent of the stock once isolated, we can presume that this is the justified approach.

Summary Statistics for [beta] variable
N: 21481
mean: 1.2417
standard deviation: 0.7829
median: 1.1043
minimum: -2.2283
1st Percentile: -0.1209
99th Percentile: 3.8351
maximum: 7.3953

Task 2 – Estimating Idiosyncratic Volatility

Ang et al. cites the following 3-factor model to model the expected returns for each stock

$$r_t^i = \alpha^i + \beta_{MKT}^i MKT_t + \beta_{SMB}^i SMB_t + \beta_{HML}^a HML_t + \epsilon_t^i$$

for security i at time t

That is, for our regression, for each stock and for each month, we run the ff3 regression, using data from both at day (t).

In our initial regression, we obtain for each security i:

- Each factor beta
- Our alpha coefficient

Then using the values from above we calculate

- our epsilon for each time t

Why is this not look-ahead bias?

- Because the epsilons are actually part of the portfolio, and
- An investor would use this data calculated from month to make their decision in what to invest in for month if they are interested in the stock's idiosyncratic risk.

Note that when we run our regression, we change our volatility from daily to monthly, meaning we will need to multiply our epsilons by the square root of the number of days in that month (the number of measurements).

We end up with the following statistics

Summary Statistics for [ivol] variable
N: 21952
mean: 0.0763
standard deviation: 0.0557
median: 0.0604
minimum: 0.0060
1st Percentile: 0.0189
99th Percentile: 0.2845
maximum: 1.0141

<u>**Task 3 – Estimating momentum**</u>

This approach taken for this task is mostly the same as the previous two, except for one important measure.

Since some periods can have less than 12 observations (10, or 11), we get the geometric mean return and then compound to 12 periods. This way, all our data points correctly showcase a 12-month compound return.

Summary Statistics for [mom] variable
N: 21565
mean: 0.2610
standard deviation: 0.7422
median: 0.1655
minimum: -0.9723
1st Percentile: -0.6760
99th Percentile: 2.7300
maximum: 26.3729

Task 4 – Winsorize

Like in assignment 2, we winsorize at the top and bottom 3 standard deviations for every month.

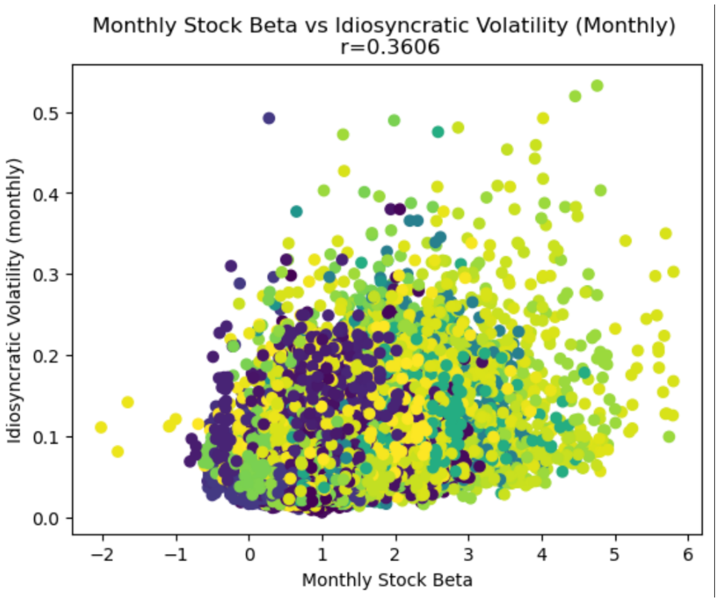
Summary Statistics for [beta] variable (winsorized)
N: 21481
mean: 1.2396
standard deviation: 0.7679
median: 1.1043
minimum: -2.0165
1st Percentile: -0.1209
99th Percentile: 3.8008
maximum: 5.8011

Summary Statistics for [ivol] variable (winsorized)
N: 21952
mean: 0.0754
standard deviation: 0.0518
median: 0.0604
minimum: 0.0060
1st Percentile: 0.0189
99th Percentile: 0.2669
maximum: 0.5325

Summary Statistics for [mom] variable (winsorized)
N: 21565
mean: 0.2436
standard deviation: 0.5649
median: 0.1655
minimum: -0.9723
1st Percentile: -0.6750
99th Percentile: 2.2863
maximum: 10.1792

Task 5 – Correlation between beta and ivol

We compare the correlation between our winsorized beta and ivol statistics. We arrive at a correlation coefficient of 0.3606, illustrated by the following graph.



This information suggests a weak positive correlation between a stock's beta and level of idiosyncratic volatility. If we follow this weak data correlation, we can conclude that the higher a stock's beta, the higher the stock's idiosyncratic volatility, the portion of the stock's risk that cannot be removed through diversification.

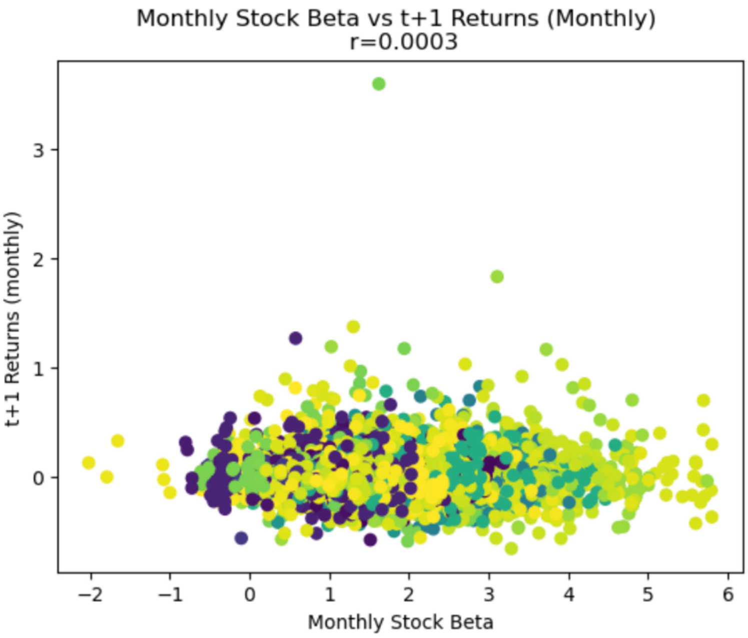
In the theory, the CAPM beta of a security is typically a measure of how much systematic risk it holds compared to its benchmark. We see in our data that this may also be true in the non-systematic (idiosyncratic) component of the risk as well to a weak level, at least following the 3-factor model used to determine our residuals.

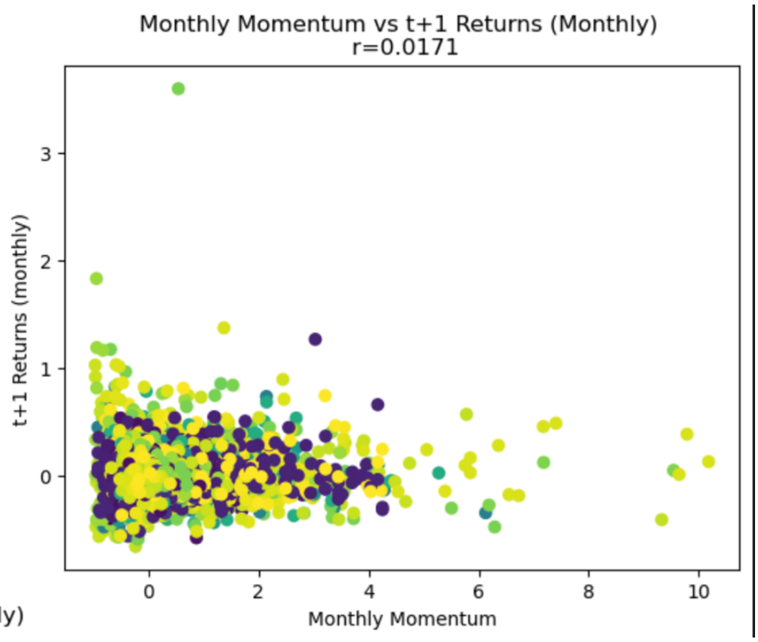
Task 6 – Looking for Correlations

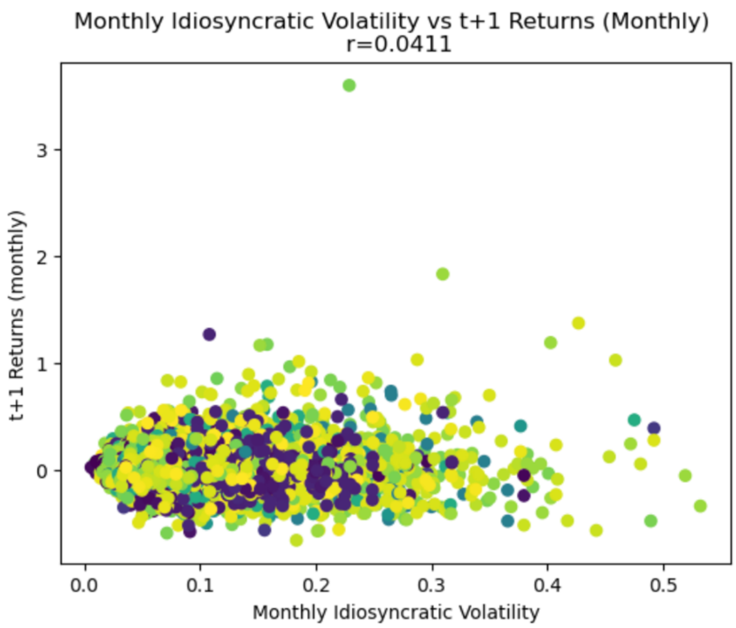
We run a series of correlation analysis between month t+1 stock return and our winsorized factors to see if there are any linear associations.

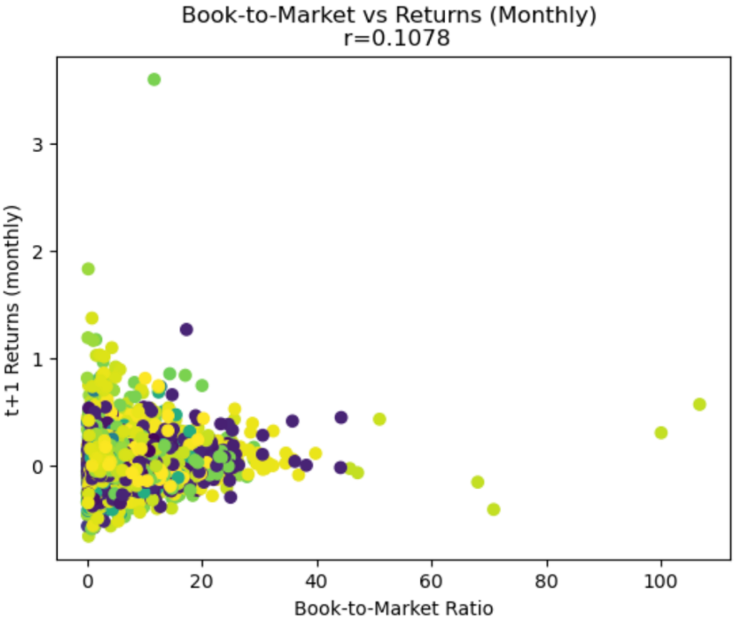
For the three variables brought in from Assignment 2, I decided to compare them to month t returns instead, since we already applied a 3-month lag to their base components (from financial reports), meaning that there won't be any look-ahead bias as investors will always be looking at financials from the past 3 months to make investment decisions.

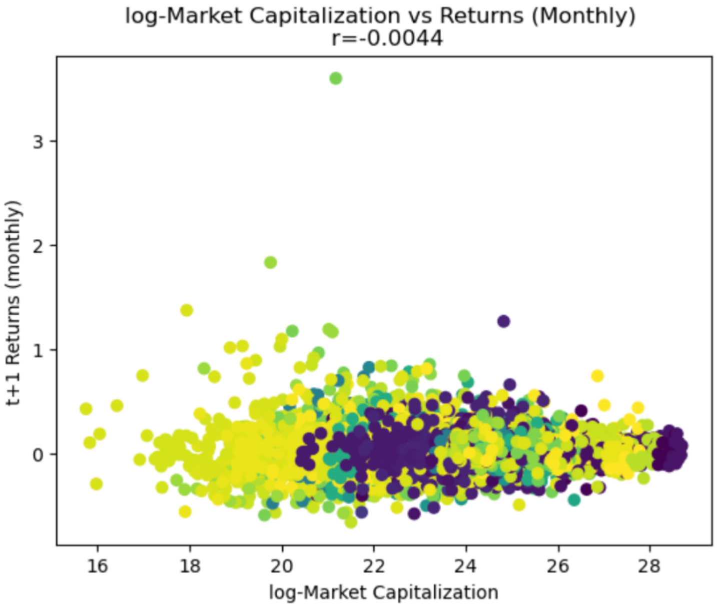
Below are the resultant scatter plots with each of the relationships being investigated.

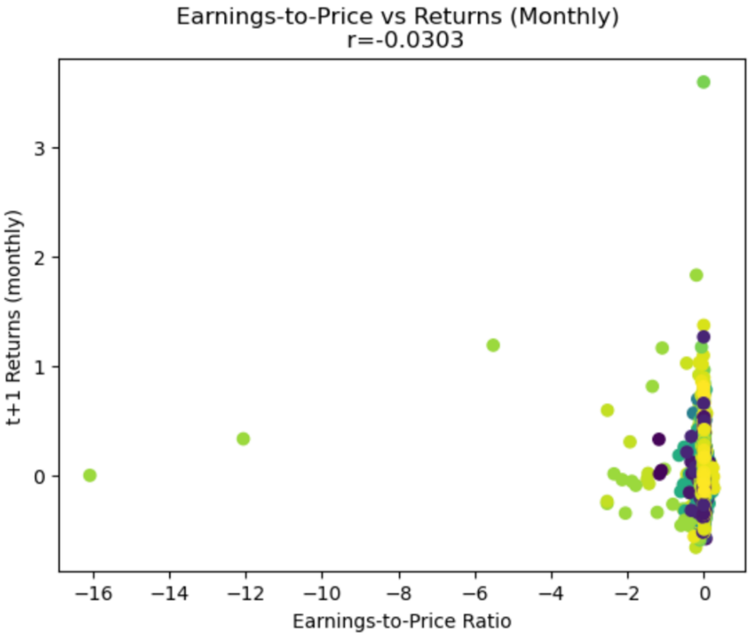

Monthly Stock Beta vs t+1 Returns (Monthly)
r=0.0003
t+1 Returns (monthly)
Monthly Stock Beta
−2
−1
0
1
2
3
4
5
6


Monthly Momentum vs t+1 Returns (Monthly)
r=0.0171
t+1 Returns (monthly)
Monthly Momentum
0
2
4
6
8
10


Monthly Idiosyncratic Volatility vs t+1 Returns (Monthly)
r=0.0411
t+1 Returns (monthly)
Monthly Idiosyncratic Volatility
0.0
0.1
0.2
0.3
0.4
0.5


Book-to-Market vs Returns (Monthly)
r=0.1078
t+1 Returns (monthly)
Book-to-Market Ratio


log-Market Capitalization vs Returns (Monthly)
r=-0.0044
t+1 Returns (monthly)
log-Market Capitalization


Earnings-to-Price vs Returns (Monthly)
r=-0.0303
t+1 Returns (monthly)
Earnings-to-Price Ratio

Correlation Matrix

	beta vs T+1 Return	ivol vs T+1 Return	mom vs T+1 Return	lnSize vs Return	bk2mkt vs Return	eP2 vs Return
r	0.0003	0.0411	0.0171	-0.0044	0.1078	-0.0303
pval	0.96	0.00	0.01	0.53	0.00	0.00

Our correlation matrix yields the conclusion that expected returns are significantly correlated ($p < 0.05$) with idiosyncratic volatility, momentum, book-to-market, and earnings-to-price ratios, albeit at rather low correlation coefficients. This mostly matches the findings and learnings from class.

- For idiosyncratic volatility, we have learned that investors are rewarded for taking on more idiosyncratic risk.
- For momentum, we have learned that a momentum portfolio generally performs well.
- For book-to-market, we have learned that value stocks (high book-to-market) tend to outperform growth stocks (low book to market).
- For earnings-to-price, we see that stocks prices tend to swing dramatically to earnings results when they either outperform or underperform expectation.

Betas and market capitalization do not appear to be significant factors in estimating returns using a linear relationship, which makes sense from our theory. First of all, we have shown that the CAPM model fails at predicting returns, and our findings here once again repeat that finding. Next, we should not expect that large-cap companies return any higher returns. In fact, we have learned from class that small-cap companies actually are the ones that return higher expected returns.