Data Assignment 4: Forming and Evaluating Trading Strategies

Financial Data Analytics

It's understood that all assignments are individual work. Failure to observe this may result in immediate failure of the course.

This assignment accounts for 12.5% of the course mark.

This assignment relies on the monthly data sets from Data Assignments 1-3. To ensure consistency for grading, I have put the final data as CA.sas7bdat. Please use this data set as the base data for this assignment. For your reference, here are the variables:

#	Variable	Type	Note
1	permno	Num	PERMNO
2	date	Num	year-month
3	ret	Num	return at date
4	ret_t1	Num	return of next month
5	TICKER	Char	TICKER
6	COMNAM	Char	COMNAM
7	PRC	Num	PRC
8	SHROUT	Num	SHROUT
9	datadate	Num	datadate
10	fyearq	Num	fyearq
11	fqtr	Num	fqtr
12	atq	Num	atq
13	ceqq	Num	ceqq
14	epspxq	Num	epspxq
15	ibq	Num	ibq
16	saleq	Num	saleq
17	lnSize	Num	lnSize
18	bk2mkt	Num	bk2mkt
19	ep1	Num	ep1
20	ep2	Num	ep2
21	lnSize_winsorized	Num	lnSize_winsorized
22	bk2mkt_winsorized	Num	bk2mkt_winsorized
23	ep1_winsorized	Num	ep1_winsorized
24	ep2_winsorized	Num	ep2_winsorized
25	lnSize_truncated	Num	lnSize_truncated

26	bk2mkt_truncated	Num	bk2mkt_truncated
27	ep1_truncated	Num	ep1_truncated
28	ep2_truncated	Num	ep2_truncated
29	lnSize_winsorized_zscore	Num	lnSize_winsorized_zscore
30	bk2mkt_winsorized_zscore	Num	bk2mkt_winsorized_zscore
31	ep1_winsorized_zscore	Num	ep1_winsorized_zscore
32	ep2_winsorized_zscore	Num	ep2_winsorized_zscore
33	gvkey	Char	
34	beta	Num	
35	ivol	Num	
36	mom	Num	
37	beta_winsorized	Num	
38	ivol_winsorized	Num	
39	mom_winsorized	Num	

This data set is not exhaustive for the purposes of this assignment. In what's following, winsorized variables are those highlighted in yellow above.

Depending on questions, you may need to calculate additional variables, either from existing variables or from raw data to be downloaded from WRDS on your own.

1. (18 marks) Use the [*m,n,l*] month rule to construct a quintile portfolio for each of the six winsorized factors from Assignment 3 (*lnSize*, *bk2mkt*, *eP, beta*, *ivol*, and *mom*, yellow-highlighted in the variable list above) for the period of January, 2000 to November, 2021, where *m* is the estimation window above (it changes for each of the factor), $n = 0$, and $l = 1$. In this case, you're forecasting one month ahead returns.

    a) Portfolio sorting. Create equal-weighted quintile portfolios for each of the factors based on the ranked values of your factor in each month (that is, in each month, you sort the stocks into five equal-weighted portfolios). Explain that your data handling contains no look-ahead bias for return prediction in this step. Output your six quintile portfolios into an excel file. Your returns range from January, 2000-November, 2021 (i.e., your last period portfolio return is Dec. 2021).

    b) Form a hedge portfolio using quintiles 1 and 5 (long one, and short the other) for each of the factors. Your ex ante goal is to generate profit, and the direction of long-short should depend on the knowledge about the factor covered in the course. What are the overall return, excess return over the market (recall that Fama-French factors has market returns), CAPM-alpha, and Fama-French four-factor (market, SMB, HML, and

MOM) alpha, and Sharpe ratio of the hedge portfolio of the difference between quintiles 5 and 1? Are they significant?

2. (6 marks) Betting against beta strategy. Assuming that you use AQR's betting against beta strategy by using quintile-1- and quintile-5-beta stocks (Frazinni and Pederson, 2014; also consult the slide set "Low Risk Anomaly" for the strategy), repeat Question 1 and compare your results with the simple *beta* strategy in Question 1 and comment why differences exist.

3. (10 marks) Assume that you create your own factor-mimicking ETF tracking one or more factors that you just constructed. For now, assume that you follow Blackrock's single-factor ETF "USMV" to focus solely on the idiosyncratic volatility factor, and assume that you're not able to short sell.

    a) If you follow Question 1 (by using only the long-leg of your hedge portfolio), what would your excess return over the market return and Fama-French four-factor alpha? Have you delivered value for your investors?

    b) What is the annual turnover of your ETF? For simplicity, if a stock appears in the portfolio next month (regardless of its weight), we assume that you don't need to trade the stock, hence monthly turnover is defined as the number of stocks that are new in the month, divided by the total number of stocks in the month. For example, in month *t* you hold 10 stocks, out of which, 4 stocks are new; in this case, your turnover for the month is 40%. Calculate the annual turnover as the average monthly turnover times 12. (For your project, you might wish to think about the implications of turnover—it means transaction costs which eat into your returns). This snippet in the following may help you code this question out in a simplified way (where *in_t* is a dummy variable if a stock is in the portfolio in month *t*, and *in_t_minus1* is a dummy variable if the stock is also in month *t-1*, and *new* is a dummy variable that the stock is new in month *t*):

PERMNO	monthid	yyyymm	in_t	lag_permno	in_t_minus1	new
10107	2	200002	1	.	.	1
11403	2	.	.	11403	1	.
11618	2	200002	1	.	.	1
13856	2	200002	1	13856	1	0
19502	2	200002	1	.	.	1
21776	2	200002	1	21776	1	0
23931	2	200002	1	23931	1	0
24109	2	200002	1	24109	1	0
48486	2	200002	1	.	.	1
59328	2	200002	1	59328	1	0
60506	2	200002	1	.	.	1
61621	2	.	.	61621	1	.
62148	2	.	.	62148	1	.
76076	2	.	.	76076	1	.
85913	2	.	.	85913	1	.
88031	2	200002	1	88031	1	0
11403	3	200003	1	.	.	1
13856	3	200003	1	13856	1	0
19502	3	.	.	19502	1	.
21776	3	200003	1	21776	1	0
23931	3	200003	1	23931	1	0
24109	3	200003	1	24109	1	0
44644	3	.	.	44644	1	.
59328	3	200003	1	59328	1	0
60506	3	.	.	60506	1	.
61621	3	200003	1	.	.	1
62148	3	200003	1	62148	1	0
76076	3	200003	1	.	.	1
85631	3	.	.	85631	1	.
85913	3	200003	1	85913	1	0
88031	3	200003	1	.	.	1

4. (6 marks) Assume instead that now can short-sell, in an unlimited way. Ignore all transaction costs. You instead wish to build a multi-factor ETF.

   a) If you equal-weight the betting-against-beta factor (as in Question 2) and the idiosyncratic volatility (i.e., *ivol*) factor, i.e., by holding an equal weight of the hedge portfolios of each of the factors at each month, what is the FF-4 alpha and the Sharpe ratio of the ETF?

   b) What if you decide to hold a fund-of-funds for your multi-factor ETF, where assume that there're two ETFs called BABF (tracking the betting against beta factor that you constructed) and USMV (tracking the idiosyncratic volatility factor) each charging a management fee of 60 bps p.a. What is the FF-3 alpha and the Sharpe ratio of your ETF then?

5. (10 marks) Conduct a Fama-MacBeth monthly cross sectional test for January 2000 to November 2021, with the following specification for the idiosyncratic volatility factor:

   - Dependent variable: month t+1 monthly stock returns.
   - Independent variable: month t idiosyncratic volatility; with the control variables of CAPM beta, log size, and book-to-market ratio.

   At each month, make sure that have winsorized the outliers as you did in Assignments 1 to 3.

   Interpret your results on the idiosyncratic volatility factor.

6. (10 marks) One drawback of the above exercise is that Nasdaq 100 firms tend to be homogeneous and that the factors may not work out as expected. Replicate your Q1 results (portfolio sorting) for **size** only with each of the following two schemes:
   a. all firms in NYSE/Nasdaq;
   b. micro-caps removed in the fashion of Hou, Xue and Zhang (2018).

   Here is the stock screening procedure to allow you to download of all NYSE/Nasdaq common stocks on CRSP:

   - exchcd of 1 and 3 (EXCHCD is a code indicating the exchange on which a security is listed. Normal exchange codes are respectively 1,2, and 3 for NYSE, AMEX and the Nasdaq Stock MarketSM.)
   - shrcd of 10, 11 and 12 (Check shrcd—share class—definition on CRSP).

   One way is to download the entire database of the monthly variables you need ("prc", "shrout", "ret", "exchcd", "shrcd") for years 1999-2021 by clicking "Search the entire database", and engage in proper filtering later on. Note the existence of negative "prc" observations in the dataset and properly deal with them. Remove potential duplicates at the permno-month level.

   - Compare your results with those in Q1.
   - Also compare your results with the 10-year performance of MSCI USA Minimum Volatility Index ending Sept. 30, 2022 (the earliest available on the website) in https://www.msci.com/factor-index-scorecard. Note that this is a rough comparison as the data horizon does not exactly match.

Please submit to Dropbox "Data Assignment 4", by 11:59 pm, March 30:

1. Datasets including all variables up to Question 5. To aid with grading, please output all your datasets in easily readable format, usually starting with the first two columns being "Permno" and "Date (or yyyymm)". For Question 6, provide a brief summary statistics (refer to Assignment 3 for the list of summary statistics) for your filtered data set instead of the full data.

2. Your codes (can be Python, SAS, Stata, R, SPSS, Matlab, etc.) in Dropbox.

3. A final output report. Please make sure your output report is easy to read. Coefficient estimates do not exceed 4 decimal places, and *t*-statistics do not exceed 2 decimal places. Any submitted work with output that is *only* embedded into codes will automatically get at least 25% off the entire mark. We grade your work on your final output "report" and only recourse to your codes and data if needed.

4. Any notes if you wish to identify problems and any thoughts in the entire process. As it goes, the key to data analytics rests on good data cleaning work. A good note that has good understanding of data issues may have 5 bonus points.