ORIE5252 Final Project Report Multi-Factor Investment Strategy with Markowitz Optimization

Cornell Master of Financial Engineering
Dongchen Yuan, Yue Han, Langkun Guo, Haocheng Han
Professor: Valmiki Prasad
Nov 27th, 2020

I. Abstract

Increasing number of funds are applying quantitative methods and machine learning in their portfolio construction and many have shown positive results over the past years.

In this project, we will start from data processing, which is the very basic but essential step for quantitative trading. We will then explore the factors that can potentially help us predict the future returns and build trading and asset allocation strategies based on our predictions. Our analysis provides insights about important factors that drive the future returns and shows the performance of different trading strategies according to investors' risk preference.

II. Methodology

A. Data Processing

1) Data Description: We extracted our data from Wharton Research Data Service CRSP Daily Stock. Specifically, our data includes stocks with the following features:

Name	Type	Description
date	date	Date
TICKER	char	Stock Ticker
VOL	num	Share Volume
PRC	num	Close Price
OPENPRC	num	Open Price
DIVAMT	num	Dividend Cash Amount
FACPR	num	Factor to Adjust Price

We selected 53 stocks from S&P500 (see in Appendix) and then checked the missing value in our selected dataset. There is no missing open price, close price and volume in our selected dataset. The weights of each company in S&P 500 index were calculated by market capitalization weighted methodology and these 53 stocks are the largest stocks by market capitalization. Therefore, the S&P 500 index is mostly driven by these 53 stocks and we believe they are representative of the market.

2) Corporate Action:

- **Stock Split:** The FACPR is the factor to adjust price. Basically, the factor is 0 when there is a dividend and is the split rate when stock split happens. We would like to use this variable for our split adjustment. We first replaced 0 with 1 and filled the NAs with 1 so that the FACPR is transformed to split rate, then we computed the SplitAdjustmentFactor as the cumulative product of split rate.
- **Dividends:** We computed the SplitAdjusted-Dividend as dividend * SplitAdjustmentFactor and computed CumulativeSplitAdjustedDividend as the cumulative sum of SplitAdjusted-Dividend.
- **Return:** We adjusted the close price of day t to the benchmark day t-1 and computed the return.

3) Factor construction: We are going to construct the following five factors and the forward return.

- **Price Factor:**

$$priceFactor(t) = log(\frac{price(t-1)}{median(price(t-1), ..., price(t-5))}) \quad (1)$$

To compute the 5-day median price, first we adjusted all the close price to the benchmark day (day 1) by multiplying the SplitAdjustmentFactor and adding the CumulativeSplitAdjustedDividend. Next we computed the median on the adjusted price, and then applied the inverse transformation of previous steps to the value to get our 5-day median price so that we are able to compute the price factor with the formula above.

- **Volume Factor:**

$$volumeFactor(t) = log(\frac{Volume(t-1)}{Volume(t-2)}) \quad (2)$$

We computed this factor by adjusting the volume of t-1 and t-2 to benchmark (current day).

- **Mean Reversion Factor:**

$$MRF(t) = -ln(\frac{P_{open}(t)}{P_{open}(t-1)}) \quad (3)$$

We calculated the factor by adjusting the open price of day t to the benchmark day t-1. We are planning to trade with the close price of day t, so there should not be any looking forward bias in using open price of day t.

- **Trending Factor:** Our definition of trending factor is slight different from what we have learned in class. The original formula used the natural log and $ln(x)$ should have $x > 0$. After we computed the return on open price of day t divided by the sum of return of past k days, where k is the window or looking back period, we found that this value could be negative. Therefore, we choose not to take the natural log and this factor is computed with the following formula:

$$trendingFactor_{k-days} = \frac{Ret(t-1)}{\sum_{j=t-k}^{t-1} Ret(j)} \quad (4)$$

In our model, we used the trending factor with $k = 2, 5$.

- **Overnight Sentiment:**

$$overnightSentiment = \frac{P_{open}(t)}{P_{close}(t-1)} - 1 \quad (5)$$

The overnight sentiment is computed with the open price of day t and the close price of day t-1. We would adjust the open price of day t to day t-1. Also, since we plan to trade with the close price of day t, there should not be any looking forward bias in using open price of day t.

- **21-day Standard Deviation:**

$$21daystd = std(Ret(t-1), ..., Ret(t-21)) \quad (6)$$

We computed the $21daystd$ by taking the standard deviation of the returns over the past 21 days.

- **Transaction Cost:** We model our transaction cost as

$$TC(t) = 0.004 \times \sigma_{ann} \times \frac{OrderSize}{MDV_{21}} \times P_{prevClose} \quad (7)$$

We assume one share ($OrderSize = 1$) in order to eliminate the effect of transaction cost from the forward return.

- **Forward Return:** We will run our factor model of the 1-day forward return with respect to the above factors. Our forward return is computed by

$$r_{close,forward}(t) = \frac{P_{close}(t+1) - P_{close}(t) - TC(t)}{P_{close}(t)} \quad (8)$$

Our adjusted close price on day $t + 1$ is computed by

$$P_{close}(t+1) \times \frac{SAF(t+1) + CumDiv(t+1) - CumDiv(t)}{SAF(t)} \quad (9)$$

where SAF is the SplitAdjustmentFactor.

4) In-sample and out-sample data: We processed our in-sample and out-sample data separately. The in-sample data has the date range from 02/04/2015 to 12/29/2016 after we applied the previous formulas to compute the factors and drop rows with NA values while the out-sample data has the date range from 02/03/2017 to 12/28/2017.

B. Factor Model

For our factor model, we would like to run a simple linear regression on the forward return with respect to all the factors we constructed. The formula is:

$$
\begin{aligned}
r_{close,forward}(t) = \beta_0 &+ \beta_1 \times priceFactor(t) \\
&+ \beta_2 \times volumeFactor(t) \\
&+ \beta_3 \times MeanReversionFactor(t) \\
&+ \beta_4 \times trendingFactor_{2-day}(t) \\
&+ \beta_5 \times trendingFactor_{5_{day}}(t) \\
&+ \beta_6 \times overnightSentiment(t) \\
&+ \beta_7 \times 21daystd
\end{aligned} \tag{10}
$$

We will run this model on our in-sample data and make predictions of the forward return on our out-sample data. We will then use the out-sample predictions to construct trading strategy and perform backtest.

C. Trading Strategy

We decided to utilize a long-only trading strategy with daily close-out. We would build a daily-rebalanced portfolio by selecting stocks with positive predicted returns out of our stock-pool and close our position by the end of the trading day.

We do not engage short selling as the stock borrowing is not guaranteed and the borrowing costs are hard to estimate.

We first tried to build our portfolios as equally-weighted and optimized weights assigned to each stocks by using Markowitz optimization.

1) Equally-weighted Model: The equally-weighted strategy will assign all the stocks in the daily portfolio with equal weights of stock value. We do this by computing the total number of selected stocks n each day and use $\frac{1}{n}$ as individual stock weights. The share of stocks is then computed based on the weights.

2) Model with Markowitz Optimization: Our Markowitz model is structured as:

$$
\begin{aligned}
&\min_{x_0,x} \sigma_p^2 = \mathbf{x^T P x} \\
&s.t. \; e^T x = 1 \\
&\quad\; x_0 \geq 0 \\
&\quad\; x_i \geq 0
\end{aligned} \tag{11}
$$

The covariance matrix we used in Markowitz Optimization is the historical covariance computed from the in-sample data. Since our predictions provided positive returns, we would like to minimize the portfolio risk in general. Therefore, We applied Markowitz Optimization to find the optimal allocation weights that minimize the portfolio variance (minimize potential risks).

III. Implementation Results

A. Factor Model

As previously mentioned, we took the 2015 and 2016 stock data as in-sample data, and 2017 stock data as out-sample data. The following table shows in-sample statistics:

Variable	Coef	p-value
priceFactor	-0.0009	0.664
volumeFactor	-0.0016	0.000
meanReversion	0.0079	0.047
overnightSentiment	0.0414	0.001
$\text{trendingFactor}_{2d}$	8.322e-18	0.930
$\text{trendingFactor}_{5d}$	-5.596e-08	0.981
21daystd	0.0318	0.000

As we can see from the regression results, most factors are statistically significant except priceFactor, trendingFactor_2d and trendingFactor_5d.

Specifically, volumeFactor has a negative relationship with forwardReturn while meanReversionFactor, overnightSentiment and 21daystd have a positive relationship with forwardReturn. Moreover, the intercept is -0.0053.

Other metrics are shown in the following table:

Name	In-sample	Out-sample
Mean of residual	3.828e-19	0.0007
SD of residual	0.0188	0.0127
Reduced Mean-square residuals	0.0004	0.0002

This shows that the residual is small both in in-sample and out-sample estimates.

B. Backtest

We performed our backtest on the out-sample data with our predicted return and trading strategy. The trading period is from 02/03/2017 to 12/28/2017 (same as out-sample). We selected stocks to hold by their predicted return and used equal weights or Markowitz to determine the weights with daily rebalance. We computed daily transaction cost with formula(7) and subtracted it from daily portfolio value. Below is the performance of our portfolio and the market (S&P500).

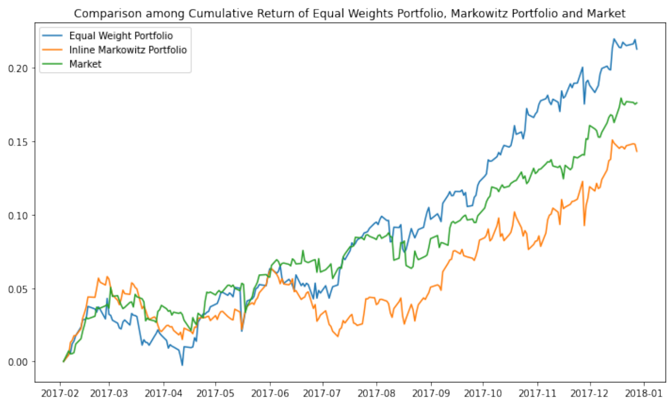

Fig. 1. Cumulative Performance

The performance metrics are shown in the following table:

TABLE I
SUMMARY STATISTICS FOR PERFORMANCE

Statistic	Equally-Weighted	Markowitz	Market
Holding Period Return	21.28%	14.32%	17.61%
Annual Return	21.87%	15.23%	18.44%
Volatility	8.48%	7.87%	6.73%
Sharpe	2.06	1.38	2.09
MaxDrawdown	24.63%	18.76%	19.53%
Skewness	-1.05	-0.81	-0.51
Kurtosis	5.11	3.92	3.11
Sortino	0.25	0.20	0.29

IV. Analysis

In general, the equally-weighted portfolio gives us the highest holding period return (and annual return) while its Sharpe ratio is slightly lower than the market. The Sortino ratio is consistent with Sharpe ratio. This proves that our prediction of forward return based on factors and our strategy of selecting stocks can work in some extent.

Meanwhile, although the portfolio with Markowitz Optimization shows a lower return, it shows a lower volatility than the equally-weighted portfolio as expected since we tried to diversify the risk of this portfolio by minimizing the variance. It can also be seen that the portfolio with Markowitz Optimization has a lower maximized-drawdown than that of the market and the equally-weighted portfolio.

While backtest is not a guarantee of future performance, we believe that our work can provide some insights that factor model can help to create a good trading strategy.

V. Conclusion

Our factor model can help to predict the forward return. Our trading strategy based on the predicted forward return is able to achieve an outperformed return without significantly driving up the volatility. Markowitz optimization helps to diversify the risk of our portfolio.

VI. Future work

- Short-selling: For the simplicity of this project, we did not allow short selling. In the future, we may include short-selling to improve the performance of our strategies as we have more accurate estimates of the borrowing costs in the future.
- Additional Factors: We may construct more factors to explore their influence on the predictions to future returns. The potential factors candidates are Momentum, Mean Reversion Sector Neutral and fundamental factors such as book to market ratio, earning to price ratio and cash-flow to price ratio. We can also incorporate alternative datasets such as job posting data, ESG-ratings and company-sepecifc transaction data.
- Other machine learning models: We can try to apply decision tree or ensemble algorithms such as random forest and XGBoost to conduct prediction. We can also use grid search to explore best parameters for these models
- Transaction Cost: In this project, we assumed that even if a stock appears in our selected stock-pool in two consecutive days, we still close our previous position and buy this stock again for our new daily portfolio. This is easier for our rebalance process. However, this will increase the transaction cost. We could improve our backtest algorithms to calculate the transaction cost more accurately.

VII. Acknowledgement

The authors would like to thank Professor Valmiki Prasad for his guidance.

VIII. Reference

References

[1] Aboody, David and Even-Tov, Omri and Lehavy, Reuven and Trueman, Brett, Overnight Returns and Firm-Specific Investor Sentiment (October 1, 2016). Journal of Financial and Quantitative Analysis (JFQA), Forthcoming, Available at SSRN: https://ssrn.com/abstract=2554010 or http://dx.doi.org/10.2139/ssrn.2554010

IX. Appendix

TABLE II

Summary for Assets

TICKER	**Company name**
CMCSA	Comcast Corp.
DGX	Quest Diagnostics
ALXN	Alexion Pharmaceuticals
KO	Coca-Cola Company
UNH	United Health Group Inc.
V	Visa Inc.
T	AT&T Inc.
IRM	Iron Mountain Incorporated
MA	Mastercard Inc.
VRSN	Verisign Inc.
PYPL	PayPal
NOV	National Oilwell Varco Inc.
CHRW	C. H. Robinson Worldwide
DIS	The Walt Disney Company
MRK	Merck & Co.
PEP	PepsiCo Inc.
ZION	Zions Bancorp
AMZN	Amazon.com Inc.
GOOG	Alphabet Inc. (Class C)
VZ	Verizon Communications
NEE	NextEra Energy
FB	Facebook Inc.
CRM	Salesforce.com
MSFT	Microsoft Corp.
AAPL	Apple Inc.
TXN	Texas Instruments
COST	Costco Wholesale Corp.
DHR	Danaher Corp.
NKE	Nike Inc.
NVDA	Nvidia Corporation
WMT	Walmart
ANSS	ANSYS
MNST	Monster Beverage
ETFC	E*Trade
PWR	Quanta Services Inc.
URI	United Rentals Inc.
ADBE	Adobe Inc.
MTD	Mettler Toledo
BAC	Bank of America Corp
CSCO	Cisco Systems
YUM	Yum! Brands Inc
PG	Procter & Gamble
NFLX	Netflix Inc.
PFE	Pfizer Inc.
INTC	Intel Corp.
XOM	Exxon Mobil Corp.
JNJ	Johnson & Johnson
JPM	JPMorgan Chase & Co.
CVX	Chevron Corp.
HD	Home Depot
ADSK	Autodesk Inc.
GOOGL	Alphabet Inc. (Class A)
ACN	Accenture plc