Project 5: How to predict future price of a security?

Group 2

Columbia University

yy2502@columbia.edu

November 28, 2018

Overview

1. Introduction
   - Background
   - Highlights
2. Mathematical Model
   - ARMA Model
   - Influence Measure
3. Analysis and Results
   - Cross Validation in Time-Series Data
   - Data
   - Results and Performance
   - Robust Portfolio
4. Conclusion
   - Summary
   - Forward Looking Statement
   - Acknowledgement
5. Appendix
6. Reference

Background: Random Walk

Security prices follow random walk. Nobel Laureate Eugene Fama and researcher Kenneth French, former professors at the University of Chicago Booth School of Business, attempted to better measure market returns and, through research, found that value stocks outperform growth stocks. Similarly, small-cap stocks tend to outperform large-cap stocks. As an evaluation tool, the performance of portfolios with a large number of small-cap or value stocks would be lower than the CAPM result, as the Three-Factor Model adjusts downward for small-cap and value outperformance.

There is a lot of debate about whether the outperformance tendency is due to market efficiency or market inefficiency. However, there is no agreement settled in this field.

Background: Asset Pricing (The Quants on the Street)

A five-factor model directed at capturing the size, value, profitability, and investment patterns in average stock returns performs better than the three-factor model of Fama and French (1993) [Fama French 1993]. The five-factor models main problem is its failure to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability. The models performance is not sensitive to the way its factors are defined. With the addition of profitability and investment factors, the value factor of the FF three-factor model becomes redundant for describing average returns in the sample we examine.

$$r = R_f + \beta_1(R_m - R_f) + \beta_2 \mathrm{SMB} + \beta_3 \mathrm{HML} + \alpha + \epsilon$$

Source: `https://www.sciencedirect.com/science/article/pii/S0304405X14002323`

In industry, traders look at the a variety of technical indicators for trading opportunties. For example, the most common one in the following (middle in the bottom line) is the flag patterns, e.g. bull flag and bear flag.

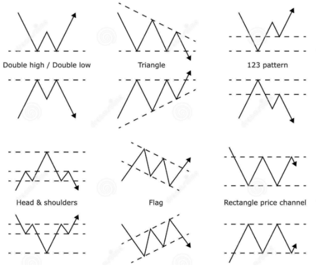

Figure: Collection of common chart patterns for professional intra-day traders.

Source: `https://www.tradingview.com/chart/0FKPiwjU/`

Motivation

- Before Columbia, I was under Novy-Marx's supervision. My research was submitted to AQR Capital Management led by Fama (Nobel Laureate). After undergraduate school, I worked as a trader on the street (licensed and to manage $1m AUM).
- We know what may explain security returns, but uncertain if they are persistent.
- Fama and French: not for the purpose of doing predictions.
- They raised the question: "is market efficient?" Despite the fact that scholars cannot agree on the answer to the question, we would go nowhere even if they do. For people who want to trade, they still trade stocks. For people who do not want to trade, they still stay away from the market.
- How to digest all these information so that we can provide prediction to investors? (e.g. What is tomorrow's stock price?)

Highlights

Highlight 1

Per stock basis, we provide analysis and explanation how the security price behaves as time move on. (A time-series story)

Highlight 2

Per analysis, we provide a baseline model and an improved model. Baseline model we simply adopt ARMA(p, q) time-series analysis. Improved model we proposed Lo and Zheng (2002, 2008, 2016) as main methodology. We present error reduction of at least 97%.

Highlight 3

We land this project on a portfolio strategy that can beat the market. Simulating from March 2016, $1000 initial investment can give you $1700 USD while S&P 500 Index Fund gives you $1400.

AutoRegressiveMoving-Average (ARMA)

Theorem (ARMA, Peter Whittle 1951)

The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms. This model contains AR(p) and MA(q). The equation follows

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Influence Measure (I-Score) in Discrete Framework

Chernoff, Lo, and Zheng (2009) [Chernoff Lo Zheng 2009] proposed the Partition Retention method to detect both marginal and high-order interaction effects based on Lo and Zheng's earlier work [Lo Zheng 2002].

Assume that $\{X_j, j = 1, ..., m\}$ taking values 0 or 1. There are 2^m possible partitions for each set of m explanatory variables.

Theorem (I-score)

Normalized influence score, I-score, as

$$I = \frac{1}{n\sigma_Y^2} \sum_{k=1}^{2^m} n_k^2 (\hat{Y}_k - \bar{Y})^2,$$

where $\hat{Y}_k$, the estimated value, is the average of the n_k observations on Y falling in the k^{th} partition cell, $\hat{Y}$ is the global mean of Y and σ_Y^2 is the variance of Y.

Influence Measure (I-Score) in Continuous Framework

Chernoff, Lo, and Zheng (2009) [Chernoff Lo Zheng 2009] proposed the Partition Retention method to detect both marginal and high-order interaction effects based on Lo and Zheng's earlier work [Lo Zheng 2002]. Related papers are [Lo Zheng 2002] [Lo Chernoff Zheng Lo 2015] [Lo Chernoff Zheng Lo 2016]. Please also see Huang, Chien-Hsun (2014) and Ding Yuejing (2008) `https://clio.columbia.edu/catalog/11876689?counter=2`.

Theorem (I-score)

Given a data set $\mathbf{X}$, for each observation i, we can define local mean by the nearest K neighborhood surrounding X_i. We can then define global mean as $\bar{Y} = \frac{1}{n}\sum Y_i$. The predictivity of this data set $\mathbf{X}$ can be measured by the following equation

$$I_C = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{K}\sum_{j\in N(i)}^{K} Y_j - \bar{Y}\right)^2$$

Influence Measure (I-Score) in Continuous Framework

In continuous framework, instead of 2^m partitions, we use k nearest neighborhood.

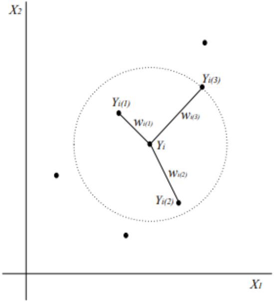

Figure: Graphical Illustration of using NN for Local Measure

Cross Validation in Time-Series Data

- Cross validation is conducted in the following manner:
- First, we cut data set into training set and test set;
- Second, for training set, one fold is defined as training and validating;
- We approach second step doing 5 folds and each fold we store the model result;
- Last, we use the optimal result on test set.

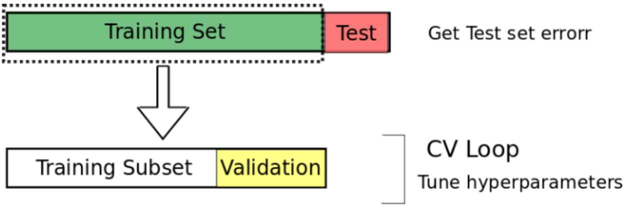

Figure: Cross-Validation in Time-Series Data

Data and Source of data

- Due to limited time and resources, we use only Dow Jones 30 Components.
- We use *quantmod* package in R console and download stock data from Yahoo/Google Finance.

`http://indexarb.com/indexComponentWtsDJ.html`

Top Weighting in Dow Jones 30 Components: Boeing (BA)

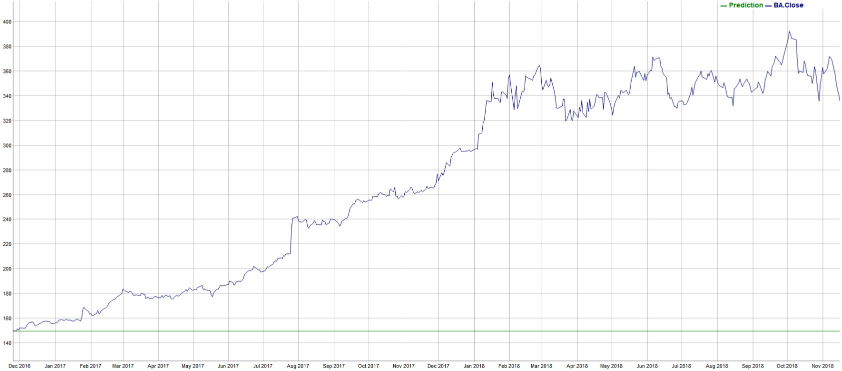

Figure: This figure presents MSE (mean square error) results of held out test set for top weighted stocks in Dow Jones 30 Components, Boeing (BA), using ARMA model.

Top Weighting in Dow Jones 30 Components: Boeing (BA)

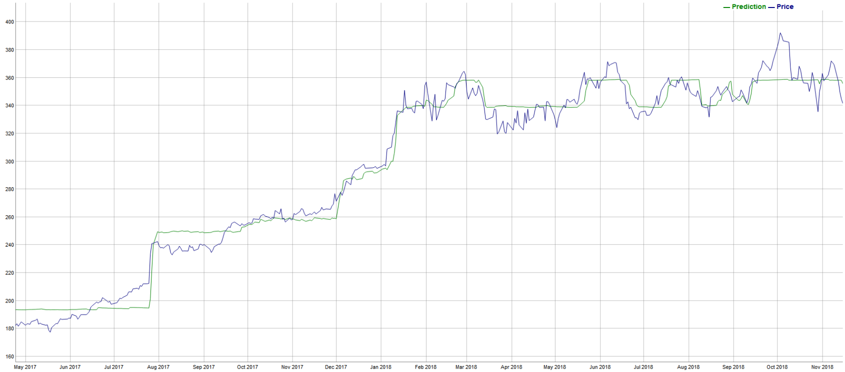

Figure: This figure presents MSE (mean square error) results of held out test set for top weighted stocks in Dow Jones 30 Components, Boeing (BA), using influence measure.

Top 3 Weightings in Dow Jones 30 Components

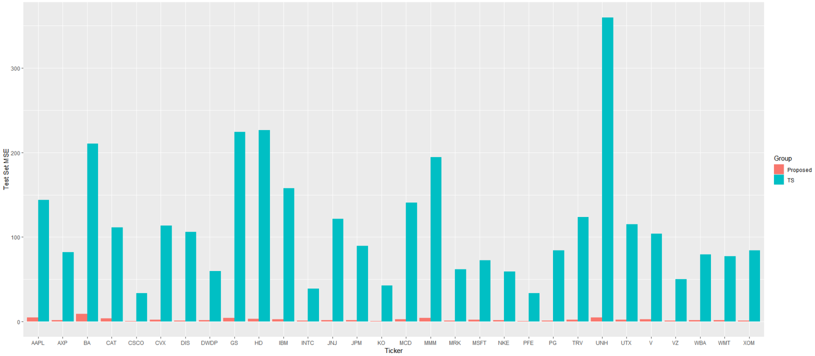

Figure: This figure presents MSE (mean square error) results of held out test set for all 30 components of Dow Jones Index. The bar charts shows MSE for both baseline model (ARMA) and improved model (I-score).

Top 3 Weightings in Dow Jones 30 Components

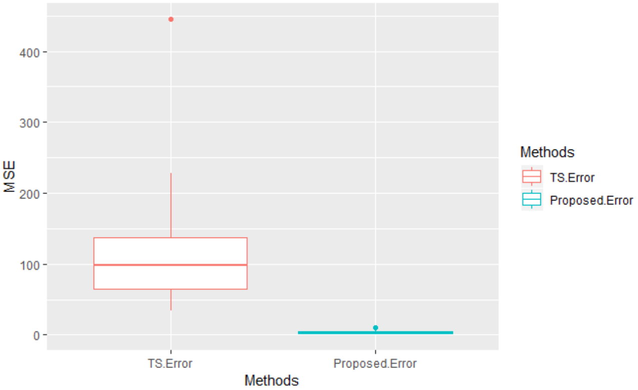

Figure: This figure presents MSE (mean square error) results of held out test set for all 30 components of Dow Jones Index. The barplot shows distribution of MSE for both baseline model (ARMA) and improved model (I-score). This is a 97% error reduction on average.

Robust Portfolio: (1) Timing and (2) Stock Picking

1. Timing is very important.
2. Check it out: `https://medium.com/@yiqiaoyin/yins-philosophy-the-dip-digger-7f732ada8fba`

Robust Portfolio: (1) Timing and (2) Stock Picking

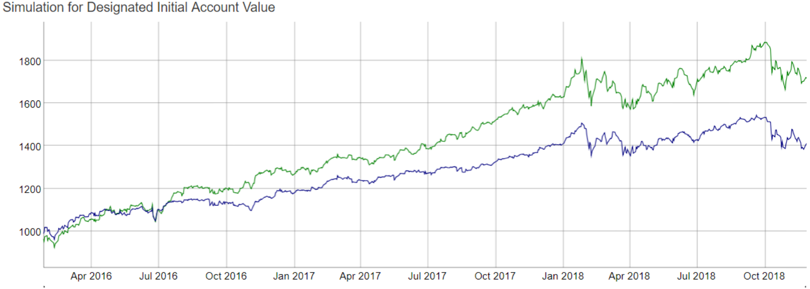

Figure: This figure presents two portfolios. The path in green presents portfolio simulated by using influence measure to pick stocks. The path in blue is portfolio invested in S&P 500 Index Fund. This simulation starts from March of 2016.

Robust Portfolio: (1) Timing and (2) Stock Picking

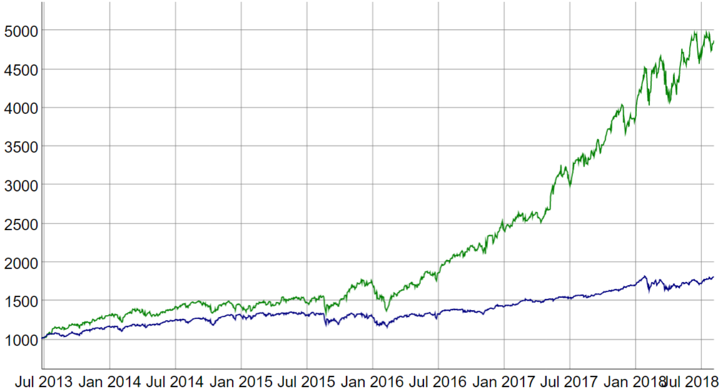

Figure: This figure presents two portfolios. The path in green presents portfolio simulated by using influence measure to pick stocks. The path in blue is portfolio invested in S&P 500 Index Fund. This simulation starts from January of 2013.

Summary

- We outperform time-series model by reducing error (MSE) by at least 97% on average for all stocks in Dow Jones 30 Components;
- Simulating from March 2016, $1000 initial investment can give you $1700 USD while S&P 500 Index Fund gives you $1400;
- Simulating from January 2013, $1000 initial investment can give you $5000 USD while S&P 500 Index Fund gives you $1700;
- We construct a portfolio that beats the market without hesitation;
- We promote a philosophy that machine and human psychology can both work together to form decision making process.

Forward Looking Statement

- End of the day, a heart trying to make money will not sustain in stock market.
- This game is more of art then science.
- A license to trade is also a license not to trade.

Acknowledgement

- We have not disclosed strategies and game planning in risk management. Hence, this project does not function as investment advises. We are not responsible for any monetary losses from third party as it is the third party's responsibility to understand his/her risk profile.
- We also want to thank Professor Ying Liu and Professor Tian Zheng for hosting lectures of Advanced Data Science this semester. It is with transcending gratitude that we announce here what an inspiration both professors have been throughout our experience of building this shiny app. Their knowledge, understanding and genuine care for others is illuminated in everything they do! We, Group 8, are in forever debt for their teachings. Moreoever, we also want to thank TA, Chengliang Tang. There is not enough we can say about how much we thank heaven that he is our teaching assistance. His patience and understanding are unsurpassed. We are grateful for being his students.

Backward Dropping Algorithm (BDA) B times based on influence measure:

Step 1: Randomly select a subset of d variables from total m variables. $\mathbf{X}_d = \{x_1, ..., x_d\}$ where x_i indicates the i^{th} variable of the selected subset. d is usually set as a moderate number such as between 5 and 20;

Step 2: Step 2.1: To backward drop noisy variables within current d-dimensional variable set $\mathbf{X}_d$, compute the score $I(\mathbf{X}_d)$ and $I(\mathbf{X}_{d[-i]})$ for all $i = 1, ..., d$ where $I(\mathbf{X}_{d[-i]})$ represents the score computed without variable x_i. Delete j^{th} var. having maximum difference $I(\mathbf{X}_{d[-j]}) - I(\mathbf{X}_d)$

Step 2.2: If there is no variable remaining in the set, stop: otherwise repeat Step 2.1 with $d = d[-j]$;

Step 2.3: Return d_1 variables that attain the highest influence score as the returned variable set in the eliminating procedure;

Step 3: Repeat Step 1 to Step 2.3 B times

Step 4: Conduct further analysis based on the returned variable sets with the highest $B_1 (B_1 << B)$ scores among the B repeat times.

Appendix: BDA, An Example

- Create an artificial data, $\tilde{X} = \{X_1, ..., X_{50}\}$ with 100 observations.
- Define $\mathbb{P}(Y = 1|\tilde{X}) = \frac{1}{1+\exp(X_1+X_2)}$ and $\mathbb{P}(Y = 0|\tilde{X}) = \frac{\exp(X_1+X_2)}{1+\exp(X_1+X_2)}$

Var. Left	X8 X7 → X4 X9 X3 X2 X1 X10	X8 X7 → X9 X3 X2 X1 X10	→ X8 X7 X3 X2 X1 X10	X7 X3 X2 X1 → X10	→ X7 X3 X2 X1	→ X3 X2 X1	X2 → X1	X2
I-score	38.28	42.71	54.69	63.54	71.18	79.95	91.67	81.34
Var. To Drop	3	3	1	5	1	1	2	0

Figure: This table explains the procedure of running one Backward-Dropping Algorithm (BDA).

Appendix: How many rounds of BDA?

The backward dropping algorithm, depending on random sampling, is required to sample as many different combinations of the variables as possible. Assume there is an l-order interaction and it will be captured only when these l variables are selected simultaneously. In general, the repeat time B should be large enough to capture the interaction effects, and it is related to the variable size of the data m, the order of interaction l and number of variable selected d for each random sample where $d << m$.

Given a data set with m variables, to capture certain l-order interaction by the algorithm with at least certain probability p, this implies the following inequality

$$\mathbb{P}(\text{capture } l - \text{order interaction}) = 1 - \left(1 - \frac{\binom{m-l}{d-l}}{\binom{m}{d}}\right)^B > p$$

Appendix: How many rounds of BDA?

We present the following table for illustration of how many times B is needed for an m-size data with l-order interaction by selecting d variables each BDA. For example, given 200 observations and to have at least 50% probability that the order of 2-way interaction being selected while letting the algorithm select $d = 30$ variables initially, we would expect at least 31 rounds of interactions (yellow highlighted cell). Notice that the notation "E+i" means $\times 10^i$ while $i \in \mathbb{Z}_+$.

m=200	p = 0.5				
d/l	2	3	4	5	6
7	6.56E+02	2.60E+04	1.28E+06	8.37E+07	8.16E+09
14	1.51E+02	2.50E+03	4.48E+04	8.78E+05	1.90E+07
20	7.20E+01	7.98E+02	9.25E+03	1.13E+05	1.47E+06
24	5.00E+01	4.49E+02	4.22E+03	4.14E+04	4.24E+05
30	3.10E+01	2.24E+02	1.64E+03	1.23E+04	9.62E+04

Appendix: Comparison between correlation and I-score

Let us conduct a more complicated experiment. We generate 200 observations artificially for two variables, say X_1 and X_2, that come from $\mathcal{N}(0, 1)$. We can define different underlying model for response variable Y. We can compare the results of correlation of (Y, X_1) and (Y, X_2), respectively, and continuous I-score (modified I-score) of (X_1, X_2). We can simulate (1) $Y = X_1 + X_2 + \epsilon$, (2) $Y = X_1 X_2$, (3) $Y = X_1^2 + X_2^2$, (4) $Y = e^{X_1 X_2}$, and (5) $Y = \sin(X1X2) + \cos(X1X2) + \epsilon$.

Underlying	(1)	(2)	(3)	(4)	(5)
cor(y,x_1)	0.55	0.14	0.09	0.11	0.08
cor(y,x_2)	0.55	-0.05	-0.06	-0.01	-0.01
$k = 1$, I-score(x_1, x_2)	2.27	1.45	3.10	5.39	1.15
$k = 3$, I-score(x_1, x_2)	1.92	0.89	2.30	4.00	0.68
$k = 6$, I-score(x_1, x_2)	1.71	0.68	1.91	2.99	0.47
$k = 12$, I-score(x_1, x_2)	1.41	0.51	1.50	1.98	0.34

Appendix: Data Processing

- For each company i at a time t, we observe a price, $p_{i,t}$
- Define SMA to be

$$\mathrm{SMA}_n = \frac{1}{n}\sum_{i=n}^{t-n} p_{i,t-n}$$

- Let the distance between price and moving average to be **D** which is defined as

$$\mathbf{D}_i := p_n - \mathrm{SMA}_n$$

while $i = n$, and then we can consider $\mathbf{D}_i$ to be i.i.d. with $\mathbb{E}\mathbf{D}_i = 0$ and $\mathbb{E}\mathbf{D}_i = \sigma^2 \in (0, \infty)$. Then

$$\sum_{m=1}^{n} \mathbf{D}_m \Big/ \left(\sum_{m=1}^{n} \mathbf{D}_m^2\right)^{1/2} \Rightarrow \chi$$

while χ is the stand normal distribution.

- But why?

From weak law we know that

$$\sum_{m=1}^{n} D_m^2 / n\sigma^2 \to 1.$$

Also note $y^{-1/2}$ s continuous at 1, then we have

$$\begin{aligned}
\left(\sigma^2 n \Big/ \sum_{m=1}^{n} \mathbf{D}_m^2\right)^{1/2} &\to 1, \text{ in prob., see } \star \\
\frac{\sum_{m=1}^{n} \mathbf{D}_m}{\sigma\sqrt{n}} \left(\frac{\sigma^2 n}{\sum_{m=1}^{n} \mathbf{D}_m^2}\right)^{1/2} &\Rightarrow \chi \cdot 1, \text{ from } \star \\
&= \chi
\end{aligned}$$

Notice that the $\star$ is because in Weak Convergence, there is a theorem stated that $X_n \Rightarrow X_\infty$ if and only if for every bounded continuous function g we have $\mathbb{E}g(X_n) \to \mathbb{E}g(X_\infty)$. Since we discussed the continuity of function $y^{-1/2}$ at 1, this line is valid.

References

Chernoff Lo Zheng (2009)
Discovering influential variables: a method of partitions
The Annals of Applied Statistics, 1335 – 1369.

Fama French (1993)
Common risk factors in the returns on stocks and bonds
Journal of Financial Economics 33(1), 3 – 56.

Lo Zheng (2002)
Backward Haplotype Transmission Association (BHTA) Algorithm a Fast Multiple-Marker Screening Method
Hum. Hered 53, 197 – 215.

References

Lo Chernoff Zheng Lo (2015)
Why significant variables aren't automatically good predictors
Proceedings of the National Academy of Sciences 112, 2015, 13892.

Lo Chernoff Zheng Lo (2016)
Framework for making better predictions by directly estimating variables? predictivity
Proceedings of the National Academy of Sciences 113, 2016, 14277.

Thank you very much!