Summer Internship Report

WallachBeth Capital

Yuefan Zhu

M.S. in Operations Research

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu

Contents

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu

Duties and responsibilities

I am doing summer internship at WallachBeth Capital in New York, which is a sell-side trading firm located on Wall Street, providing clients with unconflicted advice and intelligent trading solutions. The internship period is from June 1st to August 31th. I am working as a desk quant intern on trading floor and my principle responsibilities include:

- Provide quantitative rationale to support trading ideas and content generated by the team
- Statistical data analysis and algorithmic problem solving to identify and solve business problems
- Evaluate trends and identify relationships in data to provide strategic advice
- Communicate findings and recommendations on critical initiatives and influence leaders to take action on those findings
- Work directly with Chief Quant and Chief Macro Strategist

Notable Projects

1. Dark Liquidity Research

The purpose of this project is to help traders capture the information in the ATSs which are operated by banks, market makers and other institutions. As a broker, our company focuses on controlling the trading cost. Our order book records also show that we put a lot of weight on dark pools trading.

The procedure of this project:

- Fetch data from FINRA and output a data frame which can be easily updated over time automatically.
- Develop a GUI in R shiny to provide volume, number of trades and trade size information for each ATS and each equity using heatmap with clustering analysis, where equities can be consolidated by sectors, market cap, asset class, security type, region, exchange, volatility, etc, and provide overall information for all ATSs.
- First determine the objective of our prediction. After some trials, I decide to use rankings of dark pools. That is for each equity, top 5 get Y=1, others get Y=0. Then I apply random forest to do supervised learning, predicting trading volume after adding beta, average spread, ADV, volatility, sector, region, market cap, exchange and other variables, then conduct backtests based on FINRA's releasing policy and get 22% overall error rate and 32% false positive error rate.

Random forest algorithm is an ensemble learning method, which corrects for decision trees' over fitting to their training set. The steps are to get n samples with replacement from training set, then train a decision tree on each sample, getting n decision trees. Then we can do classification by taking the majority vote of these n trees. It can be easily implemented in R using 'randomforest' package. But if you look at the ranking graph of ATSs, you may find that

top 5 most liquid ATSs on some symbols will remain top 5 for a long time, which addresses the importance of looking back to historical info and market conditions. Although sometimes the forecasting result doesn't matter, it can show which variable significantly influence the dark volume. It can also provide meaningful information to predict when there will be a big move of rankings on some symbols.

2. Volatility Prediction

The purpose of this project is to predict volatility and provide trading advice. Popular methods of predicting volatility of individual equities include moving average, EWMA, ARMA-GARCH models, Implied Standard Deviation models, etc. For this internship, I apply ARMA-GARCH models with calendar adjustment.

The procedure of this project:

- Get price data from yahoo, calculate daily log return and calculate monthly return and sd
- Fit seasonal ARMA model for monthly return in ' auto.arima' with 'ts(r,frequency=3)'
- Fit EGARCH(1,1) model with student t distribution on the residuals of ARMA, with external regressors include dummy variables standing for earnings calendar, expected EPS and real EPS
- Conduct backtests and compare prediction effectiveness of different methods, which include naïve method, EWMA, naïve GARCH with different parameters, EGARCH, EGARCH with earnings adjustment, etc.

Seasonal ARMA Model

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

$$\text{ARIMA } (p, d, q) \times (P, D, Q)\ S$$

With p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\varphi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t$$

The non-seasonal components are:

$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_p B^p$$

$$\text{MA: } \theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$$

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu

The seasonal components are:

$$\text{Seasonal AR: } \Phi(B^S) = 1 - \Phi_1 B^S - \ldots - \Phi_P B^{PS}$$

$$\text{Seasonal MA: } \Theta(B^S) = 1 + \Theta_1 B^S + \ldots + \Theta_Q B^{QS}$$

Note that on the left side of the formal equation, the seasonal and non-seasonal AR components multiply each other, and on the right side of the equation, the seasonal and non-seasonal MA components multiply each other.

EGARCH (1, 1) with Earnings Adjustments

After we de-center the returns of equities using seasonal ARMA model, we get a white noise series with time-varying volatility as shown below.

$$r_t = \mu_t + a_t, \quad a_t = \sigma_t \epsilon_t,$$

Then we use EGARCH (1, 1) to model the volatility clustering effect.

$$\log \sigma_t^2 = \omega + \sum_{k=1}^{q} \beta_k g(Z_{t-k}) + \sum_{k=1}^{p} \alpha_k \log \sigma_{t-k}^2$$

$$g(Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|))$$

With p=1, q=1 and Z as a standard normal variable in our model. The external regressors are added at the end of the above equation.

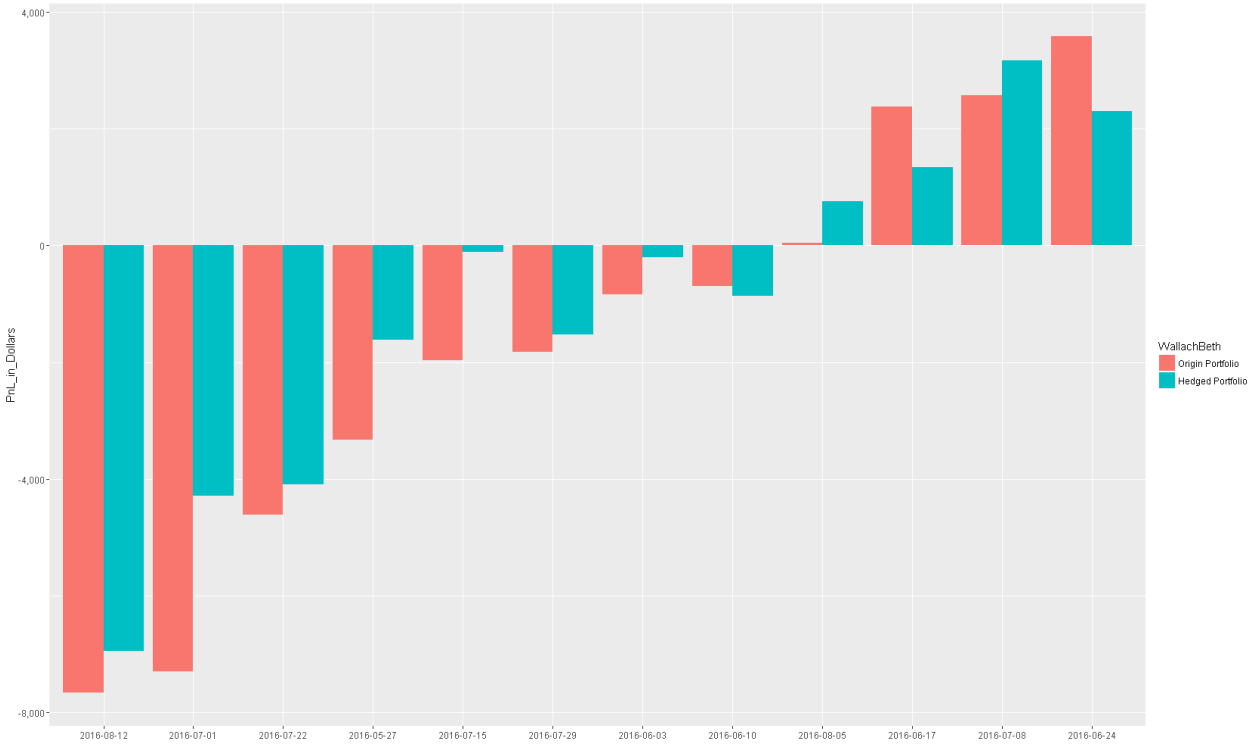
3. Equities Hedging (in progress)

The purpose of this project is to meet our clients' requirements to protect their equity asset, particularly when recent stock rally is not convincing the managers that U.S. economics is better. And we provide solutions using ETFs and ETNs.

The procedure of this project:

- Get price data from Bloomberg API in excel, then import it to R
- Apply t-Copula to give a symbols' ranking that matches the spread of long/short portfolio
- Conduct backtests and compare hedging effectiveness of different methods
- The methods include regressing on top 2 to top N tickers, then pick the best one based on backtest; conducting PCA on the tickers first, then use the major components as regressors; playing with optimizing different loss functions with constraints on the positions…(I am still working on it)
- Prepare presentations include nice charts and graphs for investors (example below)

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu



A Typical Day

7:00 AM ~ 8:00 AM	• Wake up and catch up with subway
8:00 AM ~ 9:00 AM	• Have breakfast and meet with my boss
9:00 AM ~ 12:30 PM	• Write codes or look at papers
12:30 PM ~ 13:30 PM	• Have lunch and hang out with friends
13:30 PM ~ 16:30 PM	• Write codes or look at papers
16:30 PM ~ 17:00 PM	• Meet with boss about tomorrow's tasks

Skills and qualifications acquired

- Strong academic background in a relevant field – Computer Science, Engineering, Physics or Mathematics
- Strong technical and computer programming skills required
- Strong quantitative / analytic reasoning and problem solving abilities
- Strong interpersonal and communication skills, both written and verbal
- Intelligence, creativity with a natural inquisitive disposition
- Motivated, a self-starter that has the ability to work independently as well as thrive in a team environment

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu

- Results oriented work ethic and the ability to handle time sensitive projects while focusing on the quality

Contribution to career plans

My current career goal is to become a quant in future and I will look for full time opportunities both in U.S. and China. I think this desk quant internship will not only highlight my resume, but most importantly, it shows me where I should focus on and what the current industry's trend is. I am also very thankful to my supervisor, who shares with me a lot of brilliant ideas and provides me many opportunities training my programming skills, which is really helpful for my future career development. For example, when I was trying to fetch the data and build a GUI for traders, I failed to provide an efficient solution to meet the traders' requirement at first. Then I studied a lot on how to speed up the program to deal with and focus on providing a user friendly interface that can help traders make decisions in seconds.

WBC gives me a good platform that broadens my professional horizon. I meet very nice and professional people here and I learn a lot of business and market senses after communicating to 'talkative' traders. So, I will not feel too surprising if there is going to be a big turn of my career path. I'd like to try interesting and challenging things any time in my life.

Relationship to coursework at Columbia

I am currently enrolled in M.S. in Operations Research, concentrated in Financial Engineering. What I do at WBC is highly correlated to the coursework at Columbia and I believe it is the coursework at Columbia that helps me to get admitted to this internship. For example, my final project of IEOR4733 Algorithm Trading is about pair trading using Naïve, OU, GARCH methods. And as you can see, the hedging project at WBC is an extension of that GARCH method, modeling volatility of returns instead of spreads. Another example is that the theories I learned in IEOR4525 Machine Learning provide me a lot of intelligent ideas on my work. Furthermore, the programming skills I learned and the training I received in IEOR4500 Applications Programming and IEOR4404 Simulation help me implement my ideas in very efficient ways.

Yuefan Zhu
M.S. Operations Research
yz2835@columbia.edu