Introduction to the Performance of My Strategies

***Claim:** In this page, I only give introduction to the performance of my quant strategies. No any exposure to the tricky data mining method to different kinds of data, self designed network structures, specially self-defined loss function and the method to make a neural network totally interpretable. I reserve all the rights to this personal page.

Basic Setting

This basic setting is suitable for all the experiment results shown in this page. If I have used different settings on some experiments, I will give it a special description.

1.About the Market

All the experiments are done on the Chinese A-share market. In this market, I assume that I should strictly obey the trading principles, such as T+1 trading; can't short single stock; can't buy the stock which hits raising limit up; in a single fund product, contains more than 10 stocks and no single stock takes more than 10% weight and etc... And there are a lot of constraints from the specific industry and products. Because at this time, I was at mutual fund, our CSI500 Quant product needs 82%'s inventory comes from CSI500, but this point is not required by the hedge fund. And for a product which is almost ready, we more or less have some constraints on the risk exposure, such as the industry exposure and single stock's exposure should be not too far away from the CSI500. Good thing and bad thing all comes from the constraints. And this is the specific flavor, which should be made clear for both the researchers and investors.

2.Rolling Forecast

All the experiments are done on the rolling forecast, including the process of generating and combining alphas. Many researchers, as far as I know, they will fit the history once, and then get the generated alphas. In this case, their gap between backtesting and paper trading will be much lager than expected. eg. today is T, [T-train_length-hoding-1, T-hoding-1] is the training set, [T, T+test length] is the testing set, normally, I set training set as 300 trading days, testing set as 21 trading days. And each neural network's loss functuon is predicting the next 3 trading days' rank return. To sum up, for each 21 trading days, before trading, I will train several neural networks at T-1 day, and then use these modes for the next 21 trading days' trading. During the trading period, I will change the inventory every 3 days. As for the predicting the next 3 trading days' rank return, this is only the loss function of my model. To simplify the descrption, I didn't mention validation set, actually I use it in my strategies, and it's very helpful and tricky. If you feel confused about this descrption, please feel free to read it again and again. This is the right and fair description, but most of people whom I talked to, all feel confused.

3.Backtesting

Normally, I apply 0.2% transaction fees on both side (Shown in the following experiment results). It's fair because the majority of institutional investors use it to do backtesting. It covers the sum of must paid fees and the assumption of market impact (if the turnover rate is much higher, the assumption will change). What's more, you can quickly calculate my performance in the suitation of 0.3% or 0.4% transcation fees on both side. Because my turnover rate is very low, thus, it bring impact but far from making my strategy doesn't work; For untradable stocks in each trading day, I will use it as a training sample, but in the testing set, I will clear it. Which means, I only trade tradable stocks in the real situation; For the trading price, I generate signals at T-1, and do trading at T day's moring. I use open price at T and T+3 day to do the trading. Also I try the close price, VWAP in the first 30 minutes, (open+high+low+close)/4 price to do the experiment, it doesn't bring significant change. (0-3% added alpha difference every year) <u>To sum up, this strategy is not sensitive to the obtained price, the main reason is that I have very little exposure to the signals from high frequency intraday data, at this moment. And this strategy aims at low turnover rate and high strategy volume, thus, my holding period is long and focused on middle-term tendency.</u>

4.The Data Vendor

I use [Wind] to serve as my data source. Till now 15 Oct, 2021, I use its daily price and volume data (Daily), 3 accounting table's data (Seasonal), Wind financial indicators data (Seasonal), wind rolling consensus data (from Chinese sell side analyst) (Seasonal), the first 30 mins and the last 30 mins money flow data (Daily).

5.Performance Expectation

Here are two kinds of expectations I want to mention. <u>First is the expectation of trading performance in the Chinese A-share market, a quant product in the mutual fund industry.</u> This is a middle or low frequency strategy, only long stocks on CSI800, T+1 trading, and no any other short books. My final middle level expectation (currently, this strategy is not complete, doesn't finding alphas from all common data source) to this quant strategy is that it can trade more than 1 billion chinese yuan (without significant performance change), more than 10% added alpha every year (compared with CSI500), sharpe ratio should be between 1.5-2.5. (Maybe currently you are not a player in the Chinese A-share market, here is some info to better help you. For the hedge fund in the mainland, their CSI500 product is not selecting stocks from CSI800, they select from the 4000 stocks. For the mutual fund player, they must select stocks from CSI800, which is much harder but much safe. For these players, they can also enjoy pre-IPO benefits. pre-IPO will only need about 1-5% inventory, but bring 5%-10% income for the entire book every year! This description applies for the product's AUM less than 0.5 billion, but the majority of active quant investment mutual fund product, currently their AUM which is less than 0.5 billion.) So you can get more info, and measure my performance. Second is the expectation of the gap between backtesting and real trading. I think the history can not be fully repeated in the future, thus, there is some gap, more or less. However, in the process of feature construction and feature selection, whether we use rolling forecast is a big difference! A lot of people do rolling forecast on the strategy, but they fit all the history data onece, for the feature construction and selection. Thus, for this expectation, I hope you can go deep into the experiment setting and measure that whether my setting is decent.

6.Metrics

I have calculated the revenue before fee, after fee, daily turnover rate, max-drawdown, net sharpe ratio and etc. More specifically, I want to stress out three metrics which are frequently used in this page. 1. The accumulated alpha curve is added net return curve, which means I add up T-2, T-1 and T day's return together, it's not multiply. 2. **Accumulated Alpha (Long VS CSI500)** means that I long the top 20% stocks in each day, equal weighted, with 100% inventory. And then minus the index return of CSI 500 to show my access return. **Accumulated Alpha (Long VS Short)** is similar, I long the top 20% with 100% inventory and short the tail 20% with 100% inventory. 3.**Turnover Rate** refers to the daily turnover rate. In this page, I get used to calculate turnover rate on both side. eg. Turnover Rate 8% means that the I sell 4% and then I buy 4%.

If you find sth shown in this page is not a fair compare, or some important settings are not made clear in this page, or you want to make a double check on some important procedures, please feel free to write an e-mail to me. I will be grateful to you. E-mail: fangx18@tsinghua.org.cn

---

Section1 @Finding Alpha from Daily Price and Volume

I design a neural network based framework to construct alphas. Each deep neural network represents a powerful alpha. It has been published on the top computer science conference KDD 2020 and the well known financial journal Quantitative Finance. (You need to use scientific way to access the internet if you are in the Mainland of China...) If you want to get more info, you can read its original arXiv version.

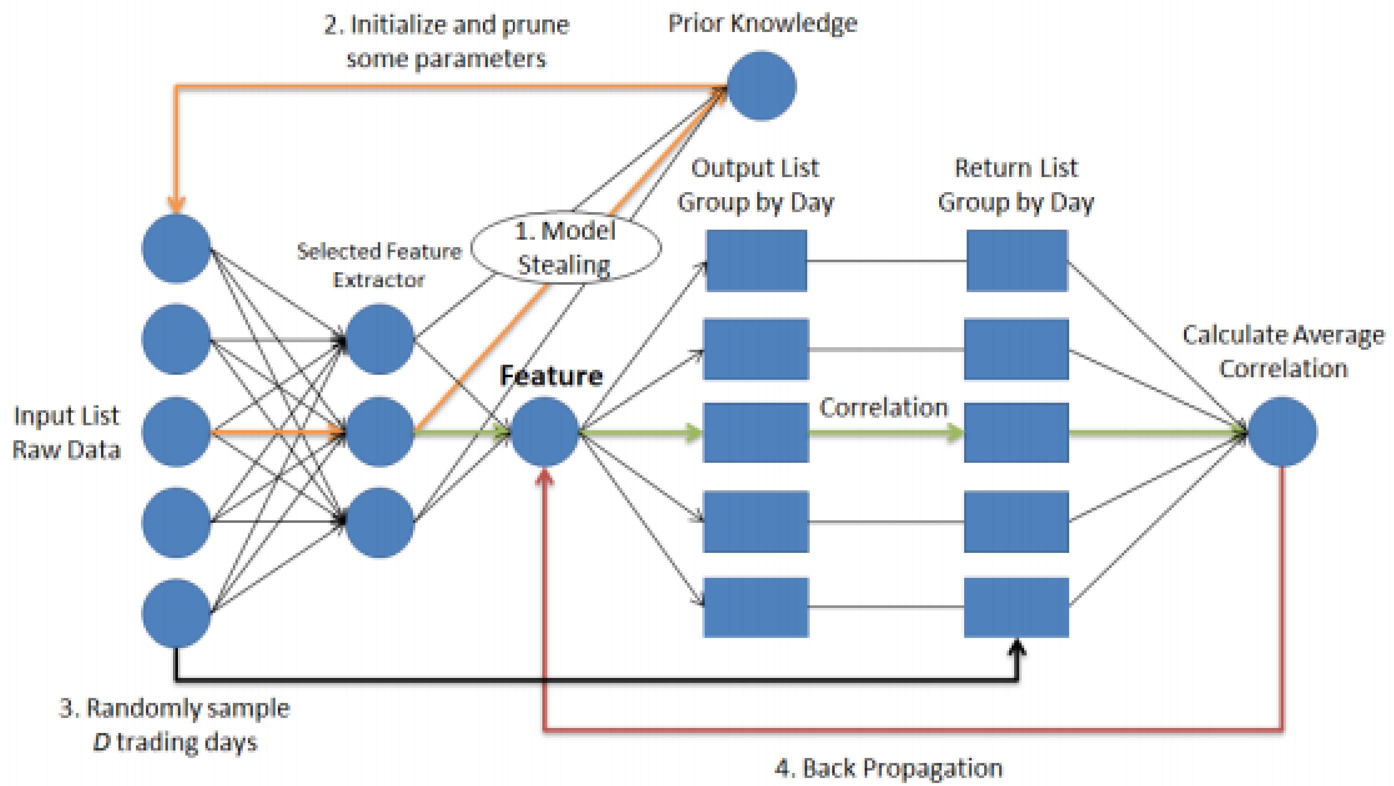

Figure 1.1 Use self-designed sampling rules, loss function and diversified method to automatically construct factors.

In my current framework, I propose a much better way to construct diversified factors. But for the loss function and basic sampling rules, I still used what I mentioned in the previous paper. Due to the NDA, I will not give more exposure about this diversified method. Deep neural network is no longer a black box, it inherits the pattern from diversified traditional factors, but for the detailed hyper-parameters and other uncertain settings, I let the neural network learn this things freely. To sum up, I settle down the pattern (computation logic), and let the neural network to learn sth uncertain but important freely. The freely learning part maybe the real meaning of aritifical intelligence. Each neural network is a single factor, and I can approximate its detailed formula.

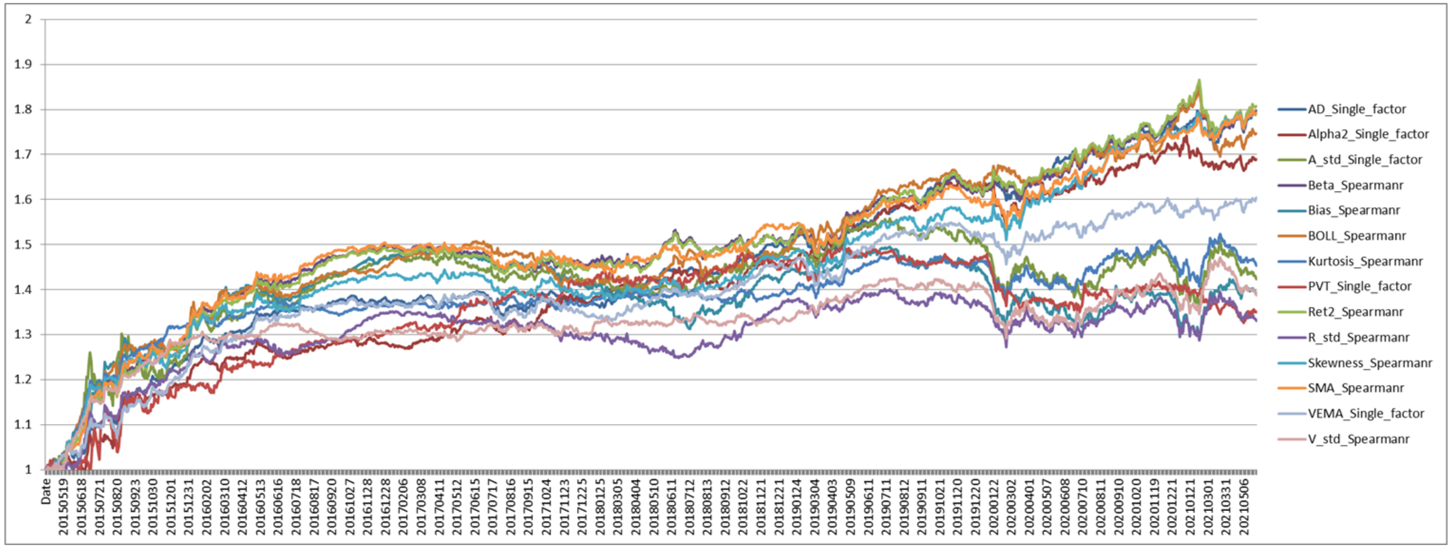

Figure 1.2 AlphaLib, alphas generated from daily price and volume, Value is Accumulated Alpha (Long VS CSI500). Notes: xxx_Spearmanr refers to the neural network factors, xxx_Single_factor refers to traditional factors.

Combine these technical indicators together, I can get the final alpha generated from daily price and volume. Because currently I only specialized in alpha construction,have some experience in factor combination, no previous experience in factor selection and portfolio optimization. Thus, I simply adds up all these factors together to do the experiment. I think it can better show the contributions of this alpha construction system.

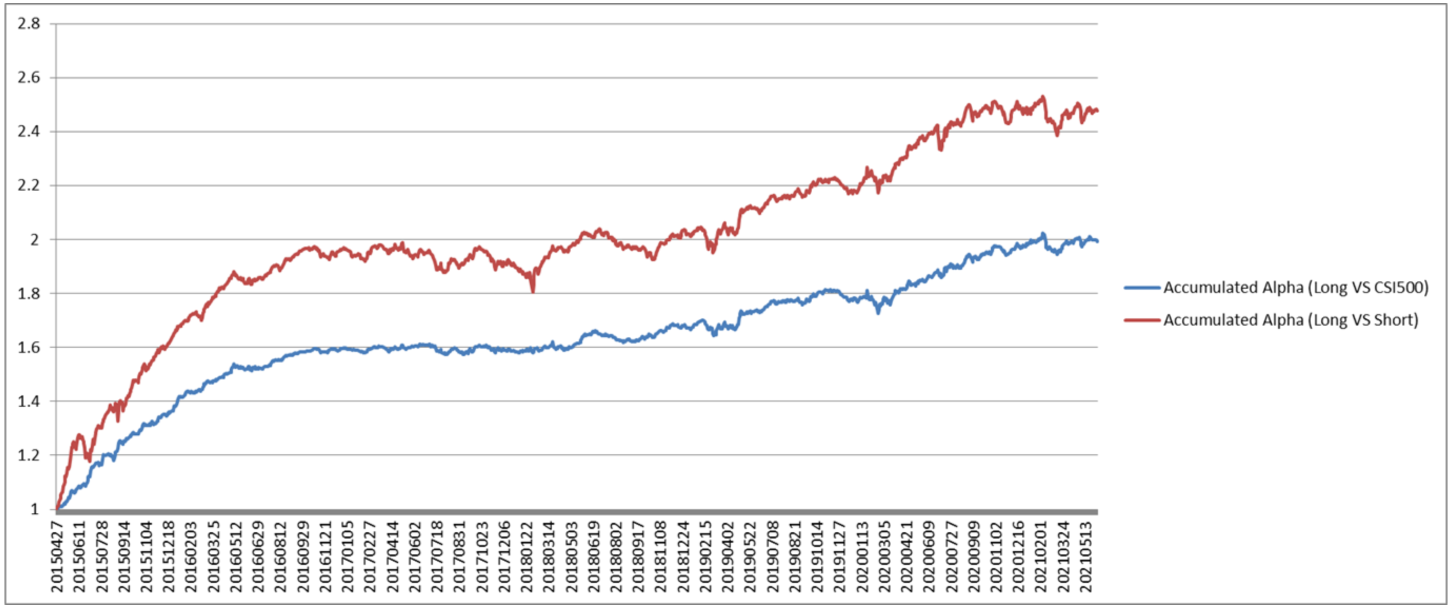

Figure 1.3 Combine all the price and volume related alphas together, into a big compounded alpha, avg daily turnover rate is 7%-12%. As for the metrics, please see Basic Setting - 6.Metrics.

Let's look at the blue curve at Figure 1.3, it perform well in the 2015 and 2016, very bad at 2017, not bad at 2018, 2019, very good at 2020. 2021 is not fully tested.

I want to talk more about this part. In 2017, this is a very tough time for Chinese A share market's price and volume data. You can earn the money easily if you use fundamental data. But look at the technical indicator, there are still some can make revenue at 2017. You may say that I am foolish, why I don't pick out them. Actually, 2017 is a big change. The technical indicators works well in 2017, which is very very weak before and after 2017. Some of them can't make money, let alone to get enough weight if I do factor selection according to the momentum. Thus, to be honest, if you don't look at the future and fully use quant method, it's hard to make daily technical alpha in 2017. To further improve this part, absolutely, using the fundamental. But if I want to stick to the daily price and volume, maybe some factor timing (more than momentum) will be a good direction. For the early 2021, MingHong (one of the biggest hedge fund in China) also faces big max-drawdown. For me, not only in price and volume, but also in my final big singal, it's the point which causes max-drawdown. But luckily, my drawdown is not as big as MingHong. I think it's mainly because I chose stocks from CSI800, but they chose from the entire market.

To sum up, the performance may looks far from perfect if I only use daily price and volume data. I think it's good enough, although there is still much and much can be done. If you interested and think your's are better, please contact, we can let all the experiment settings be the same and fairly compare.

	AD_Single_factor	Alpha2_Single_factor	Beta_Spearmanr	Bias_Spearmanr	BOLL_Spearmanr	Ret2_Spearmanr	R_std_Spearmanr	SMA_Spearmanr	V_std_Spearmanr
AD_Single_factor	1.00000	0.42426	0.74091	0.20249	0.58602	0.76625	0.16967	0.73980	0.32269
Alpha2_Single_factor	0.42426	1.00000	0.48381	0.17207	0.43652	0.47356	0.19833	0.47322	0.19686
Beta_Spearmanr	0.74091	0.48381	1.00000	0.24169	0.84001	0.95925	0.17074	0.71224	0.19774
Bias_Spearmanr	0.20249	0.17207	0.24169	1.00000	0.19111	0.20381	0.63228	0.59936	0.69997
BOLL_Spearmanr	0.58602	0.43652	0.84001	0.19111	1.00000	0.84864	0.06723	0.52240	0.10282
Ret2_Spearmanr	0.76625	0.47356	0.95925	0.20381	0.84864	1.00000	0.14775	0.72435	0.20050
R_std_Spearmanr	0.16967	0.19833	0.17074	0.63228	0.06723	0.14775	1.00000	0.53092	0.67828
SMA_Spearmanr	0.73980	0.47322	0.71224	0.59936	0.52240	0.72435	0.53092	1.00000	0.59338
V_std_Spearmanr	0.32269	0.19686	0.19774	0.69997	0.10282	0.20050	0.67828	0.59338	1.00000

	AD_Single_factor	Alpha2_Single_factor	Beta_Spearmanr	Bias_Spearmanr	BOLL_Spearmanr	Ret2_Spearmanr	R_std_Spearmanr	SMA_Spearmanr	V_std_Spearmanr
AD_Single_factor	1.00000	0.07203	0.30154	0.15500	0.47296	0.57531	0.02051	0.45836	0.28335
Alpha2_Single_factor	0.07203	1.00000	0.23267	0.05696	0.21943	0.24450	0.01791	0.18367	-0.06312
Beta_Spearmanr	0.30154	0.23267	1.00000	0.24493	0.89138	0.80375	0.01134	0.90693	-0.57179
Bias_Spearmanr	0.15500	0.05696	0.24493	1.00000	0.37113	0.33753	0.05627	0.30782	0.05524
BOLL_Spearmanr	0.47296	0.21943	0.89138	0.37113	1.00000	0.78193	0.02196	0.97083	-0.40345
Ret2_Spearmanr	0.57531	0.24450	0.80375	0.33753	0.78193	1.00000	0.02023	0.76594	-0.03420
R_std_Spearmanr	0.02051	0.01791	0.01134	0.05627	0.02196	0.02023	1.00000	0.01765	0.00972
SMA_Spearmanr	0.45836	0.18367	0.90693	0.30782	0.97083	0.76594	0.01765	1.00000	-0.43196
V_std_Spearmanr	0.28335	-0.06312	-0.57179	0.05524	-0.40345	-0.03420	0.00972	-0.43196	1.00000

Figure 1.4 The spearmanr correlation of different factor value (up), and the spearmanr correlation of same day's long short return (down), measured as average value of the testing sets.

As we can see from Figure 1.2 and Figure 1.3, simply add up all the technical indicators can make its overall performance better. Because I have calcualte the covariance metrics of these factors, both the factor value and daily long short return's covariance matrix. Their correlation ranges from 0.5-0.9, shown in Figure 1.4. Now, we have got a big compounded alpha generated from daily price and volume.(I want to make a difference between daily price and intraday price, because their function and conclusion is very different). Because I have very little experience in risk control and portfolio optimization. Thus, I do some simple experiment. First, I did a industry balance, let the portfolio's industry exposure 100% same at the CSI500's industry exposure. Second, I did the marke value balance, use market value to do linear regression with my signal, and keep the residual. Third, due to some contracts' strict requirement, I let 90% of inventorr put into CSI500, and the other 10% inventory goes into CSI300, which is index balance.

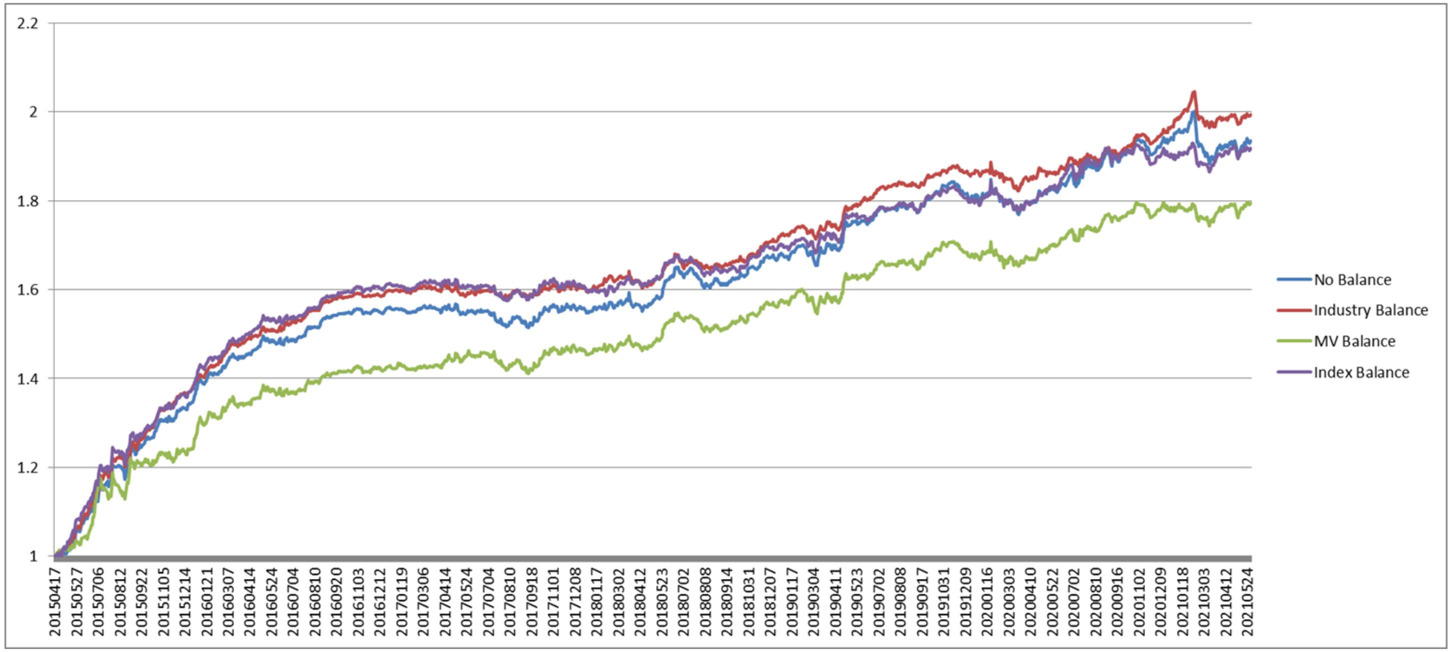

Figure 1.5 Control some risks in the big compounded alpha.

I have realized the importance of this part. First it can help me to conrtol some risks. Second, after get rid of some risk, I can measure whether my alpha is pure or not. For future improvement, I will learn the Chinese Barra risk model, CNE5.

Here is a short end for my daily price and volume factors. In the following part, I will give brief introduction to the performance of other factors.

Section2 @Finding Alpha from Seasonal Fundamental Data

The seasonal fundamental data can be divided into three groups, which are growth, quality and value. I didn't just use the raw data, but construct some financial indicators, according to some reprensentative sell side equity report.

For some of these factors, I direclty combine them by using neural network. For some of them, I get rid of the size and industry, and then keep the residual. And more importantly, factor timing is very important in fundamental data. To sum up, the data minning in fundamental data is limited. But it doesn't mean that we can do nothing. <u>So far, there are three types of operators, I think, which is useful for fundamental data. 1. Linear adds up. 2.Do regression, get the residual. 3.Timing different fundamental factors, and then</u>

combine it. I am still looking for more reasonable operators, I think the current AI can't live without the domain knowledge, which is the human knowledge. After all, the stock market is played by human. Here is only the backtesting result, I don't want to exposre more details according to the NDA.

One kind of factors can be looked like this:

	x	y	Weight	Return（过去六年滚动预测，6年费前总超额收益，累加）	Turnover（每天的总换手率，含买卖）	Return（过去六年滚动预测，6年费后总超额收益，累加）
0	Current, Operate_imcone_revenue, NCFP, YOY_profit_raw, Sales_G_t, Gross_profit_margin_raw, Net_profit_raw, Price_dps_raw, Debt_assets_raw, EP	EP	-0.004, -0.004, -0.005, -0.021, 0.04, 0.046, -0.05, -0.091, -0.167, 0.184	0.409365049	0.033434901	0.309060347
1	OCF_sales, Profit_G_t, Q_yoysales_raw, OCFTOOR_raw, NCFP, FCFF_raw, Asset_turn_raw, PB_raw, Gross_profit_margin_raw, Operate_imcone_revenue	Operate_imcone_revenue	-0.005, -0.012, 0.02, -0.043, 0.045, 0.076, 0.088, -0.092, 0.093, 0.263	0.539139713	0.039200475	0.421538287
2	T_yoysales_raw, PEG, OCF_sales, Total_assets_raw, Total_debt_raw, Operate_imcone_revenue, Profit_G_q, PE_raw, EP_cut, ROA	ROA_raw	-0.002, 0.003, -0.005, -0.007, -0.007, -0.022, -0.061, -0.069, 0.187, 0.342	0.446041821	0.036974285	0.335118967
3	Market_value_raw, Net_assets_raw, YOY_profit_raw, PEG, PB_raw, PE_raw, Profit_revenue_raw, FCFF_raw, Net_profit_raw, ROE_G_t	T_yoyprofit_raw	-0.013, 0.015, -0.02, 0.029, -0.042, -0.056, 0.059, 0.059, 0.08, 0.083	0.747666907	0.033888944	0.646000075
4	PEG, OCFTOOR1_raw, Deducted_profit_raw, ROE_G_t, Total_debt_raw, SP, PE_raw, EP, ROA_raw, ROIC	ROIC	0.0, -0.002, -0.008, -0.012, -0.036, 0.038, -0.049, 0.15, 0.184, 0.275	0.446609794	0.03712532	0.335233835
5	Sales_G_t, Deducted_profit_raw, PEG, Gross_profit_margin, ROE_raw, ROE, Operate_income_profit_raw, Q_yoyprofit_raw, ROA_raw, Debt_assets_raw	EP_cut	0.001, 0.003, -0.005, 0.032, -0.05, -0.05, -0.074, -0.087, 0.103, -0.139	0.441613654	0.039365179	0.323518117
6	DP_p, OCF_sales, OCFTOOR1_raw, Net_assets_raw, Deducted_profit_raw, Current, NCFTTM_raw, NCFP, YOY_profit_raw, ROA_raw	Operate_income_profit_raw	-0.012, -0.013, -0.019, 0.028, 0.033, -0.04, -0.042, 0.073, -0.082, 0.217	0.332787698	0.03450096	0.229284818
7	Current_raw, Q_yoysales_raw, Sales_G_q, T_yoysales_raw, OCFTTM_raw, Profit_revenue_raw, FCFF_raw, Current_debt_debt, Price_dps_raw, Deducted_profit_raw	Deducted_profit_raw	0.002, 0.014, 0.014, 0.017, -0.042, 0.065, 0.072, 0.1, -0.199, 0.268	0.488940059	0.032530906	0.391347343
8	OCFTOOR1_raw, Market_value_raw, T_yoysales_raw, Gross_profit_margin, Current_debt_debt, OCFTTM_raw, Asset_turn_raw, FCFF_raw, PB_raw, Price_dps_raw	Net_profit_raw	0.001, -0.001, 0.031, 0.034, 0.038, -0.046, 0.053, 0.058, -0.062, -0.132	0.665541338	0.038624535	0.549667734
9	Gross_profit_margin, Net_profit_raw, T_yoysales_raw, ROE_G_t, T_yoyroe_raw, FCFF_raw, ROIC, ROA, EP_cut, ROE	ROE	0.005, -0.01, 0.011, -0.018, -0.018, 0.024, 0.029, 0.144, 0.181, 0.378	0.500322148	0.040404471	0.379108734
10	NCFTTM_raw, Gross_profit_margin_raw, NCFP, Profit_G_t, Market_value_raw, SP, ROA_raw, ROE_raw, Debt_assets_raw, PB_raw	DP_p	-0.021, 0.034, 0.041, -0.055, 0.064, 0.076, 0.077, 0.077, -0.096, -0.098	0.576806669	0.037318627	0.464850789
11	OCF_ratio, T_yoyprofit_raw, Profit_G_t, ROE_G_t, Sales_G_q, FCFP, PE_raw, Debt_assets_raw, Asset_turn_raw, FCFF_raw	OCFTOOR_raw	0.014, -0.017, -0.017, 0.036, 0.041, 0.065, -0.098, -0.104, 0.122, 0.128	0.186945582	0.040363208	0.065855958
12	OCF_ratio, PEG, Operate_imcone_revenue, ROIC_raw, Profit_G_t, Current_debt_debt, PE_raw, ROA, EP, Total_assets_raw	Total_assets_raw	-0.007, -0.009, -0.013, -0.018, -0.027, 0.035, -0.053, 0.089, 0.188, 0.31	0.512755808	0.034801883	0.408350159
13	OCFTOOR1_raw, PEG, NCFTTM_raw, Gross_profit_margin, T_yoysales_raw, ROE, SP, Debt_assets_raw, EP, T_yoyprofit_raw	T_yoyprofit_raw	0.003, 0.006, -0.011, 0.019, 0.02, 0.047, 0.065, -0.123, 0.16, 0.311	0.439824959	0.036527296	0.330243071
14	Current_debt_debt, ROE_G_t, Profit_revenue_raw, Asset_turn, Asset_multiply_raw, FCFF_raw, Debt_assets_raw, Price_dps_raw, SP, ROA_raw	Gross_profit_margin	0.002, -0.009, 0.026, -0.035, 0.046, 0.059, -0.11, -0.144, 0.147, 0.166	0.285042306	0.03991191	0.165306574
15	Deducted_profit_raw, Asset_turn, OCFTOOR1_raw, Net_assets_raw, FCFP, NCFP, Current_debt_debt_raw, ROIC, Price_dps_raw, EP	ROE_G_t	-0.0, 0.006, -0.006, -0.007, -0.022, -0.029, 0.043, 0.072, -0.13, 0.23	0.637327443	0.034127746	0.534944205
16	Current, NCFP, Asset_turn, Gross_profit_margin_raw, ROE_G_t, OCFTTM_raw, FCFF_raw, Asset_multiply_raw, Price_dps_raw, EP_cut	ROE	-0.001, 0.006, 0.023, 0.031, -0.051, -0.058, 0.069, -0.08, -0.159, 0.232	0.408777355	0.040051129	0.288623967
17	Current, Asset_turn_raw, NCFP, Profit_G_t, NCFTTM_raw, Total_assets_raw, ROE, ROA_raw, ROA, EP	OCFTOOR_raw	-0.013, -0.016, -0.031, -0.032, -0.035, -0.037, -0.047, 0.088, 0.088, 0.235	0.572113494	0.034404136	0.468901085
18	PEG, Current_debt_debt_raw, Q_yoysales_raw, Operate_income_profit_raw, DP_p, Revenue_raw, OCFTOOR_raw, Deducted_profit_raw, SP, ROE_raw	FCFP	0.002, 0.007, -0.011, -0.011, -0.023, -0.051, -0.07, 0.074, 0.114, 0.153	0.277961566	0.040603884	0.156149914

Figure 2.1 One kind of fundamental data's alpha lib.

Here are 3 types of big compounded fundamental alphas, each one consists of 20-300 small factors.

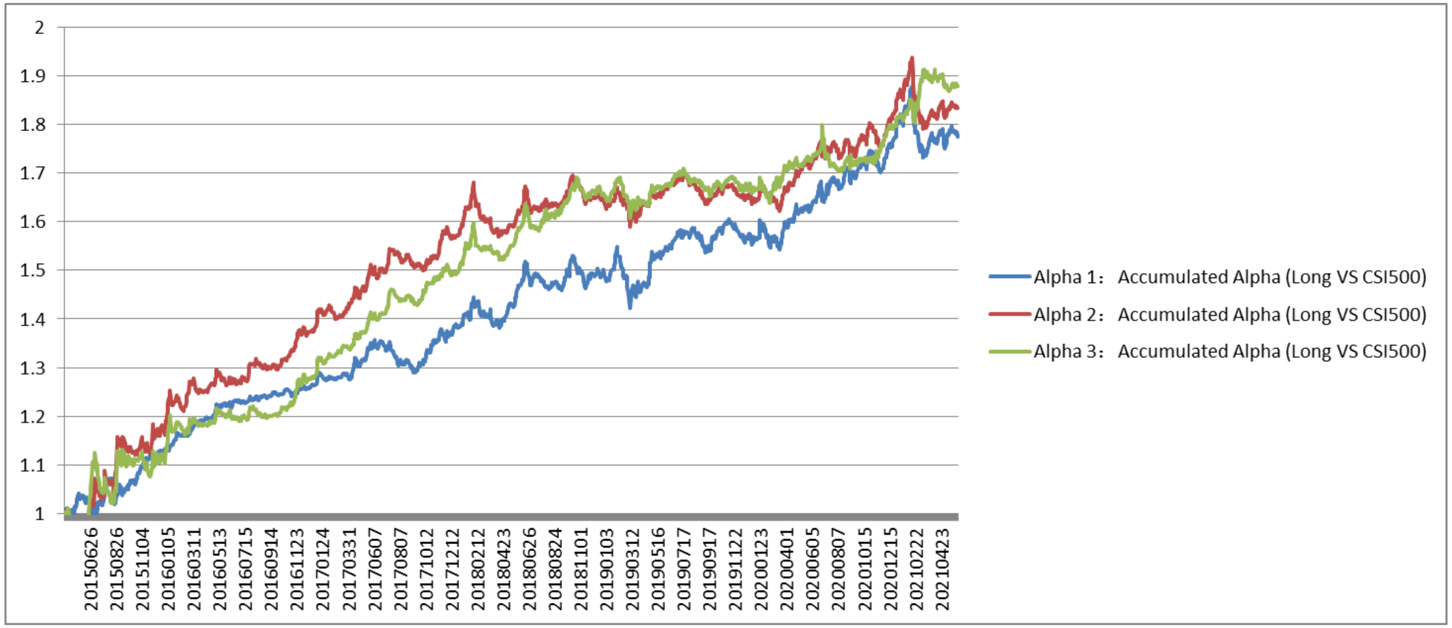

Figure 2.2 Seasonal_fundamental_data.

The correlation of technical indicator and fundamental factors is very low. They ranges from -0.1 to 0.3, in my experiment. Fundamental data is very important to middle and low frequency strategies, because they are very helpful for enlarging the size of our portfolio and provide extra info to my current alpha lib. However, its drawback is also clear. The alpha from fundamental data has larger volatility, maxdrawdown, and their immediate booming performance is not as big as technical indicators.

This is a short end for the structured fundamental data from Wind. For further research, I will use the trained NLP model like BERT and GPT-3, combined with few shot learning, semi-supervised learning, unsupervised learning, to find alphas in unstructured data. What's more, currently I do factor timing operation for this fundamental data, but the input of this sub algorithm is still not fully researched. I think I can make it better.

Section3 @Finding Alpha from Daily Money Flow Data

Money Flow is one kind of intraday price and volume data. It tells how much money and orders come in and come out during certain period of time. Wind gives the first 30 mins and last 30 mins money flow data, and big order, middle order, small order's flow data. They are very helpful, however, the data missing problem is a little annoying. Btw, this is not level2 data, they only give a total value in the first 30 mins but not give each mins' data. I think things can be much improved with level 2 data, especially the order book data. Previously, I have experience in high frequency market making strategies.

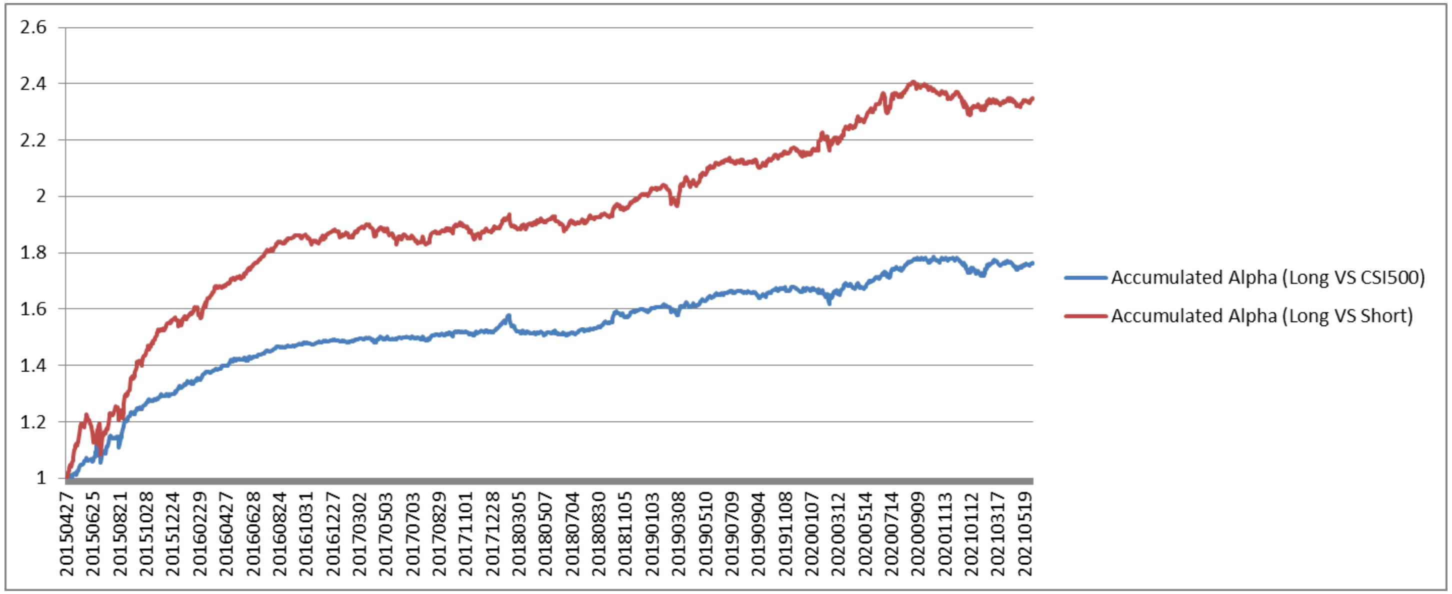

Figure 3.1 Big compounded alpha generated from money flow data, avg daily turnover rate is 15%-25%.

As we can see above, the money flow is intraday data, it can bring in a lot of new alpha but with high turnover rate. If I directly use it, is turnover can be as high as 25%, if I penalized the inventory change, it can be lowered down without significantly hurting the revenue. I keep open mind to this kind of high frequency, high turnover alpha source. Because after certain threshold, the revenue must come from trading's turnover. It's okay to let this kind of factors to join in, and at the end, the overall turnover rate is not that high. Because there are much low frequency factors, and even if two factors have same turnover rate, their combination's volatility and turnover will decrease with high possibility.

For further research, the level 2 data can make it better.

Section4 @Final Signal (No portfolio optimization and risk control yet)

Okay, here let's put all the factors together. Each factors is a big compounded alpha, which consists of 15-300 small factors.

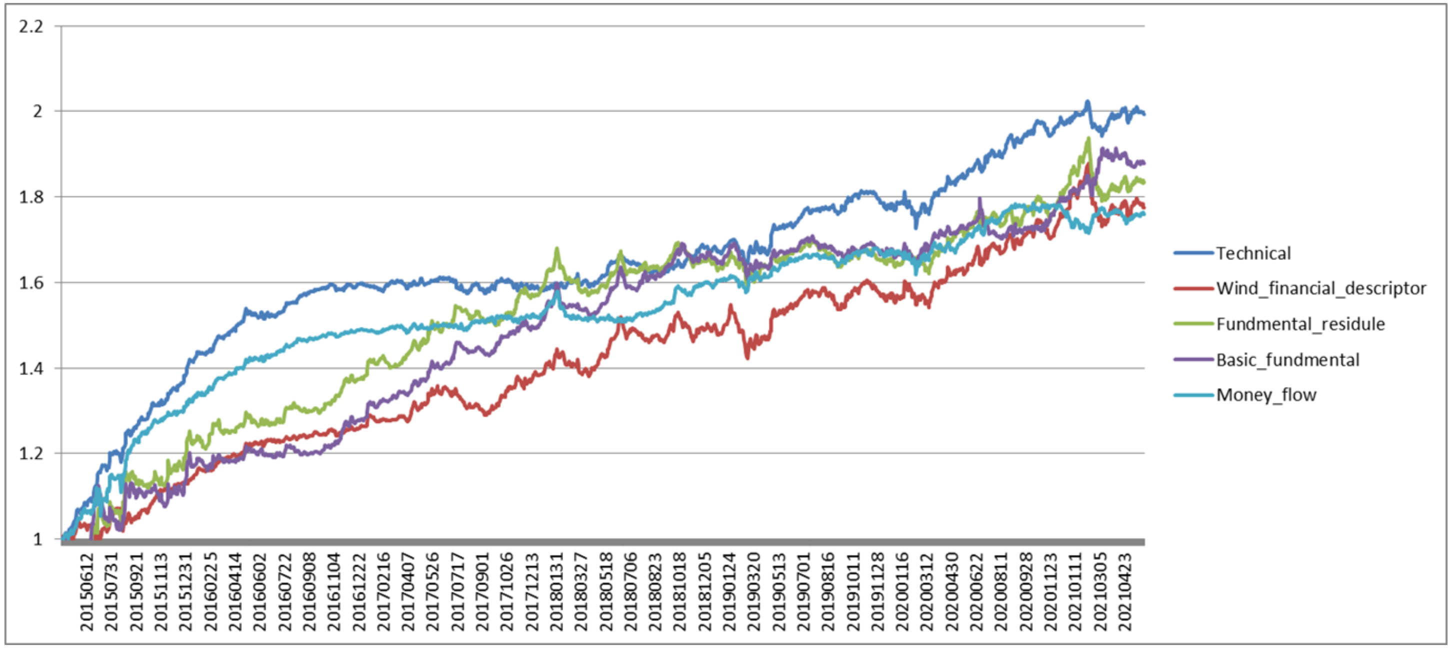

Figure 4.1 Final alpha curve, the value shown in this picture is **Accumulated Alpha (Long VS CSI500)**, avg daily turnover rate is 10%-13%.

Shown in Figure 4.1, during different peirod of time, there is a main actor. In the Chinese A share market, 2017 is hard to technical indicators, but fundamental factors can take care of it. In the late 2018 and beginning of 2019, fundamental data is suffering, maybe the Trade War lauched by Trump. But at this time, the technical indicator can pay back. In 2020, they all the shinning star, but in my lib, technical indicator is more powerful and easy to attract money's attention.

Combine all these compounded alphas together into the final signal. I simply adds up all the factors together, equally weighted. Here are two things I tried. First, the factor timing neural network I designed for fundamental data doesn't bring significant improve. I need do more experiment. I suspect that the input data is the key. The thing can timing the fundamental data may far from enough to timing the technical data, due the frequency and info flow. Second, I have tried use simle neural network to ocmbine them together. For this method, its can improve but still not impress me. If the performance can not impress me, I think the linear regression is much better, in this case. Still, I need more experience in combination and portfolio optimization.

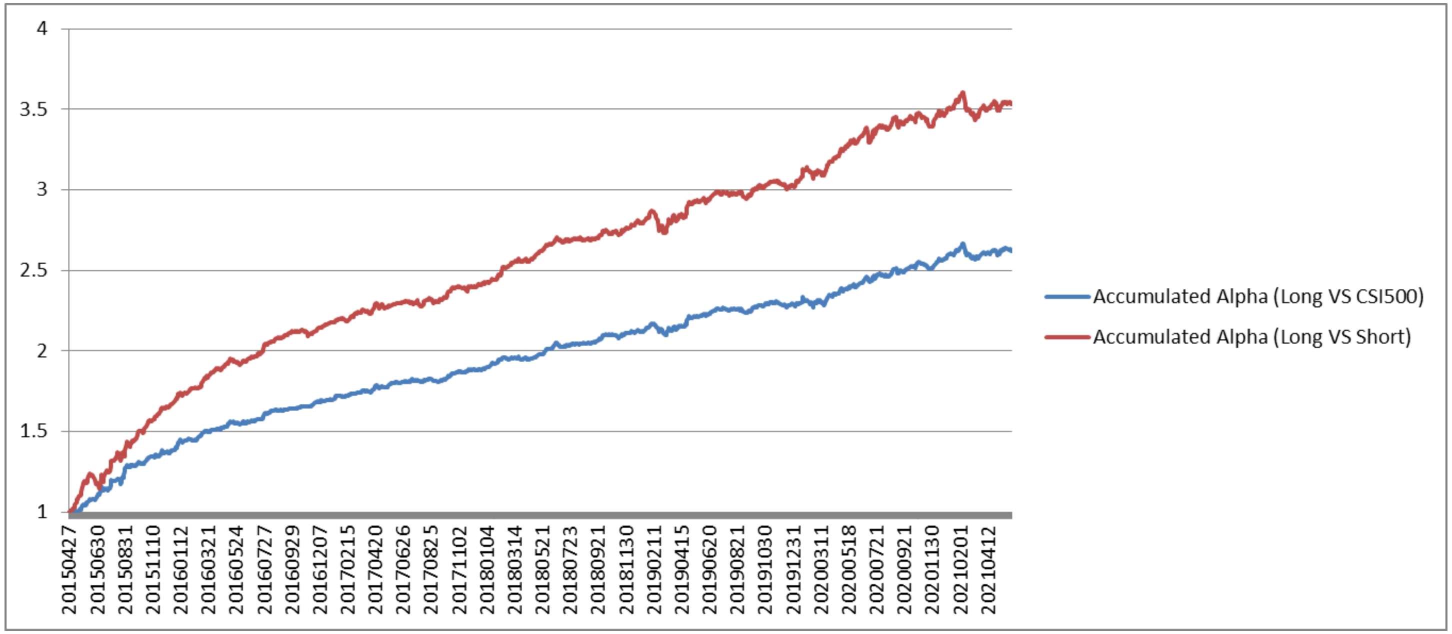

Figure 4.2 The added alpha curve. (+)

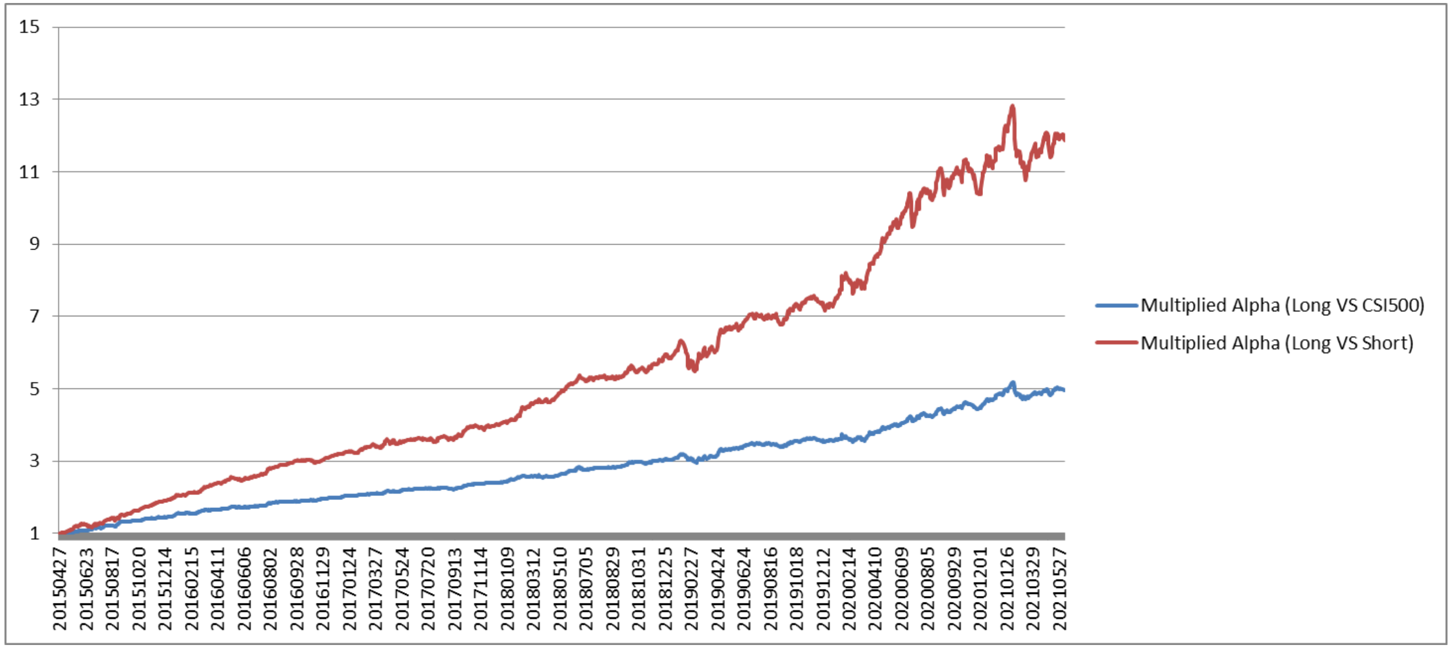

Figure 4.3 The multiply alpha curve. (*)

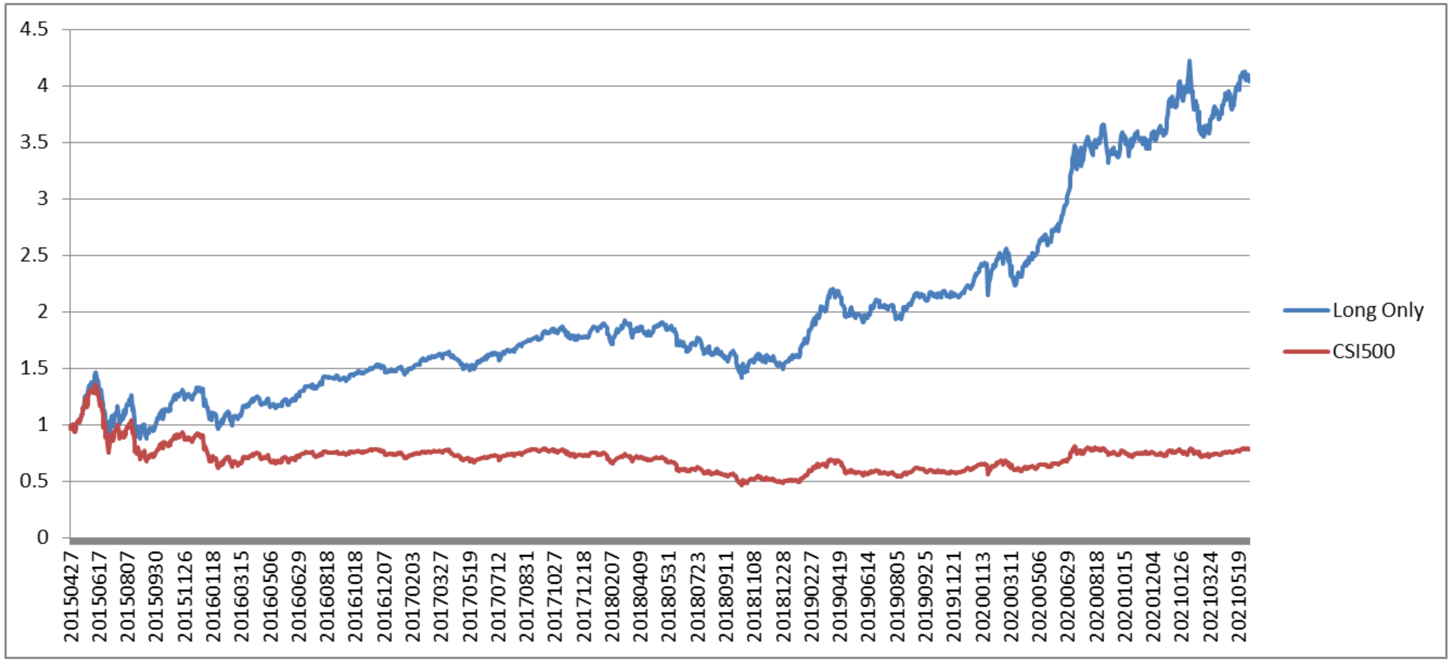

Figure 4.4 The multiply net value curve. (*)

Here is the backtest table:

Table 4.1 Backtest Result Table.

	A	B	C	D	E	F	G	H
1		Total Return (Added)	Net Return (Added)	Volatility	Net Sharpe	Net Maxdrawdown	Daily Total Turnover rate	Net Calmer
2	2015	0.402097307	0.353665049	0.100252292	3.527750255	-0.036815928	0.142447818	9.606305365
3	2016	0.296072226	0.222906345	0.060240315	3.700285195	-0.021874154	0.149930084	10.19039829
4	2017	0.193472745	0.122299668	0.054197902	2.256538785	-0.023882862	0.145846469	5.120812804
5	2018	0.233849953	0.165324884	0.059786788	2.765241084	-0.027108563	0.140998085	6.09862211
6	2019	0.166227315	0.104566657	0.083193396	1.256910555	-0.074421763	0.126353807	1.405054826
7	2020	0.282642207	0.232925226	0.106388952	2.189374192	-0.061011904	0.102298315	3.8177013
8	2021(1-6)	0.046730036	0.024491042	0.073489159	0.333260613	-0.097349048	0.10795628	0.251579677
9	Total	1.621091789	1.226178871	0.210288234	5.830943783	-0.097349048	0.132432233	12.59569449

In the past 6 years, the max-drawdown happenes in Feb 2021. But at 2021 Q2, I earn all the loss back and win new money. So far, I know that 2021 Q3 is easy for finding alpha but 2021 Q4 is another big pit. Later I will find a time to update my data and show the latest backtesting result.

To deal with this max-drawdown, I have several ideas. Listed according to the priority.

1. I have designed experiment (this idea is from a very senior colleague specialized in trading), what if I construct a single factor by using the data of 2021 Q1 and Q2, and then add this factor into the rolling forecast system. I found that before 2020, add this new factor can bring significant imrpove, in 2020 don't add this factor is a wise decision. Thus, I suspect that the training set 300 trading days is the main cause of this maxdrawdown. 2021 and 2018,2019 maybe similar, but 2021 is very different from 2020! Maybe, I should repeat and enlarge my alpha libs, use 150/300/600 as training length! But for this moment I am crying for decent infrastructure support...64G CPU, 20G GPU (the more the better), 32 Cores will be temporarily enough...

2. Shown in Table 4.1, this maxdrawdown is very distinct, compared with the max-drawdown in the other years. Thus, I can calculate the value at risk, and design a stop lossing strategy. If the loss is bigger than VaR, then I follow the index for N days. Of course, N and VaR are important hyper-parameters, I need do sensitivity analysis, if they are not robust on some para-space, this method will be regarded as failure.

3. Common barra risk control may be helpful, but I am not sure whether it will hurt the revenue, and how much?

These methods are so far I can think, if possible, I hope to get more idea.

@Last Update, 19 Oct 2021.

Section5 @Some Big Changes*

1.About the CSI800 Strategy's Big Max-drawdown Problem

I do the experiment mentioned above, which tries to enlarge the trainset. The results are very interesting. (a) In the past 6 years, if we enlarge the trainset from 200 days to 300 days, the overall revenue is the same. It shows that enlarging the train length will not make reasonable improvements. Somewehere we win, and we will lose in the other place. (b) Besides, if we focused on the Feb 2021, enlarging the train length will help a lot. But dose this move really comes from strength but not the lucky? Actually we add any other period of time except 2020, will also help a lot. (c) Thus, I draw a very personal conclusion, aboslutely this is not for sure, welcome idea sharing. I think enlarging the trainset is not a very reasonable method, because different periods have its own patterns. Thus, enlarging the trainset too much may even bring short-comings, it may learn average line and can't learn specific patterns. Thus, I think the big pits in Feb 2021 shows the drawback of quant stragegy. <u>We should overcome it via risk conrtol, or some risk control strategy that can trace the other competitors' risk control degree. And the timing strategy may be another solutions.</u>

2.Backtesting the Strategy Together with a very Senior PM, Get New Directions for Further Improvement

Previously, I only do the backtesting according to my own code. Although the alpha generation part has been tested by other senior people during my previous internship. Now, I got the help of a very senior portfolio manager who manages more than 1B Chinese Yuan ,to do backtest for my strategy. If he use the top 20% in CSI800, his backtest result almost have no difference with mine. However, for a normal CSI500 Enhanced Index Strategy, there are many other constraints, such as 82% inventory should be in CSI500, this is required by the mutual fund product, the equity holding should be lower than 93%, and reasonable tracking error should be considered. I didn't take this constraints into account, thus, if he runs my strategy under the strict product requirement, the results are quite different. And the differences mainly come from the constraints that I need to buy at least 82% CSI500 all the time. In 2018 and 2019, this constraint make me loss a lot alpha. Thus, I begin to rethink my alpha. Maybe my previous alpha is not very pure. Sometimes it makes money from betting the relative strength of CSI300 and CSI500.

3*.Make Sure the Specific Flavor of My CSI500 Enhanced Index Strategy, and Meet all the Product's Requirements (Mutual Fund)

Currently, I am still at mutual fund. There are some special characteristics about the Chinese mutual fund business. The trading environment is much worse than the hedge fund, which encourages the mutual fund to develope low turnover rate but large volume strategies. Normally hedge fund's cost is 20bps on sell side, but mutual fund's cost can range from 20bps-40bps on both side. 【Thus, I will penalize the turnover rate harder and make sure the specific flavor of my Enhanced CSI500 Index Quant Strategy, super low turnover rate, mainly based on price and volume, stable and only comes from csi500, high strategy volume.】.

A Ready CSI500 Enhanced Index Strategy

In this strategy, I rethink my ways to construct alphas more thoroughly, make some improvements on my self-designed 20 neural networks. But I should mentioned that I only slightly change the networks designed for price and volume data. For the fundamental factors and consensus factors, so far, I didn't make some change. And I strictly **select all the stocks from CSI500**, no matter the training process nor the testing process. Here are some results about the nearly-ready CSI500 Enhanced Index Strategy.

Here is the factors constructed from price and volume. Although I still use the traditional name, eg. The SMA factors are normally constructed from close price, but now, I apply SMA' pattern into a lot of technical indicators' time series, thus, this is a compounded alpha.

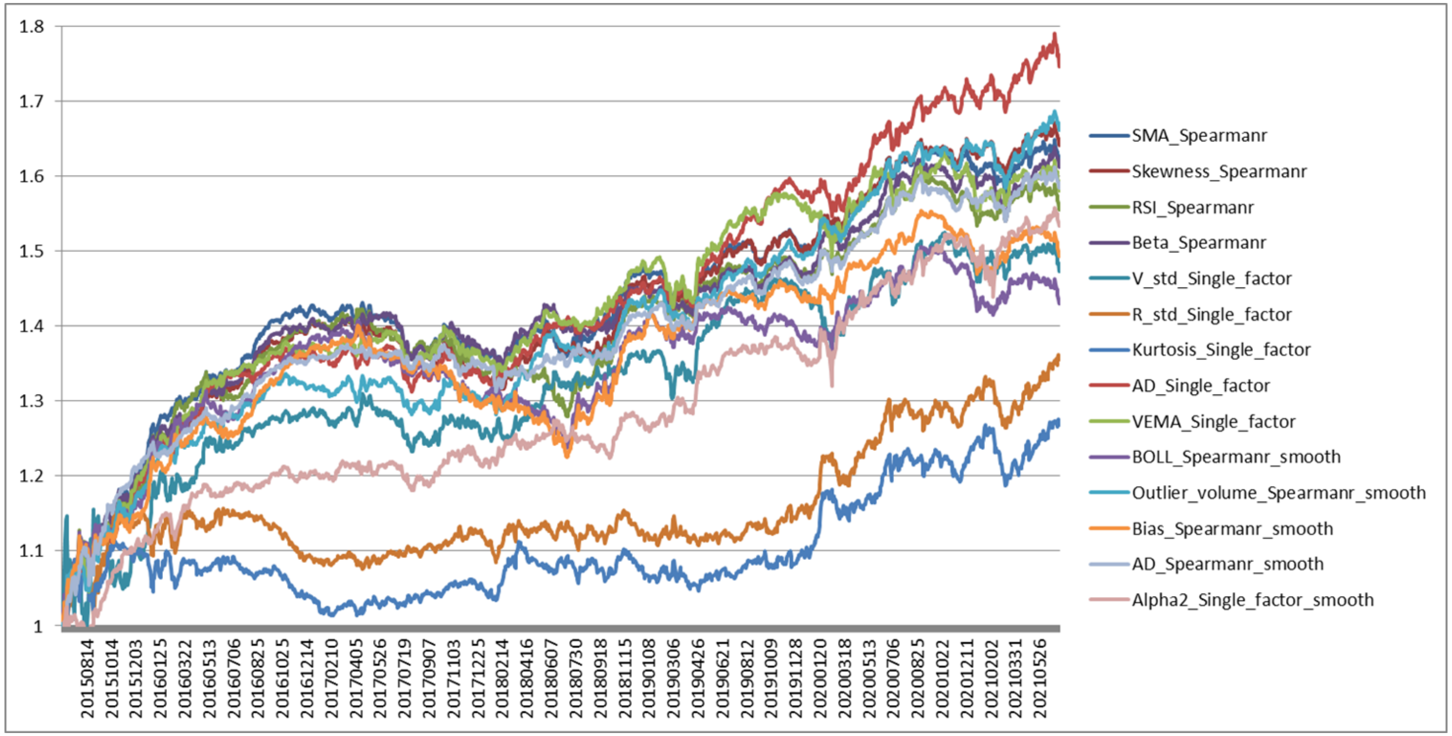

Figure 5.1 All factors constructed from daily price and volume.

Here is the final signal from daily price and volume data. Long 20% stocks in CSI500, with equal weight, the curve below is the added alpha, before fee.

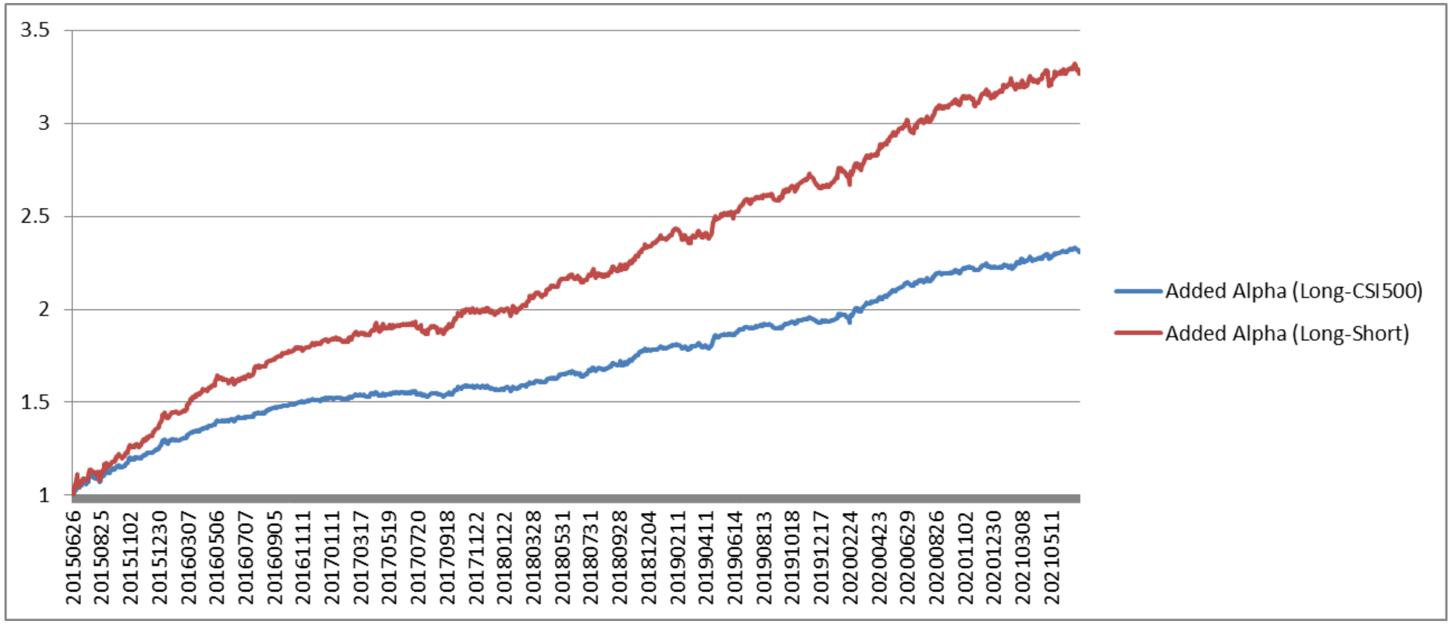

Figure 5.2 Final signal from daily price and volume data.

As we can see from the figure above, in 2017, we still have no sufficient alphas. I have explained this reason in the previous paragraph. In 2017, there is a very special pattern, which 10% stocks go up and 90% stocks go down. This pattern is so special that few technical indicators can work in this year. However, for the technical indicators work well in this year performed very baddly in the other year. Thus, we still need to construct factors from more data sources. In 2017, fundamental data can easily construct some alpha.

Here is the final signal from all data. <u>Long 20% stocks in CSI500, with equal weight, the curve below is the added alpha, before fee.</u>

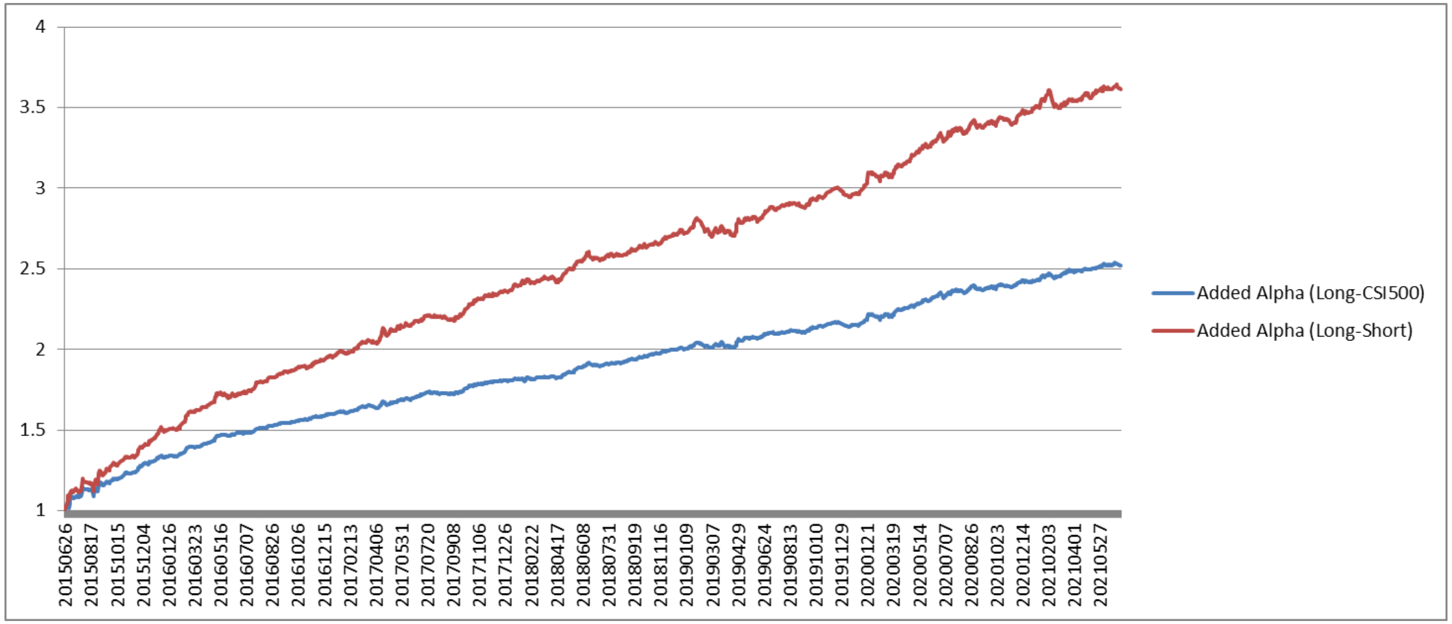

Figure 5.3 Final signal from all data.

Compared figure 5.2 and figure 5.3, the total alpha didn't become larger, because my technical factors are strong and diversified, but my fundamental factors are not so strong. But the curve at figure 5.3 is more stable and smooth, it has alpha in 2017. Thus, I still think the involvement of more diversified data souce and pattern is very helpful.

Here is the final signal from all data. Long 20% stocks in CSI500, with equal weight, the curve below is the added alpha, before fee.

Table 5.1 Final signal from all data, summary of the backtesting result.

	Total Alpha (Long20%-CSI500)	Net Alpha (Long20%-CSI500)	Alpha Volatility	Annual Sharpe (VS CSI500)	Maxdrawdown (Net Alpha)	Daily Turnover rate (Buy+Sell)
2015(6-12)	30.86%	29.17%	0.82%	4.383610117	-5.03%	6.6%
2016	27.95%	24.93%	0.28%	5.833579762	-1.47%	6.19%
2017	20.97%	17.57%	0.31%	3.661653421	-2.22%	6.97%
2018	20.29%	17.78%	0.31%	3.77099389	-2.15%	5.17%
2019	14.34%	11.28%	0.38%	1.90716311	-3.6%	6.27%
2020	26.22%	23.16%	0.55%	2.737273157	-3.81%	6.28%
2021(1-6)	9.48%	7.73%	0.42%	2.327491278	-2.81%	7.08%
Average	150.74%	132.25%	0.44%	3.517394962	-5.03%	6.29%

I summarized the results, shown in the table above. All the indicators except the Total Alpha, are net value, with the deduction of transaction fees, 20bps on buy side and 20bps on sell side.

And the backtesting results match the senior PM's own backtesting system, here are the results based on different conditions, from my PM.

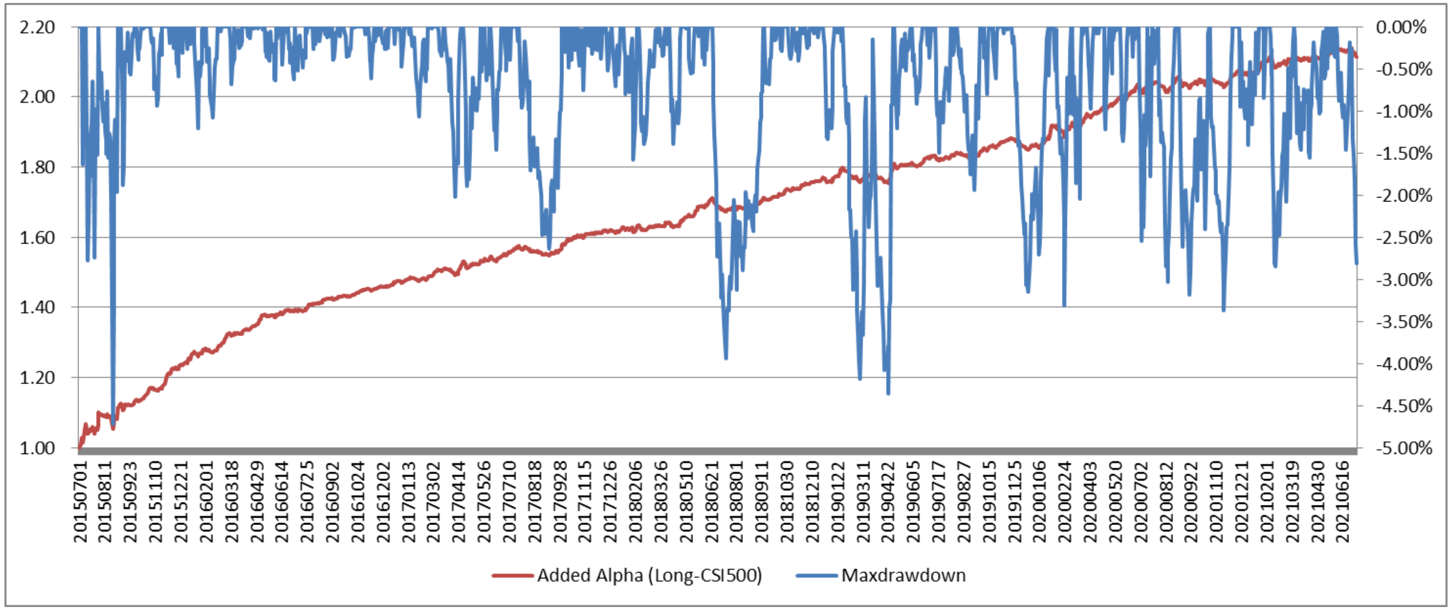

Figure 5.4 top 20% with equal weight, 20bps transaction fees on both side.

My code has been checked, and I am required to do the backtesting by loading part of the data, and then save my portfolio to a folder. And then rolling to the next month's data. The PM's backtesting results match my backtesing result.

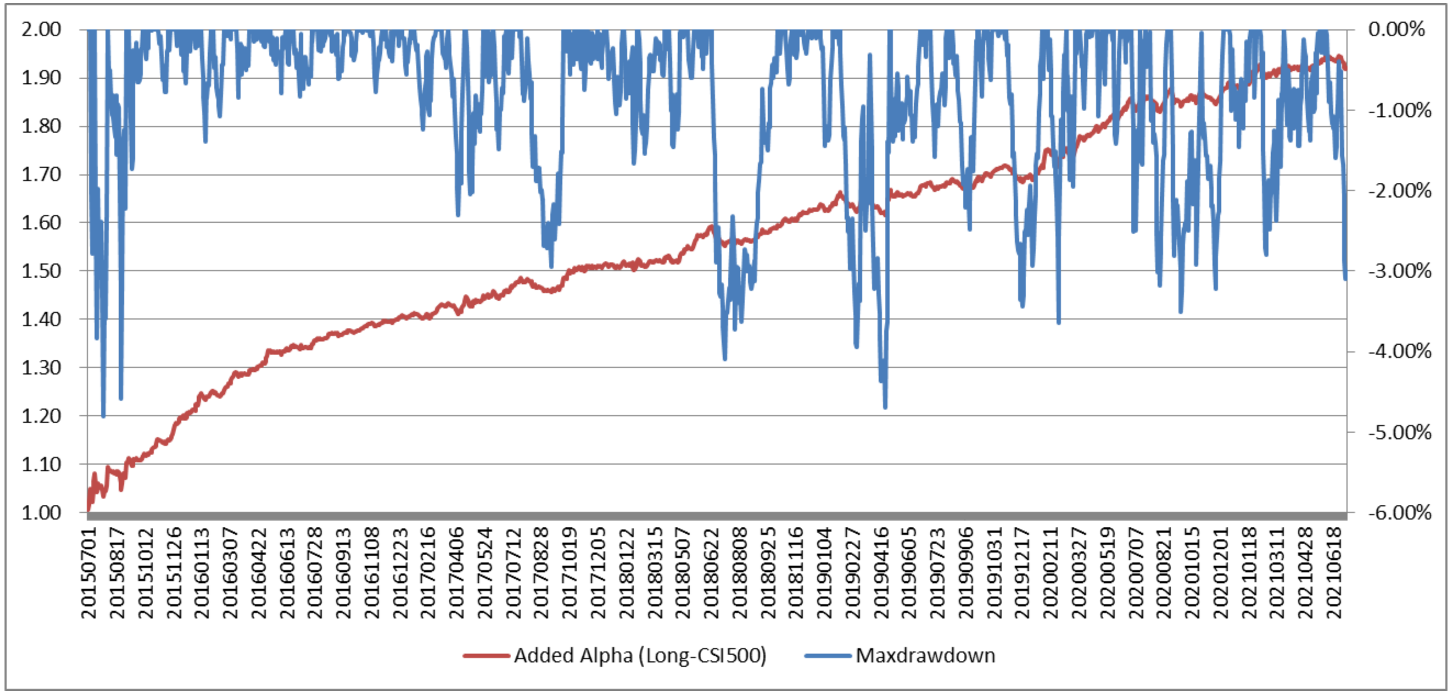

Figure 5.5 top 20% with equal weight, 40bps transaction fees on both side.

40bps transaction fees on buy and 40bps transaction fees on sell is a very very strict condition to test the robustness of the quant strategies. Especially for the strategies which mainly relies on price and volume data. But this test is needed in the mutual fund industry, and needed to be test if I want to label my strategy as a high volume strategy.

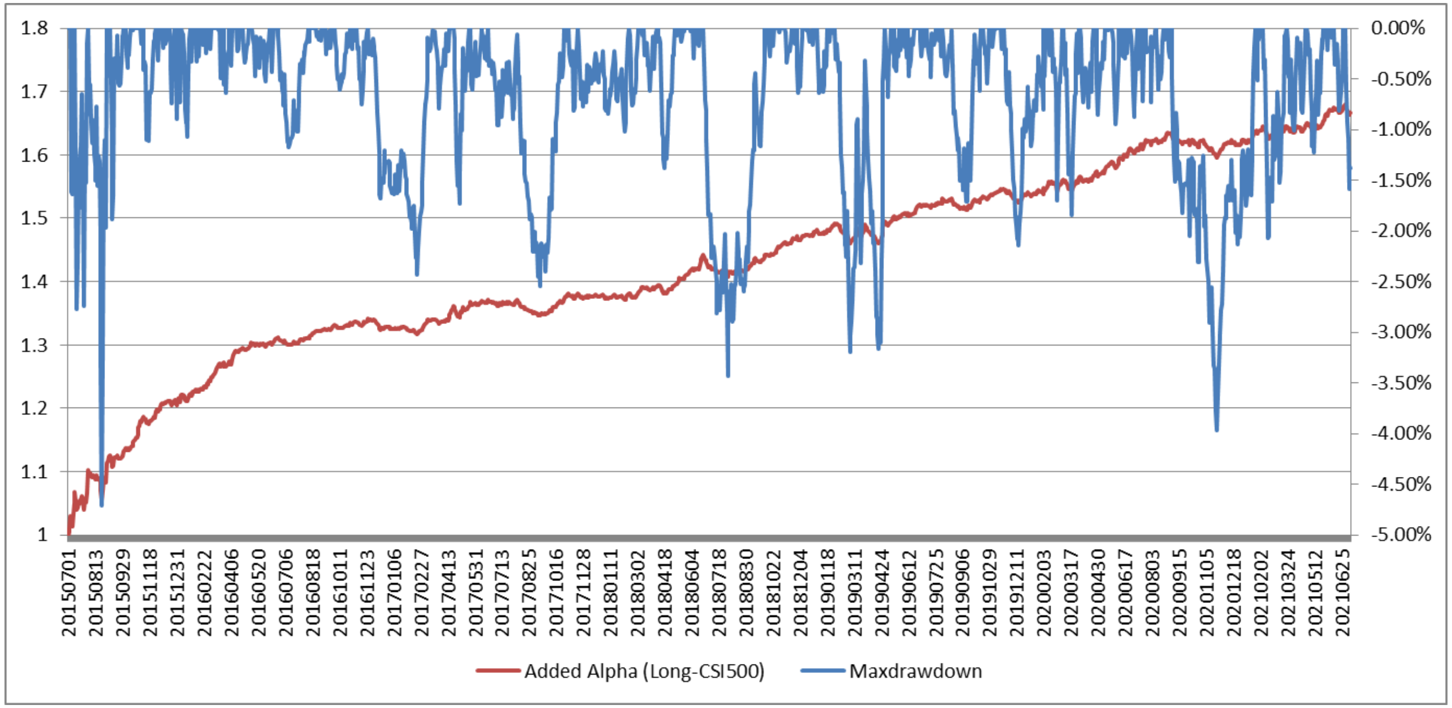

Figure 5.6 Portfolio optimization under strict risk control, 40bps transaction fees on both side.

This is the most strict sensitivity analysis and risk control condition. Although the maxdrawdown has been cutted down from 5% to 4%, we loose too much revenue. Thus, risk control is another very important and difficult problem in asset management. We don't think there are people always running under the most strict risk condition, it can't make long-term performance. But in some points, appropriate risk control can save people's lives. And sometimes, other competitors are running freely, these people will bring some impact to the other players in the short-term, this is a very difficult timing problem. So I found this is an interesting direction to be researched. And what I can do now is to give you some exposure about my alpha generation ability, under the ideal situation, difficult situation and the most terrible situation.

@About the Strategy Capacity.

It's also very important to know how much money this strategy can trade. **Here is a very common way to measure the strategy capacity, given by my PM. It was applyed for the Chinses mutual fund business. There is 5% constriction on the traded amount, although this number is not applyed to other business, I think test this case can help us to see how my strategy perform with the consideration of market impact.** Everyday, I will have some stocks to buy and some stocks to sell, **for these stocks, I calculated its maximum traded amount by using its first 30 minutes' amount*5%.** In this way, I can calculate the exact traded amount by selecting the small one of ideal traded amount and maximum traded amount. Then, I calculate the strategy's alpha by changing the total AUM of my strategy. I assume that 100% money is put into the equity assets, of course, in reality, 80%-95% of money will be put into the equity assets.

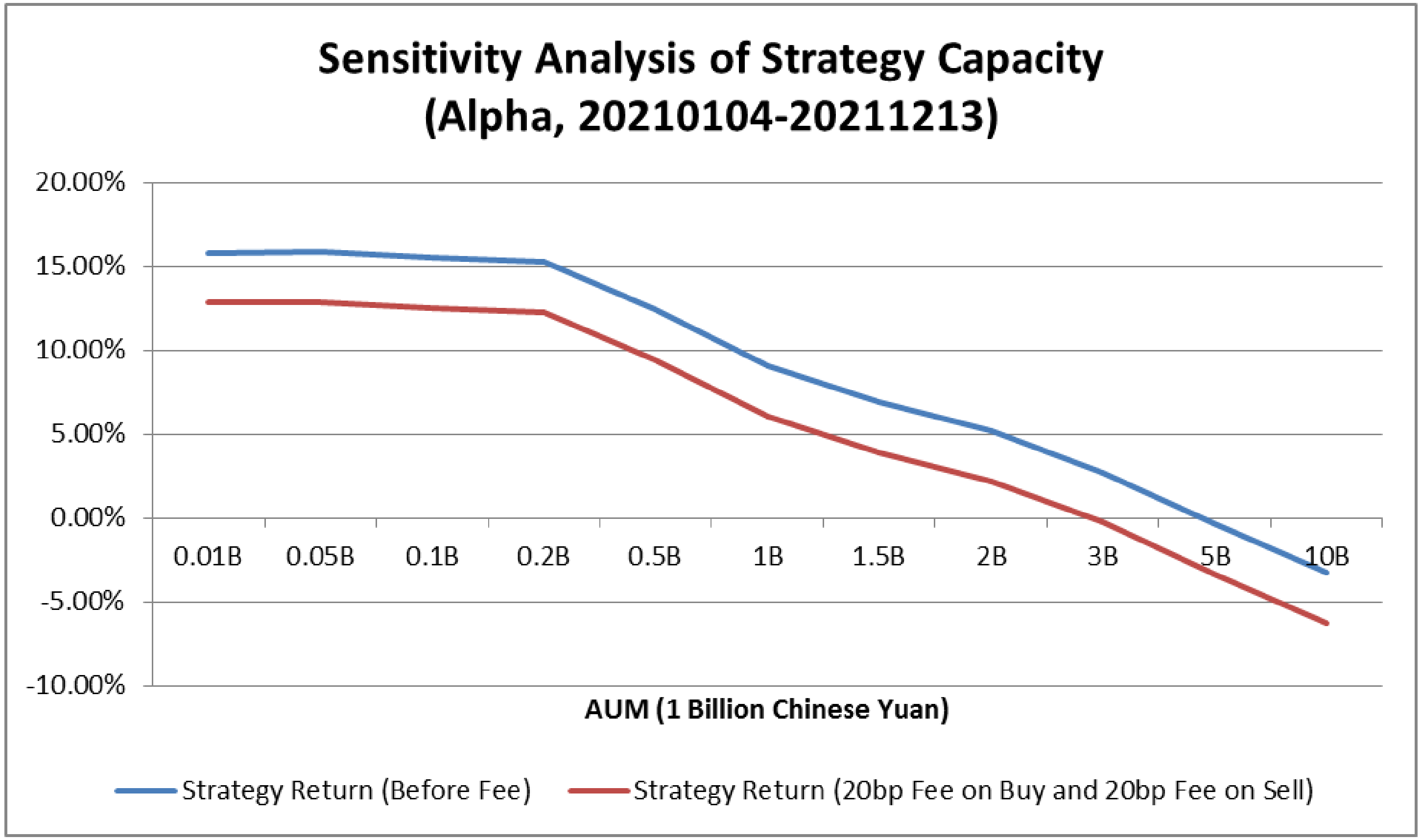

Figure 5.7 Sensitivity Analysis on Strategy Capacity.

The alpha shown above is calculated by long 20% of CSI500 and minus the CSI500 itself. I didn't take the hedge cost into account, because it was designed for Index Enhanced Strategy. As for the 20bps on buy and 20bps on sell, it's the fair transaction fees in China mutual fund business. If you are in other industry, this fee can be much lower, more than 50% cheaper. Thus, this is the root reason why the mutual fund player like low turnover rate and high capacity strategy.

As for the strategy capacity, if the AUM is smaller than 0.2 Billion Chinese Yuan, it can show its best performance. From 0.2B-3B, it suffers linear revenue decay. When it comes to 3B or 5B, this single strategy can't make alpha.

For further solution, on the one hand, if this strategy is run in hedge fund or prop trading desk, it may have larger capacity. On the other hand, I can try to construct more alphas and from more data source. Combining the diversified alphas into different sub-strategies, which may help enlarge the strategy capacity.

@Last Update, 11 Jan 2022.

@Final Backtesting.

Previously, I haven't done backtest on Q3 and Q4 of 2021. The biggest reason is that I know that Q3 and Q4 is hard for quant strategy, especially for the deep learning backed strategy. I hope to do test on this period of time just like real trading. But now, I want to make some personal change and look for new opportunity. Thus, I decide to give the full picture of my strategy's performance.

The backtesting period is 20160104-20211231.

Here is the final signal from all data. Long 20% stocks in CSI500, with equal weight, the curve below is the added alpha (Long20% CSI500 - CSI500 Index), before fee.

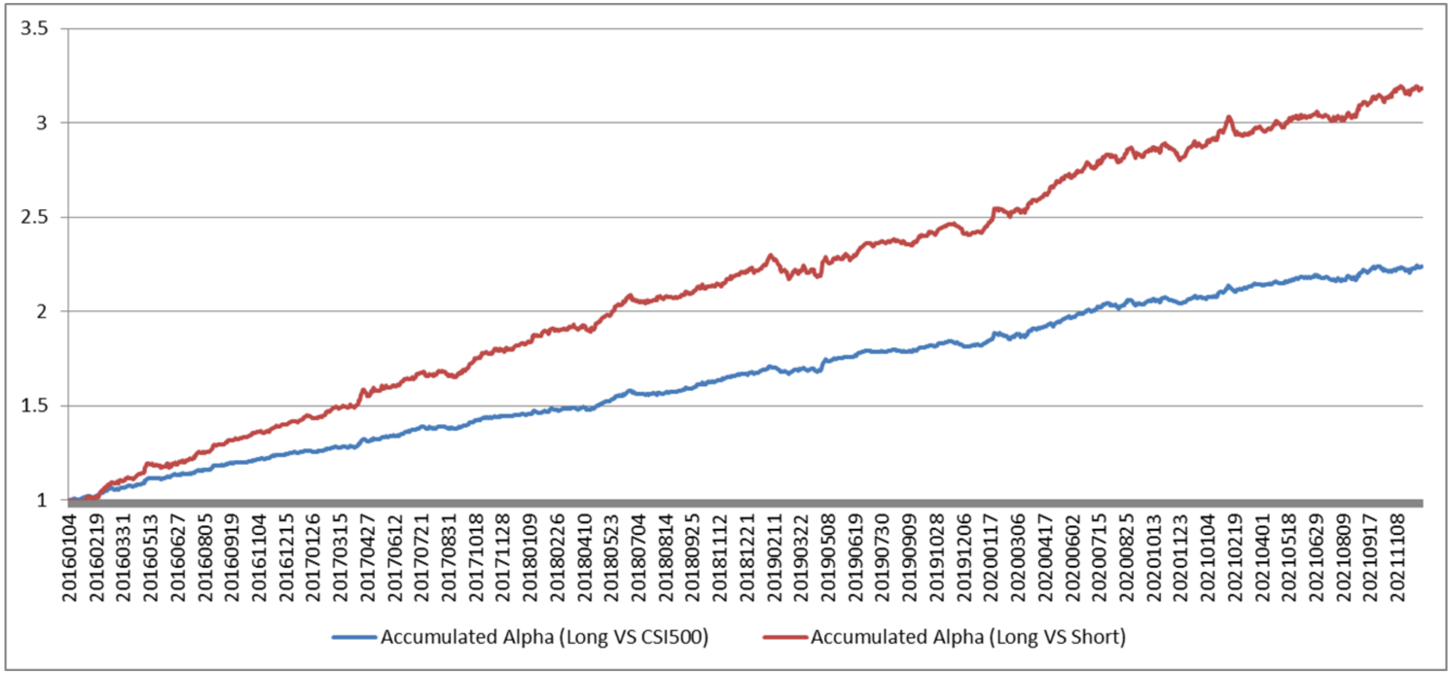

Figure 6.1 Final signal from all data.

Table 6.1 Final signal from all data, summary of the backtesting result.

	Total Alpha (Long20%-CSI500)	Net Alpha (Long20%-CSI500)	Daily Alpha Volatility	Annual Sharpe (VS CSI500)	Maxdrawdown (Net Alpha)	Daily Turnover rate (Buy+Sell)
2016	25.51%	22.58%	0.25%	5.942844589	-0.9%	6.0%
2017	20.24%	16.71%	0.29%	3.776906452	-1.46%	7.23%
2018	22.12%	19.19%	0.3%	4.12317624	-2.3%	6.03%
2019	14.43%	11.15%	0.34%	2.115081798	-3.97%	6.71%
2020	24.55%	21.68%	0.48%	2.917489056	-3.59%	5.92%
2021	16.32%	12.93%	0.5%	1.773613241	-3.35%	7.33%
Average	21.39%	18.13%	0.37%	3.441518563	-3.97%	6.53%

I summarized the results, shown in the table above. All the indicators except the Total Alpha, are net value, with the deduction of transaction fees, 20bps on buy side and 20bps on sell side.

My strategy perform very well in 2021 Q3 and Q4, becasue during this period of time. The majority of players sufferred from max-drawdown. And in mutual fund business, for CSI500 enhanced product, if people just follow the index or having extremely strict risk control, will rank 1 during these two seasons. For me, my strategy is not following the index or doing strict risk control. **I think my strong comings is that all my neural network backed factors are not complicated network, like transformer, Bidirection LSTM, and etc… I just use the pattern in traditional technical indicators like SMA, BOLL, I use the structures like attention, mask embedding and more, to mimic and design these technical inidcator's logic. Thus, many of factors' revenue is not amazing, about they are far from each other and very robust.**

***To sum up, in my current design, all the self-desined networks are easy and simple, just try to mimic the trational logic in technical indicators. For the majority of the factors, if you really wants its explicit formula, I can calculate if for you. But the way to predict and infer some middle-steps parameters is done by the neural network, eg. use sub networks to decide the size of window, to decide the ideal time series used for SMA structures. The way to settle down domain knowledge in finance (especially for the non-stationary finance problem, don't be too greedy) and let the neural network to learn some part freely, this is my cuurent answer about the Quant+AI.**

I fully held the IP of this framework, the loss function and how to input diversity is from my paper at KDD2020. The way to design network structures has not been disclosed, and there is no any line's programming assist from other people. For further research, I think I can do more about factor generation from alternative data, risk control research and timing strategy. For future career development, I am determined to look for new opportunity. If you agree with my research ability and contents, have faith in the development of AI and Finance, please contact me via fangx18@tsinghua.org.cn, you can also make a copy for my secondary e-mail, 346225716@qq.com, in case of the strict spam detection. Thanks a lot.

@Last Update, 8 Feb 2022.

@Thanks for Your Attention.

I got solid training in deep learning and classical machine learning algorithms when I were a student at Tsinghua University. Also, thanks for the great exposure when I were an intern at the financial institutions. These two types of experience help me a lot, which trains my way to thinking, and put the domain knowledge in finance into the design of new algorithms. Near the graduation, I look at my Linux Server, and found that I have more than 1000 day's experience in quant research and deep learning. I am always running against my time, thus, more or less, I hope you can treat my experience and research more seriously. Thanks for your great attention to read here, sincerely.