

Summary

This report studies the possibility of establishing a security selection strategy in Chinese A-share market. The aim is to find a strategy being able to provide stable yet profitable payoff in a long run.

In order to achieve that, an alpha factor strategy is studied in this report. It simultaneously holds positions in both stock market and stock index futures market. By holding short position in stock index future, it hedges the risk exposure of long positions held in stock share market. However, a perfect hedging leaves no profit. Therefore, this report tries to find profitable alpha factors and leave appropriate risk exposure to those factors. These discussions of alpha factors are based on the Arbitrage Pricing Theory of Stephen Rose's.

The first part of the report discusses factor performance. Criterion of choosing profitable alpha factors is being established and several factors are being examined by the criterion. Factor Momentum, Expectation and Value are found to be profitable and stable alpha factors. Cross-section regression and sectioned backtests are applied to examine factor performance.

The second part of the report establishes a security selection strategy as well as the hedging strategy. It uses an optimization model to determine the position to hold in each stock so that the portfolio's total payoff is maximized while risk exposures are constrained within containable levels.

Basic assumptions of this report

According to CAPM model, the yield of a security can be divided into two parts: the risk-free interest and the systematic risk premium. However, Merton H. Miller's studies as well as others have shown that this classification of security yield might miss something valuable. That's the alpha factors. Alpha factors, unlike beta, are not part of the general market risk. Its indications varies from stocks to stocks, and who believe in alpha believe that those factors along with beta contribute to the risk premium. To some extent, alpha is generated because of the flaw of beta. In traditional CAPM model, the only systematic risk factor is the performance of the whole market, in Fama French three factors model however, systematic factors include totalshare, book-to-market rate and evaluation level. Whatever factor excluded from systematic risk factors which does affect the stock performance is an alpha factor. They are generated because a large proportion of participants in this market act in a certain way without realizing this pattern themselves. Discovering this pattern and thus forecasting the behaviour of those investors can lead to profit. This report tries to find those alpha factors who persistently offer stable excess earnings.

Part One: The Factor Performance Report

Topic of the Factor Performance Report

This report serves as the first part of the Multi-factor Model report. It mainly discusses the characters of several factors including fundamental, financial, technical analyzation factors and so on. This report tries to answer the following questions:

- How much does each factor contribute to the stock performance
- During how many periods does each factor maintains effective
- Is there a pattern of the factor's performance

Data Procedure

- Sample space

This report sets the sample space to all component stocks of [CSI 800 Index](#) (*000906.SH*) which comprehensively includes most common types of stock in Chinese stock market. This report covers all data from the first trade day in 2013. The updated sequence will be gathered in a monthly sequence. The data is read from the [Wind](#) database through *Wind API* for R. (May demand certain membership to access the API)

- Standardization

In order to make different measurements with different dimensions comparable, data standardization is deployed for all observations before analyzation. There has three steps to obtain standardized and tidy data.

1. Deal With Missing Value

In order to reduce the impact of missing values on the liability of my analyzation, I set three rules in dealing with them. First, the proportion of missing values is one of my concerns while choosing factors. I try to avoid taking in those factors whose data is hard to obtain. Second, since most of factor data is time series data, missing spaces are filled with the nearest historical data that does exist. Finally, those missing spaces with no available historical data will be remained missing in this step and being taken care of in the third step below.

2. Deal With Extreme Value

Despite of the satisfying application of MSCI Barra Model^[1] in the use of moving smoother as a way to move extremes. I find it inevitably erasing too much of valuable information when being applied to Chinese stock markets. Therefore, I decide to simply use the straightforward triple standard difference method in order to remove extremes.

3. Remove dimension

We divide original data into cross-section data, which means we put together all the observations from different stocks at the same timepoint and remove their dimension as a whole. The common standardization method is applied: $\frac{x-\mu}{\sigma}$. (In which μ is the mean of each cross-section data set and σ is the standard variation of it). After this standardization, all data should has the same mean zero and a common variance of one. Those missing values being left in step one can now be filled with zero.

Factors

I divide all factors into two big categories: the industrial factors and the trait factors.

- **Industrial Factors**
Industrial factors identify a very basic feature of stocks since stocks within the same industry are widely acknowledged for their similar movements and identical reactions towards external stimulations.
- **Trait Factors**
Trait factors try to explain why certain types of stocks perform better than others during a particular time period. This report tries to point out those contributive factors and the circumstances in which they remain effective and thus deducing a guide in how to correctly use the factors.
- **Evaluation Criterion Of Factors**

Each factor will be judged in three ways:

1. **The significance level between the factor indications and the stock performance**
This significance level can be measured if we apply regression model to the cross-section data of a certain factor at each and every timepoint. The regression model should be a single-factor linear model with the factor indications of each stock as the independent variable and the stock performance within that period as the dependent variable. The t-value of the independent variable generated by the regression model is the significance level between the factor indications and the stock performance. And within our research, the t-value of a factor can be achieved from regressions at each timepoint and thus form a time-series sequence. The mean of that sequence may serve as the average significance level of that factor.
2. **The proportion of effective time periods**
Since the t-value of a factor forms a time-series sequence, we can determine the proportion of times when this factor significantly affects the performance of the stock. I define “significantly affect” as the t-value at that time point exceeds two (or less than minus two).
3. **The performance of stocks with high or low factor indications**
Within each period, I divide all stocks into five groups with the factor indications from low to high and thus generating five portfolios for each timepoint. Then I set up the strategy always picking up the same group at each period and thus generating a dynamic portfolio being changed at each time point to always hold for example, the 20% of all stocks which have the highest 20% of factor data indications. Then I back testing the strategy’s performance. This could demonstrate whether a higher level of factor indication leads to a better performance or a lower one does.(Or maybe this factor has shifty influence on the performance and thus making it a bad factor.)

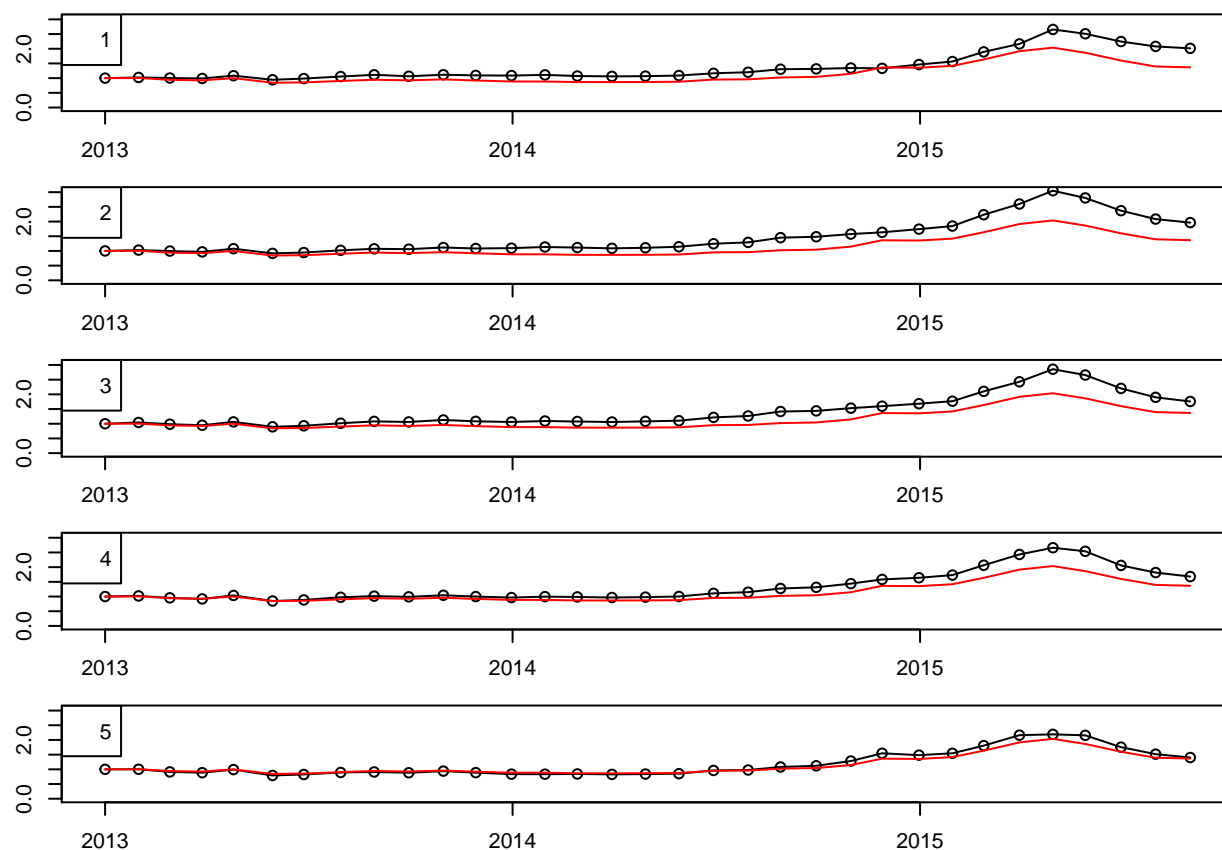
Trait Factors: Beta

What is beta

Beta is the correlation between a certain stock and the reference index. Here we choose the *CSI 800 Index* as the reference and the time span is set to be the past 100 weeks. The beta indication is calculated basing on the time series of stock performance of daily frequency.

sectioned portfolio backtest result

[1] "beta"



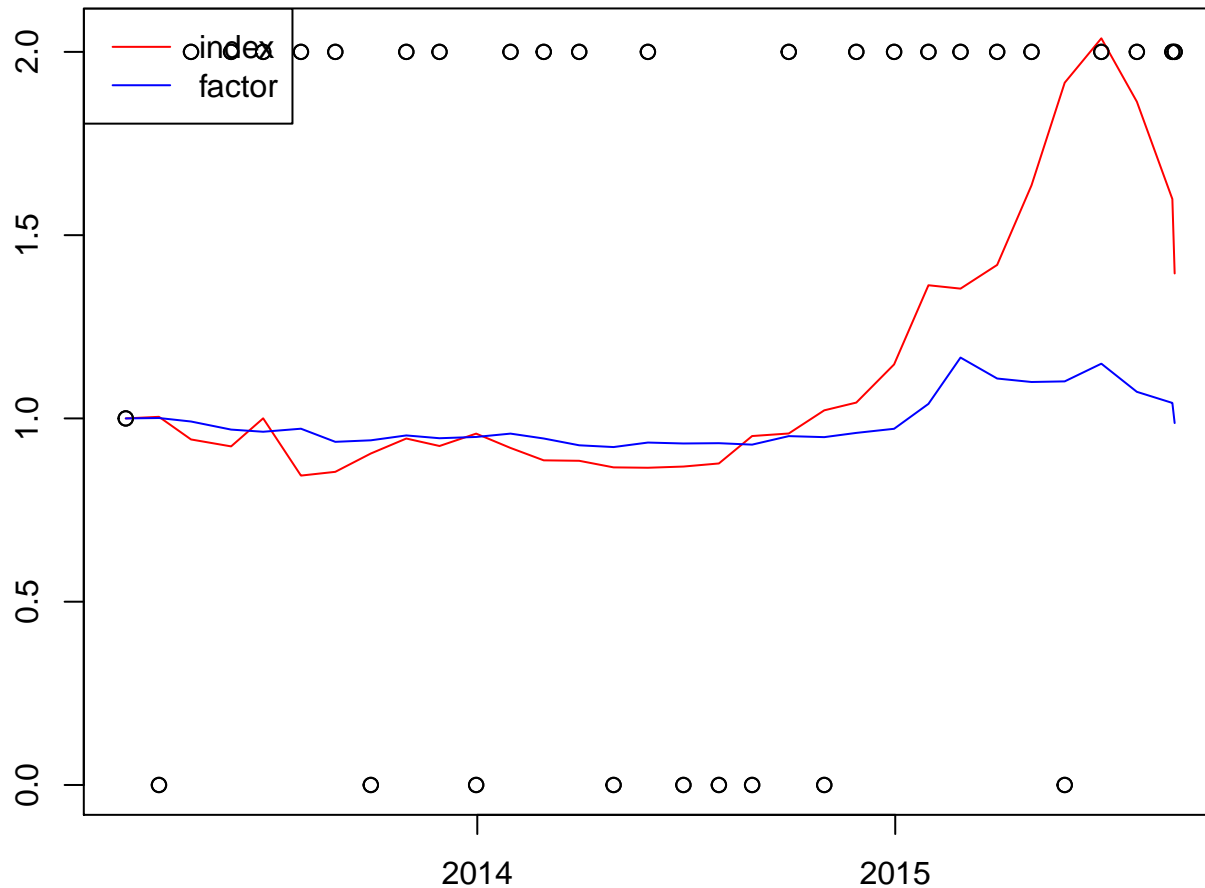
	[,1]	[,2]	[,3]	[,4]	[,5]
gain	2.013654	1.9636477	1.757396	1.6794272	1.4021931
	0.484900	0.5886386	0.530553	0.5179784	0.4300799

What does the plot mean These plots shows the performance of five portfolios. The No.1 portfolio always includes those stocks with the lowest factor indications at last month and the No.5 has the highest. The red line is the reference index performance.

The table below the plots shows the performance of each group. The first row shows the final value of each portfolio whose initial value is set to be one and the second row shows the standard deviation of each portfolio.

From sectioned portfolio backtest result we can easily find out that those with high beta usually perform better than those with lower ones. Also, I should point out that **unlike the reference index (CSI 800) which sets the weights of each component stocks according to their market capitalizations, here I set the weights to be equal among all components within one portfolio.**

The cross-section regression result



	average_t_value	times_larger_than_threshold	rsquare
1	5.168851	0.7096774	0.05242268

	average_factor_gain_year	volatility_factor_gain
1	0.1926867	12.04279

What does the plot mean This plot shows the regression result. The red line is the performance of our reference index and the blue line shows the performance of factor return, that is the coefficient of the factor generated by the regression model. Those points indicates whether the factor's t-value is larger than the threshold or not. If it locates at the top of the plot than t-value exceeds threshold during that period and vice versa.

The table below the plot shows detailed information including the average t-value, the proportion of t-value exceeding threshold, the total explanatory power of the regression model, the average annual yield of factor and the volatility of factor return.

It's obvious to see that beta has a large t-value and a high proportion of times when its t-value exceeds threshold. Therefore, beta should be a rather effective factor.

Conclusion

Beta seems to be a decent factor but being different from an alpha factor, beta actually enlarge the gain and the loss, therefore, it makes no sense to allow risk exposure on beta.

Beta			
WindCode:	beta_100w	t-value:	5.2
Effective:	YES	Effective%	71%
Way Of Effect:	The smaller the better	FactorGain	19.3%
MissingData%:	193/26334		

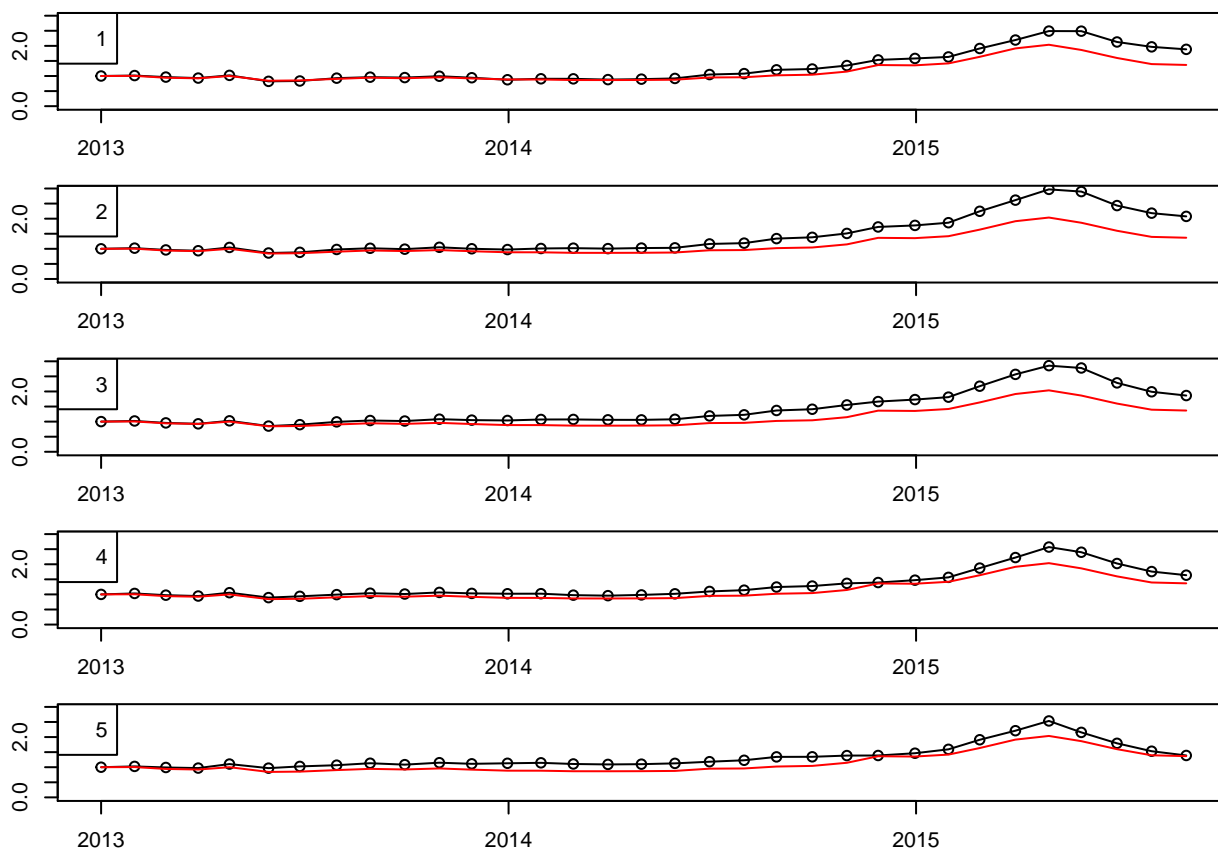
Trait Factors: Momentum

What is momentum

Momentum measures the historical performance of this stock, it's the average annual yield for the last 100 weeks. This factor may suggest whether those good stocks tend to maintain good (commonly known as momentum effect) or tend to become weak (commonly known as turnover effect)

sectioned portfolio backtest result

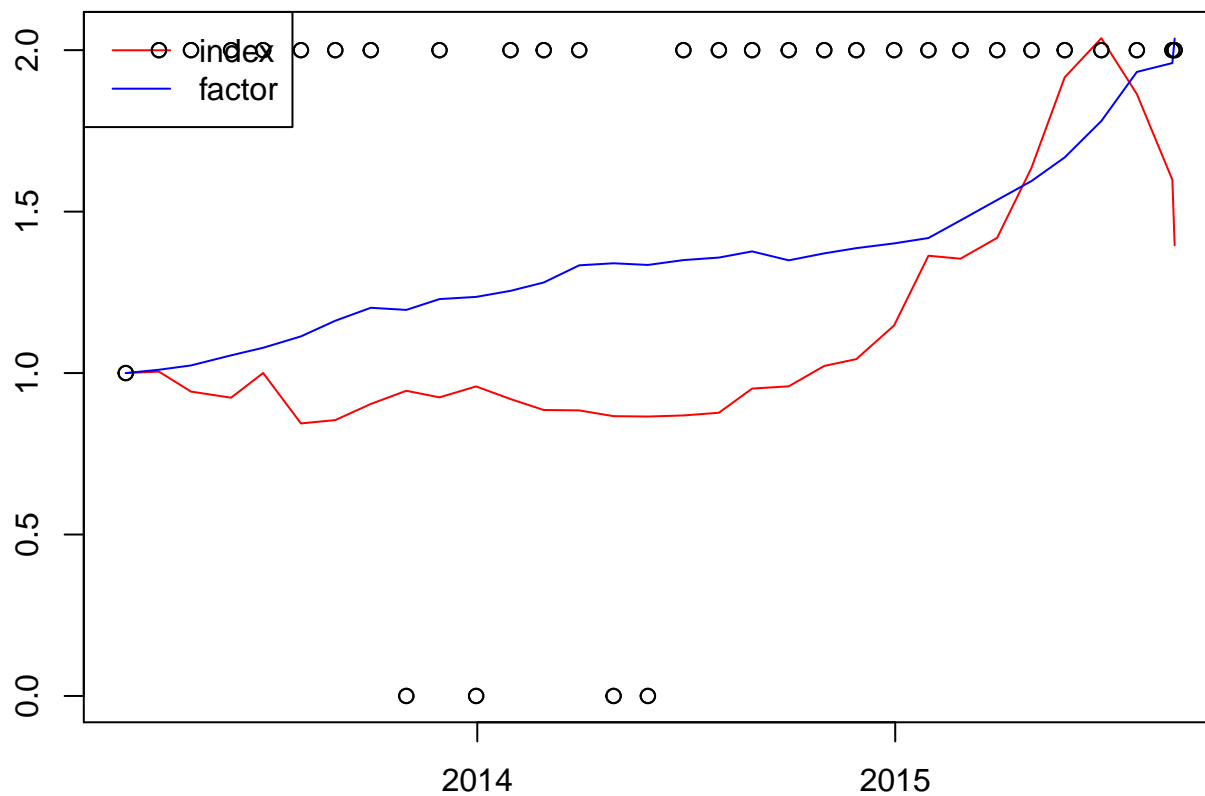
```
[1] "momentum"
```



	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.8844672	2.0753914	1.8647006	1.6380351	1.3893224
	0.5110121	0.6292815	0.5738793	0.4578942	0.3900905

Stocks with low momentum tend to perform better, however, the group with the lowest indication is not the best. Those ranked in the bottom twenty to forty percent perform the best. This indicates that the reversal effect in A-share market tends to be more decisive than the momentum effect.

The cross-section regression result



```

average_t_value times_larger_than_threshold    rsquare
1          6.54379                0.8709677 0.06354431
average_factor_gain_year volatility_factor_gain
1          32.02111                7.480909

```

Regression result suggests that momentum is a highly effective factor with an average t-value as high as 6.6 and remains effective in up to 88% of time. This is a good factor and contribute positively to the stock performance.

Conclusion

Momentum			
WindCode:	annuallyield_100w	t-value:	6.5
Effective:	YES	Effective%	87%
Way Of Effect:	The bottom 20% to 40% is the best	FactorGain	32%
MissingData%:	193/26334		

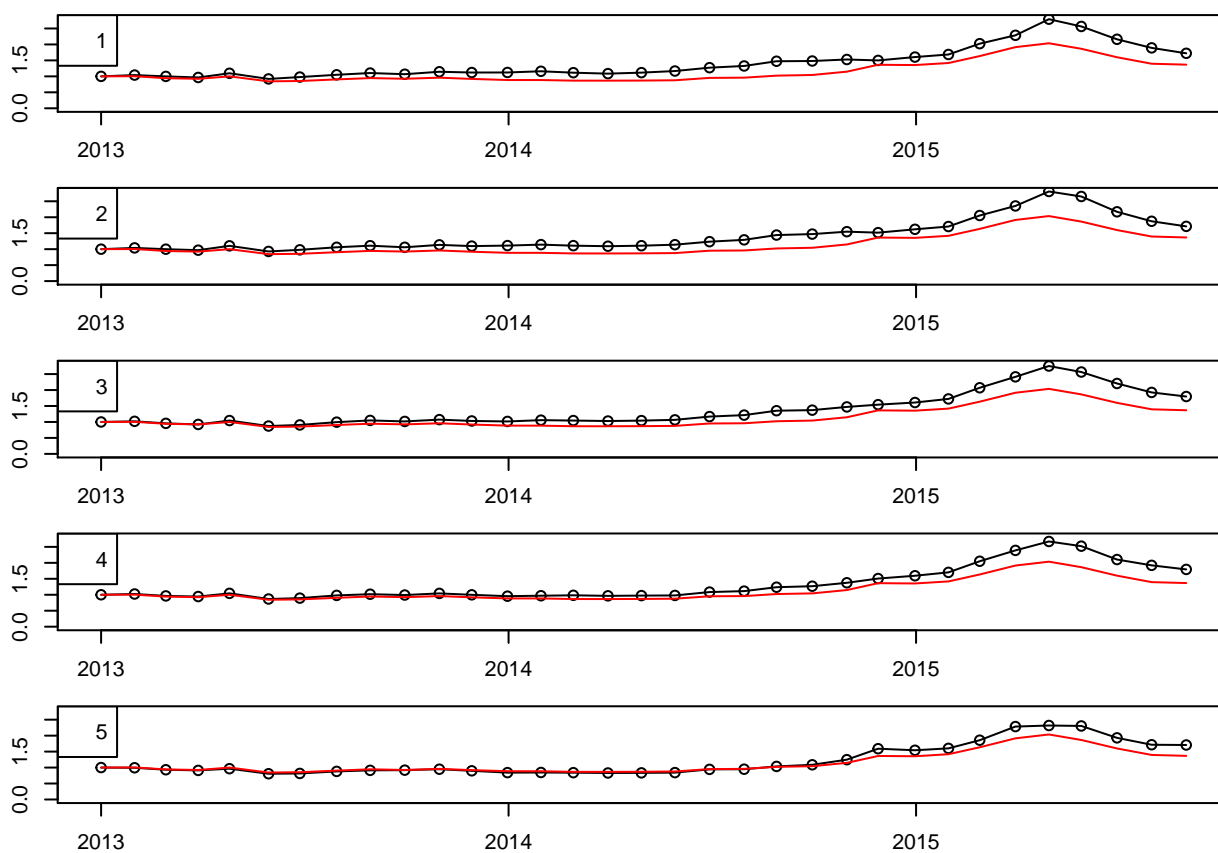
Trait Factors: Size

What is size

Size factor is a rather fundamental character of a stock. It indicates the total shares of this company and thus showing the scale of this stock which is a decisive but complicated factor in Chinese A-share market.

sectioned portfolio backtest result

[1] "size"

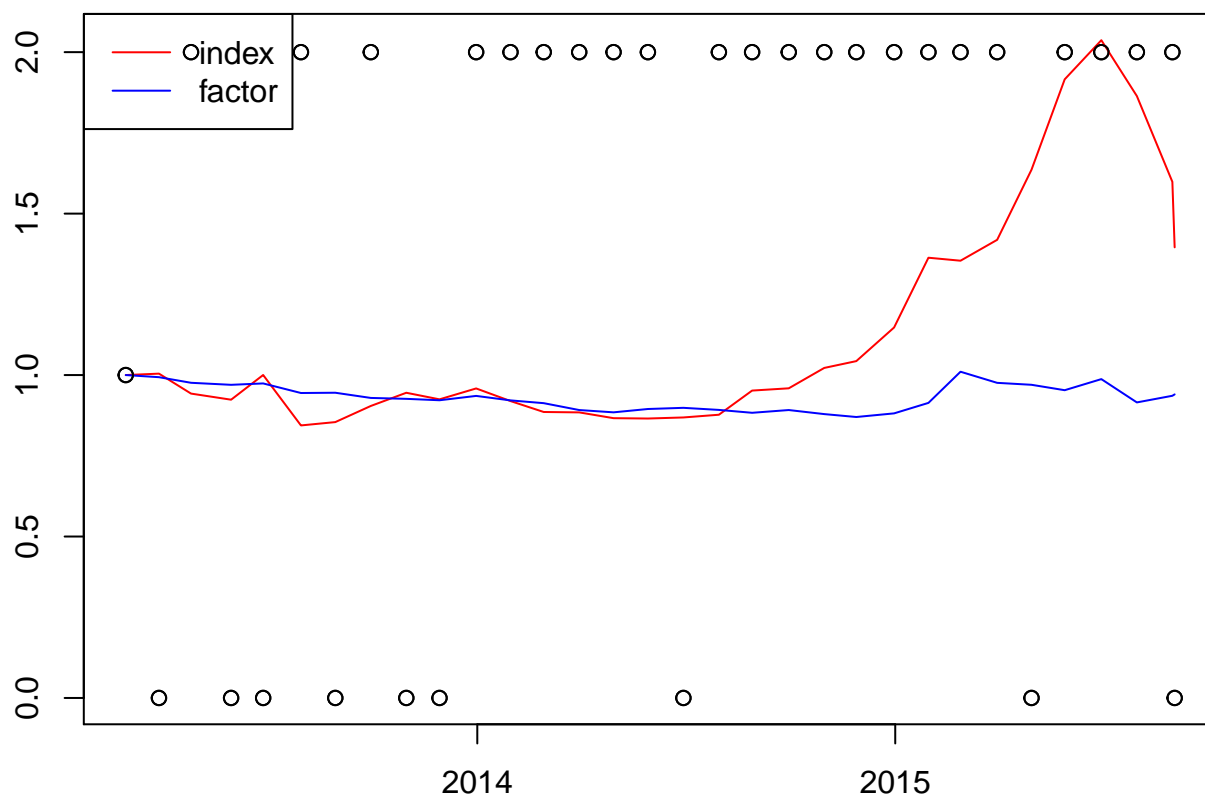


	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.7228097	1.7140251	1.7968161	1.7941650	1.7069841
	0.4881725	0.5032814	0.5247315	0.5200379	0.4848826

A-share market has some really big companies. For example, those commercial banks occupy almost half of the total weight in the stock index.

We can see that stocks with smaller size are preferred by market. However this conclusion may vary from time to time which and thus making this factor unreliable.

The cross-section regression result



```

average_t_value times_larger_than_threshold    rsquare
1         4.347548                0.7096774 0.03347399
average_factor_gain_year volatility_factor_gain
1          -1.899973                9.953076

```

As we can see from the regression result that the size factor does influence stock performance. But may not in a clear pattern and this factor may not generate ideal payoff. This result suggests to leave no risk exposure to size factor.

Conclusion

This factor has ambiguous effect on the stock payoff which makes it too dangerous to play with. Therefore, I think it's better to limit the risk exposure on this factor to be 0.

Size			
WindCode:	total_shares	t-value:	4.3
Effective:	YES	Effective%	71%
Way Of Effect:	Unclear	FactorGain	-2%
MissingData%:	0/26334		

Trait Factors: Consensus Expectation

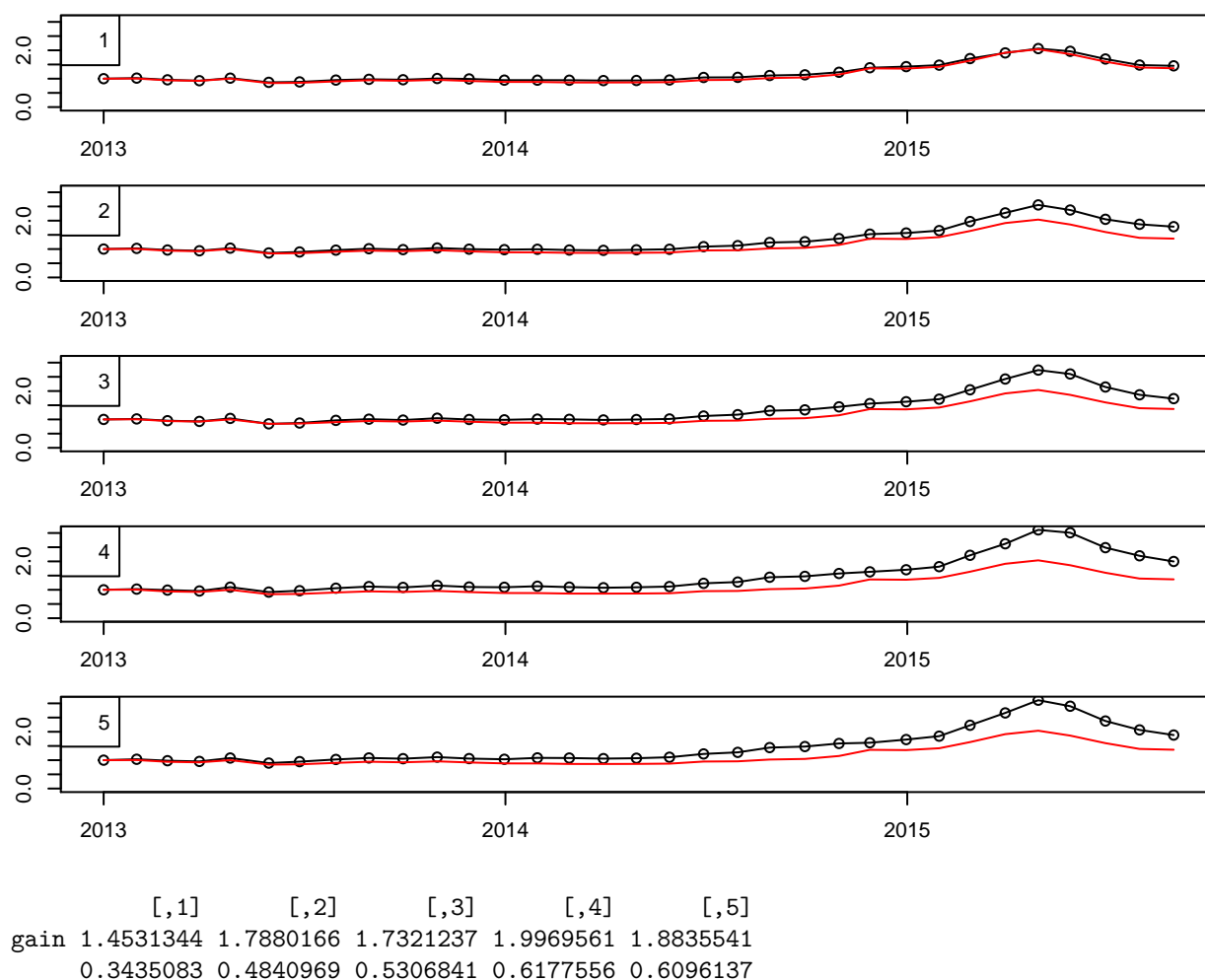
What is Consensus Expectation

Consensus Expectation is the unified opinion towards a stock from thousands of stock analysts. This factor is consisted of three sub-factors: the expected roe(Rate of Return on Common Stockholders' Equity),the expected eps(Earnings Per Share) and the average rating from analysts.

The factor is a linear combination of sub factors and the weight of each subfactor is determined by an optimization model whose target is to maximize the explanatory power of the new generated factor.

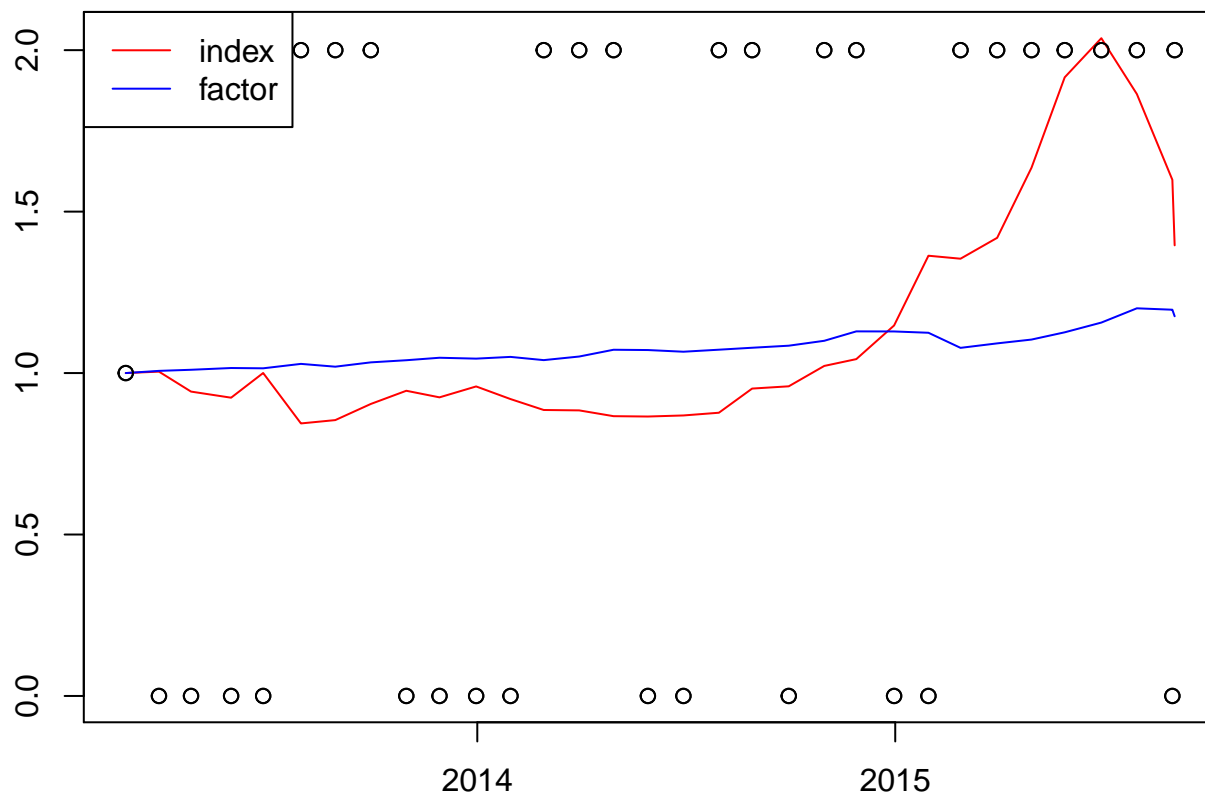
sectioned portfolio backtest result

[1] "Consensus Expectation"



Backtest suggests that those stocks preferred by stock analysts tend to be preferred by the market as a whole. However, the most welcomed by analysts are not those perform the best in market.

The cross-section regression result



```

average_t_value times_larger_than_threshold    rsquare
1          2.72611                0.5483871 0.01308858
average_factor_gain_year volatility_factor_gain
1              6.600953                5.018281

```

Regression suggests that consensus expectation, as a factor is normally stable and profitable. However, it might go very wrong sometime. For example, the start of 2015 witnesses a Waterloo of stock analysts.

Conclusion

My suggestion is to use this factor with care. On one hand, it's stable and profitable at most time. On the other, it may lead to big disaster. Consequently, we shall use this factor. With great care, though.

Expectation			
WindCode:	west-avgroe-FY1,west-eps-FY1,rating-avg		
Effective:	YES	Effective%:	55%
Way Of Effect:	The higher is better	FactorGain:	6.6%
MissingData%:	3380/26334	t-value:	2.72

Trait Factors: Volatility

What is Volatility

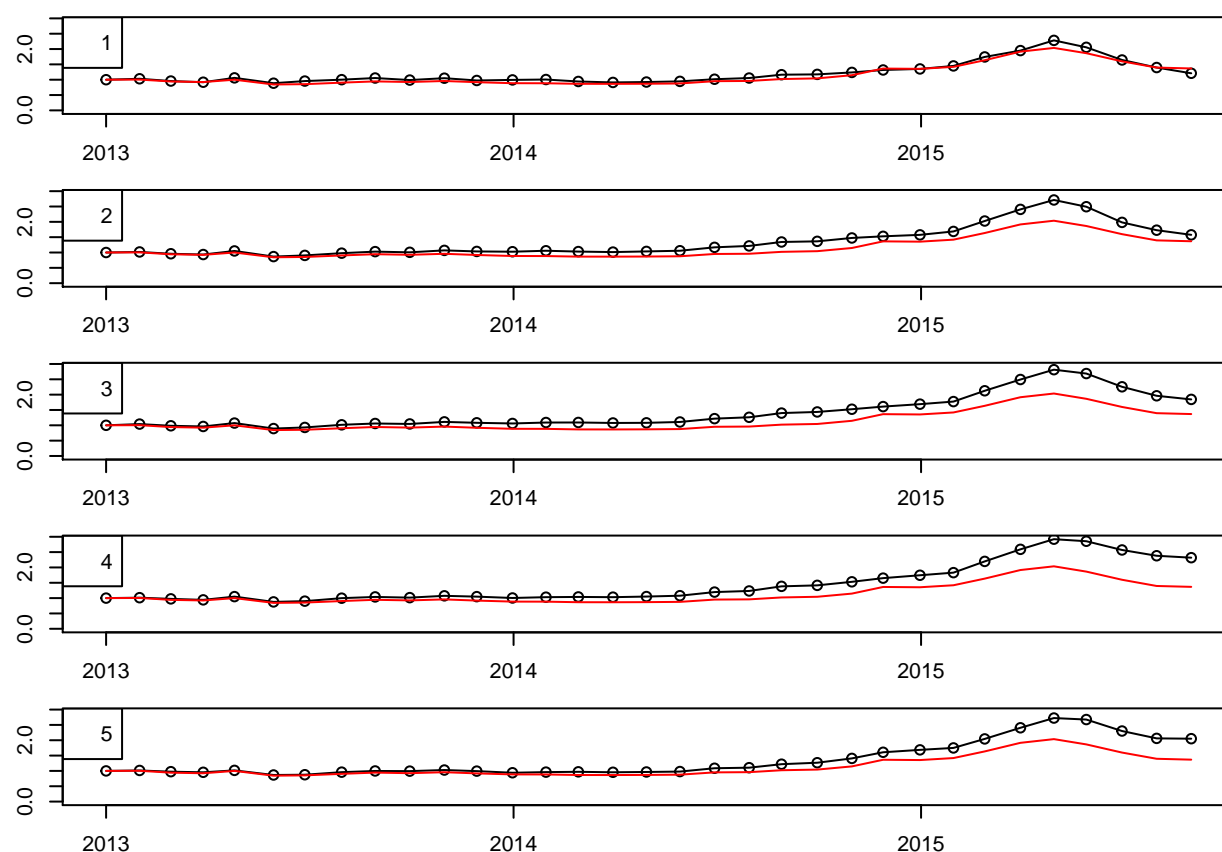
Volatility shows how severe does this stock swing in the past. In a broader way, I define volatility as the risk factor. It indicates how dangerous and active this stock is.

It is consisted of three sub-factors including the annual standard deviation of the monthly yield sequence for the last one hundred weeks and the last three years. Also, I include the turnover rate factor to indicate how actively is this stock being traded.

The factor is a linear combination of sub factors and the weight of each subfactor is determined by an optimization model whose target is to maximize the explanatory power of the new generated factor.

sectioned portfolio backtest result

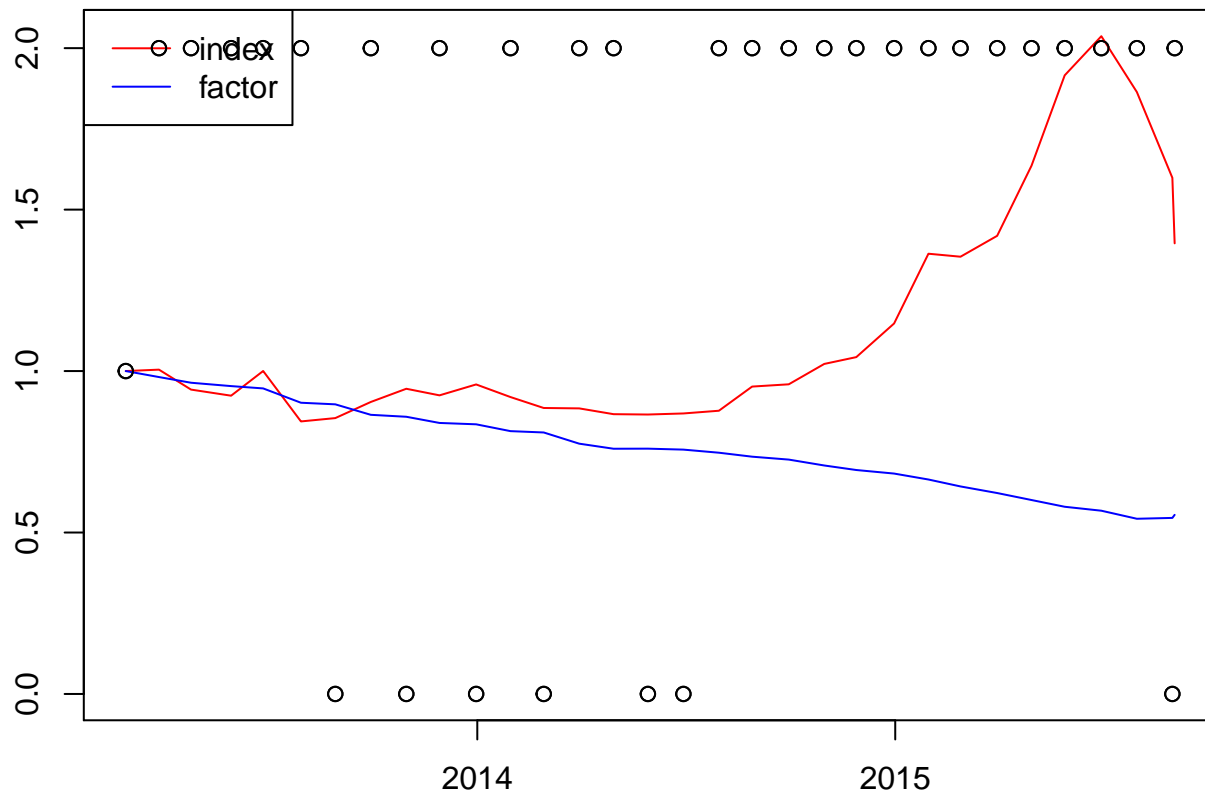
[1] "volatility"



	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.2097871	1.5768101	1.8447268	2.3184581	2.0508528
	0.3575024	0.4937885	0.5409802	0.6329991	0.5666599

The plot suggests that the higher the volatility, the better the performance and top 20% to 40% give the best return. So this result suggests that more active stocks may have larger risk yet better payoff.

The cross-section regression result



```

average_t_value times_larger_than_threshold  rsquare
1          5.439729                0.7741935 0.0465274
average_factor_gain_year volatility_factor_gain
1          -20.28525                5.204972

```

Regression model helps us to decide whether risk is a good thing. It's quite clear that taking too much risks may end up broken. Therefore we should strictly limit the risk exposure on volatility to zero while generating our factor strategy.

Conclusion

Applying volatility as an alpha factor is too risky. I suggest to limit its exposure to zero.

Volatility			
WindCode:	annualstdevr-100w,annualstdevr-36m,turn		
Effective:	YES	Effective%:	77%
Way Of Effect:	Not a good factor	FactorGain:	-20%
MissingData%:	197/26334	t-value:	5.43

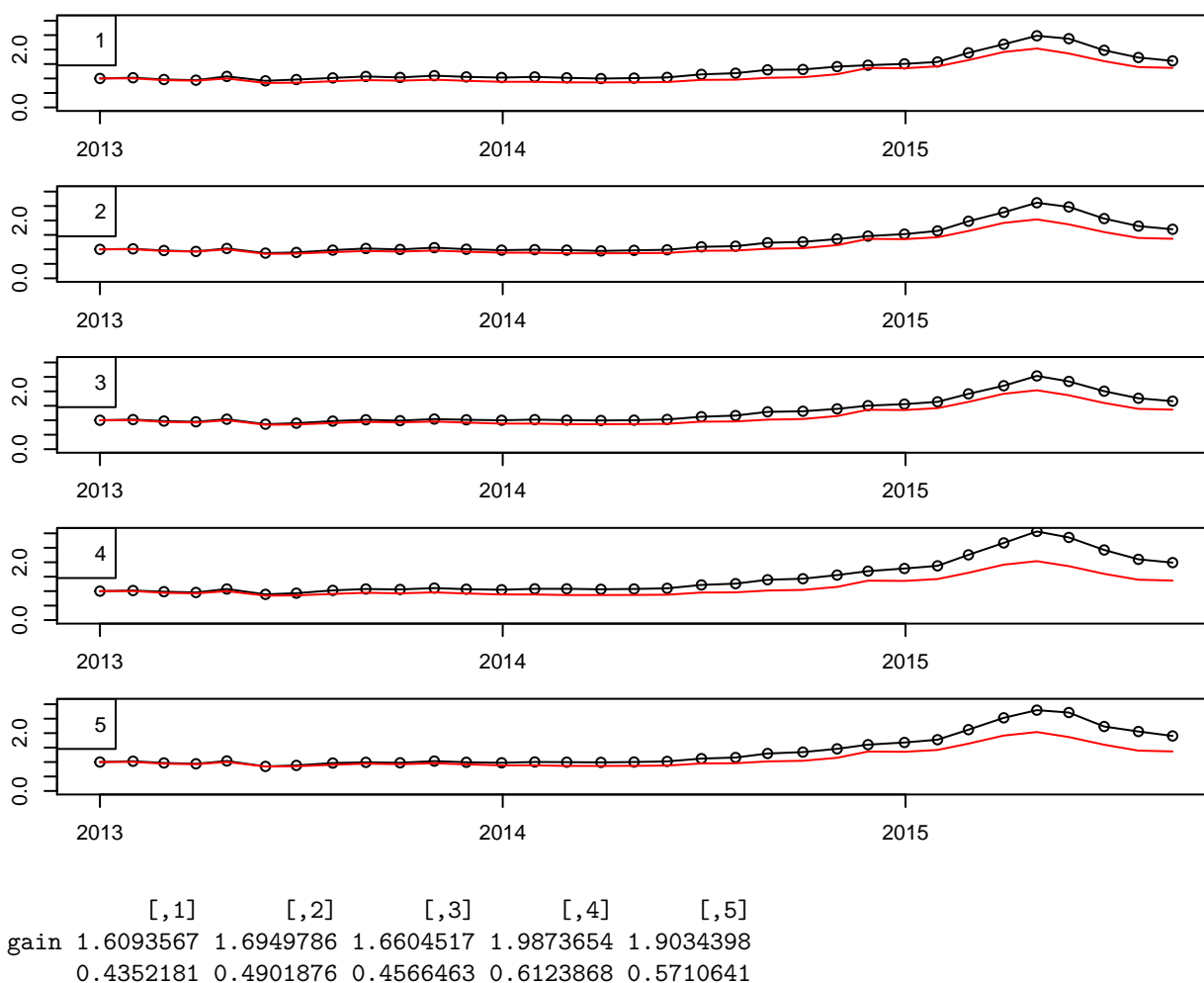
Trait Factors: Growth

What is Growth

Growth factor contains two aspects, the historical growth of this company and the estimated growth of it. Therefore, this growth factor is consisted of four sub factors: the estimated net profit growth, the estimated sales growth, the roe growth for last time period and the gross revenue growth rate. The factor is a linear combination of sub factors and the weight of each subfactor is determined by an optimization model whose target is to maximize the explanatory power of the new generated factor.

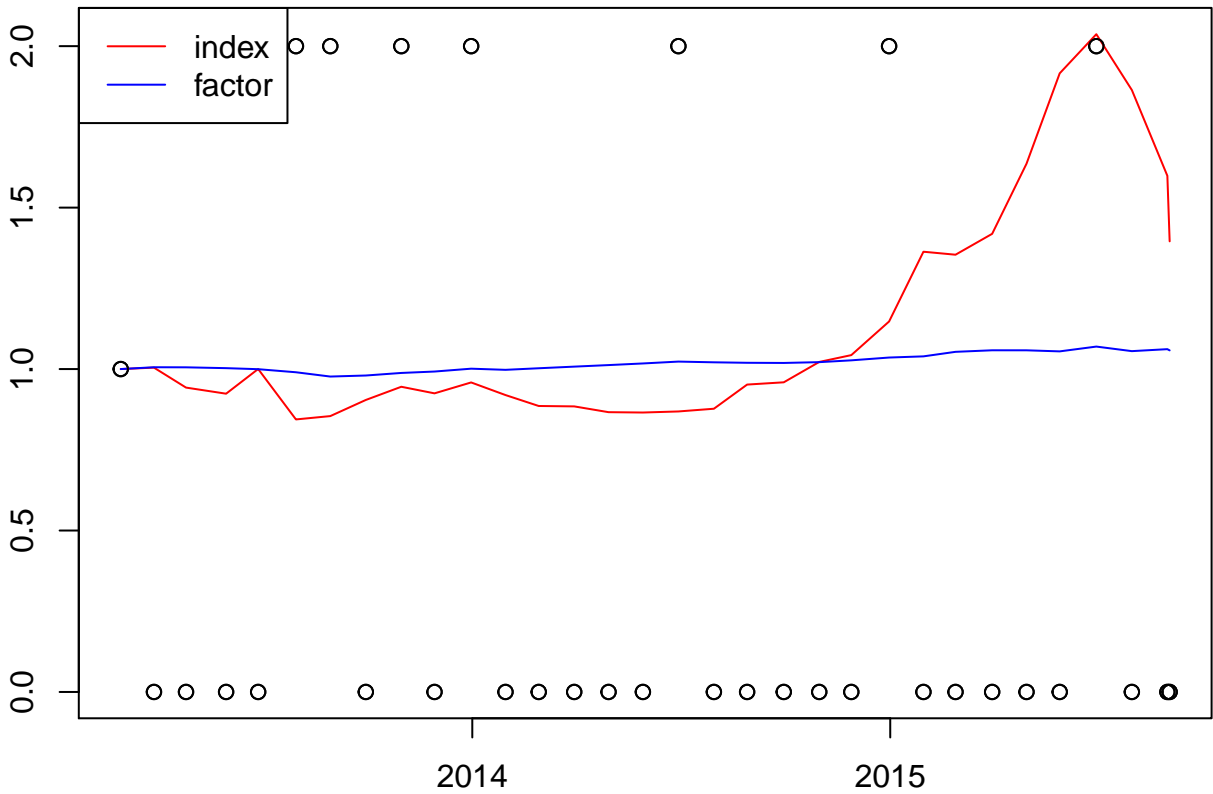
sectioned portfolio backtest result

[1] "growth"



We can see from the plot that companies with high incidents in growth factor tend to perform better. However, it's not clear whether it's the higher the better or not. What definitely can be told is that a decent growth promises a better performance.

The cross-section regression result



	average_t_value	times_larger_than_threshold	rsquare
1	1.391899	0.2258065	0.003410064
	average_factor_gain_year	volatility_factor_gain	
1	2.226708	2.272134	

The regression model result leads to a big surprise. The growth factor actually has a very small positive outcome. But this may be the consequence of regression data choice. We now use regression model to explain the performance with the data at the same time period. **However, growth data tends to influence the stock performance at the next period or even later. Besides, the time point we choose is not exactly the timepoint when the raw data is released. After all, these factors are mainly financial indicators and tend to be released in quarterly and yearly reports.** Consequently, I do not want to rush to a conclusion toward whether the growth factor is a good one or not, it will be discussed with more detail in the following reports.

Conclusion

This factor's effect remains unclear. My suggestion is to leave no risk exposure on it.

Growth			
WindCode:	west-net-profit-CAGR,west-sales-CAGR,growth-roe,growth-gr		
Effective:	NO	Effective%:	23%
Way Of Effect:	Unclear	FactorGain:	2%
MissingData%:	3166/26334		

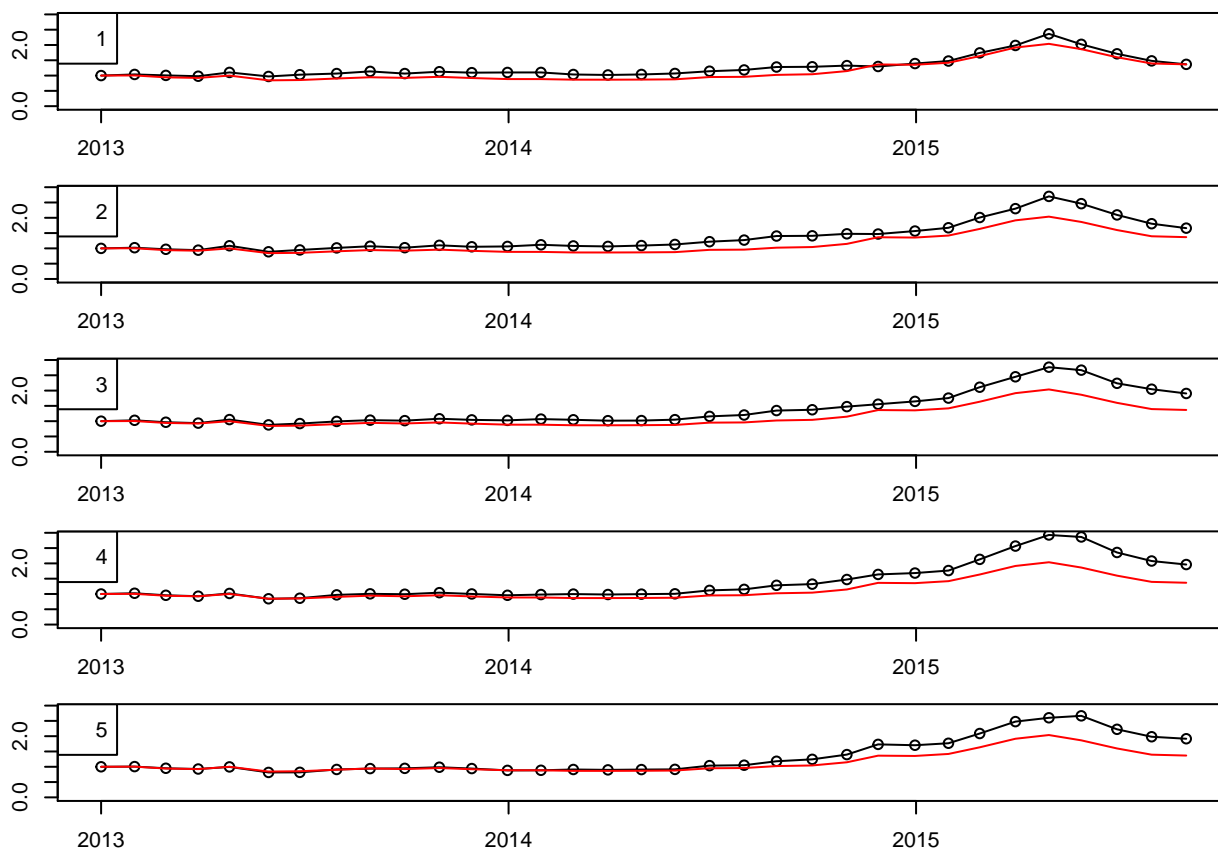
Trait Factors: Value

What is Value

Value indicates the valuation level of the stock. Despite the fact that lots of indicators can be applied to demonstrate the valuation level, I find them all too much similar that causes collinearity. Actually, the single PB-TTM factor is well qualified to serve as the value indicator. Here we actually use the reciprocal of PB.

sectioned portfolio backtest result

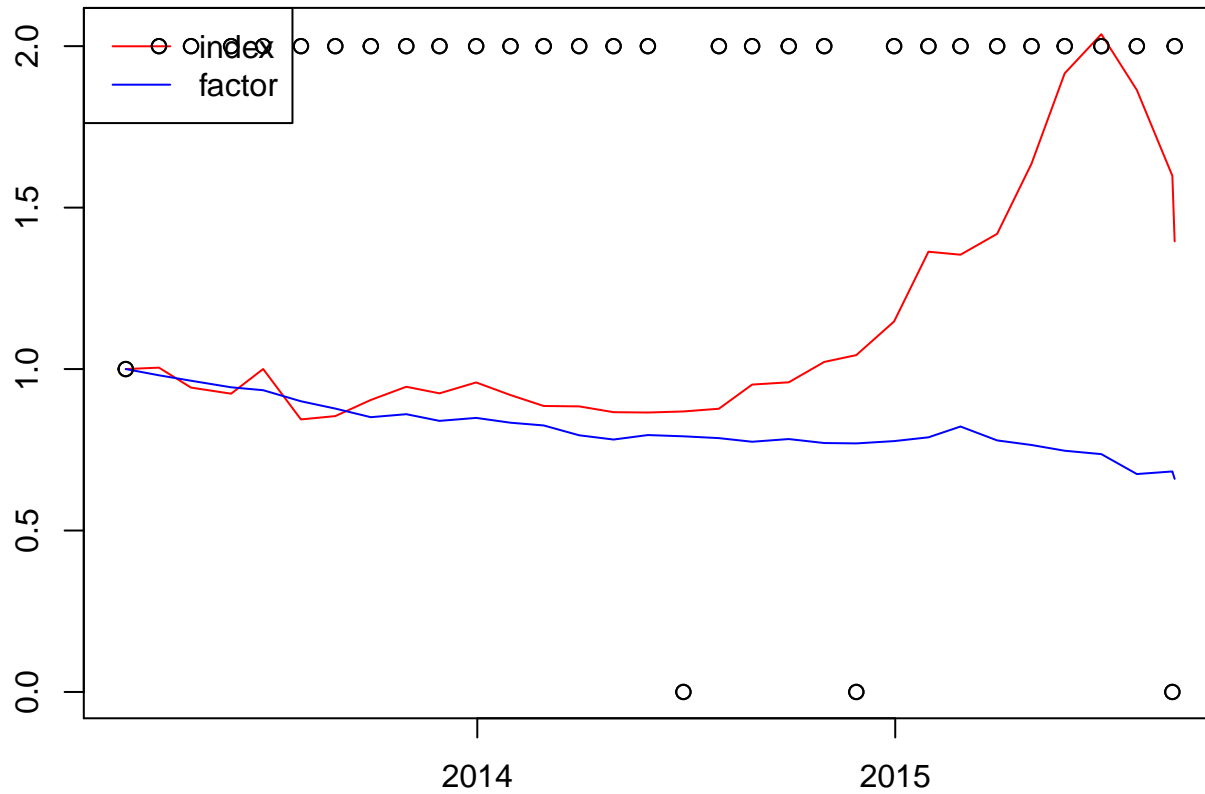
[1] "value"



	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.3670905	1.661747	1.9057346	1.9625849	1.9163527
	0.3403885	0.477320	0.5466945	0.6072014	0.5744059

These plots indicates that those stocks with low evaluation level tend to become profitable while those overestimated shares will devalue. However, it's best to rank in the bottom 20 to 40 percent rather than the lowest 20% in order to perform best.

The cross-section regression result



```

average_t_value times_larger_than_threshold    rsquare
1          5.451005                0.9032258 0.04391627
average_factor_gain_year volatility_factor_gain
1          -14.57273                8.214382

```

The regression result supports the result of portofolio simulation. This indicator offers a steady nagetive yield which suggests that the inverse of it (the PB) offers a positive one.

Conclusion

The value factor is effective and should be applied as an profitable alpha factor. However, we should notice that value factor is defined as the reciprocal of PB. Therefore its value being smaller suggests a better stock performance.

Value			
WindCode:	$\frac{1}{PB}$	t-value:	5.5
Effective:	YES	Effective%:	90%
Way Of Effect:	The higher PB the better	FactorGain:	-15%
MissingData%:	0/26334		

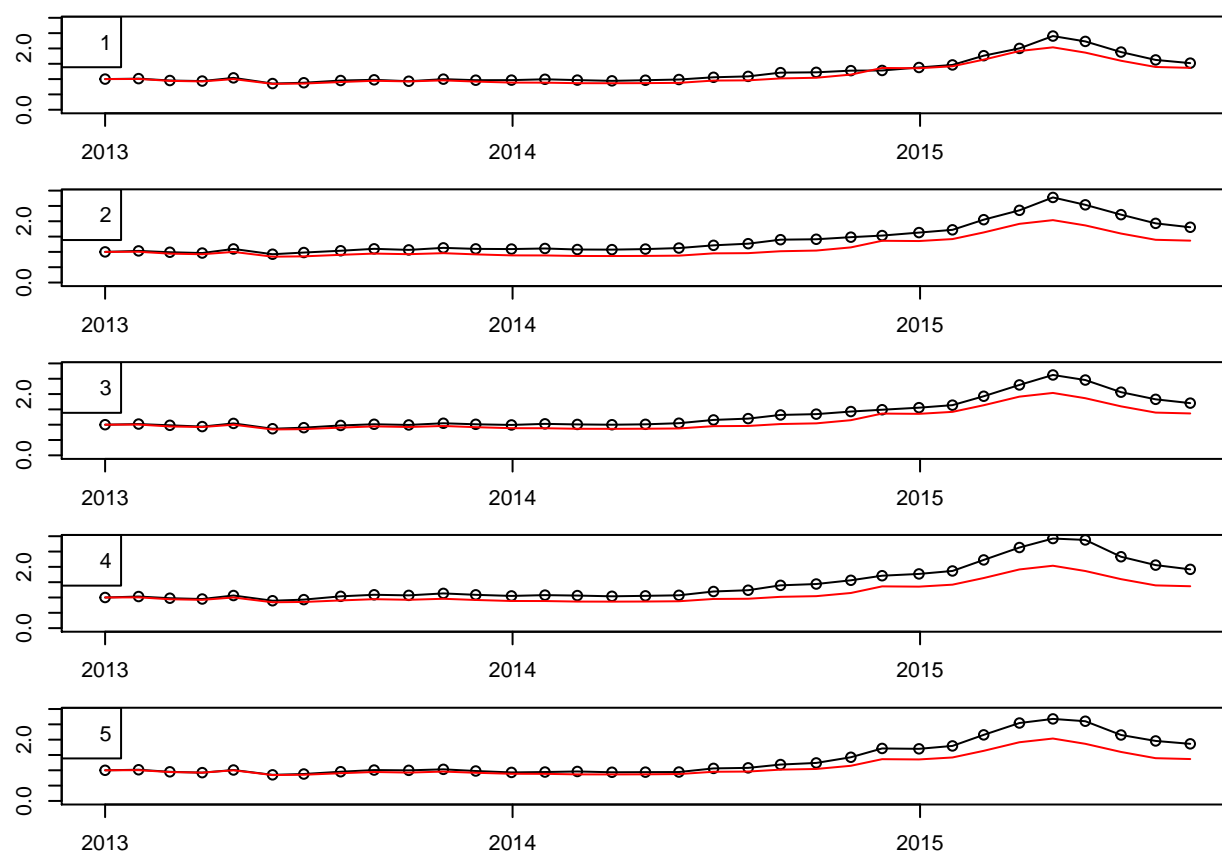
Trait Factors: Leverage

What is Leverage

Leverage is the factor showing the debt burden of the company. It is consisted of four sub factors: debt ratio, long time debt ratio, debt paying ability, long time debt paying ability. Here we actually use the reciprocal of these factors. The factor is a linear combination of sub factors and the weight of each subfactor is determined by an optimization model whose target is to maximize the explanatory power of the new generated factor.

sectioned portfolio backtest result

[1] "leverage"

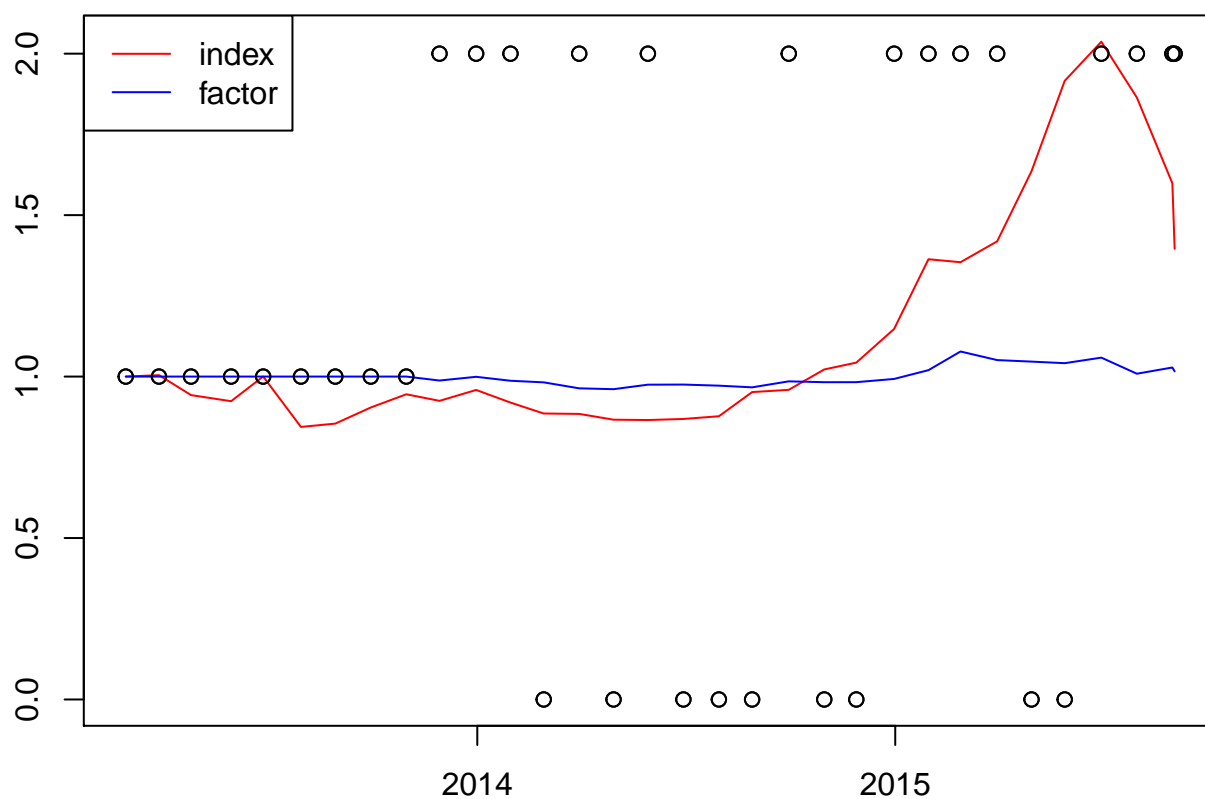


	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.521425	1.804271	1.7040369	1.918482	1.8634170
	0.410194	0.502542	0.4831758	0.592107	0.5626551

It's hard to tell any conclusion for sure from this plot.

The cross-section regression result

[1] "leverage"



```

average_t_value times_larger_than_threshold    rsquare
1          3.146083                0.6086957 0.01748063
average_factor_gain_year volatility_factor_gain
1              1.078987                7.054694

```

The regression result doesn't clarify much ambiguity of this factor. In consideration of the remarkable difficulty of obtaining the raw data for this factor, it doesn't qualified to be a decent factor.

Conclusion

Leverage factor has uncertain effect on stock performance. Besides, its data is hard to obtain. Therefore I suggest leaving no risk exposure to this factor.

Leverage			
WindCode:	longdebttoworkingcapital,netdebttoev,longdebttolongcapital,debttoassets		
Effective:	Unclear	Effective%:	61%
Way Of Effect:	No significant influence	FactorGain:	1%
MissingData%:	7590/26334	t-value:	3.1

Trait Factors: News

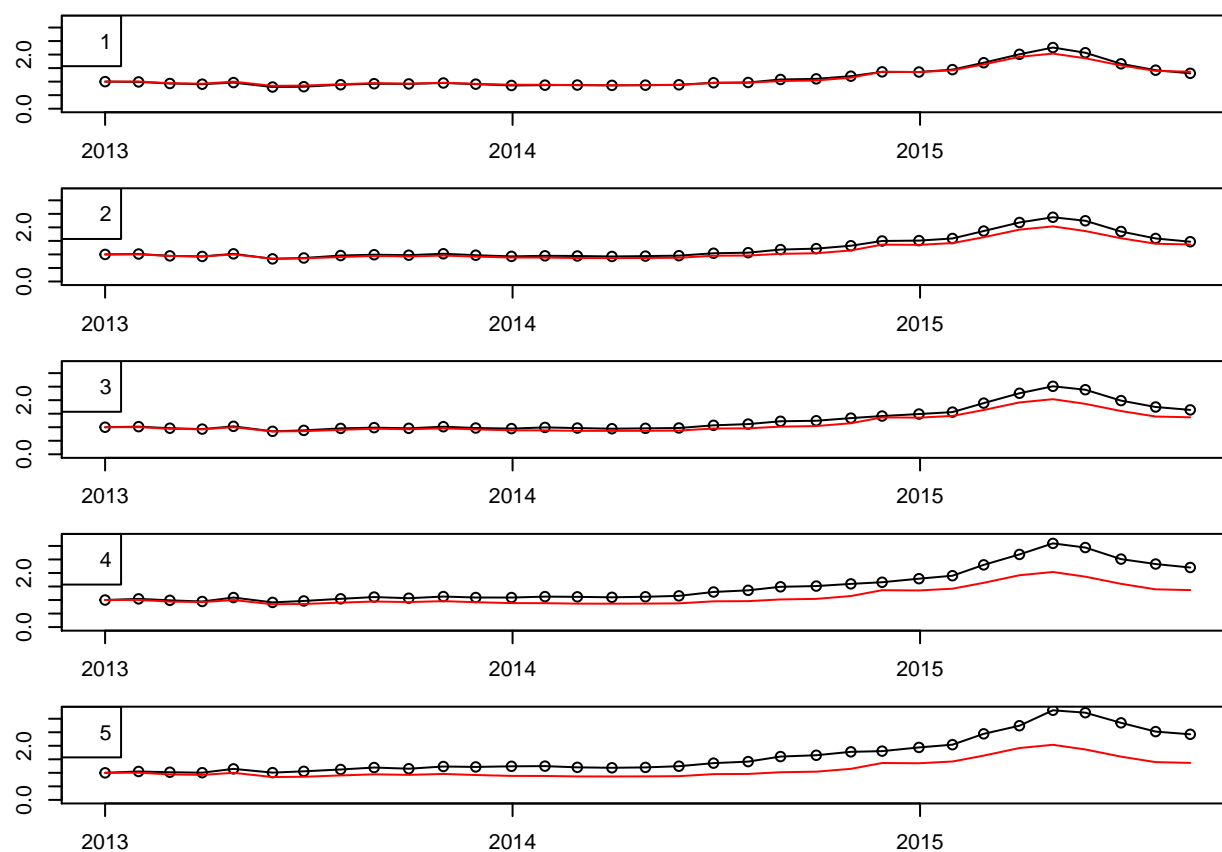
What is News

News is the factor measuring the market heat level of a particular stock. It includes data from [XueQiu](#), a widely accepted and frequently used online forum to discuss stocks in Chinese stock market. The News factor is consisted of four sub factors including the newly increased population who focuses, discusses and shares trade history about this stock during last observation period. Also, the average inventory per account is taken into consideration since it has been proved by much experience as a useful factor.

The factor is a linear combination of sub factors and the weight of each subfactor is determined by an optimization model whose target is to maximize the explanatory power of the new generated factor.

sectioned portfolio backtest result

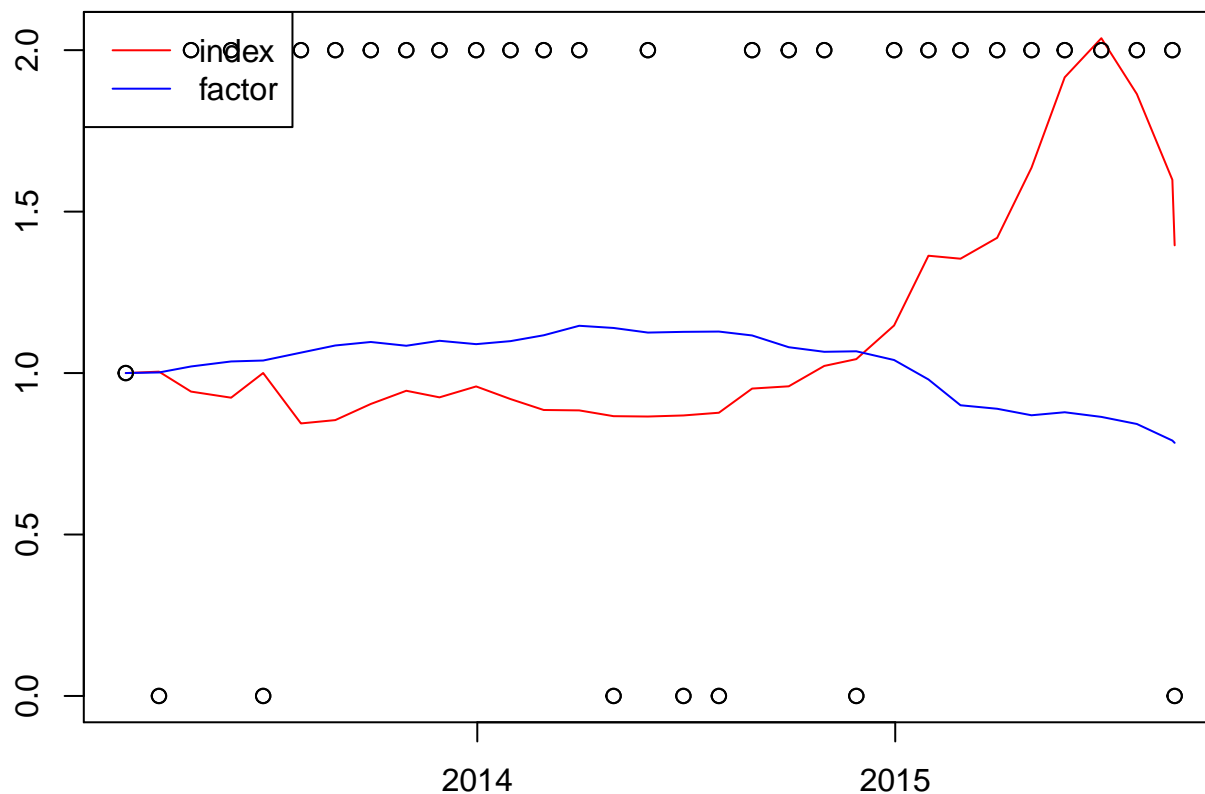
[1] "news"



	[,1]	[,2]	[,3]	[,4]	[,5]
gain	1.3070402	1.4695820	1.641385	2.2022952	2.4247232
	0.3914615	0.4321049	0.466475	0.6313889	0.6861911

The portfolio backtest result indicates that high marks in News factor obviously lead to better payoff. What's even better is that the higher the factor indication, the more the payoff.

The cross-section regression result



```

average_t_value times_larger_than_threshold    rsquare
1         4.676988                0.7741935 0.03862131
average_factor_gain_year volatility_factor_gain
1             -8.645025                8.722447

```

Conclusion

This factor is a rather new one, the data offered is somehow messy and its real effect remains much unclear. Since the backtest and the regression model offer opposite conclusions, I put my final conclusion as unclear for this factor.

However, its limited history does promise some potential of it becoming a profitable alpha factor. Therefore I suggest to carefully monitor this factor while currently leave no risk exposure to it.

News

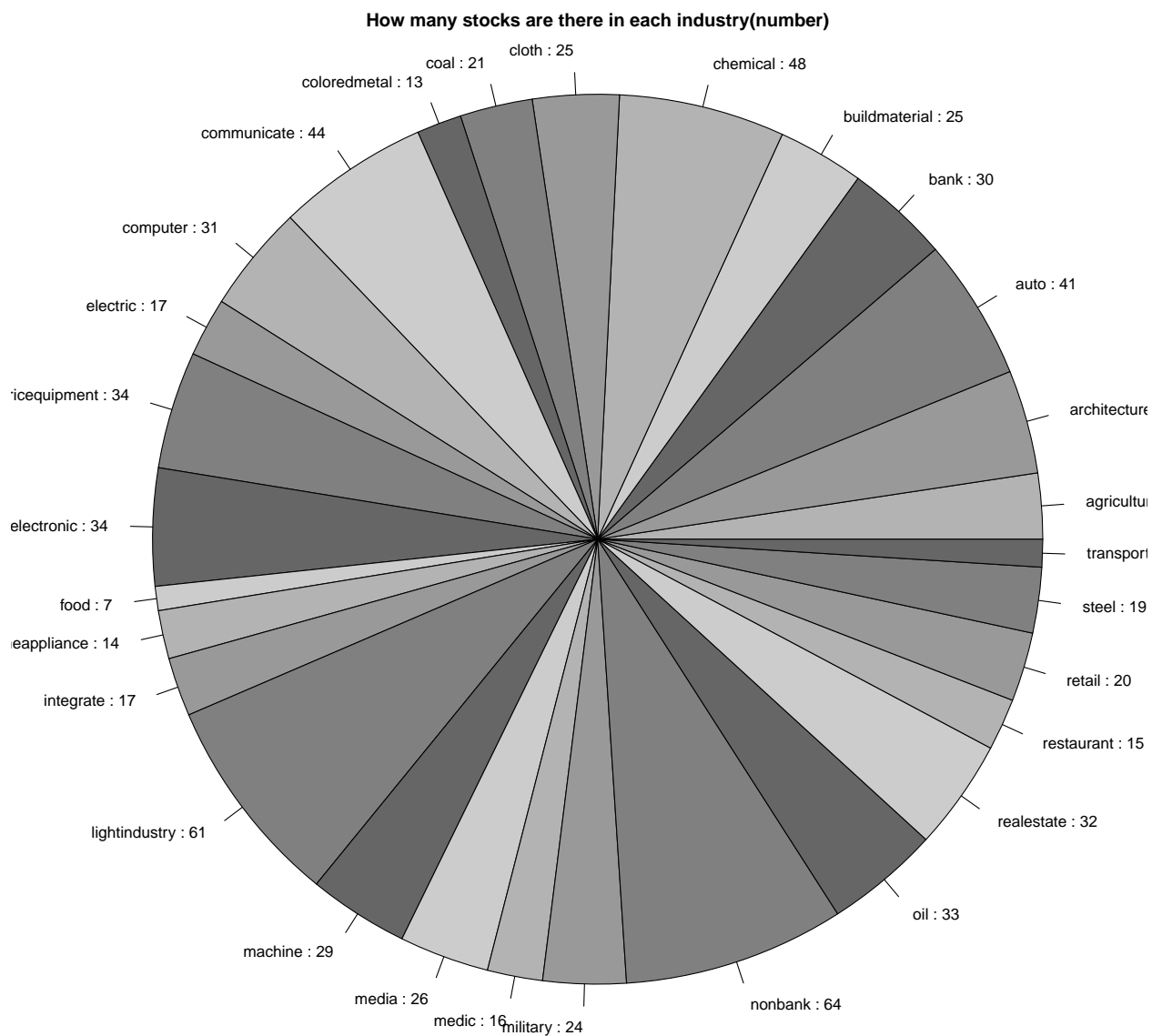
WindCode:	xq-WOW-focus,xq-WOW-comments,xq-WOW-shares,holder-avgpct	Effective%:	77%
Effective:	Unclear	FactorGain:	-9%
Way Of Effect:	Unclear	T-value:	4.7
MissingData%:	14400/26334		

Industry factors

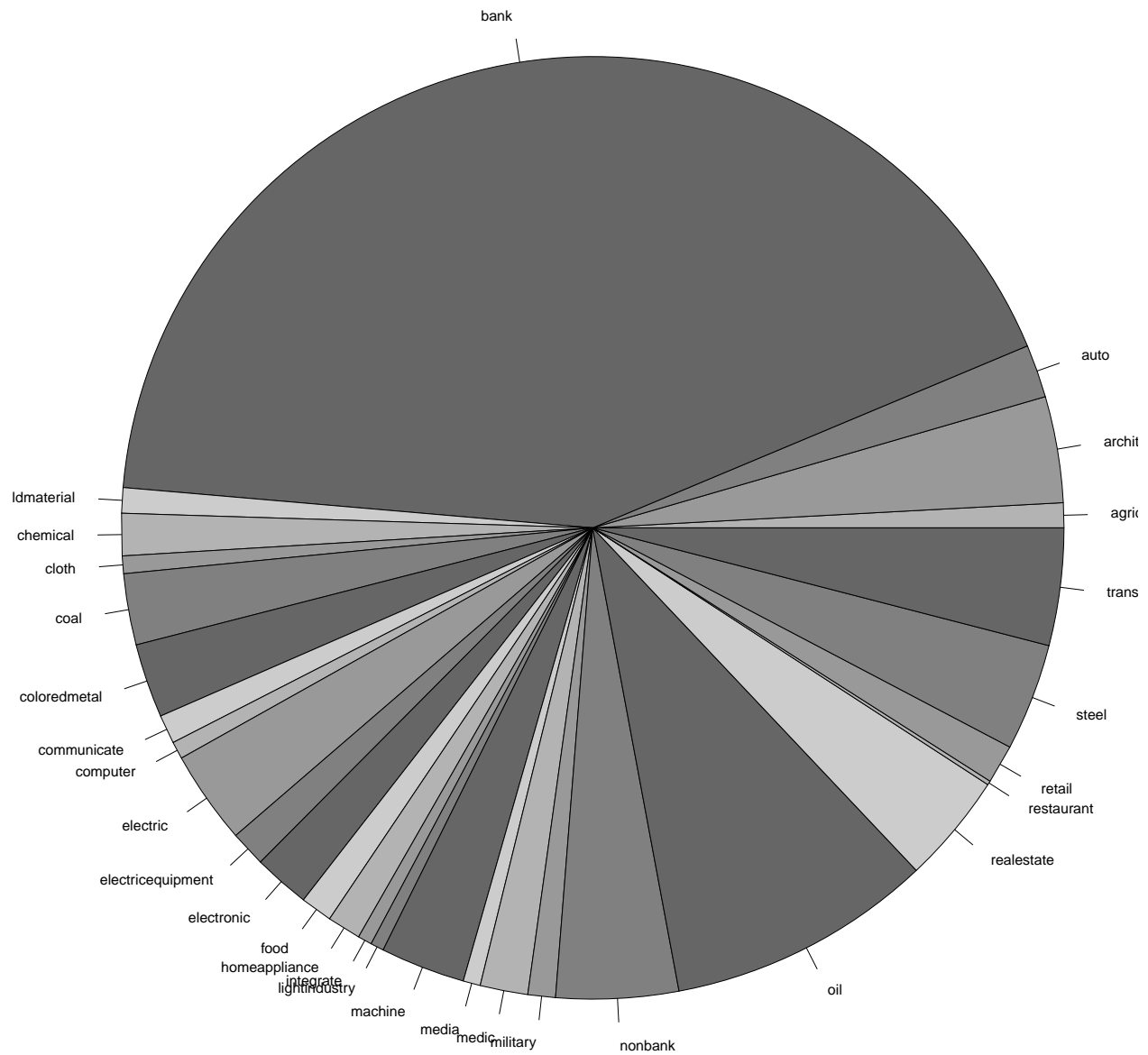
What is industry factors

Industry factors are set on the reference of the CITIC primary industry classification, which includes 29 industries. I treat it as a factor variable. That is, if this stock belongs to the first industry class than it will get “yes” for the first industry variable and “no” for the rest 28 of them.

The first glance at industry factors

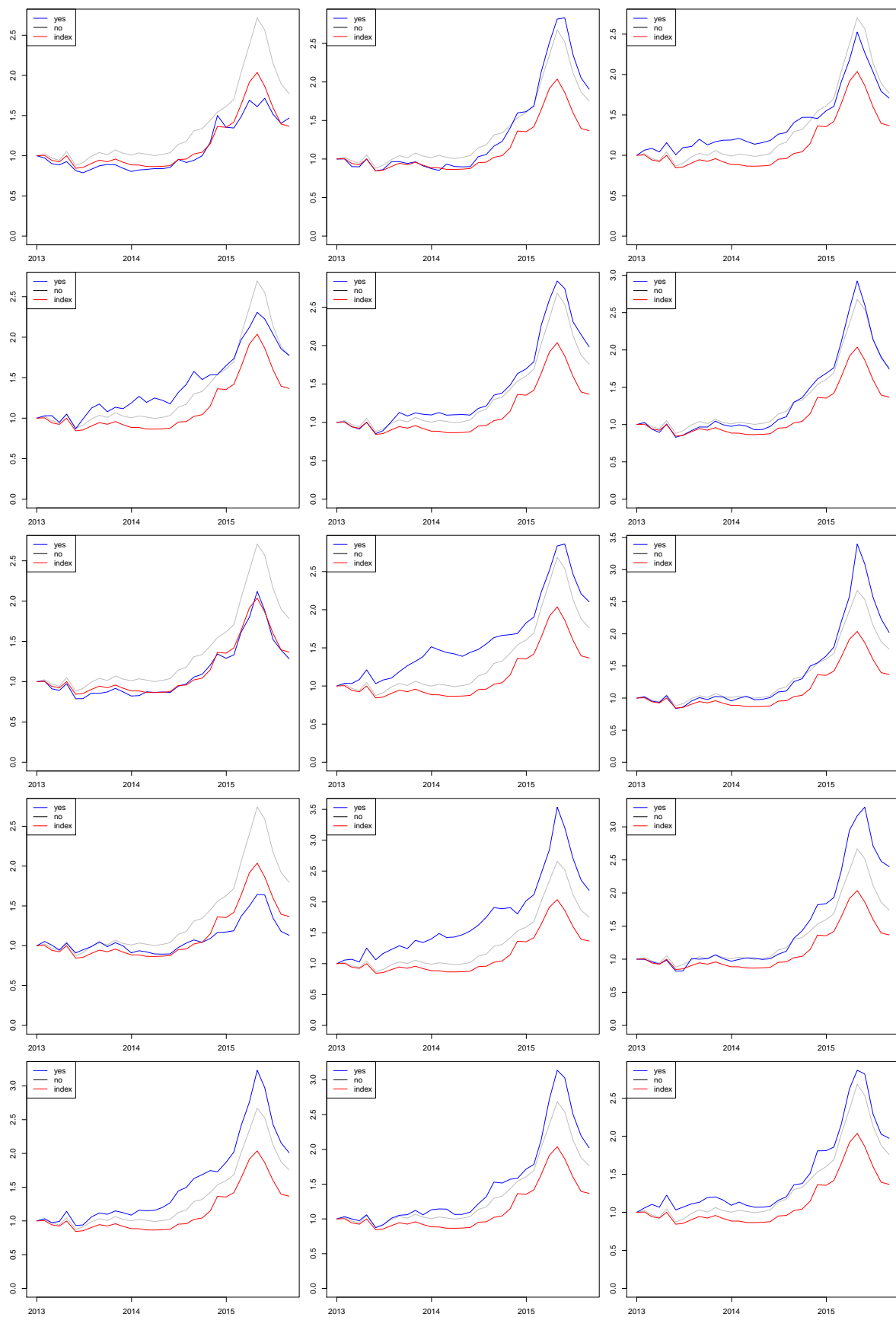


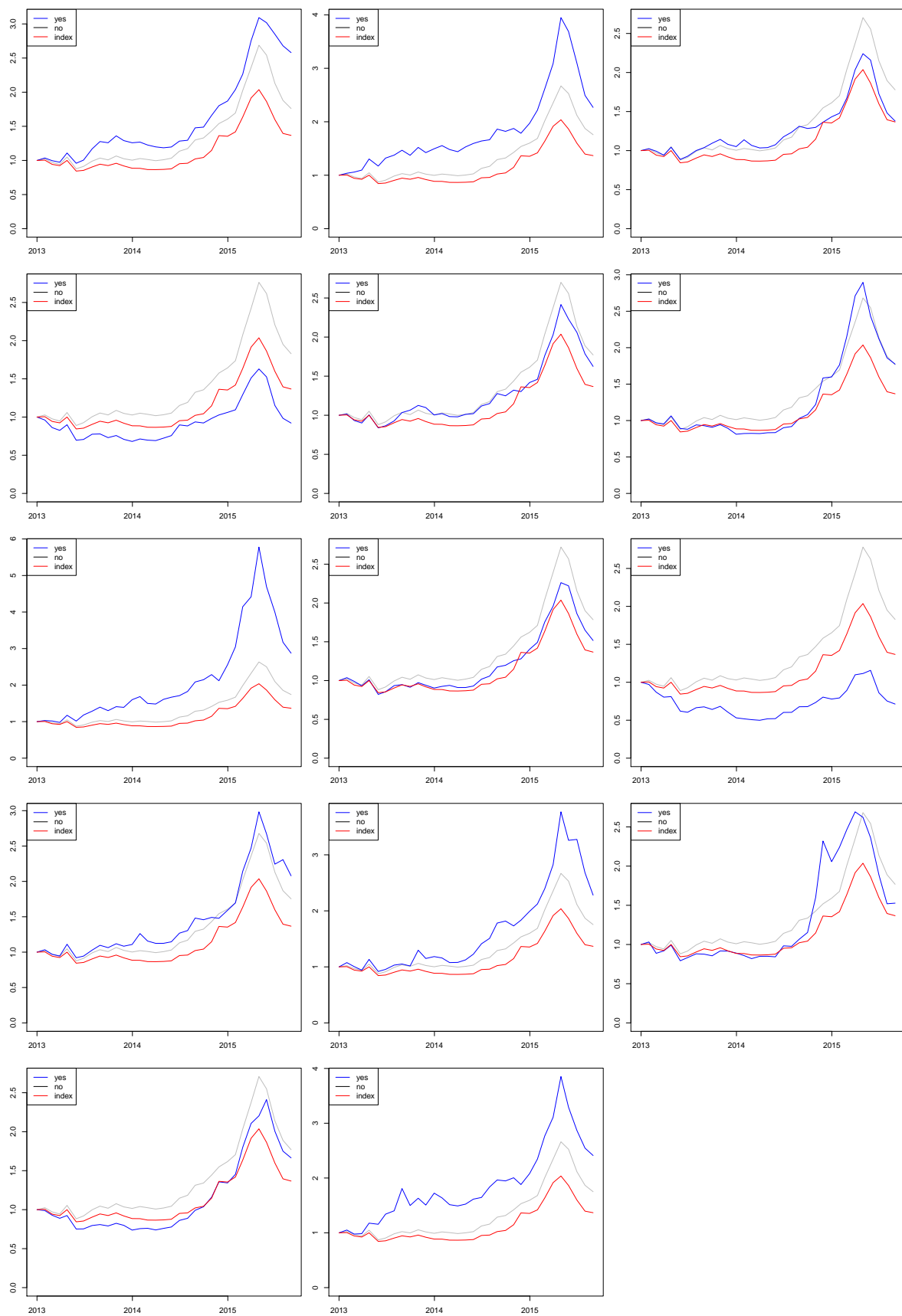
How many stocks are there in each industry(Market Value)



sectioned portfolio backtest result

For each factor, we group our samples into two portfolios: those stocks in this industry and all the others. And we observe their performance.



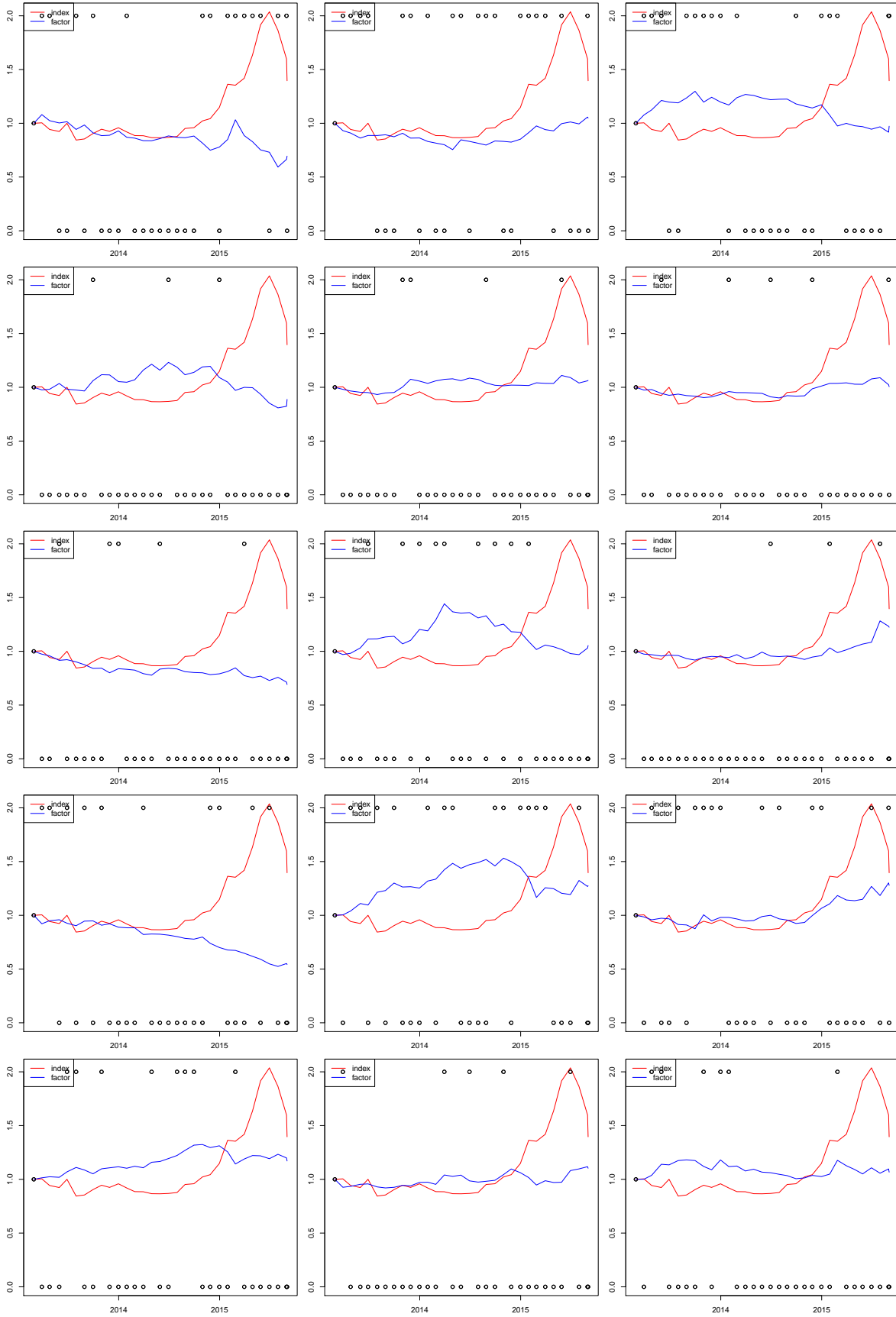


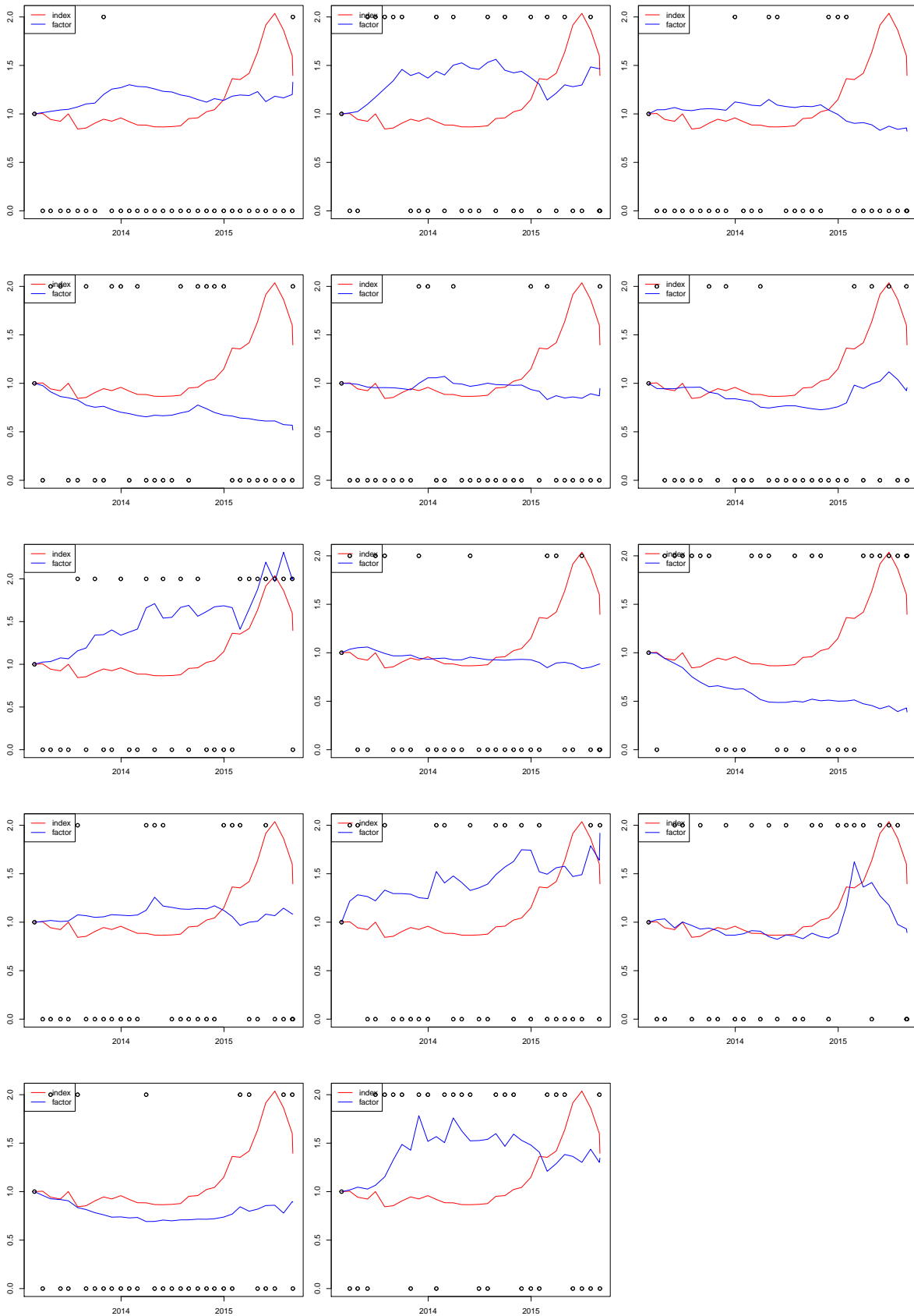
What does the plot mean The blue line is the portfolio consist of stocks in that particular industry while the grey one is those outside this industry. The red line is the reference index of CSI 800. **Unlike CSI 800 which determines each component stock's weight according to their market value, the portfolios of mine give equal weights to each and every component stock. Hence the influence of size factor can be removed.**

As we can tell from all plots, the industry factors do have obvious effect on stock payoff. However the way of influence varies with time and industries. It suggests that industry is a powerful yet not tame factor. The best option might be trying to exclude the influence of it rather than manipulating it.

The cross-section regression result

We then use single factor regression model to precisely determine the significance level of industry factor.





What does the plot mean The blue line is the performance of the factor yield generated from the regression model. The red line shows the performance of the referece index. Those points indicate whether this factor is effective at that time period or not. A point on the top edge suggests effective while a point at bottom suggests that the factor doesn't significantly influence the stock performance. A point in the middle suggests a missing value.

It's clear to see that industry performance vary largely from industry to industry. It agrees with the result of the backtest result that industry factor is too dangerous to manipulate. The best option is to maintain an industry-neutral portofolio.

Conclusion

Industrial factor is an important component of factor model. However, it's too dangerous to play with. I strongly suggest to remain industry-neutral at all time.

Industry			
WindCode:	Citic	T-value:	2(avg)
Effective:	Unclear	Effective%:	50%(avg)
Way Of Effect:	Unclear	FactorGain:	Unclear
MissingData%:	0/23142		

Part Two: The Neutral Strategy Report

What does this part try to do

Establishing a profitable neutral strategy is the main pursuit of this report. In order to achieve risk-neutral, a short position in stock index future market is taken. However, due to the rather limited choice of available futures to choose from, it's hard to find a perfect future. Actually, only three futures are available. They are IF whose underlying index is [HS 300](#), IH whose underlying index is [SSE 50](#) and IC whose underlying index is [CSI 500](#). In order to offer abundant stocks to choose from, I use both IC and IF to hedge and therefore set the initial stock pool to choose from as all components of **CSI 800**, which is the combination of **HS 300** and **CSI 500**. The weight distributed on each future contract is determined by their weight in **CSI800**. I then calculate the indications of **CSI 800 Index** on each and every factor. After that, an optimization model is established to decide which stock to hold and how much position should be held. Historical data is used to solve this model by backtesting the strategy generated by this model. Optimization goal is to maximize the payoff and constraints include matching each factor's risk exposure of chosen long positions to be equal to that of short position's (**CSI 800's**). The exposure of industry factors and most factors are set to be zero while the risk exposure of those factors considered as profitable alpha factor by the first part of this report are allowed to deviate in a limited degree.

Portfolio establishment strategy

Unlike most research within multi factors area which put most of effort in picking factors, I emphasize the importance of portfolio weight distribution. Only by carefully distributed portfolio weight can the strategy achieve a decent hedge effect and thus limit the risk.

In order to hedge the risk in portfolio's long positions, we actually need an independent risk model. I have started to study this model on the basis of [Barra Risk Model](#). However, it will not be covered until next report. **In this report, we don't estimate the variance of factors, instead, we use the factor loadings as its risk exposure and try to match the factor loadings of our portfolio with the short position's.** (That's CSI 800 in our example.)

Method used in this model

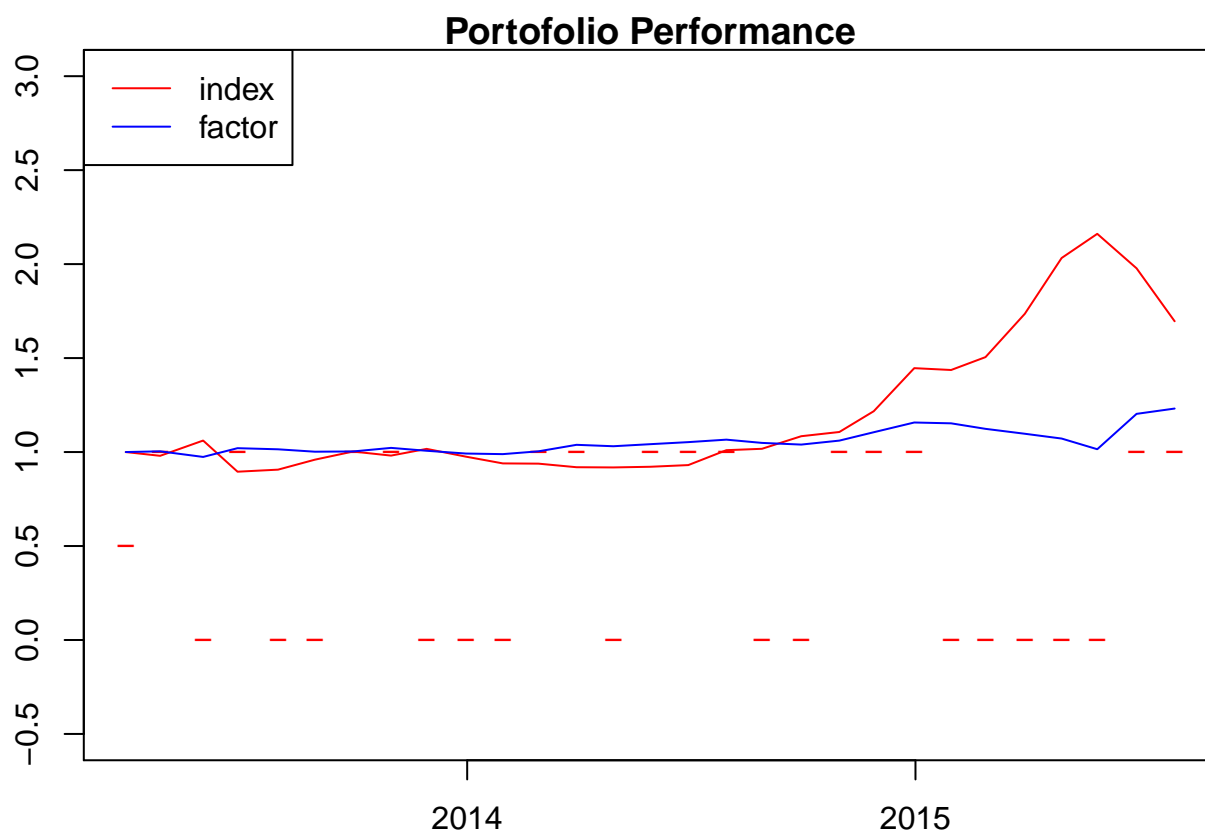
Optimization method is applied to determine the weight distribution of assets in our portfolio. In this report, I am only going to cover a rather naive optimization which optimizes the weight distributed to each stock and set the optimization goal as to maximize the portfolio return while maintaining neutral in industry factor and trait factors. However, those effective alpha factors are allowed to deviate certain level from completely neutral.

Single factor exposure backtesting

First, I'm only going to show several examples in which all factors but one is being set to be neutral. The blue line in plot shows the performance of the portfolio determined by the optimization method. The "What does the plot mean" section will offer more details about the plots and how to interpret the outcomes.

Backtest Result

Example: Value Exposure Model

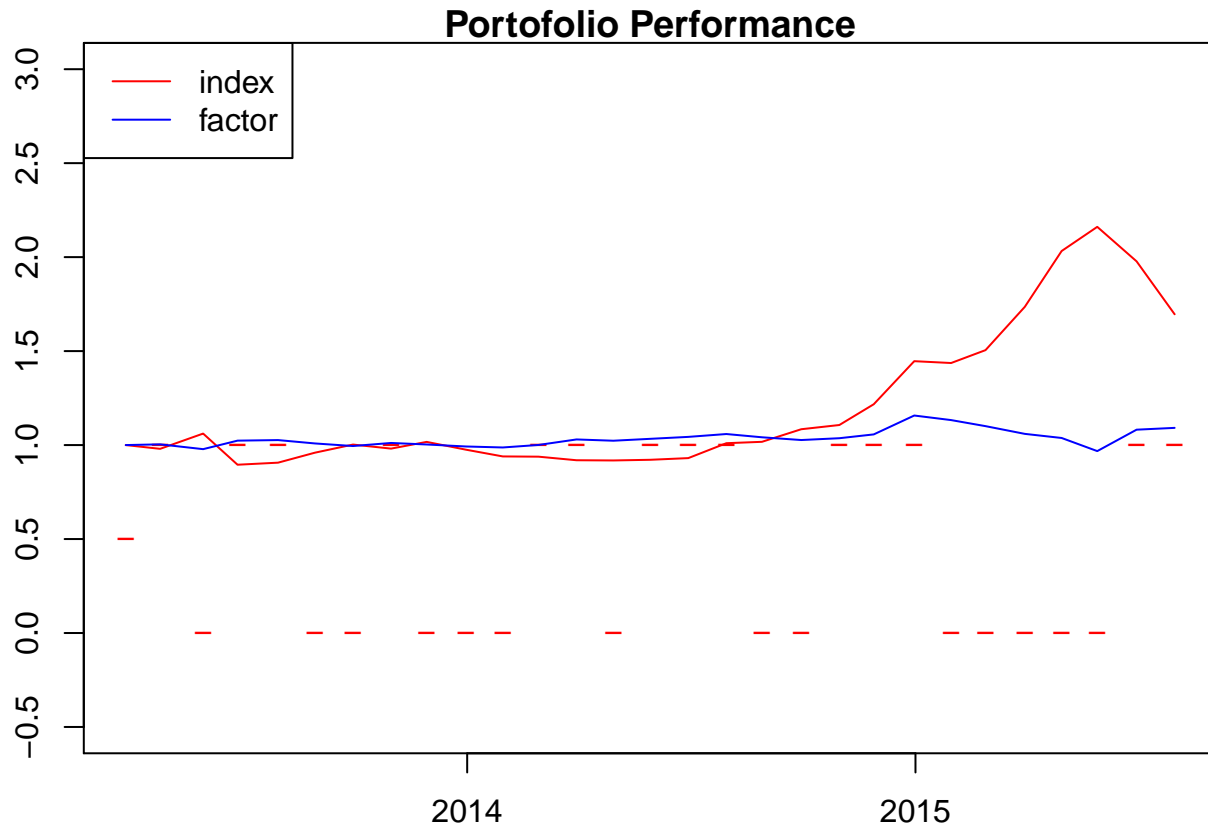


	gain_extra	win_ratio	final
1	0.008414188	0.5172414	1.244011

What does the plot mean These plots shows the performance of our portfolios. The blue is the performance of our portfolio. The redline is the reference performance of index. The red dot indicates whether our portfolio performs stronger or weaker than index at that particular month. A dot above means porto stronger than index, a dot below means porto performs weaker than index and a dot in middle means porto performs exactly the same as index. The table below the plots shows the performance in number. Gain_extra is the average monthly extra gain of our porto. Win-ratio indicates for how many periods is our porto performing stronger than index. The last row shows the final value of our porto.

Value Exposure Portfolio let the factor value to have a risk exposure within minus one to one standard deviation. The result shows that this porto doesn't perform so good. Despite the relative small fluctuation, it fails to get enough profit and the win ratio is not satisfying. The turning rate of this porto is 279% with in 32 months.

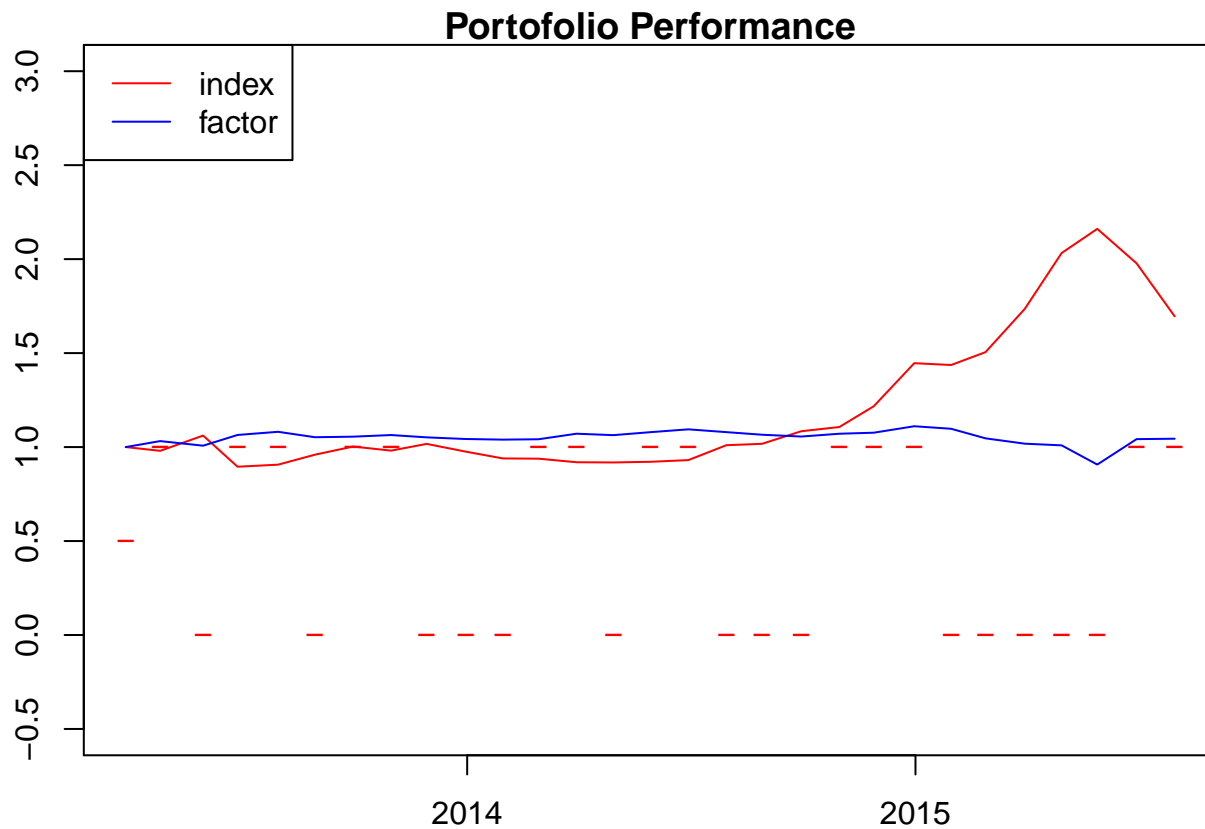
Example: Expectation Exposure Model



```
gain_extra win_ratio    final
1 0.002994053 0.4827586 1.086828
```

Expectation Exposure Portfolio let the factor expectation to have a risk exposure within minus one to one standard deviation. The result shows that this porto doesn't perform so good. Despite the relative small fluctuation, it fails to get enough profit. The turning rate of this porto is 209% with in 32 months.

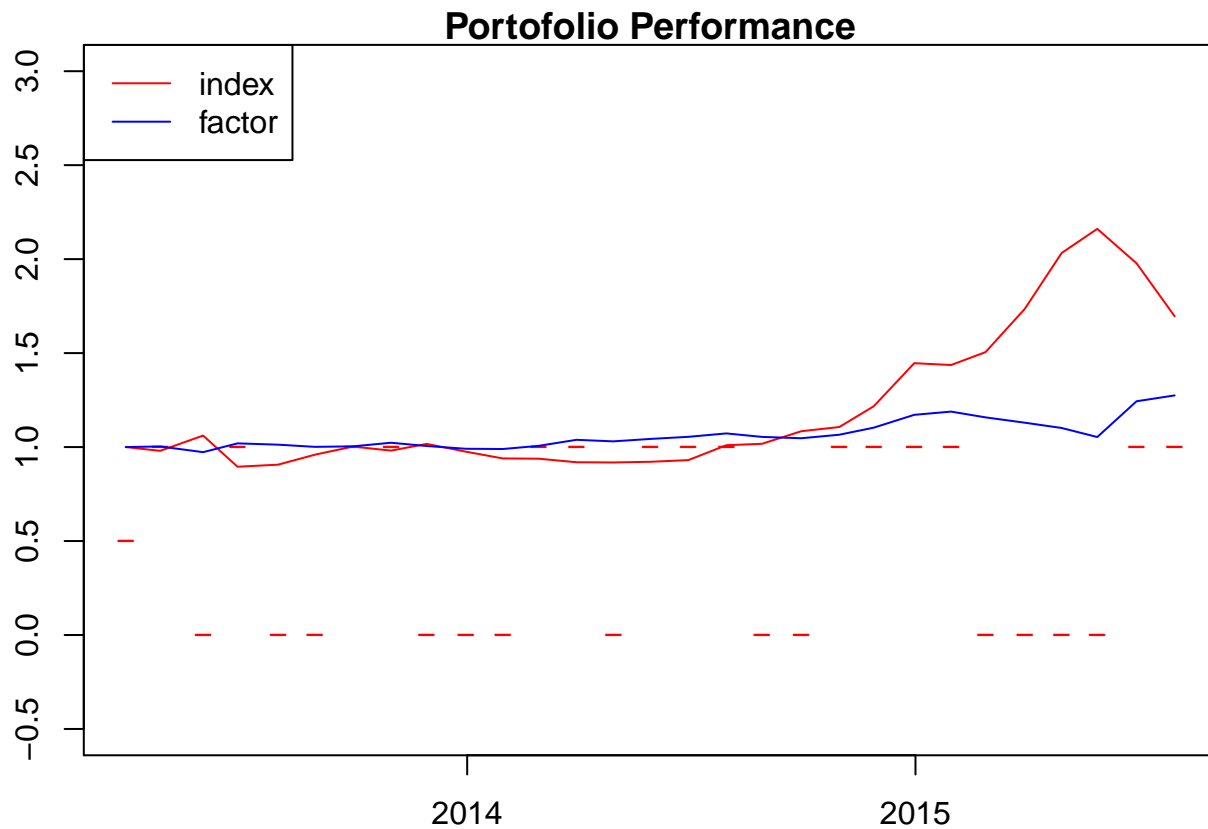
Example: Momentum Exposure Model



```
gain_extra win_ratio final
1 0.001601034 0.5172414 1.04643
```

Momentum Exposure Portfolio let the factor momentum to have a risk exposure within minus one to one standard deviation. The result shows that this porto performs OK. It generates 54% of profit in 33 months and the win ratio is close to 60%. However, the turning ratio of this porto is up to 737% which is too high.

Comprehensive Model Performance



```
gain_extra win_ratio final
1 0.009814558 0.5517241 1.284622
```

This Portfolio is the final outcome of this report. It lets the effective alpha factors we determined in the first part to have a risk exposure within minus one to one standard deviation. The result shows that this porto performs rather well. It has created almost 30% profit in 32 months. And the turing rate is 354% in 32 months which is acceptable.

Conclusion and further research plan

This report is the first report of the multi-factor research. It mainly establishes the very foundation of multi-factor analyzation system. The strategy offered here hasn't been carefully studied and those factors used aren't those most up to date.

The following research will focus on three aspects:

- First, a structured risk model is going to be included in our system. I will build a risk model and thus distinguishing between risk factors and profit factors. By applying the risk model, the constraint of risk exposure will be much better and it shall give us more confident in amplify alpha gain.
- Second, new and rarely discussed alpha factors are going to be discovered. Since the profit of this model mainly comes from effective alpha factors. And those factors being studied by lots of people will quickly lose the ability to generate profit. Consequently, the search for alpha factors will always be necessary.
- Third, strategy is going to be studied in a more detailed way. I may study the pattern of factors being effective and try to conclude the suitable market situations for each factor.