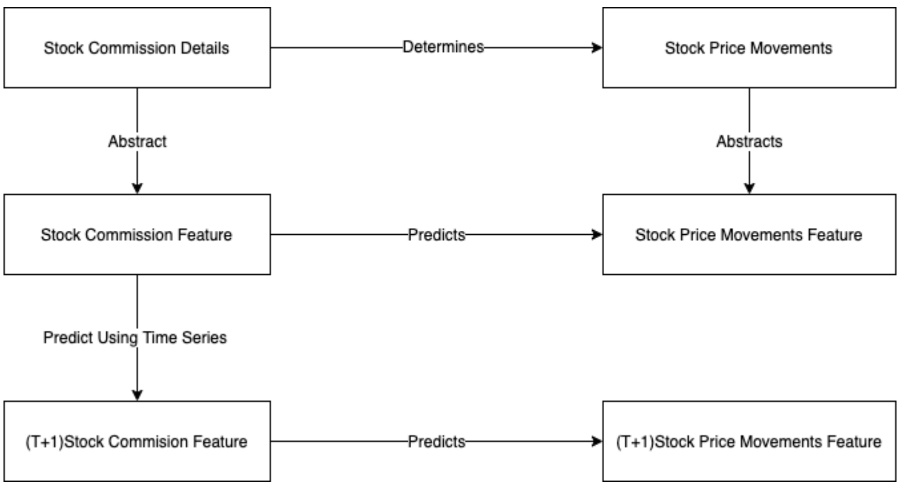
Stock timing strategy based on commission characteristics

1. A framework of stock timing algorithm based on commission characteristics



The algorithm framework is shown in the figure above: stock commission is the essential cause of stock price trends. The stock commission details combined with the stock market matching mechanism can restore the stock price trend completely and accurately, but the stock commission details themselves have a large amount of data and there are a lot of noisy data, which is difficult to handle directly Therefore, the design and establishment of the stock commission feature framework, the commission feature is an abstract summary of the commission details, so the commission feature has sufficient interpretation and regression capabilities for the stock price trend on the day. After experimentation, it is found that the commission characteristics have time series correlation, that is, the historical commission characteristics have certain predictive ability for the future commission characteristics. Based on this characteristic, the stock timing algorithm based on the commission characteristics is established. First, the relevant commission characteristics are designed and the commission characteristics are established. The time series forecasting model predicts the characteristics of future commissions, and further predicts future stock price fluctuations based on the characteristics of future commissions, so as to realize the algorithmic ability of timing.

2.Commissioned Feature Framework

A stock commission or entrustment is an entrusted transaction application sent by a stock trader to the stock exchange, including entrusted transaction information such as stock code, buying and selling direction, entrusted price, entrusted shares, etc. After entrusted to enter the stock exchange, according to the exchange's matching mechanism, if there is a counterparty order, the order will be traded with the counterparty, and the latest stock transaction price will be generated; if there is no counterparty, the order will enter the waiting transaction queue, reflecting the change in the number of shares waiting to be traded on level 2.

Orders can be divided into active orders and passive orders according to whether they are executed immediately. Active orders are placed at a price higher than the market price (buy higher than sell or sell lower than buy), and they can be executed immediately after placing the order, which has an immediate impact on the price; passive orders are commissioned at a price inferior to the market price (buy at a\\ price lower than the selling price or sell at a higher price), and enter the transaction waiting queue after placing the order, which will have a potential impact on the price in the later period.

Active orders have a more direct impact on prices than passive orders. We mainly use active orders to establish an order feature framework to analyze and predict stock price fluctuations. Active orders are divided into active buy orders and active sell orders. Statistics The time unit is day (it is also possible to establish intraday features according to the same logic to predict intraday stock price fluctuations). The statistical stock group is CSI 800 index's stocks which have listing time earlier than January 1, 2016, in order to exclude new stocks/sub-new stocks influences.

The design of the commission feature framework is as follows:

I. Active commission frequency, from the overall distribution, is divided into total active (buy/sell) frequency; according to the commission price relative to the average price of the day, it is divided into high price active (buy/sell) frequency/low price active (Buy/Sell) frequency: According to the price trend of the day, it is divided into active (buy/sell) frequency in the rising phase and active (buy/sell) frequency in the falling phase.

II. Active entrusted volume, according to the same design logic, is divided into total active (buy/sell) entrusted volume; active entrusted volume at high prices (buy/sell) and active entrusted volume at low prices (buy/sell) Entrusted volume; active (buy/sell) entrusted volume in the rising phase; active (buy/sell) entrusted volume in the falling phase.

III. The average value of active entrusted quantity is divided into the mean of total active (buy/sell) entrusted volume according to the same design logic; the mean of active entrusted volume at high price (buy/sell), and the average entrusted volume at low price (buy/sell). Selling) mean value of commissioned volume; mean value of active (buy/sell) commissioned volume in the rising phase; mean of active (buy/sell) commissioned volume in the falling phase.

IV. Other distribution statistics, such as standard deviation, kurtosis, skewness, etc., can also be designed as commissioned features. There are currently no statistics in this article.

3. The predictability test of actively commissioned characteristic time series

The commission feature needs to have the predictability of the time series. Only the feature has the predictability of the time series, can the future characteristics be predicted by the historical characteristics, and the future price trend can be predicted by the future characteristics. We test the predictability of the time series of active orders from two aspects: the statistical indicators ACF and PACF, and the time series regression effect test.

I. Time series statistical indicators ACF and PACF

ACF is an autocorrelation function that describes the autocorrelation between one observation and another observation. That is, it describes the degree of correlation between the current value of the sequence and its past value. PACF is a partial autocorrelation function or partial autocorrelation function. Intuitively, PACF only describes the direct relationship between the observed value and its lag term and adjusts the influence of other shorter lag terms (the intermediate term between the description term and the observation term). The ACF and PACF values whose absolute value exceeds 10% are considered statistically significant, that is, there is a time series correlation.

The following table shows the PACF value and ACF value of the active commission.

Table 1: Active commission feature distribute time series attributes overall.

		pacf1	pacf2	pacf3	pacf4	pacf5	acf1	acf2	acf3	acf4	acf5
ActBid	Avg	0.54	0.26	0.18	0.12	0.09	0.54	0.48	0.45	0.43	0.41
	Freq	0.76	0.23	0.14	0.10	0.08	0.76	0.68	0.63	0.60	0.57
	Sum	0.69	0.21	0.14	0.09	0.07	0.69	0.60	0.55	0.51	0.48
ActOffer	Avg	0.58	0.27	0.19	0.13	0.10	0.58	0.52	0.49	0.47	0.46
	Freq	0.79	0.23	0.12	0.09	0.08	0.79	0.72	0.67	0.63	0.60
	Sum	0.77	0.21	0.12	0.08	0.07	0.77	0.68	0.63	0.59	0.56

Table 2: The feature of active commissions distributed time series attributes according to high and low price points.

		pacf1	pacf2	pacf3	pacf4	pacf5	acf1	acf2	acf3	acf4	acf5
ActBid	LowerAvg	0.48	0.26	0.18	0.13	0.10	0.48	0.43	0.41	0.39	0.37
	LowerFreq	0.60	0.28	0.18	0.12	0.09	0.60	0.54	0.50	0.48	0.45
	LowerSum	0.61	0.24	0.15	0.10	0.07	0.61	0.53	0.49	0.45	0.43
	UpperAvg	0.45	0.22	0.16	0.11	0.09	0.45	0.38	0.36	0.34	0.32
	UpperFreq	0.51	0.24	0.16	0.12	0.09	0.51	0.44	0.41	0.39	0.37
	UpperSum	0.50	0.23	0.15	0.10	0.08	0.50	0.43	0.39	0.36	0.34
ActOffer	LowerAvg	0.47	0.24	0.18	0.13	0.10	0.47	0.42	0.40	0.38	0.36
	LowerFreq	0.57	0.27	0.18	0.11	0.09	0.57	0.51	0.48	0.45	0.43
	LowerSum	0.58	0.27	0.17	0.11	0.08	0.58	0.52	0.48	0.45	0.43
	UpperAvg	0.51	0.27	0.19	0.14	0.11	0.51	0.46	0.45	0.43	0.41
	UpperFreq	0.55	0.27	0.17	0.12	0.09	0.55	0.50	0.46	0.44	0.42
	UpperSum	0.57	0.24	0.15	0.11	0.08	0.57	0.49	0.45	0.42	0.40

Table 3: Active commission features distribute time series attributes according to intraday trend distribution

		pacf1	pacf2	pacf3	pacf4	pacf5	acf1	acf2	acf3	acf4	acf5
ActBid	LowerAvg	0.48	0.26	0.18	0.13	0.10	0.48	0.43	0.41	0.39	0.37
	LowerFreq	0.60	0.28	0.18	0.12	0.09	0.60	0.54	0.50	0.48	0.45
	LowerSum	0.61	0.24	0.15	0.10	0.07	0.61	0.53	0.49	0.45	0.43
	UpperAvg	0.45	0.22	0.16	0.11	0.09	0.45	0.38	0.36	0.34	0.32
	UpperFreq	0.51	0.24	0.16	0.12	0.09	0.51	0.44	0.41	0.39	0.37
	UpperSum	0.50	0.23	0.15	0.10	0.08	0.50	0.43	0.39	0.36	0.34
ActOffer	LowerAvg	0.47	0.24	0.18	0.13	0.10	0.47	0.42	0.40	0.38	0.36
	LowerFreq	0.57	0.27	0.18	0.11	0.09	0.57	0.51	0.48	0.45	0.43
	LowerSum	0.58	0.27	0.17	0.11	0.08	0.58	0.52	0.48	0.45	0.43
	UpperAvg	0.51	0.27	0.19	0.14	0.11	0.51	0.46	0.45	0.43	0.41
	UpperFreq	0.55	0.27	0.17	0.12	0.09	0.55	0.50	0.46	0.44	0.42
	UpperSum	0.57	0.24	0.15	0.11	0.08	0.57	0.49	0.45	0.42	0.40

It can be seen from Table 1 to Table 3 that the characteristic values basically have significant time series correlation, acf1~acf5 are all significantly effective, and pacf value is less than 0.1 after 5 lag units. Therefore, these active commission features can pass time Sequence modeling prediction.

II. Actively commissioned characteristic time series modeling and forecasting effects

Based on the above-mentioned active commission feature time series attributes, we build a time series regression model for each feature to test its regression effect. Considering that the PACF5 of each characteristic value is basically close to or less than 0.1, we adopt the AR(5)(Autoregressive mode) model.

Taking into account that different stocks often have different characteristic scales, such as large-cap stocks with large characteristic value and small-cap stocks with large volatility, we first normalize the characteristic value of individual stocks. The normalization method is (Original value-minus rolling mean)/rolling standard deviation. After normalization, all stocks have the same mean and standard deviation for the same feature, and then build a time series AR5 model. For Subsequent optimization, we can consider more complex models, such as LSTM, etc. Deep learning model.

The modeling effect of each active commission feature time series is as follows (Correlation represents the correlation coefficient between the predicted value and the true value, and Score represents the model regression coefficient R2):

		Correlation	Score
ActBid	Avg	0.370	0.128
	Freq	0.527	0.259
	Sum	0.521	0.233
ActOffer	Avg	0.360	0.126
	Freq	0.556	0.297
	Sum	0.581	0.324
ActBid	LowerAvg	0.330	0.095
	LowerSum	0.443	0.189
	UpperAvg	0.318	0.093

	UpperSum	0.333	0.111
ActOffer	LowerAvg	0.289	0.079
	LowerSum	0.396	0.164
	UpperAvg	0.314	0.093
	UpperSum	0.393	0.154
ActBid	DecreaseAvg	0.397	0.151
	DecreaseSum	0.539	0.268
	IncreaseAvg	0.328	0.102
	IncreaseSum	0.402	0.146
ActOffer	DecreaseAvg	0.325	0.103
	DecreaseSum	0.460	0.214
	IncreaseAvg	0.372	0.134
	IncreaseSum	0.535	0.270

It can be seen from the results in the above table that the models basically have significant effects. The regression coefficient R2 of the model is between 10%-30%, which can prove the effectiveness of the model. But for the algorithm, the index with a better reference effect is the regression coefficient between the predicted value and the true value. Its value is basically between 30%-50%, which is significantly effective.

4. Building positions based on features and feature earnings statistics

I. Feature Position Conversion

Based on the prediction results of the time series model, we have obtained the predicted values of the active commission features in the next period. Now we need to predict the stock positions of the next period based on the predicted values of these features. We assume that there are two positions, namely long positions and short positions. If the stock price fluctuation in the next period is positive, the long position will be profitable and the short position will lose. If the next period of stock price fluctuation is negative, the long position will lose and the short position will gain. Shorting is currently not allowed in the market, so there is no reasonable short position. Therefore, the current short position is only used to verify the validity of the feature and will be excluded in the subsequent strategy construction.

Because the strategy type is a timing strategy, it is assumed that the proportion of long positions and short positions is 50% to 50%, that is, the long and short positions remain the same. Therefore, according to the distribution of predicted characteristic value, the part of predicted characteristic value higher than the average represents bull positions , The part of the predicted characteristic value below the mean represents the bear distribution. Considering that the characteristic values are normalized before the prediction, it is assumed that the average characteristic value is 0, that is, the predicted characteristic value greater than 0 indicates that the next period is a long position, and the predicted characteristic value is less than 0 indicates that the next period is a short position. In this way, a long-short combination of individual stocks is constructed.

II. Feature validity test and benefits

According to the above rules, each feature constructs a time series AR(5) model, predicts the feature value of the next period based on the historical feature value, and then regards the predicted feature value greater than 0 as a long position according to the feature position conversion principle in the previous section. The predicted feature value with a value less than 0 is regarded as a short position, a long-short combination of individual stock dimensions is constructed, and the annual yield and Sharpe rate of individual stocks are calculated. The time period covers March 2017 to December 2020. The average return and Sharpe rate data of each feature on the stock group (CSI 800 stock group) are shown in the following table.

		Annualized income	Sharpe Ratio
(ActBid,	Sum)	16.23%	1.16
(ActBid,	Freq)	15.32%	1.01
(ActBid,	Avg)	13.75%	1.04
(ActOffer,	Sum)	16.45%	1.10
(ActOffer,	Freq)	14.51%	0.93
(ActOffer,	Avg)	14.34%	1.07
(ActBid,	LowerSum)	14.11%	1.04
(ActBid,	LowerFreq)	12.03%	0.82
(ActBid,	LowerAvg)	13.39%	0.99
(ActBid,	UpperSum)	15.12%	1.08
(ActBid,	UpperFreq)	14.37%	0.98
(ActBid,	UpperAvg)	12.12%	0.90
(ActOffer,	LowerSum)	12.63%	0.89
(ActOffer,	LowerFreq)	10.59%	0.72
(ActOffer,	LowerAvg)	13.24%	1.00
(ActOffer,	UpperSum)	16.27%	1.13
(ActOffer,	UpperFreq)	15.03%	1.03
(ActOffer,	UpperAvg)	13.14%	0.99
(ActBid,	DecreaseSum)	16.40%	1.11
(ActBid,	DecreaseFreq)	14.32%	0.92
(ActBid,	DecreaseAvg)	14.14%	1.07
(ActBid,	IncreaseSum)	14.57%	1.07
(ActBid,	IncreaseFreq)	13.82%	0.96
(ActBid,	IncreaseAvg)	13.38%	0.99
(ActOffer,	DecreaseSum)	15.19%	1.03

(ActOffer, DecreaseFreq)	13.98%	0.86
(ActOffer, DecreaseAvg)	13.59%	1.03
(ActOffer, IncreaseSum)	14.80%	1.09
(ActOffer, IncreaseFreq)	13.15%	0.90
(ActOffer, IncreaseAvg)	13.13%	1.01

It can be seen from the above table that all features have achieved positive feature returns from 2017 to 2020, the average return is about 10%, and the average Sharpe rate is about 1. Because it is the time series of individual stocks, we believe that the Sharpe rate 1 can prove that the feature has a certain timing validity

5. strategy construction and Backtesting

I. Multi-dimensional feature position combination

At present, this strategy has counted a total of about 30 commission characteristics. According to the conversion principle of characteristic positions in the previous section, each characteristic will predict a position in the next time period. The strategy uses the algorithm of comprehensive average positions to calculate the positions after all features are converted and set the long position threshold (0/0.25/0.33/0.5). If the average value of the position is greater than the long position threshold, the position in the next period is predicted to be long, otherwise it is regarded as a short position.

II. Time series rolling forecast

We conduct model rolling training in a 6-month cycle. For the current month, take the feature data of the first 6 months including the current month as the training data to establish a time series model of commissioned features, for the commission of a certain month (verification of data) in the future, the commissioned features are predicted and a long-short portfolio is constructed to test the return of individual stocks' long portfolio. Take 3 or 6 months in the next month (to ensure sufficient inspection data).

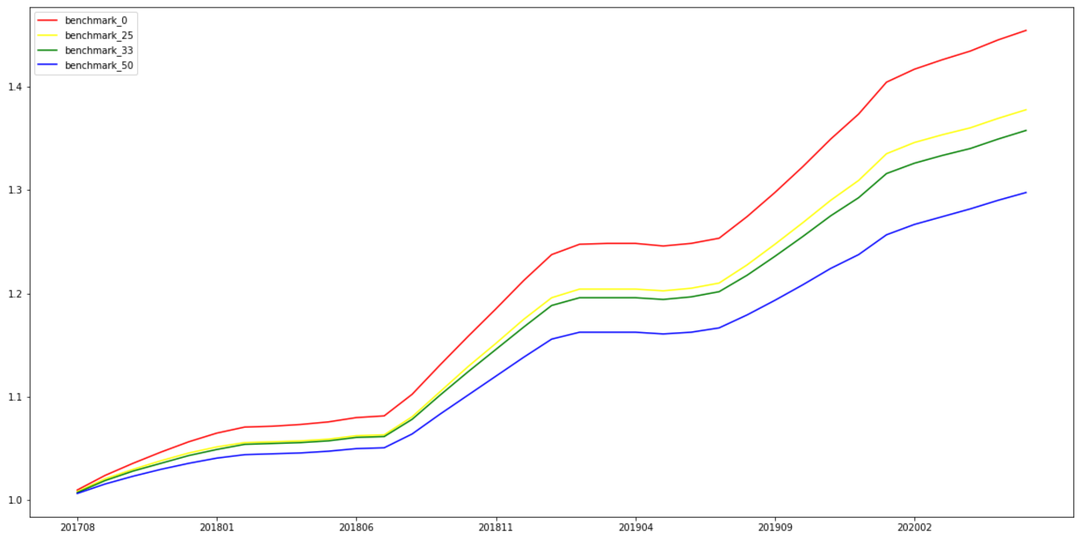
Taking June 2020 as an example, the training data includes the daily commission characteristic data of the CSI 800 stock pool (stocks listed before 2016) from January 2020 to June 2020, and the time series model is constructed one by one based on the commission characteristics. Taking the six month verification cycle as an example, the verification data is the data from July 2020 to December 2020. According to the generated time series model, the commissioned characteristics of the verification data set are predicted, and the predicted position is further determined according to the characteristic position conversion principle and the position average principle, construct a long position portfolio, and verify the return of the long portfolio within the validation set.

The backtesting results of the strategy's historical return with 6 months as the test data cycle are shown in the following table, where the long threshold is 0/ 0.25/0.33/ 0.5, and the return rate is the annualized return rate:

month	profit0	profit25	profit33	profit50
201708	0.12	0.10	0.09	0.08
201709	0.17	0.15	0.14	0.11
201710	0.14	0.11	0.11	0.09
201711	0.13	0.10	0.09	0.08
201712	0.12	0.09	0.09	0.07
201801	0.10	0.07	0.07	0.06
201802	0.07	0.05	0.06	0.04
201803	0.01	0.01	0.01	0.01
201804	0.02	0.01	0.01	0.01

201805	0.03	0.02	0.02	0.02
201806	0.05	0.04	0.04	0.03
201807	0.02	0.01	0.01	0.01
201808	0.25	0.21	0.20	0.16
201809	0.34	0.29	0.28	0.23
201810	0.33	0.29	0.27	0.22
201811	0.32	0.27	0.26	0.22
201812	0.33	0.28	0.26	0.22
201901	0.30	0.25	0.25	0.21
201902	0.12	0.10	0.09	0.08
201903	0.01	0.00	0.00	0.00
201904	0.00	0.00	0.00	0.00
201905	-0.03	-0.02	-0.02	-0.02
201906	0.03	0.03	0.03	0.02
201907	0.06	0.06	0.06	0.05
201908	0.25	0.21	0.19	0.15
201909	0.28	0.24	0.22	0.17
201910	0.30	0.25	0.23	0.18
201911	0.32	0.26	0.24	0.19
201912	0.29	0.23	0.21	0.16
202001	0.37	0.31	0.28	0.23
202002	0.15	0.13	0.12	0.12
202003	0.11	0.09	0.09	0.09
202004	0.10	0.08	0.08	0.09
202005	0.13	0.11	0.11	0.10
202006	0.11	0.10	0.10	0.09
Average Value	0.16	0.13	0.12	0.10
standard deviation	0.12	0.10	0.10	0.08
Sharpe Ratio	1.29	1.27	1.29	1.30

The annualized net value trend is shown in the figure below, in which we could conclude that the net value has maintained an overall upward trend.

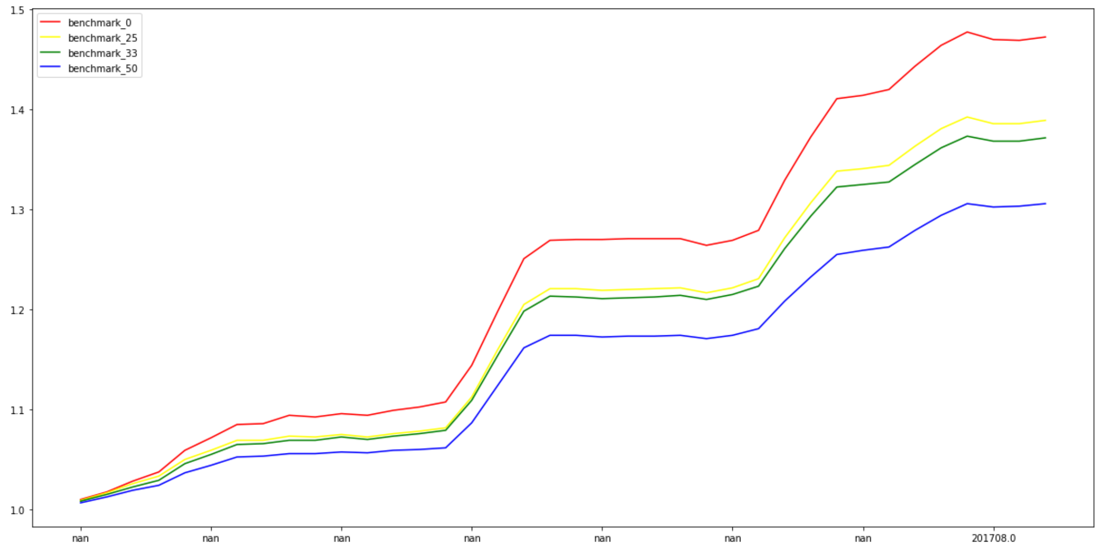


The backtest results of the strategy's historical return with 3 months as the test data cycle are shown in the following table, and the long threshold is 0/0.25/0.33/0.5:

month	profit0	profit25	profit33	profit50
201708	0.12	0.11	0.10	0.08
201709	0.09	0.09	0.08	0.07
201710	0.13	0.11	0.09	0.08
201711	0.11	0.09	0.08	0.06
201712	0.26	0.20	0.20	0.15
201801	0.15	0.11	0.11	0.09
201802	0.16	0.12	0.12	0.10
201803	0.01	0.00	0.01	0.01
201804	0.10	0.05	0.04	0.03
201805	-0.02	-0.01	0.00	0.00
201806	0.04	0.03	0.04	0.02
201807	-0.02	-0.03	-0.03	-0.01
201808	0.06	0.04	0.04	0.03
201809	0.04	0.03	0.03	0.01
201810	0.06	0.04	0.04	0.02
201811	0.44	0.37	0.36	0.30
201812	0.65	0.57	0.54	0.45
201901	0.63	0.54	0.53	0.45
201902	0.22	0.19	0.18	0.15
201903	0.01	0.00	-0.01	0.00

201904	0.00	-0.02	-0.02	-0.02
201905	0.01	0.01	0.01	0.01
201906	0.00	0.01	0.01	0.00
201907	0.00	0.01	0.02	0.01
201908	-0.08	-0.06	-0.05	-0.04
201909	0.06	0.06	0.06	0.04
201910	0.12	0.11	0.10	0.08
201911	0.60	0.49	0.45	0.33
201912	0.52	0.42	0.39	0.29
202001	0.46	0.38	0.35	0.27
202002	0.04	0.03	0.03	0.05
202003	0.07	0.04	0.03	0.04
202004	0.28	0.23	0.21	0.20
202005	0.25	0.21	0.20	0.18
202006	0.16	0.14	0.14	0.14
202007	-0.09	-0.08	-0.06	-0.04
202008	-0.01	0.00	0.00	0.01
202009	0.04	0.04	0.04	0.03
Average Value	0.15	0.12	0.12	0.10
standard deviation	0.20	0.17	0.16	0.13
Sharpe Ratio	0.75	0.74	0.74	0.76

The net worth trend chart is shown below:



It can be seen from the above two backtesting results that the annual average value of the strategy can reach about 15%, but with the increase of the long-position threshold, that is, the stricter long-position standard, the reduction in the number of long-position positions also leads to a decline in the return rate, but it can Achieve an annualized rate of return of more than 10%. The Sharpe rate reached about 1.3 in the 6-month verification backtesting period, and dropped to about 0.75 in the three-month backtesting period, because the reduction in the verification backtesting period led to an increase in the fluctuation of the verification samples, all of which are normal.