



High Frequency Trading Strategy Evaluation System Based on Grey Relational Analysis and Power Spectral Estimation

Youwei Chen^{*a, b}, Jianfeng Guo^{a, b}, Chao Wang^c, Nan Feng^d

^a College of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an, China, 710121

^b Computational Finance and Risk Management Research Center, Xi'an University of Posts and Telecommunications, Xi'an, China, 710121

^c School of Computer, Xi'an University of Posts and Telecommunications, Xi'an, China, 710121

^d Xi'an Medical University, Xi'an, China, 710021

chenyouwei@163.com

In order to make a comprehensive evaluation of the profit, risk prevention, profit and loss coefficient and other performance parameters of high frequency trading strategies, this paper proposes High Frequency Trading Strategy Evaluation System (SES) model based on gray relational analysis and power spectral estimation algorithm. This model uses gray relational analysis to determine the main strategy factors and power spectral estimation algorithm for strategy evaluation. Based on historical data simulation test, SES provides strategy performance evaluation report, and it also has functions of data management, strategy development, regression testing and clearing implementation.

1. Introduction

Currently, high-frequency tradings are mainly highly-integrated procedural algorithmic trading (Zou and Zhang, 2014). The strategy principles used in high-frequency trading have no fundamental difference with common trading principles, they are just models constructed through careful quantification by mathematicians or physicists who have a high level of mathematical logic ability, and then packed into strategy pools one by one through sequencing in order to establish objective criteria for trading strategy by IT engineers according to the special requirements of high-frequency trading hardware and software performance environment. But this does not mean that investors can fully abandon their subjective judgment, after all, a lot of things are difficult to quantify (Hu, 2010)(Hendershott, 2008).

Since the "flash crash" of 2010, global regulators are increasingly concerned about algorithms and high-frequency trading strategies. Recently, the US Securities and Exchange Commission (SEC) sued two high-frequency trading firms, accusing them of setting a large number of instructions which will not be executed in transaction, resulting in a false market demand and luring investors to trade under manipulation rather than the price determined by market supply (Ji, 2010; Wang, 2014). High-frequency trading distorts the market and frauds investors through advanced computer equipment and malicious trading strategies. Therefore, for regulators, strategy developers and investors, it is of great significance to design and implement a testing and evaluation system for market trading strategies.

2. Model Structure and Workflow

High-frequency trading is a kind of computer programming trading, that is to say, it is through computer and high-speed data connection technology to handle the large number of orders at a high speed, such as analyzing the small changes between bid and ask price of certain securities, or the tiny price change of a certain stock between different exchanges, quick implementation of orders, earning substantial profits in milliseconds, microseconds or even nanoseconds (Yang and Paddrik, 2012). The high-frequency trading strategy evaluation

system based on gray relational analysis and power spectral estimation algorithm (Strategy Evaluation System, SES), its function modules and performance flowchart are shown in Figure 1:

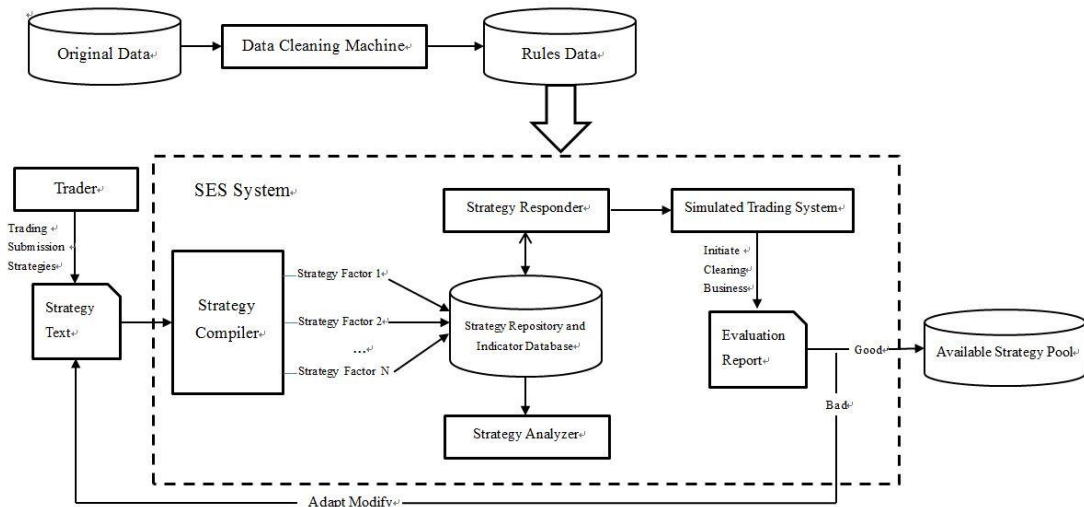


Figure 1: Evaluation system model of high frequency trading strategy

- 1) Strategy compiler module does the natural language processing of trading strategies, and analyzes and describes the strategy factors.
- 2) Strategy analyzer module identifies and discover strategy execution timing from historical data based on power spectrum estimation algorithm timely.
- 3) Strategy responder module makes operations such as placing orders, withdrawing orders according to the rules set in the rule database, and issues clearing command to simulation trading system.
- 4) Event and rule database module stores a variety of log information of storage strategy execution, including storing the strategy's primary factor information selected by gray correlation analysis algorithm, as well as the preset strategy rules.

Strategy evaluation system workflow: 1) transform the trading ideas of traders into trading strategies which can be identified by the system; 2) clean and organize the historical data, and format it into neat uniform analysis data for system use; 3) Simulate matchmaking deals; 4) give strategy evaluation analysis report. Specifically shown in Figure 2:

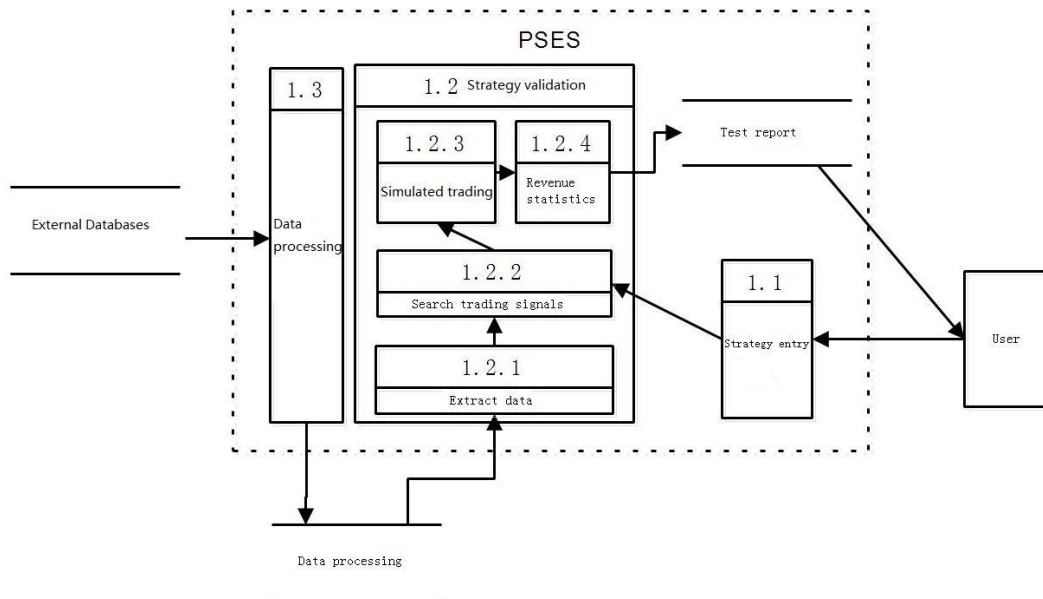


Figure 2: Work data flow diagram of SES

3. Key Algorithms

3.1 Data preprocessing and data modeling

The data SES model uses are from high-frequency trading market monthly historical data and needs to be processed through the data washer before using. It uses auto regression moving average (Auto Regression Moving Average, ARMA) algorithm to fill the default data, and sets the price threshold to exclude some stocks of higher prices (Camilo, Rostoker, 2011).

3.2 Determine the main strategy factors based on gray correlation analysis

Extract strategy factors such as key policy parameters of transaction costs, bid-ask spread, placing manner and transaction speed from the strategy description text through strategy compiler which is based on semantic rules. This paper uses gray correlation analysis to analyze and determine the main factors of input strategy, and determines the association degree by gray correlation analysis based on the similarity degree of parameter geometric curves in order to quantify the continuously changing situation problems (Yang and Wu, 2008; Wang et al., 2010). Set strategy reference sequence $X_0=(X_1, X_2, \dots, X_n)$, $Y_j=(Y_{j1}, Y_{j2}, \dots, Y_{jm})$, to indicate the m comparison sequences at the moment j, then the gray correlation coefficient is:

$$\xi_j(t_i) = \frac{\min_j \min_i |x_i - y_{ji}| + \rho \max_j \max_i |x_i - y_{ji}|}{|x_i - y_{ji}| + \rho \max_j \max_i |x_i - y_{ji}|} \quad (1)$$

ρ is the distinguish coefficient within the interval (0.1) which is generally set as 0.5. The smaller ρ is, the greater the distinguish particle. However, excessive refinement will lead to a waste of system resources, thus one reasonable approach is to adjust at any time according to the system requirements. The association degree is:

$$\gamma(x, y_j) = \frac{\sum_{k=1}^n \xi(k)}{n} \quad (2)$$

3.3 Strategy estimation algorithm based on power spectral estimation

Power spectral density estimation algorithm is mainly used to study the various changeable objects in frequency domain; it can extract the constant change states from the noise according to the limited original information, which makes it very suitable for the estimation and forecasting of the development of events. In practical application, the strategy power spectrum is the sum of the respective frequency components of specific information, the power size of each component reflects the effect ratio (Zhang Hui, 1979)]. SES model tracks all strategy event elements, and strategy analyzer will evaluate the strategy based on the power spectral density estimate after the completion of basic information screening, filtering and pretreatment, the specific process is as follows:

$\Phi(m)$ represents the autocorrelation function of event features $x(n)$ of a certain strategy; $P(m)$ indicates its power spectral density, the relationship between the two is as follows:

$$\varphi(m) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n)x^*(n+m) \quad (3)$$

Substitute and average, then we can get:

$$p(\omega) = \lim_{N \rightarrow \infty} E \left[\frac{1}{2N+1} \left| \sum_{n=-\infty}^{\infty} x(n)e^{j\omega n} \right|^2 \right] \quad (4)$$

The actual strategy implementation is a continuous process, but no system can get the whole sequence of events except subsequences, so our system uses the discrete Fourier transform. There are:

$$\hat{\varphi}(m) = \frac{1}{N} \sum_{n=-\infty}^{\infty} x(n)x_x(n+m) \quad (5)$$

$$\hat{p}(\omega) = \sum \hat{\varphi}(m)e^{-j\omega m} \quad (6)$$

For strategy elements, take A (Action) as an example: when a strategy is activated, there will be plenty of "action" instructions accumulated to the above two components, then the strategy analyzer can send commands to the strategy responder.

4. Test Results and Analysis

SES extracts and filters effective information from the strategy information source, uses Oracle database to store evaluation data, calculates each report data through statistical performance data, and does a longitudinal comparison to give the strategy evaluation analysis results.

4.1 Test index

The strategy return evaluation is relatively simple, so we consider the performance of different strategies mainly around risks: 1) The maximum retracement gains; 2) The maximum retracement duration; 3) The maximum continuous time during which no profit is made; 4) Return volatility; 5) The skewness and kurtosis of returns. As shown in figure 3:

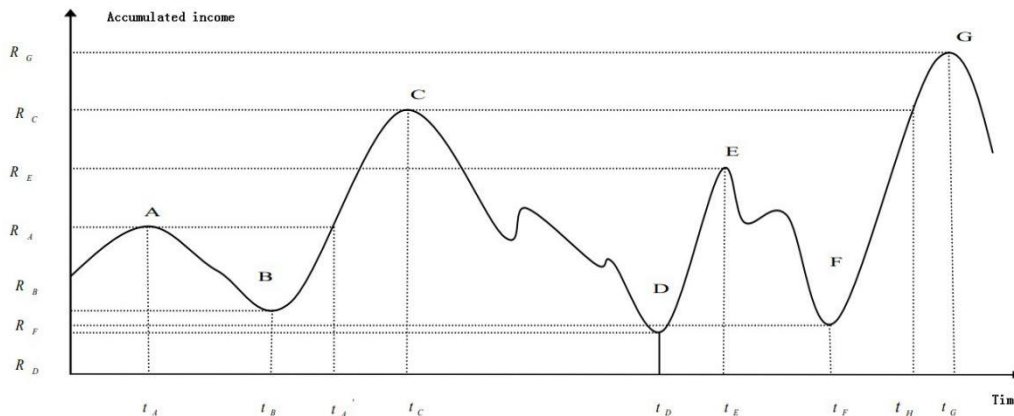


Figure 3: Trading strategy revenue curve

4.2 Analysis of test results

The data set used for model test is from the intraday high frequency data (40 ms / T) of Tokyo Stock Exchange (Tokyo Stock Exchange, TSE), and processes them into minute data. The period is from January 4, 2014 to May 25, 2014, a total of 95 days, with data of 11 days missing. Based on this data set, the following are tested:

- 1) High-frequency trading strategies based on polynomial fitting (PF);
- 2) High-frequency trading strategies based on double averages system (DA) under different parameters.

4.2.1 High frequency trading strategies based on polynomial fitting

The traditional strategy needs analysis of continuous form (for example, triangle, square flag, round bottom / top, etc.) and reversal patterns (the cup handle, head and shoulders, etc.) according to the practical experience of traders, and these morphological indexes are hard to quantify. Using polynomial fitting stock price model can better get the stock price change pattern (Guo Jianfeng, 2012), and its benefit lies in describing the trend of price changes over a period of time by using finite-dimensional data, the transaction manner is as follows:

- 1) Do not make transactions during the half an hour of everyday opening quotation, and estimate future changes by polynomial fitting to determine whether or not to open positions, long or short positions;
- 2) Forecast every 15 minutes once again to determine whether the original way of the position building is correct;
- 3) if it is right, continue to hold; If wrong, close out the original position and then open a reverse position until the closing of the day, close the position to end the cash position.

4.2.2 High frequency trading strategies under double averages system

Double averages trading system analyzes the moving averages of price and volume simultaneously, considers the price-volume factors, and avoids overlooking important factors (Wu, 2012). There are three common types: Simple Moving Average (SMA), Linear Weighted Moving Average (LWMA), Exponential Moving Average (EMA) (Abioye and Olukorede Eliza, 2012), and this paper uses SMA whose function representation format is MA (price property, the number of cycles). The function format of double averages system is MA (p, q) whose trading manner is as follows:

- 1) Calculate the two long and short averages MA (p) and MA (q) based on minute line price data;
- 2) Build more long positions when short average cross long average, and build more short positions otherwise;
- 3) Close the long position held at the following short signal, otherwise hold the long position to the closing of the day and close the long position to end transaction;
- 4) Close the short position held at the following long signal, otherwise hold the short position to the closing of the day and close the short position to end transaction.

The test results of polynomial fitting and double averages are shown in Table 1 and Table 2 below respectively:

Table 1: PF strategy simulation results under different parameters (trading commissions 0.008%, execution costs 10 yen, period: 2014-01-04 to 2014-05-25)

	Rate of Return	Winning Rate	Profit and Loss Ratio	Maximum Single Profit Rate	Maximum Single Loss Rate	Maximum Drawdown Rate	Average Daily Transaction number
f(x,3)	-85.2	21.6	1.36	3.2	-2.1	-90.1	12.2
f(x,4)	-54.0	25.9	1.45	2.4	-1.9	-91.4	15.1
f(x,5)	-13.6	27.6	1.57	4.1	-1.8	-76.2	17.8
f(x,6)	12.9	30.0	1.68	2.8	-1.9	-45.8	20.4
f(x,7)	43.7	48.4	1.73	3.5	-1.9	-32.6	22.3
f(x,8)	24.5	35.0	1.66	4.7	-2.1	-20.1	25.0
f(x,9)	20.4	32.7	1.57	5.1	-1.9	-12.1	28.9

Table 2: AD policy simulation results under different parameters (trading commissions 0.008%, execution costs 10 yen, period: 2014-01-04 to 2014-05-25)

	Rate of Return	Winning Rate	Profit and Loss Ratio	Maximum Single Profit Rate	Maximum Single Loss Rate	Maximum Drawdown Rate	Average Daily Transaction number
MA(15,30)	-24.6	35.5	1.75	5.4	-1.4%	-73.8	15.3
MA(20,40)	-11.0	34.1	1.88	5.3	-0.9%	-30.3	9.9
MA(25,50)	43.5	37.4	1.93	4.8	-1.9%	-15.4	7.2
MA(30,60)	74.9	38.3	1.84	5.2	-2.3%	-17.9	5.8
MA(35,70)	107.2	40.4	1.98	5.1	-2.4%	-10.3	4.9
MA(40,80)	69.2	41.3	1.77	5.7	-1.9%	-9.9	3.6
MA(45,90)	34.1	42.0	1.74	5.2	-2.1%	-6.2	2.4

As can be seen from Table 1, the low-order polynomial can capture the major trend of price changes, but it is easy to overlook the subtle price changes over time, so the yield rate is not high; and although high-order polynomial can capture more subtle price changes, the frequent fluctuations will reduce the final strategy benefits for they are easy to make the trading system issue too many invalid trading signals. As can be seen from Table 2, when the short-term average time is less than 20, the strategy trade frequency is too high and results in loss; while when the short-term average time is greater than 45, the strategy response is too slow and lower the strategy earnings. Therefore, in actual modeling process, we need to weigh therefrom.

5. Conclusion

This paper describes in detail the model building, workflow and main algorithm of high-frequency trading strategy evaluation system and verifies the validity of SES strategy evaluation functions through test. Meanwhile, SES has a complete analog system, it can take market influence factors into account and generate a new directive according to the current spreads, it also has a very fast "product" matching engine which makes it possible to simulate multiple markets (Guo et al., 2015). Future research will focus on the test of "flash crash" mode and introduce mass market events to observe the application of algorithm for further improvement.

Acknowledgments

(1) National Natural Science Foundation of China, Tianyuan Fund for Mathematics

Project Name: Path-dependent Option Pricing Research Based on Markov Chain Approximation Stochastic Volatility under Lévy Model.

Project Number: 11426176

(2) Project of soft science research project of Shaanxi Province

Project Name: Study on the development countermeasures of the Silk Road Economic Zone supported by science and technology.

Project Number: 2015KRZ001

(3) Xi'an Science and Technology Program

Project Name: Financial Risk Management Research Based on Big Data Network

Project Number: SF1504

(4) Shaanxi Provincial Department of Education Philosophy and Social Science Focused Research Base Project

Project Name: Research on the innovation and development of financial informationization in Shaanxi Province

Project Number: 14JZ048

References

- Abioye O.E., 2012, Managing Risk in Financial markets: A market simulation approach [C], 2012-3rd International Conference on Services in Emerging Markets, 75-86, DOI: 10.1109/ICSEM.2012.18.
- Camilo R., 2011, A Parallel Workflow for Real-time Correlation and Clustering of High-Frequency Stock Market Data [J], Computational Statistics & Data Analysis, 48: 431-442, DOI: 10.1109/IPDPS.2007.370216.
- Guo J.F., 2012, Ultra-short Average Rise [M], Shandong: Shandong University Press, 127-136.
- Guo J.F., Wang C., Chen Y.W., 2015, Design of a fingerprint keyboard based on RSCI [J], 2015, V.20, No.112 (01) 64-69
- Hendershott T., 2008, Algorithmic Trading and Information. NET. Institute Working Paper, No. 09-08.
- Hu H., 2010, Understanding of Quantitative Trading [Z], Beijing: Capital Market.Ji J., 2010, US HFT Regulatory Measures and Its Enlightenment [Z], Beijing: Financial World.
- Wang J.C., Chu W.Z., Zhang Y.B., 2010, Improved Gray Relational Degree Algorithm Based on Area [J], Systems Engineering and Electronics, 2010, 32(4): 777-779.
- Wang Q., 2014, How Germans Regulate High-frequency Trading [N], Shanghai Securities News, (1231).
- Wu X.X., 2012, High-frequency Trading Strategy Report: Decrypt The "Black Box" of High-frequency Trading Strategy [R], Shanghai: Haitong Securities.
- Yang S., Paddrik M., 2012, Behavior Based Learning in Identifying High Frequency Trading Strategics [C], 2012 IEEE Conference on Computational Intelligence for Financial Engineering and Economic, 133-140, DOI: 10.1109/CIFEr.2012.6327783.
- Yang W.H., Wu J.H., 2008, A New Gray Correlation Coefficient Expression [J], Journal of Hohai University: Natural Science Edition, 36(1): 40-43.
- Zhang H., 1979, Maximum Entropy Method of Power Spectral Density Estimation [J], Journal of Shaanxi Astronomical Observatory, 1979. 01, DOI: 10.13875/j.issn. 1674-0637. 1979.01.001
- Zou J., Zhang H., 2014, High-frequency financial statistics with parallel R and Intel Xeon Phi coprocessor. 2014 IEEE International Conference on Big Data [C]:2014. 61 – 69, DOI: 10.1109/BigData.2014.7004414.