

# Study on Financial Time Series Prediction Based on Phase Space Reconstruction and Support Vector Machine (SVM)

Hong Zhang, Li Zhou, Jie Zhu

School of Information, Beijing Wuzi University, Beijing, China

## Email address

514050209@qq.com (Hong Zhang)

## To cite this article:

Hong Zhang, Li Zhou, Jie Zhu. Study on Financial Time Series Prediction Based on Phase Space Reconstruction and Support Vector Machine (SVM). *American Journal of Applied Mathematics*. Vol. 3, No. 3, 2015, pp. 112-117. doi: 10.11648/j.ajam.20150303.16

---

**Abstract:** Analyzing and forecasting the financial market based on the theory of phase space reconstruction of support vector regression. The key point of the phase space reconstruction is to choose the optimal delay time, and to find the optimal embedding dimension of space. This paper proposes the use of false nearest neighbor method to construct the error function for all the variables to determine the appropriate embedding dimension combinations. Kernel function in the SVR is an important factor for algorithm performance. Experiments show that the theory of phase space reconstruction based on support vector regression has a certain degree of predictive ability of market value at risk.

**Keywords:** Phase Space Reconstruction Theory, Support Vector Regression, Financial Time Series Prediction

---

## 1. Introduction

In general, Time Series model used in analyzing and forecasting financial markets can be divided into single variable model and multivariate model according to the selection of variables, according to the structure of the model and can be divided into linear and nonlinear model. It is now widely used now that linear method such as exponential smoothing, moving average, moving average (ARIMA), multiple regression and autoregressive conditional heteroscedasticity. Evaluation of a Time Series model is needed to study the fitting validity and forecasting effectiveness. A good model must not only be able to have a good understanding of phenomena that occur within the Time Series, finding the internal law, but also to analyze the extent and scope of the impact and the impact of external factors, and also need to be extrapolated by model, correctly predicting the trend of Time Series. Nonlinear prediction method based on phase-space reconstruction theory<sup>[1-3]</sup> is a new forecasting method developed in the last 20 years. The method applied to financial Time Series with nonlinear and non-stationary characteristics and uncertainty, it can accurately predict the short-term behavior of financial markets. Based on the principle of Takens embedded, if the selection of embedding dimension and delay time appropriate, single variable Time Series can obtain more ideal prediction effect. Because in practical application, there always exist noise in time series

data limited in length. It is difficult to ensure the univariate Time Series contained sufficient information to reconstruct the power system. To resolve this issue, Cao and others propose method of phase space reconstruction of multivariate Time Series. The method of phase space reconstruction of chaotic model and support vector machine in recent years in the field of artificial intelligence is very popular, nonlinear sequence can be mapped into a high dimensional space, the information of nonlinear dynamic characteristics in the sequence can be revealed. Therefore, using the theory of phase space reconstruction of chaotic theory determine the training samples. Finally, applying support vector regression to model the training samples, realize the prediction of financial time series.

## 2. The Phase Space Reconstruction of Multivariate Time Series

Assuming that a system with  $n$  measurable variables, Time Series corresponding to each variable is  $z_i = (z_{i1}, z_{i2}, \dots, z_{iN}), i = 1, 2, \dots, N$ , Based on the phase space reconstruction theory, it can be reconstructed as follows.

$$V_n = (z_n, z_{n-r}, \dots, z_{n-(d-1)r}) \quad (1)$$

Among them, the  $\tau_i, d_i$  respectively is time delay of Time Series  $i$  and embedding dimension.  $n = (d-1)\tau + 1$ . If  $d$  or every  $d_i$  is large enough,  $d = d_1 + d_2 + \dots + d_M > 2D$ ,  $D$  is the attractor dimension. When  $n = 1$  for single variable Time Series, it is a special case of multivariate Time Series.

Auto-correlation function and the mutual information method is usually used to selected delay time respectively for every single variable Time Series that makes up multivariable Time Series. Calculated as follows:

$$I(\tau) = \sum_{n=1}^N P(x(n), x(n+\tau)) \times \log \left[ \frac{P(x(n), x(n+\tau))}{P(x(n)) \cdot P(x(n+\tau))} \right] \quad (2)$$

$I(\tau) \geq 0$ ,  $P(\bullet)$  is the probability  $I(\tau)$  represents the information that a sequence is obtained from another sequence. Fraer recommends that we could regard  $\tau$  which is reached the first minimum correspond to the  $I(\tau)$  as the embedding time delay.

### 2.1. To Determine the Embedding Dimension

Another issue of concern is how to select appropriate embedding dimension for multivariable Time Series. The false nearest neighbors method, Kennel proposed, is a common method for single variable and multivariable Time-Series. The basic idea of the method is: The Time Series is normalized, using auto-correlation function and mutual-information to determine the delay time of each single variable Time Series, given a set of embedding dimension to obtain delay vector. For every vector, finding its nearest neighbor vector, making that:

$$\eta(n) = \arg \min \{ \|V_n - V_j\| : j = \max(dj-1)\tau_i + 1, \dots, N, j \neq n \} \quad (3)$$

Calculating one step prediction error:

$$E(d_1, d_2, \dots, d_M) = \frac{1}{N - J_0 + 1} \sum_{n=J_0}^N |x_{1,n+1} - x_{1,\eta(n)+1}|, \quad (4)$$

$$J_0 = \max_{1 \leq i \leq M} (d_i - 1)\tau_i + 1$$

$$a(M, d_1, d_2, \dots, d_M) = \sum_{i=1}^M \frac{\|V_n(d_1, d_2, d_{i+1}, d_M) - V_{\eta(n)}(d_1, d_2, d_{i+1}, d_M)\|}{\|V_n(d_1, \dots, d_i, \dots, d_M) - V_{\eta(n)}(d_1, \dots, d_i, \dots, d_M)\|} \quad (6)$$

Further definition,

$$E(d_1, \dots, d_i, \dots, d_M) = \frac{1}{N - J_0 + 1} \sum_{i=J_0}^M a(n, d_1, \dots, d_i, \dots, d_M) \quad (7)$$

Making,

$$d_1, \dots, d_i, \dots, d_M = \arg \min (E(d_1, \dots, d_i, \dots, d_M) : (d_1, \dots, d_i, \dots, d_M) \in Z) \quad (8)$$

The algorithm in the global phase space reconstruction can determine the appropriate embedding dimension combination, is an ideal method of determining multi variable embedding

The measurement of error  $(d_1, d_2, \dots, d_M)$  depends on the choice of  $d_1, d_2, \dots, d_M$ . Making  $E(d_1, d_2, \dots, d_M)$  a set of minimum embedding dimension as the embedding dimension of phase space reconstruction, i.e.

$$d_1, d_2, \dots, d_M = \arg \min \left( E(d_1, d_2, \dots, d_M) : (d_1, d_2, \dots, d_M) \in Z, \sum_{i=1}^M d_i \neq 0 \right) \quad (5)$$

The reason of the choice of embedding dimension that makes prediction error is minimum is that when the embedding dimension is too small, will not be able to get a proper embedding, the predicted results will be very poor, so the prediction error will be large; when embedding dimension is the optimal embedding dimension, suitable embedding and may determining the mapping  $iF$  reduce the prediction error; and when the embedded dimension is more than the optimal embedding dimension, especially when the system exists positive Lyapunov index or noise, the increase of the prediction error will be far beyond the acceptable range, because the historical data used in the prediction process is associated very small with now or even completely unrelated.

### 2.2. Improved Method of Determining Embedding Dimension

Construction of error function:  $E(d_1, d_2, \dots, d_M)$  just for the variable  $z_1$ , the appropriate embedding dimension finally got may only applies to the variable  $z_1$  to the reconstruction of phase space, and it does not necessarily apply to other variables. Therefore, this article will build error function for all variables. First, for each vector, finding the adjacent vector, and we will define the following error function which based on this definition.

dimension. It can be used to calculate the phase space dimension of single variable Time Series, is also applicable to the single variable case.

### 2.3. The Calculation of the Maximum Lyapunov Exponent

Getting the time delay and embedding dimension can calculate the maximum Lyapunov index, to test the existence of chaotic phenomena of Time Series, if the Lyapunov index is the regular indicates the existence of chaotic phenomena. The Lyapunov index estimation expressions are as follows:

$$\lambda_1(i, k) = \frac{1}{k\Delta t} \frac{1}{M-k} \sum_{j=1}^{M-k} \log \frac{d_j(i+k)}{d_j(k)} \quad (9)$$

where  $k$  is a constant,  $\Delta t$  is the sample period,  $d_j(k)$  is the first  $j$  distance of the adjacent points on the basic track after  $i$  discrete time step,  $M$  is the number of reconstructed phase space point.

## 3. Study on Nonlinear Prediction Algorithm of Multivariate Time Series

Stochastic process followed a large sample theory, and require research sample size tends to infinity, and obtain statistical conclusions in the conditions of mass. In fact, we are unable to get an infinity of sample, even for large-scale analysis of sample data processing also need larger price, usually, in the condition of limited, we wish get accurate predictions as much as possible. The Hurst index and Lyapunov index can't be used as the main basis of regression prediction.

In data driven, neural network as a nonparametric weakening model<sup>[4]</sup>, fault tolerance is excellent, for the incomplete data of complex system; it has the strong learning capability, and dominate the earnings prediction of Time Series. Used for Time Series prediction of neural network can be divided into two kinds: one kind is Jordan and Elman network, building the space-time mechanism in the additional layer, storage sample values in the past time point.<sup>[5]</sup> Another is to use time window for processing the internal structure of the neural network does not change, the input sample values in the past, neural network gives the predicted value of the next point in time.<sup>[6-7]</sup>

The neural network can effectively predict earnings Time Series, but, because this approach has inherent flaws, making their learning generalization performance degradation, such as over-fitting problem cannot be solved. Vapnik proposed a new neural network, a better solution to this problem. This neural network is called support vector machines (SVMs), follow the structural risk minimization principle, you can get the global optimal solution. With the introduction of non-sensitive loss function, SVMs extended to the areas of return and it is not previously could only deal with classification of primary issues, so called support vector regression (SVR). SVR was introduced to Time Series to predict stock returns. Her research proved that SVR Time Series forecast earnings are feasible and effective.

Kernel function mainly translate nonlinear problem into a linear problem through a certain way, then treated, the specific

linearization method is to map the input data from low-dimensional space to a high dimensional feature space.

Therefore the kernel function plays a decisive role to prediction performance of support vector, its construction or selection mode is particularly important. In the different time granularity, wavelet function can describe arbitrary position of the Income Time Series, so it can be used to analyze the fluctuation characteristics of return Time Series. The SVMs and wavelet theory combined with each other to build wavelet kernel function used to predict stock returns Time Series is worth studying.<sup>[8-9]</sup>

### 3.1. Construction of Wavelet Kernel Function

This section mainly introduces some basic theory of multi-resolution analysis and wavelet transform research foundation, and to construct a wavelet kernel function.

Defined the continuous wavelet transform of  $f(z)$  as follows,

$$WT_f(a, b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx, \quad (10)$$

$a \neq 0, f(z) \in L^2((R))$

$\psi(x)$  to meet the permit conditions:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \quad (11)$$

where  $\psi(\omega)$  is the Fourier transform of  $\psi(x)$ , then the inverse transformation is

$$f(x) = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_{a,b}(t) WT_f(a, b) db \frac{da}{|a|^2}, \quad (12)$$

where  $C_\psi$  is a constant,  $\psi(\omega)$  is a continuous function of  $R$ , (11) shows that the value of  $\psi$  at the origin is zero, that is:

$$\hat{\psi}(0) = \int_{-\infty}^{\infty} \psi(x) dx = 0, \quad (13)$$

### 3.2. Regression Analysis

Setting the training set  $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$ , in which  $x_i \in R^h$ ,  $y_i \in R$ , the regression function of support vector machine can be expressed as:

$$f(x) = \langle w \cdot \psi(x) \rangle + b, \quad (14)$$

$w$  and  $b$  can be obtained by minimizing the

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (15)$$

In the end can be converted into

$$\begin{cases} y_i - \sum_{i=1}^n w_i \psi(x_i) - b \leq \varepsilon + \xi_i \\ \sum_{i=1}^n w_i \psi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (16)$$

in which  $\xi_i, \xi_i^*$  are slack variables, and  $C$  is punish parameter. By introducing Lagrange multipliers, finally get the decision function:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) k(x_i, x) + b, \quad (17)$$

$$k(x_i, x) = \psi(x_i) \psi(x) \quad (18)$$

Among them,  $\alpha_i$  and  $\alpha_i^*$  is the Lagrange multiplier,  $k(x_i, x)$  is the kernel function. Training the pre-treatment sample data by SVMs regression method to calculate the support vectors  $x_i$ , and the parameters  $\alpha_i$ ,  $\alpha_i^*$  and  $b$ . The obtained results into the formula (3.13), and then put pre-treatment test set samples into the model. Obtaining the prediction value mapped to high dimension interval, and map the resulting inverse to the original range to obtain the final result. In the prediction of the actual process, the main prediction steps are as follows:

- 1) according to the theory of phase space reconstruction of chaotic system to obtain optimal embedding dimension and time delay of training historical data to generate training samples.
- 2) normalized generated training samples to improve the convergence speed, shorten training time.
- 3) Select the applicable SVM and kernel function.
- 4) training generated training samples using support vector machine to obtain prediction model.
- 5) using the trained support vector machine forecasting model to forecast.

### 3.3. The Experimental Data and the Preprocessing

This experiment using two simulated stock returns Time Series, each Time Series contains the same number of samples, collected 3120 samples. Autoregressive integrated moving average model (ARIMA) is a mainstream method of the financial Time Series prediction, using the difference operation to complete the Time Series smoothing, the choice of the model to simulate the generation first sequence. As shown in Figure 3.1 for the ARIMA model of the Time Series waveform, the following expression describes the model of ARIMA (1, 1, 1)

$$(1 - \phi_1 B)(1 - B)Y_t = (1 + \theta_1 B)\xi_t, \{\xi_t\}_t, N(0, 1) \quad (19)$$

Mackey-Glass delay differential equation is a common mathematical model of producing simulation data. Mackey-Glass equation was first used to describe white blood

cells produce process.<sup>[10]</sup> In the field of financial chaos phenomenon is widespread and the chaotic Time Series on the time axis is neither convergence nor divergence, besides the motion trajectory is difficult to describe, because it is affected by the initial condition is very serious. because Mackey-Glass data set has certain ability to describe the chaos phenomenon, gradually at an important position in the field of standard in the financial research.

The definition of Mackey-Glass delay differential equations as follows:

$$\frac{dx}{dt} = ax_t + \frac{bx_{t-d}}{1 + x_{t-d}^c} \quad (20)$$

in which the autoregressive parameter is  $\phi_1$ . the moving average parameters is  $\theta_1$  and the lag operator denoted as  $B$ . According to Figure.2 that the ARIMA sequence has the characteristics of non periodic arrangement form, also can be observed violent decay of autocorrelation curve.

Figure 3 describes the waveform curve of Mackey-Glass data sets in simulation experiment. Figure 3 describes Autocorrelation characteristics of Mackey-Glass data sets in simulation experiment. Parameter is set to  $a = 0.2$ . Compared with autocorrelation characteristics of ARIMA Time Series, Mackey-Glass Time Series has a certain periodicity, and there is no attenuation of zero point of autocorrelation. In summary, Mackey-Glass Time Series makes it relatively easy to regression prediction. The correctness of this conjecture has been verified in the experiment, from Table 1 it is not difficult to see that the minimum standard mean square error are appeared in the sequence.

There are five components of the input vector, 4 of which are yield; another is the closing price after conversion. Because the data have a relatively uniform and symmetry distribution approaching normal distribution after processing, So that the prediction performance is effectively improved, in the experiment, the closing price of the original data is converted into a corresponding return rate. Subtracting the mean can eliminate the trend term in the price, in the original closing price conversion process, also can avoid the loss of useful information.

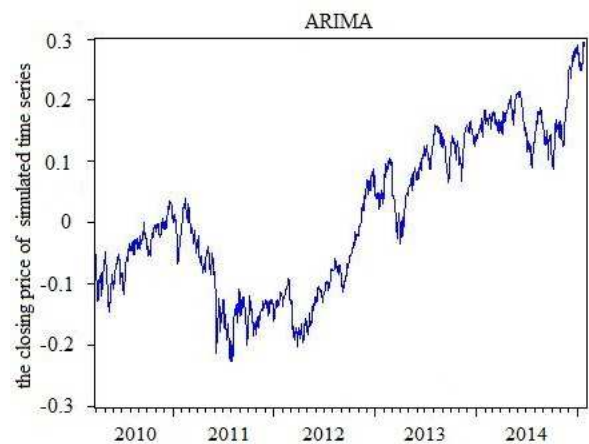


Fig. 1. The ARIMA time series.

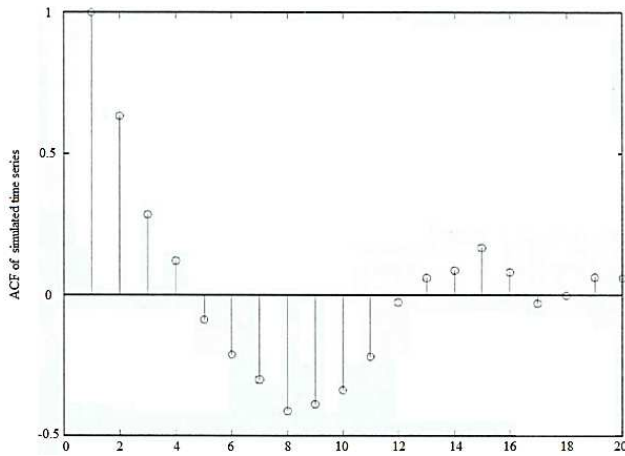


Fig. 2. Autocorrelation function of ARIMA time series.

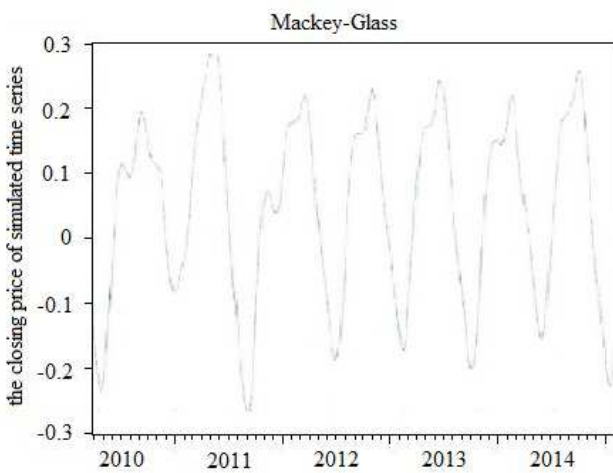


Fig. 3. The Mackey-Glass time series.

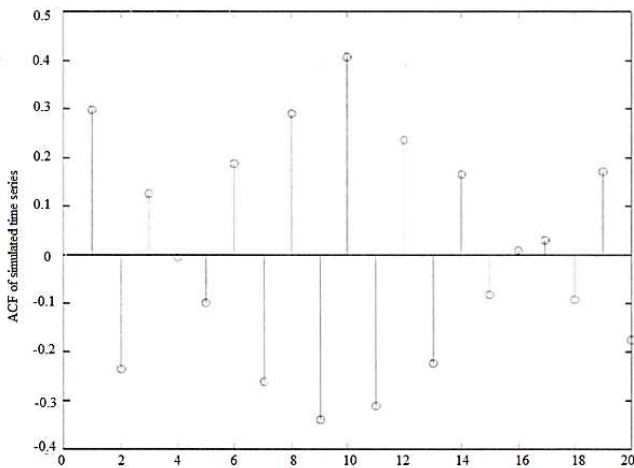


Fig. 4. Autocorrelation function of Mackey-Glass time series.

### 3.4. Performance Indicators and Results Analysis

This paper uses regular mean square error, mean absolute error, root mean square error square and several other statistical indexes as a measure of the prediction and evaluation of performance standards. Before the training, the mean standard deviation, skewness and kurtosis of the indexes

see Table 1. After processing the data, get the statistical performance, according to these statistical properties can be obtained conclusion. The experimental results shown in this article: each index all have mean close to zero, compared with the normal distribution has some deflection. The best prediction results are from March 2011 to March 2015 for the fourth time sliding. As shown in Table 2, in five real stock returns, there all exist the minimum of three indicators (RMSE, NMSE, MAE) in the wavelet kernel. Through further comparison, the prediction performance of each kernel in addition to AXJO, indicators of other real index (SSMI, CAC40, FTMIB and DAXINDEX) are inferior to the wavelet kernel. Therefore, wavelet kernel was compared with the Gaussian kernel can be as the key. For the simulated data, perhaps due to the dynamic nature of ARIMA Time Series is weaker, its Gauss nuclear is optimal, and so the good performance of wavelet kernel cannot be fully reflected. On the whole, the optimal is wavelet kernel, followed by the Gauss kernel, third is polynomial kernel, and the worst performance is the linear kernel. There is another view of the special circumstances need to distinguish between: the comparison results can be drawn based on the prediction set NMSE paired t test. The t value for the one-sided test proved that the wavelet kernel is better than the Gauss kernel at 0.1 significant level.

Table 1. Sample statistical characteristics.

Stock indices	Mean	S.D.	Skewness	Kurtosis
ARIMA	-0.2415	0.4855	0.4187	2.6954
MGLASS	0.1548	0.4875	-0.4471	2.1547
FTMIB	-0.2145	0.4365	1.3025	3.2548
SSMI	-0.2589	0.4521	1.1054	3.2154
CAC40	0.0874	0.3218	0.8444	2.5481
AXJO	0.1547	0.4196	-0.4852	2.9654
DAXINDEX	-0.3218	0.5698	0.6487	2.7542

Table 2. Prediction set results.

Stock indices	Linear kernel		
	NMSE	RMSE	MAE
ARIMA	2.5652	0.0985	0.0548
MGLASS	0.3452	0.2964	0.1584
FTMIB	1.7545	0.8543	0.2857
SSMI	1.4522	0.2857	0.2874
CAC40	0.9874	0.3878	0.1587
AXJO	0.9852	0.2859	0.8574
DAXINDEX	1.1485	0.3848	0.5845
Stock indices	Polynomial kernel		
	NMSE	RMSE	MAE
ARIMA	0.5412	0.0854	0.3332
MGLASS	0.1574	0.1658	0.3214
FTMIB	1.4952	0.2841	0.5241
SSMI	1.0542	0.2541	0.2145
CAC40	0.9854	0.1985	0.1452
AXJO	0.8745	0.8412	0.8541
DAXINDEX	1.3255	0.4412	0.2541
Stock indices	Gaussian kernel		
	NMSE	RMSE	MAE
ARIMA	0.5421	0.9512	0.1285
MGLASS	0.7007	0.4702	0.2686
FTMIB	0.3410	0.5844	0.0656
SSMI	0.0010	0.3130	0.7025
CAC40	0.7197	0.2335	0.3291

Stock indices	Linear kernel		
	NMSE	RMSE	MAE
AXJO	0.3185	0.3260	0.0361
DAXINDEX	0.1086	0.2011	0.6611
Stock indices	Wavelet kernel		
	NMSE	RMSE	MAE
ARIMA	0.5533	0.8553	0.1086
MGLASS	0.4322	0.2339	0.9551
FTMIB	0.4353	0.1252	0.3199
SSMI	0.1288	0.2402	0.7868
CAC40	0.0493	0.5928	0.4421
AXJO	0.8696	0.6229	0.9996
DAXINDEX	0.0239	0.4784	0.5587

## 4. Summary

Financial Time Series has unique features, such as dynamic characteristics and nonlinear characteristics, although the study of its dynamic characteristics are not mature, inconclusive, but the nonlinear characteristics of Financial Time Series has been widely recognized, Therefore, based on the nonlinear Time Series modeling principle, studies the application of neural network and support vector machine in financial Time Series Prediction. In order to make the results fully authentic, the research object selected in this paper is mainly for the foreign large stock market, selecting the following five stock index data sets: FTMIB, SSMI, CAC40, AXJO and DAXINDEX. Two simulated stock, first modeling using ARIMA model, through the Mackey - Glass chaos differential equation to get the required data sets, and then select the dimension of input and the data preprocessing, so that the data in the conversion process most likely to avoid losses. Select the most appropriate parameter by dividing the data set and some other methods, making the prediction performance of the model is the most outstanding, Using selected statistical indicators to measure whether the prediction performance is the best of the model. Finally, we make a summary of the experimental data analysis and draw the conclusion that wavelet kernel compared with other gaussian kernel kernel function, more close to the real situation of earnings forecast curve, so you can think it has better performance of earnings forecasts.

## Acknowledgements

This project (Empirical research on Stock index investment risk model, No.68) is funded by the "2014-2015 school year,

Beijing Wuzi University, College students' scientific research and entrepreneurial action plan project". And by Beijing Wuzi University, Yunhe scholars program (00610303/007). And by Beijing Wuzi University, Management science and engineering Professional group of construction projects. (No. PXM2015\_014214\_000039)

## References

- [1] MA Jun-hai, Wang Zhi-qiang, and Chen Yu-shu. Prediction Techniques of Chaotic Time Series and Its Applications at Low Noise Level. *Applied Mathematics and Mechanics*, 27(1):8-12, 2011.
- [2] Kugiuntzis D, Lingjirde O C, and Christophersen N. Regularized Local Linear Prediction of Chaotic Time Series. *Physica D*, 112(3-4): 345-359, 1998.
- [3] Qingfang neng and Yuhua Peng. A New Local Linear Prediction Model for Chaotic Time Series. *Physics Letters A*, 370(5-6): 468-489, 2011.
- [4] Chi-Jie Lu, Tiar Shyug Lee, and Chih-Chou Chiu. Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Machine. *Decision Support systems*, 47(2009):118-124, 2009.
- [5] Lam Hong Lee, Rajparsaci Rajkumar, and Dino Isa. Automatic Folder Allocation System Using Bayesian-Support Vector Machines Hybrid Classification Approach. *Applied Intelligence*, Vol. 36, No. 2, 298-305, 2012.
- [6] Patrick Koch, Bernd Bisch], and Oliver Flasch. Tuning and Evolution of Support Vector Kernels. *Evolutionary Intelligence*, Vol. 5, No. 3, 3 59-150, 2012
- [7] Masayuki Karasuyama, Maoyuki Harada, Masashi Sugiyama, and Ichiro Takouchi. Multi-parametric Solution-path Algorithm for Instance-weighted Support Vector Machines. *Machine Learning*, Vol. 88, No. 3, 299-310, 2012.
- [8] S. Amari. *Differential-geometrical Methods in Statistics*. Springer-Verlag New York, 1985.
- [9] S. Amari. Neural Learning in Structured Parameter Spaces Natural Rjcmannian Gradient. *Advances in Neural Information Processing Systems*, (9):128 - 132, 1997.
- [10] MC Mackey and L. Glass. Oscillation and Chaos in Physiological Control Systems. *Science*, 197(4300): 287, 1977.