Journal of Mathematics Research; Vol. 10, No. 5; October 2018
ISSN 1916-9795 E-ISSN 1916-9809
Published by Canadian Center of Science and Education

Multi-factor Stock Selection Model Based on Kernel Support Vector Machine

Ru Zhang[1], Zi-ang Lin[2], Shaozhen Chen[1], Zhixuan Lin[2] & Xingwei Liang[2]

[1] Finance Department of International Bussiness School, Jinan University, China

[2] Financial Management Department of International Bussiness School, China

Correspondence: Shaozhen Chen, Finance Department of International Business School, Jinan University, Qianshan Road 206#, Zhuhai City, Guangdong Province, Post No. 519070, China. E-mail: 1813012994@qq.com

Received: May 30, 2018 Accepted: June 14, 2018 Online Published: June 28, 2018

doi:10.5539/jmr.v10n5p9 URL: https://doi.org/10.5539/jmr.v10n5p9

Abstract

In recent years, the combination of machine learning method and traditional financial investment field has become a hotspot in academic and industry. This paper takes CSI 300 and CSI 500 stocks as the research objects. First, this paper carries out kernel function test and parameter optimization for the kernel support vector machine system, and then predict and optimize the combination of market-neutral stock selection strategy and stock right strategy. The results of the experiment show that the multi-factor model based on SVM has a strong predictive power for the selection of stock, and it has a difference in the predictive power of different nuclear functions.

Keywords: multi-factor model, support vector machine, quantitative investment

1. Introduction

Based on the computer technology, quantitative investment provides a meaningful direction for investment decision-making by using reasonable algorithms. And in combination with traditional investment concepts, it can obtain higher excess returns, for which it is gradually accepted by investors. In the field of stock investment, multi-factor stock selection model is widely used. At the earliest, the research of Fama et al (1970) proved that the stock price was jointly determined by multiple factors, and the single factor influence was insufficient to accurately describe the internal value of listed companies. Asness et al. (1997) further explored more representative factors in fundamental factors to construct stock selection portfolio and obtain excess returns. Recently, Wang (2016) used the scoring model based on equal weight rating to carry out factor selection, establish investment portfolio and conduct empirical analysis. Su (2018) further constructed the multi-factor stock selection model by removing redundant factors through fuzzy clustering algorithm. Vapnik (1995) proposed the support vector machine algorithm and later it became the most widely used machine learning model in the financial field. Sai (2013) used genetic algorithm and particle swarm optimization algorithm to optimize the kernel function of support vector machines, and then used the optimized SVM to predict the price of stock index futures, achieving good results. Han (2016) showed that the initial prediction results of time series models including GARCH or AR can improve the prediction accuracy of SVM. Huang (2017) combined the support vector machine with the traditional Fama-Fench three-factor model to construct A new stock selection strategy. After an empirical analysis of a-share, the new strategy model proved to be more profitable.

From what has been discussed above, the existing research mostly focused on the use of machine learning algorithm to optimize the traditional time series prediction model and optimizing the traditional factor to choose a single strategy, more empirical analysis of the strategy combination is less. Therefore, this paper will test various kernel functions of the kernel support vector machine system. Based on the test results, a portfolio model of stock selection strategy applied to CSI 300 and CSI 500 will be constructed. At the same time, through comparative analysis of two different portfolio strategies, market-neutral strategy and stock equity strategy, we further build a more profitable and more robust multi-factor stock selection model, providing new ideas for the application and development of machine learning methods in the financial investment field.

2.1 Theoretical Model

2.1.1 Classical Multifactor Model

The expression of classical multifactor model is:

$$\tilde{r} = \sum_{k=1}^{K} X_{jk} * f_k + \mu_j \tag{1}$$

The multi-factor model is essentially a linear regression model between current factor exposure and future earnings.

2.2 Kernel Support Vector Machine Model

2.2.1 Nonlinear Classification

The core idea of kernel support vector machines is to transform nonlinear classification into linear classification. Firstly, the original data is transformed into high-dimensional eigenspace by nonlinear mapping, and then the data in high-dimensional space are classified by linear support vector machine, so as to solve the nonlinear classification problem:

$$x \longrightarrow \varphi(x) = (\varphi_1(x), ..., \varphi_k(x), ...) \tag{2}$$

In practical application, the calculation amount of the objective function of the dual optimization problem is too large in the high dimensional space, so people use the technique of kernel function to avoid the explicit expression of the high dimensional feature, so as to avoid the problem of "dimension disaster" skillfully.

2.2.2 Kernel Function

After transforming original data x into high dimensional data $\varphi(x)$ by nonlinear mapping, the objective function of the dual problem of linear support vector machines is:

$$\max_{a} [\sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j a_i a_j \varphi(x_i)^T \varphi(x_j)] \tag{3}$$

$$s.t. 0 \le a_i \le C, \sum_{i=1}^{n} a_i y_i = 0$$

For a new sample, we can calculate the discriminant function

$$f(\mathrm{x}) = w^T x + \hat{b} = \sum_{i=1}^{n} a_i y_i K(x_i, x) + \hat{b} \tag{4}$$

And then we can determine which class the sample belongs to base on the value of the discriminant function is greater than (or less than) zero. Note that the target function does not contain an explicit expression of the low-dimensional to high-dimensional mapping, and only relates to the selection of the kernel function, so this classifier is called the kernel support vector machine (Kernel SVM). Common kernel functions are shown in table 1.

Table1. Common functions of kernel support vector machines

Kernel function	**Expression**
Linear kernel function	$K(x_i, x_j) = \langle x_i, x_j \rangle = \sum_{k=1}^{p} x_i^{(k)} x_j^{(k)}$
Polynomial kernel function	$K(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + 1)^d = (\gamma \sum_{k=1}^{p} x_i^{(k)} x_j^{(k)} + 1)^d$
Sigmoid kernel function	$K(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + 1) = \tanh(\gamma \sum_{k=1}^{p} x_i^{(k)} x_j^{(k)} + 1)$
Gaussian kernel function	$K(x_i, x_j) = \exp(-\gamma (\sum_{k=1}^{p} (x_i^{(k)} - x_j^{(k)})^2))$

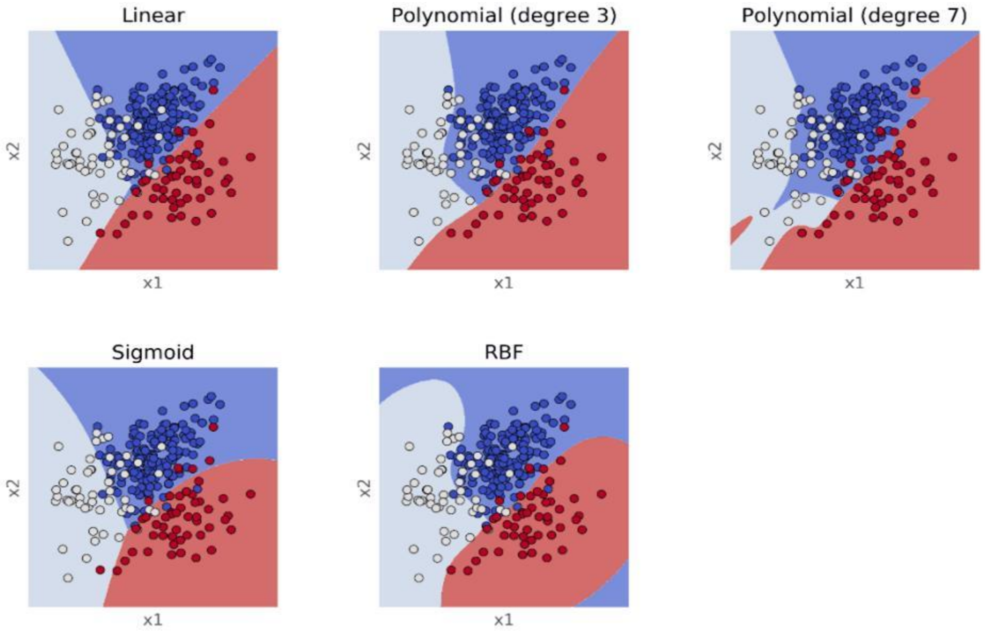


Figure 1. Classification renderings of support vector machines with different kernel functions

The classification performance and classification boundary of different kernel functions are different. Figure 1 shows the classification of data using different kernel functions for the same set of data.

2.3 Multi-Factor Stock Selection Model Based on Kernel Support Vector Machine

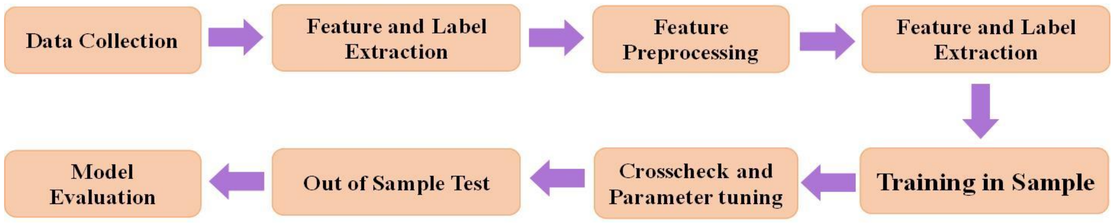


Figure 2. Schematic diagram of multi-factor stock selection model based on kernel support vector machine

As shown in figure 2, the construction method of multi-factor stock selection model based on kernel support vector machine includes the following steps:

(1).Data acquisition:

a).Stock pool: 300 stocks in Shanghai and Shenzhen and 500 stocks in China securities exchange. The ST shares were excluded, the stocks suspended on the next trading day of each cross-section period were excluded, and the stocks suspended within 3 months of listing were excluded. Each stock was regarded as a sample.

b).Within the sample interval: there are 72 end section periods from 2005-01-31 to 2011-12-31.

c).External sample interval: there are 76 end section periods from 2011-01-31 to 2017-04-28.

(2).Feature and label extraction:

On the last trading day of each natural month, the exposure degree of 70 factors in 13 general categories was calculated as the original characteristics of the samples. The excess earnings of individual stocks for the next full natural month (based on the CSI 300 index) are calculated as the sample label. Large class factor pools are shown in table 2.

Table 2. Broad class factors involved in the stock-picking model

Type of factor	Type of factor
Valuation	Share price
Growth	Beta
Financial quality	Turnover rate
Gearing	Mood
Market value	Shareholder
Momentum of the inversion	Technology
Volatility	

(3).Feature pretreatment:

a).Eliminating the extremum:Let the exposure sequence of a factor in the first period on all stocks be , D_i D_M Is the median of this sequence, D_{M1} Is the median of the sequence $|D_i - D_M|$. Reset all the Numbers greater than $D_M + 5D_{M1}$ in sequence D_i to $D_M + 5D_{M1}$ and all Numbers less than $D_M - 5D_{M1}$ Qin sequence D_i to $D_M - 5D_{M1}$;

b).Missing value processing: After obtaining the new factor exposure sequence, the value that the factor exposure is missing is set as the average value of the same stock in the citic primary industry

c).Neutral industry market value: The factor exposure degree after filling the missing value makes linear regression to the industry dummy variable and the market value after logarithm, and the residual value is taken as the new factor exposure degree.

d).Standardization: The neutrogenized factor exposure sequence is subtracted from its current mean and divided by its standard deviation to obtain $N(0,1)$ a new sequence that is approximately distributed.

e).Principal component analysis: In order to avoid the collinearity between features, principal component analysis was conducted on the exposure degree of 70 standardized factors, and new features were obtained after the transformation of 70 dimensions.

(4).Training set and cross validation set together:

a).Classification problem: for the support vector machine model (hereinafter referred to as SVM), in the cross-section period at the end of each month, the stocks that rank first and last 30% of next month's earnings are selected as positive example $(y=1)$ and negative example $(y=-1)$ respectively. The 72 month samples were merged. 90% samples were randomly selected as the training set, and the remaining 10% as the cross-validation set.

b).Regression problems: for support vector regression model (hereinafter referred to as “SVM"), directly to the 72 months within the sample merged with sample data, also according to the proportion of 90% and 10% divided training set and the cross validation set.

(5).In-sample training:

The training set is trained with SVM.SVM selects five different types of:kernel functions: linear nucleus, three order polynomial kernel, 7 order polynomial kernel, nuclear and Gaussian kernel, use 12 months rolling back to the measurement of linear regression model as a unified control group.

All models are shown in table 3

Table 3. Test model overview diagram

Broad category of methods	**Kernel function**	**Parameter setting**		
		The CSI 300 stock selection	**CSI 500 stock selection**	**A-Share**
SVM	Linear kernel function	$C = 1e-4$	$C = 0.003$	$C = 3e-4$
	Third order polynomial kernel function	$C = 0.003$ $Y = 0.03$	$C = 0.001$ $Y = 0.03$	$C = 0.01$ $Y = 0.01$
	Seventh order polynomial kernel function	$C = 0.03$ $Y = 0.01$	$C = 3e-4$ $Y = 0.001$	$C = 1e-4$ $Y = 0.003$
	Sigmold Kernel function	$C = 3$ $Y = 3e-5$	$C = 0.03$ $Y = 0.01$	$C = 10$ $Y = 3e-5$
	Gaussian kernel function	$C = 1$ $Y = 3e-5$	$C = 0.1$ $Y = 0.003$	$C = 1$ $Y = 0.01$
Linear regression	-	-	-	-

(6).Cross-check the parameters:

After model training, the model is used to predict cross validation sets. Select the set of parameters with the highest AUC (SVM)values in the cross validation set as the optimal parameters of the model.

Partial model test results are shown in figure 3:

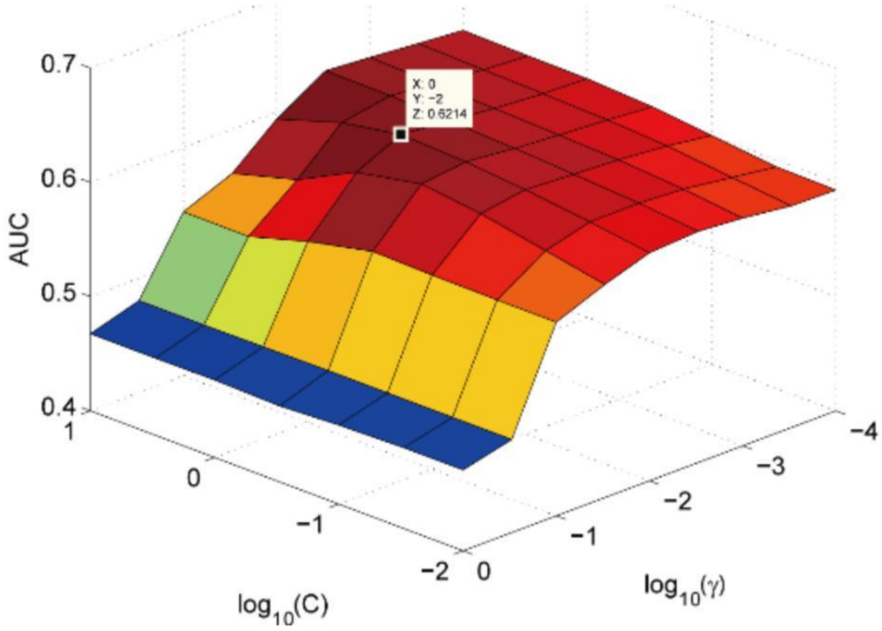


Figure 3. Cross validation results AUC of Gaussian kernel model parameters —C and Y

(7).Out-of-sample testing:

After determining the optimal parameters,the pre-processing characteristics of all samples (i.e., individual stocks) at the end of the month T were taken as the input of the model,and the predicted value of month $T+1$. $f(x)$ (discriminant function value, i.e. the distance from samples to classification hyperplanes) of each sample were obtained. Policy combinations can be constructed based on this predictive value

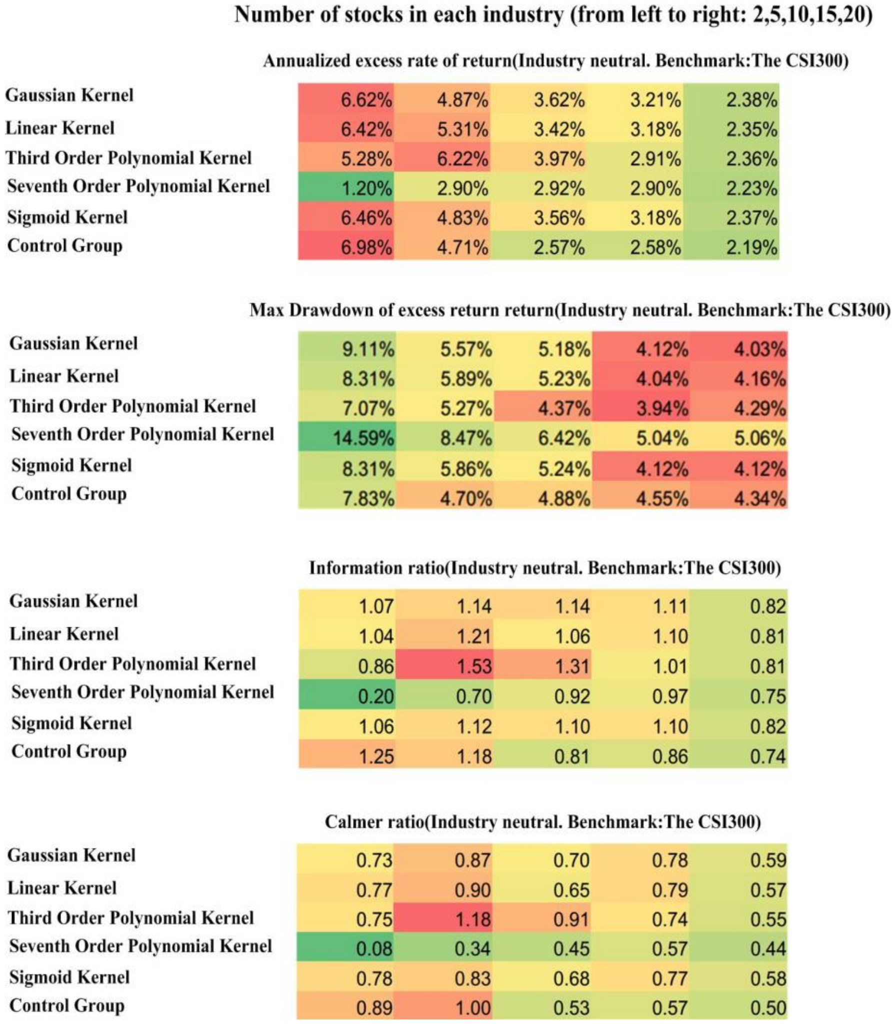
(8).Model evaluation:

Evaluation indicators include two aspects. First, the accuracy of test set and AUC and other indicators to measure the performance of the model; Second, the performance of the strategic portfolio constructed in the previous step (including annualized excess returns, information ratios, etc.)

3. Empirical Analysis

This study constructs the stock selection strategy of CSI 300 and China Certificate 500 and carries out the test. The stock selection strategy can be divided into two categories. One is the industrial neutral strategy — the industry configuration of the strategy combination is consistent with the benchmark (CSI 300, China Certificate 500). Select N stocks in each level of industry with equal weight configuration ($N = 2,5,10,15,20$). The other is individual-share equal-weighting strategy, which is to directly select N stocks in the ticket pool without distinguishing industry with equal weight configuration ($N = 20,50,100,150,200$)with the comparison datum as the indexes of 300 equal weight, 500 equal weight. The two types of strategies are both monthly frequency transfer positions, the order of stock selection is the order of the predicted values for the month they are in the SVM model.

The results of the test are as shown from Figure 4 to Figure 7:



Number of stocks in each industry (from left to right: 2,5,10,15,20)

Annualized excess rate of return(Industry neutral. Benchmark:The CSI300)

Gaussian Kernel	6.62%	4.87%	3.62%	3.21%	2.38%
Linear Kernel	6.42%	5.31%	3.42%	3.18%	2.35%
Third Order Polynomial Kernel	5.28%	6.22%	3.97%	2.91%	2.36%
Seventh Order Polynomial Kernel	1.20%	2.90%	2.92%	2.90%	2.23%
Sigmoid Kernel	6.46%	4.83%	3.56%	3.18%	2.37%
Control Group	6.98%	4.71%	2.57%	2.58%	2.19%

Max Drawdown of excess return return(Industry neutral. Benchmark:The CSI300)

Gaussian Kernel	9.11%	5.57%	5.18%	4.12%	4.03%
Linear Kernel	8.31%	5.89%	5.23%	4.04%	4.16%
Third Order Polynomial Kernel	7.07%	5.27%	4.37%	3.94%	4.29%
Seventh Order Polynomial Kernel	14.59%	8.47%	6.42%	5.04%	5.06%
Sigmoid Kernel	8.31%	5.86%	5.24%	4.12%	4.12%
Control Group	7.83%	4.70%	4.88%	4.55%	4.34%

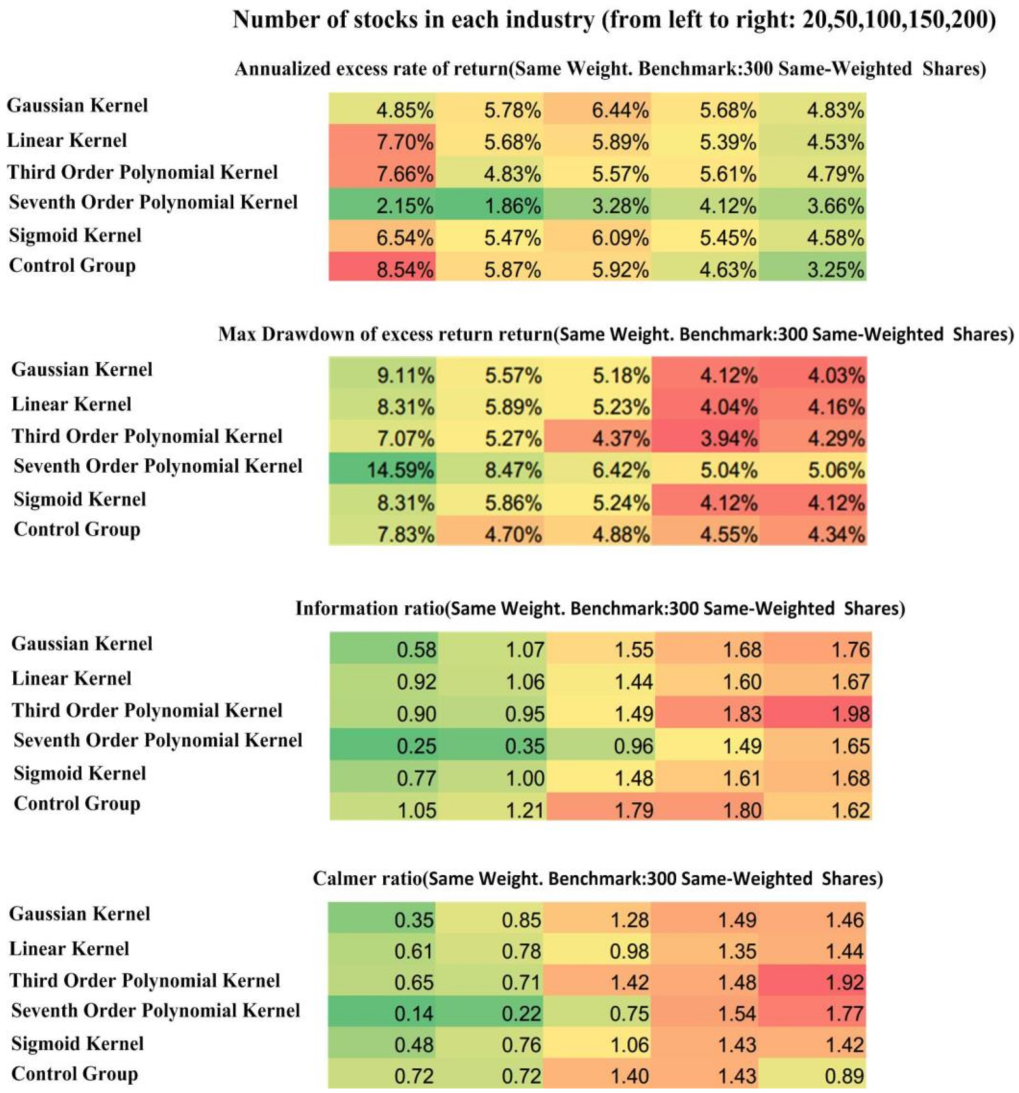
Information ratio(Industry neutral. Benchmark:The CSI300)

Gaussian Kernel	1.07	1.14	1.14	1.11	0.82
Linear Kernel	1.04	1.21	1.06	1.10	0.81
Third Order Polynomial Kernel	0.86	1.53	1.31	1.01	0.81
Seventh Order Polynomial Kernel	0.20	0.70	0.92	0.97	0.75
Sigmoid Kernel	1.06	1.12	1.10	1.10	0.82
Control Group	1.25	1.18	0.81	0.86	0.74

Calmer ratio(Industry neutral. Benchmark:The CSI300)

Gaussian Kernel	0.73	0.87	0.70	0.78	0.59
Linear Kernel	0.77	0.90	0.65	0.79	0.57
Third Order Polynomial Kernel	0.75	1.18	0.91	0.74	0.55
Seventh Order Polynomial Kernel	0.08	0.34	0.45	0.57	0.44
Sigmoid Kernel	0.78	0.83	0.68	0.77	0.58
Control Group	0.89	1.00	0.53	0.57	0.50

Figure 4. Comparison of the important indexes of the SVM model with different kernel functions (market neutral, stock selection in the Shanghai and Shenzhen 300 sub-unit)



Number of stocks in each industry (from left to right: 20,50,100,150,200)

Annualized excess rate of return(Same Weight. Benchmark:300 Same-Weighted Shares)

Gaussian Kernel	4.85%	5.78%	6.44%	5.68%	4.83%
Linear Kernel	7.70%	5.68%	5.89%	5.39%	4.53%
Third Order Polynomial Kernel	7.66%	4.83%	5.57%	5.61%	4.79%
Seventh Order Polynomial Kernel	2.15%	1.86%	3.28%	4.12%	3.66%
Sigmoid Kernel	6.54%	5.47%	6.09%	5.45%	4.58%
Control Group	8.54%	5.87%	5.92%	4.63%	3.25%

Max Drawdown of excess return return(Same Weight. Benchmark:300 Same-Weighted Shares)

Gaussian Kernel	9.11%	5.57%	5.18%	4.12%	4.03%
Linear Kernel	8.31%	5.89%	5.23%	4.04%	4.16%
Third Order Polynomial Kernel	7.07%	5.27%	4.37%	3.94%	4.29%
Seventh Order Polynomial Kernel	14.59%	8.47%	6.42%	5.04%	5.06%
Sigmoid Kernel	8.31%	5.86%	5.24%	4.12%	4.12%
Control Group	7.83%	4.70%	4.88%	4.55%	4.34%

Information ratio(Same Weight. Benchmark:300 Same-Weighted Shares)

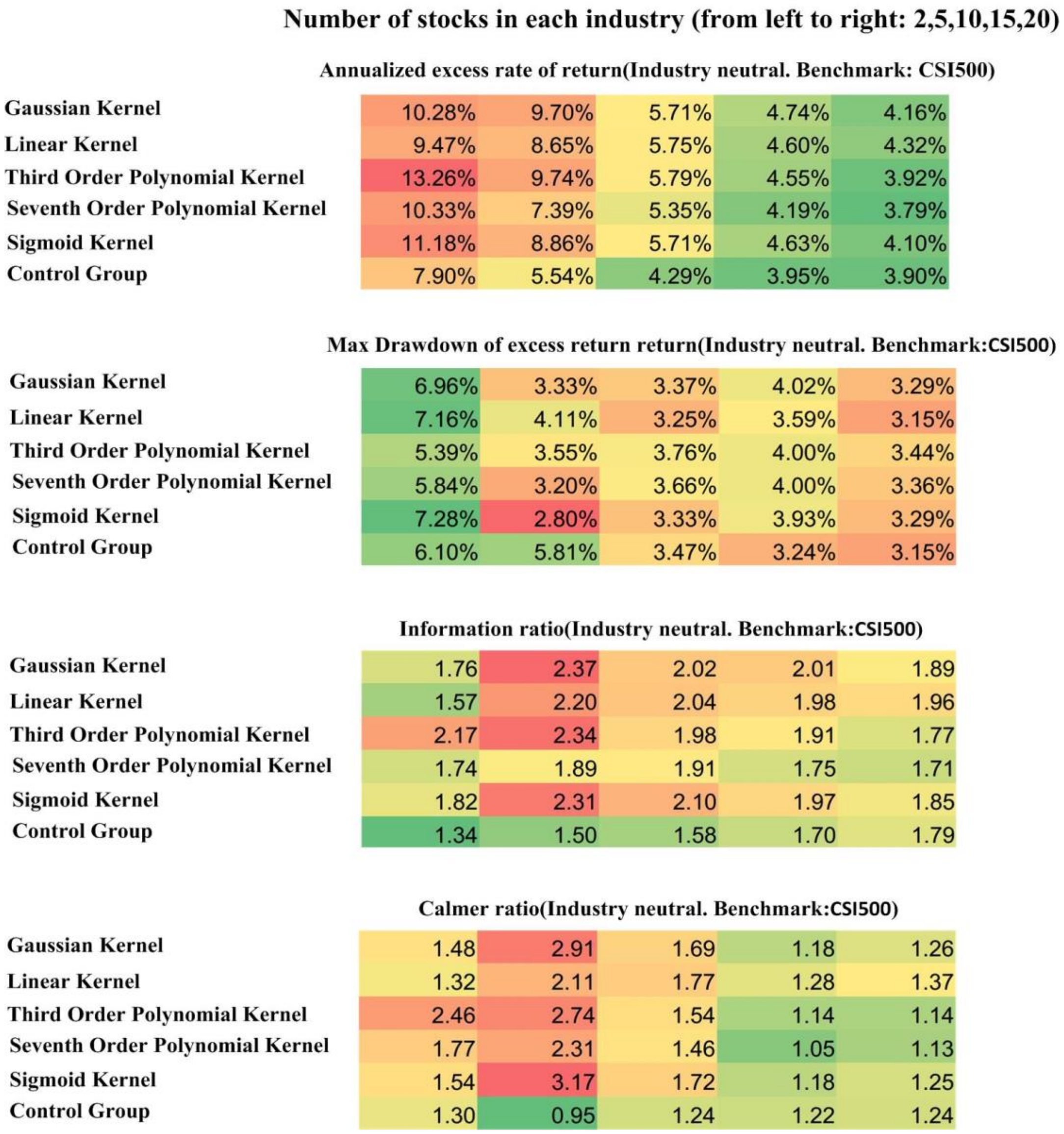
Gaussian Kernel	0.58	1.07	1.55	1.68	1.76
Linear Kernel	0.92	1.06	1.44	1.60	1.67
Third Order Polynomial Kernel	0.90	0.95	1.49	1.83	1.98
Seventh Order Polynomial Kernel	0.25	0.35	0.96	1.49	1.65
Sigmoid Kernel	0.77	1.00	1.48	1.61	1.68
Control Group	1.05	1.21	1.79	1.80	1.62

Calmer ratio(Same Weight. Benchmark:300 Same-Weighted Shares)

Gaussian Kernel	0.35	0.85	1.28	1.49	1.46
Linear Kernel	0.61	0.78	0.98	1.35	1.44
Third Order Polynomial Kernel	0.65	0.71	1.42	1.48	1.92
Seventh Order Polynomial Kernel	0.14	0.22	0.75	1.54	1.77
Sigmoid Kernel	0.48	0.76	1.06	1.43	1.42
Control Group	0.72	0.72	1.40	1.43	0.89

Figure 5. Comparison of the important indexes of the SVM model with different kernel functions (individual share equally weight, Shanghai and Shenzhen 300 sub-unit stock selection)

For the industry-neutral strategy of the stock selection of Shanghai and Shenzhen 300 constituent stocks, when the number of stocks in each industry is greater than or equal to 10, in addition to the 7-order polynomial core, The annual excess returns, information ratios and Calmar ratios of the remaining SVM models are higher than the linear regression models of the unified control group, the maximum return of excess returns is less than linear regression, in which the Gaussian core and the 3-order polynomial have the best nuclear performance. For the individual-share equal-weight strategy of Shanghai and Shenzhen 300 constituent stocks, when the total number of stock options is greater than or equal to 100, the annual excess yield and information ratio of Gauss kernel and 3-order polynomial kernel are higher than those of linear regression.



Number of stocks in each industry (from left to right: 2,5,10,15,20)

Annualized excess rate of return(Industry neutral. Benchmark: CSI500)

Gaussian Kernel	10.28%	9.70%	5.71%	4.74%	4.16%
Linear Kernel	9.47%	8.65%	5.75%	4.60%	4.32%
Third Order Polynomial Kernel	13.26%	9.74%	5.79%	4.55%	3.92%
Seventh Order Polynomial Kernel	10.33%	7.39%	5.35%	4.19%	3.79%
Sigmoid Kernel	11.18%	8.86%	5.71%	4.63%	4.10%
Control Group	7.90%	5.54%	4.29%	3.95%	3.90%

Max Drawdown of excess return return(Industry neutral. Benchmark:CSI500)

Gaussian Kernel	6.96%	3.33%	3.37%	4.02%	3.29%
Linear Kernel	7.16%	4.11%	3.25%	3.59%	3.15%
Third Order Polynomial Kernel	5.39%	3.55%	3.76%	4.00%	3.44%
Seventh Order Polynomial Kernel	5.84%	3.20%	3.66%	4.00%	3.36%
Sigmoid Kernel	7.28%	2.80%	3.33%	3.93%	3.29%
Control Group	6.10%	5.81%	3.47%	3.24%	3.15%

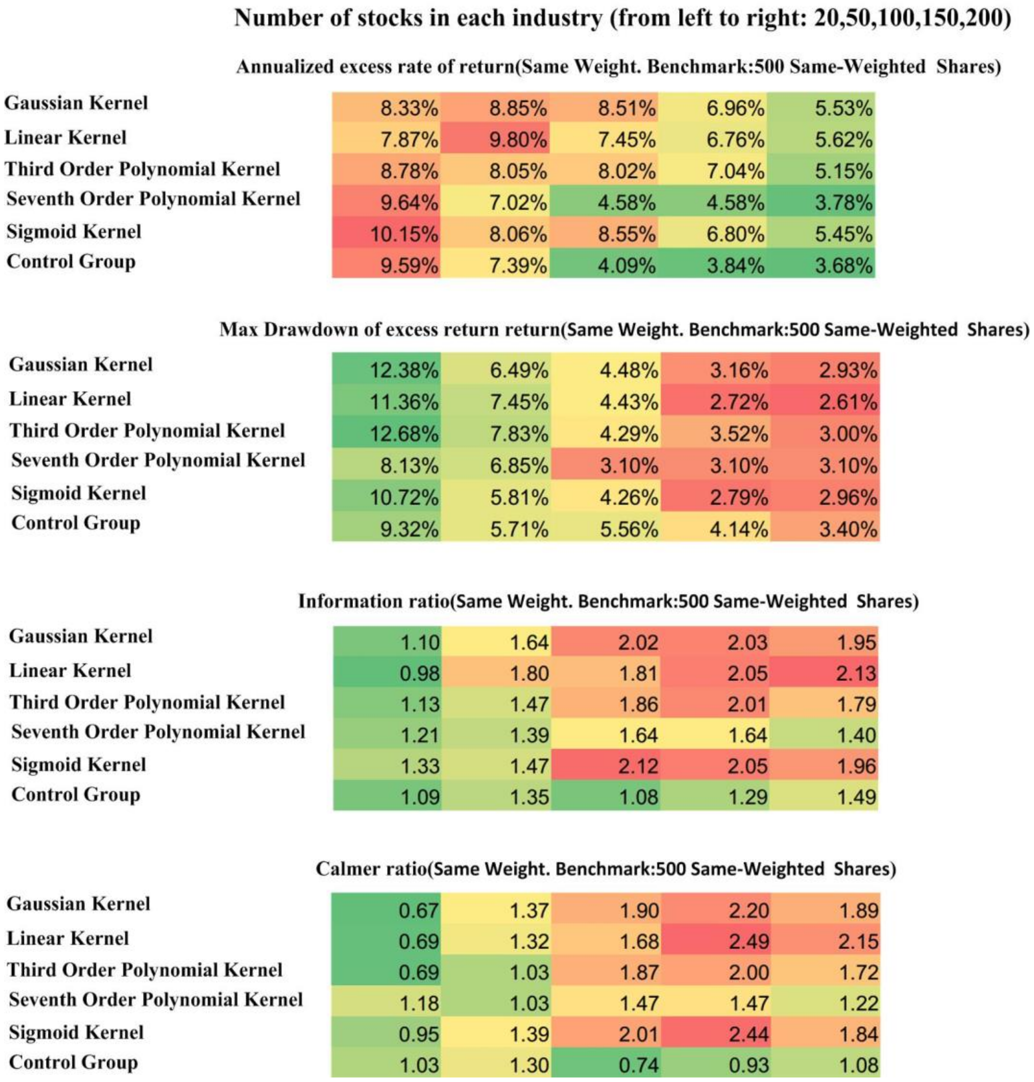
Information ratio(Industry neutral. Benchmark:CSI500)

Gaussian Kernel	1.76	2.37	2.02	2.01	1.89
Linear Kernel	1.57	2.20	2.04	1.98	1.96
Third Order Polynomial Kernel	2.17	2.34	1.98	1.91	1.77
Seventh Order Polynomial Kernel	1.74	1.89	1.91	1.75	1.71
Sigmoid Kernel	1.82	2.31	2.10	1.97	1.85
Control Group	1.34	1.50	1.58	1.70	1.79

Calmer ratio(Industry neutral. Benchmark:CSI500)

Gaussian Kernel	1.48	2.91	1.69	1.18	1.26
Linear Kernel	1.32	2.11	1.77	1.28	1.37
Third Order Polynomial Kernel	2.46	2.74	1.54	1.14	1.14
Seventh Order Polynomial Kernel	1.77	2.31	1.46	1.05	1.13
Sigmoid Kernel	1.54	3.17	1.72	1.18	1.25
Control Group	1.30	0.95	1.24	1.22	1.24

Figure 6. Comparison of the important indexes of the SVM model with different kernel functions (market neutral, the stock selection of 500 sub-unit)



Number of stocks in each industry (from left to right: 20,50,100,150,200)

Annualized excess rate of return(Same Weight. Benchmark:500 Same-Weighted Shares)

Gaussian Kernel	8.33%	8.85%	8.51%	6.96%	5.53%
Linear Kernel	7.87%	9.80%	7.45%	6.76%	5.62%
Third Order Polynomial Kernel	8.78%	8.05%	8.02%	7.04%	5.15%
Seventh Order Polynomial Kernel	9.64%	7.02%	4.58%	4.58%	3.78%
Sigmoid Kernel	10.15%	8.06%	8.55%	6.80%	5.45%
Control Group	9.59%	7.39%	4.09%	3.84%	3.68%

Max Drawdown of excess return return(Same Weight. Benchmark:500 Same-Weighted Shares)

Gaussian Kernel	12.38%	6.49%	4.48%	3.16%	2.93%
Linear Kernel	11.36%	7.45%	4.43%	2.72%	2.61%
Third Order Polynomial Kernel	12.68%	7.83%	4.29%	3.52%	3.00%
Seventh Order Polynomial Kernel	8.13%	6.85%	3.10%	3.10%	3.10%
Sigmoid Kernel	10.72%	5.81%	4.26%	2.79%	2.96%
Control Group	9.32%	5.71%	5.56%	4.14%	3.40%

Information ratio(Same Weight. Benchmark:500 Same-Weighted Shares)

Gaussian Kernel	1.10	1.64	2.02	2.03	1.95
Linear Kernel	0.98	1.80	1.81	2.05	2.13
Third Order Polynomial Kernel	1.13	1.47	1.86	2.01	1.79
Seventh Order Polynomial Kernel	1.21	1.39	1.64	1.64	1.40
Sigmoid Kernel	1.33	1.47	2.12	2.05	1.96
Control Group	1.09	1.35	1.08	1.29	1.49

Calmer ratio(Same Weight. Benchmark:500 Same-Weighted Shares)

Gaussian Kernel	0.67	1.37	1.90	2.20	1.89
Linear Kernel	0.69	1.32	1.68	2.49	2.15
Third Order Polynomial Kernel	0.69	1.03	1.87	2.00	1.72
Seventh Order Polynomial Kernel	1.18	1.03	1.47	1.47	1.22
Sigmoid Kernel	0.95	1.39	2.01	2.44	1.84
Control Group	1.03	1.30	0.74	0.93	1.08

Figure 7. Comparison of important indexes of different kernel function models (individual share equally weight, China Certificate 500 sub-unit stock)

For the industry-neutral strategy of the stock selection in the 500 component stocks, when the number of stock options in each industry is between 5 and 10, In addition to the 7-order polynomial kernel, the annual excess returns, information ratios and Calmar ratios of SVM models are significantly higher than those of the linear regression models, in which The Gaussian core, the 3-order polynomial nucleus and the Sigmoid nucleus are best performed. For the individual-share equal-weight strategy in the stock selection of the 500 constituent stocks, when the total number of stock options is greater than or equal to 100, in addition to the 7-order polynomial kernel, the annual excess returns, information ratios and Calmar ratios of SVM models are significantly higher than the linear regression models, the maximum return of excess returns is less than linear regression.

4. Conclusion

This study tested the support vector machine system of many kernel functions including linear kernel, polynomial kernel, Gaussian kernel and sigmoid kernel and used the support vector machine model to construct the stock selection strategy of Shanghai 300, China Certificate 500 to prove that the multi-factor model based on SVM has strong ability to predict stock-picking income through the back-test analysis, and analyze the differences of model prediction ability of different kernel functions, providing further theoretical basis for the application of machine learning in the field of quantitative investment.

References

Malkiel, B. G., & Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(2), 383-417. https://doi.org/ 10.2307/2325486

Asmess, C. S. (1997). The Interaction of Value and Momentum Strategies. *Finance Analysis Journal*, (3), 29-36. https://doi.org/10.2469/faj.v53.n2.2069

Chen, N., & Zhang, F. (1998). Risk and Return of Value Stocks. *Journal of Business, 71*(4), 501-535. https://doi.org/ 10.1086/209755

Mohanram, P. S. (2005). Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis. *Review of Accounting Studies, 10*(2-3), 133-170. https://doi.org/10.1007/s11142-005-1526-4

Wang, R. (2016). *Research on Multiple-factor Quantitative Stock Selection in A-share Market* [D].Shanxi University of Finance & Economics.

Su, J., & Fang, H. (2018). Research on Multiple-factor Quantitative Stock Selection Strategy Based on CSI 300 Stocks. *Journal of Fujian Business University*, (1). https://doi.org/ 10.19473/j.cnki.1008-4940.2018.01.003

Vapnik, V. N. (2002). The Nature of Statistical Learning Theory. *IEEE Transactions on Neural Networks, 8*(6), 1564-1564. https://doi.org/10.1007/978-1-4757-3264-1

Sai, Y., Zhang, F., & Zhang, T. (2013). Research of Chinese Stock Index Future Regression Prediction Based on Support Vector Machines. *Chinese Journal of Management Science, 21*(3), 35-39. https://doi.org/ 10.16381/j.cnki.issn1003-207x.2013.03.002

Han, Y., & Liu, S. (2016). Based on GARCH-SVM and AR-SVM Stock Fluctuation Forecast. *Journal of Dalian Maritime University (Social Sciences Edition), 15*(3), 25-30.

Huang, Q. (2017). Research on the Application of Support Vector Machines in China A-share Market Quantitative Strategy -- Based on Fama-fench Three-factor Model [J]. *Times Finance,* (11), 172-173.3.

Copyrights