International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 2, June 2017,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com

Empirical research on commodity futures based on SVM

Ziqiang Wang[1],Xiaoping Ren[2]

(Northeast Petroleum University school of mathematics and statistics Daqing City Heilongjiang Province China 163318)

Abstract

Commodity futures market plays an important role in the national economy. Forecasting the price of commodity futures is conducive to investors grasping the regular of price change. However, there are some problems in the effect of traditional forecasting methods. This paper researches statistical learning theory and support vector machine (SVM) theory, glass commodity futures is the research object, then establishes price forecasting model based on support vector machine (SVM),Finally, the model and prediction results are analyzed and evaluated.

Keywords:Statistical learning theory; Support vector machines; Commodity futures

1.Theoretical introduction

Support vector machine (SVM) is a kind of general effective machine learning method, it used to solve the small sample, nonlinear and high dimensional pattern recognition show many unique advantages, and can be applied to the function fitting other machine learning problems, are used in pattern recognition and time series forecasting and function estimating data mining problems widely. In essence is a kind of based on VC dimension theory, empirical risk minimization (ERM), and structural risk minimization (SRM), the study of the theory of statistical learning methods, such as learning algorithm question boils down to is a constrained quadratic programming problem.

Statistical learning theory is a theory of machine learning in the case of small sample. The theory for small sample set up a new set of theory system, under the system of statistical inference rules not only consider the asymptotic performance requirements, and the pursuit of the existing limited information under the conditions of the optimal results are obtained. In order to study the learning process of uniform convergence speed and generalization, statistical

[1] Graduate of school of mathematics and statistics of northeast petroleum university

[2]Northeast Petroleum University Youth Fund Project, <Research on quantitative investment strategy based on R-Couple-Beta and software development of stock program trading>, Item number:NEPUQN2015-1-19

learning theory defines a series of related function set performance indicators, one of the most important is the VC dimension, and another core of statistical is learning theory. The VC dimension reflects the learning ability of the function set. The larger the VC dimension, the more complex the learning machine (the stronger the learning ability).The actual risk consists of two parts of statistical learning, empirical risk (training error) and confidence, confidence range reflects the risks brought by the complex structure, it and learning machine VC dimension and the training sample. To control the process of minimizing the risk function based on a fixed number, can be minimized empirical risk (ERM), or minimize the empirical risk and confidence risks (SRM), the number of the sample with the VC dimension as control variable, is applicable to the small number of samples.

Solving the problem of support vector machine is a process of solving convex quadratic programming problems. For a given sample set:

$G = \{(x_i, y_i) : x_i \in R^n, y_i \in \{-1,1\}\}_{i=1}^{l}$, x_i^+ and x_i^- ,the problem of quadratic programming is related to the size of the training set and how to solve it and optimize the quadratic programming problem is the key to solve the optimal hyperplane.

Remember to isolate the hyperplane:

$$W \cdot X + b = 0, \qquad (1\text{-}1)$$

This paper selects the weight vector:

$$W = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\},$$

Scalar:

$$b = \omega_0$$

As an additional weight:

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_4 x_4 + \omega_5 x_5 + \omega_0 = 0. \qquad (1\text{-}2)$$

The optimal hyperplane model is:

$$\min \frac{1}{2} \| \omega \|_2^2$$
$$s.t. \quad y_i((\omega, \Phi(x_i)) - b) \geq 1, i = 1,2 \cdots, 5 \qquad (1\text{-}3)$$

The corresponding quadratic programming problem is:

$$\max W(\alpha) = -\frac{1}{2}\alpha^T Q\alpha + e^T\alpha$$
$$s.t. \quad y^T\alpha = 0$$
$$0 \le \alpha_i, i = 1,2,\cdots,5$$

$$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_5)^T \qquad (1\text{-}4)$$

α_i is Lagrange multiplier in constraint conditions,Hazen matrix Q ,Semi - positive, the two classifier can be obtained by solving:

$$u(x) = \sum_{j=1}^{l} \alpha_j y_j (\Phi(x_j), \Phi(x)) - b$$
$$y(x) = \text{sgn}(u(x)) \qquad (1\text{-}5)$$

if $\alpha_i > 0$,The corresponding sample x_i is support vector,if $\alpha_i = C$,Corresponding sample x_i Is a non-support vector.We need to use the inner product in the (1-5) formula.The kernel has the gauss RBF kernel $K(x_i, x_j) = \exp(-\frac{\| x - y \|^2}{2\delta^2})$.

2.Empirical analysis

The data selected from the data of 1000 sets of data from December 4th, 2012 to September 14th, 2016. The data stores as FG. CSV. The data derived from the TB data master. In the process of modeling, this paper chooses R software to implement SVM model.

The energy tide (OBV) predicted by the trend of statistical turnover. R language code: OBV(Cl(FG),Vo (FG)), and the results are shown in figure 2-1.

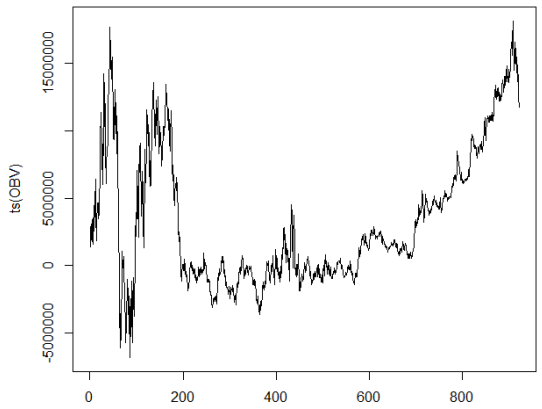

Figure 2-1 OBV Value

The exponential moving average EMA is the exponential decreasing weighted moving average, which is also a trend indicator. This paper selects 10, 30 and 50 daily. R language compute: EMA(Op(FG),n= 10). EMA10 is shown in figure 2-2.

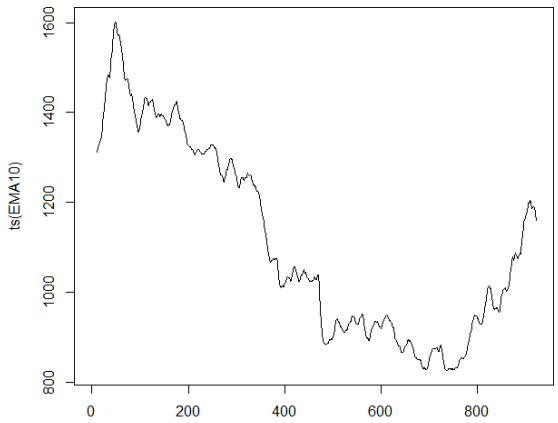

Figure 2-2 EMA10

According to the trend analysis of the opening price and EMA for 10, 30, and 50 days, the 10-day trend is shown in figure .

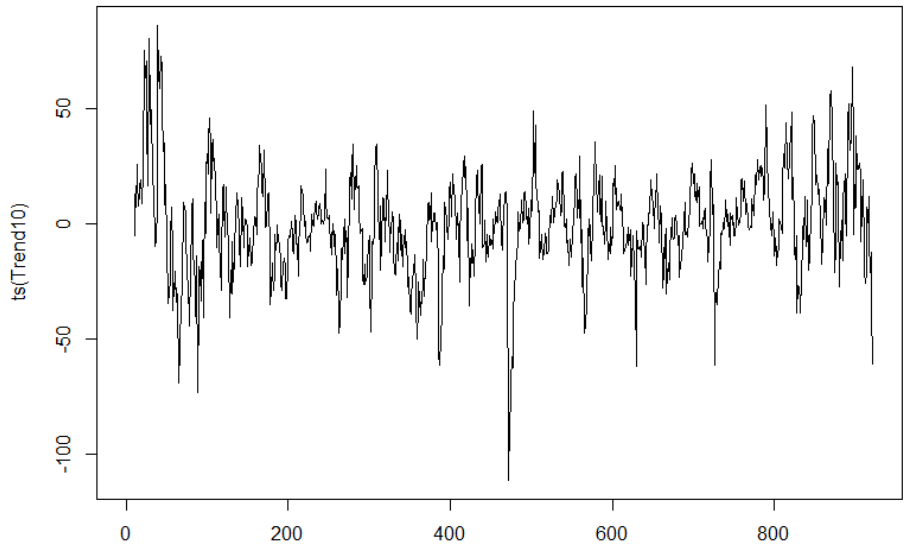

Figure 2-3 The 10-day line fluctuation

Define the pointer pv to take the Boolean value, when the closing price of the closing price >, pv = 1, the opening price is bigger than closing price, pv = 0. According to OBV, trend, 10-day momentum and pointer Pv, a new set of data ND is obtained, as shown in Table 2-4.

Table 2-4.ND data

	OBV	Trend10	Trend30	Trend50	Mom10	Positive
50	11771548	-8.9465618	65.75605354	142.96000000	13	0
51	9420818	-34.3199142	30.64275976	105.64784314	-27	0
52	11741414	-34.6253843	21.18193655	93.81851596	-49	1
53	9376014	-29.9662236	17.94439226	88.21778984	-80	0
54	11491612	-24.5178193	16.78668953	84.75826867	-91	1
55	13065538	-5.3327612	32.54238698	98.72853264	-65	1
56	11221706	7.0913772	43.53965234	108.30780587	-40	0
57	8280904	-13.0161459	19.21451348	81.96240172	-74	0
58	11073908	-37.6495740	-12.89610030	47.04230754	-82	1
59	8911082	-25.8951060	-6.45119060	50.96221704	-89	0

The NA data of the first 49 days was removed, and the training group data was 50-780. 781-811 was the data of the test group. Establish support vector machine model, kernel function select sigmoid neural network kernel function. The definition of accuracy ac is the quotient of Pv>1 in the prediction group and the number of prediction groups. The result is ac= 0.5806452. Indicates an accuracy of 58%.

Reference:

[1]Regularized least squares fuzzy support vector regression for financial time series forecasting[J] . Reshma Khemchandani,Jayadeva,Suresh Chandra.Expert Systems With Applications . 2007 (1)

[2]Using support vector machine with a hybrid feature selection method to the stock trend prediction[J] . Ming-Chi Lee. Expert Systems With Applications . 2009 (8)

[3]Financial time series forecasting using independent component analysis and support vector regression[J] . Chi-Jie Lu,Tian-Shyug Lee,Chih-Chou Chiu.Decision Support Systems . 2009 (2).