Procedia Computer Science

Volume 29, 2014, Pages 1801–1815

ICCS 2014. 14th International Conference on Computational Science

Research on the construction of macro assets price index based on support vector machine

Ping Liu[1], Jianmin Sun[2], Liying Han[1], and Bo Wang[3]*

[1] China Appraisal Society, Beijing, China
liuping@cas.org.cn, hly@cas.org.cn
[2] Pan-China Assets Appraisal Co., LTD, Beijing, China
sunjm@ccafm.com.cn
[3] University of Chinese Academy of Sciences, Beijing, China
wangbo10b@mails.ucas.ac.cn

Abstract
In this paper, a new macro assets price index (MAPI) is constructed based on support vector machine. In fact, 12 indicators, which can represent the macro economy well in both economically and statistically, are chosen to build our new index. Here, different from traditional econometric method, a novel machine learning method support vector regression machine (SVR) is employed to product the predictor of consumer price index (CPI) in China. In addition, in the experiment part, we also compare the result of SVR with that of least square regression (LSR) and vector autoregressive (VAR) impulse response analysis. The comparison shows that the latter two methods are hard to satisfy the requirement in both economically and statistically. On the contrary, SVR gives a good predictor of CPI and exhibits a manifest leading of CPI. In other words, our index can forecast the trends by 4 to 6 months, which is useful for investment and policy making.

Keywords: Macro assets price index, Financial conditions index, regression, support vector machine

1 Introduction

In the last twenty years, the research of financial conditions index (FCI), which can show the financial condition, has received more and more attention. FCI derived from Monetary Conditions Index (MCI), which was studied and constructed by Bank of Canada (BOC) in 1990s. This concept was firstly posed by Freedman (1994) [5], which pointed out that some variables representing monetary conditions were very important to monetary policy making. After the emergence of MCI, Goodhart and Hofmann (2001) [7] noticed the important role that asset price played in monetary policy conduction. For example, the change of stock price

*Corresponding Author

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.165

and real estate price influenced the demand of money obviously. Consequently, we could build composite index to represent financial conditions. Two important steps in the construction of FCI are feature selection and weight determining. Different ways of dealing with this two steps produce different FCIs. Mayes (2001) [10] constructed the FCI of 11 European countries by IS curve function. English et al. (2005) [4] tested the ability of macro economy prediction of diverse financial variables. Then, using Stock-Watson model, they built the FCI of Germany, UK and US, basing on factor analysis. Also, there are many other different methods of FCI construction [6][1]. However, the econometric methods of FCI construction neglect the nonlinear relationship between the input and output variables, which doesn't agree with the reality. In fact, the relationship among the economic indicators changes over time [8][7].

Actually, in machine learning, the construction of FCI can be treated as a standard regression problem. As we have explained above, traditional econometric methods are not suitable for this kind of regression, because of strong assumptions they need. To overcome this flaw, support vector regression (SVR) is employed in this paper. SVR is derived from support vector machine (SVM), which is firstly studied by Vapnik (1998, 2000) [15][16]. Based on the principle of the structural risk minimization, SVM algorithm focuses on not only empirical risk minimization, but also the VC dimension of the function. This principle promises a good generalization ability of SVM, and shows a widely application. In SVM, regression problem is changed into a classification problem. Kernel technique makes SVM suitable for many nonlinear problem in reality [2][3]. As a result, SVR can be applied for forecasting. For some economical and financial time series dealing problems, comparing with other econometric methods, it promises to yield a quite reliable prediction [9]. In addition, among many different machine learning methods, SVR also provides a competitive result. Trafalis et al. (2004) [14] studied the results of using different machine learning methods to forecast stock prices. The regression result showed that SVR is much more suitable for this kind of financial time series analysis than Back propagation and RBF networks. The research of Wang et al. (2012) [17] showed SVR was suitable for constructing FCI in China.

This paper proposes a brand new method to construct FCI, which is named as Macro Asset Price Index (MAPI), by taking advantage of SVR. In addition, we compare the result of SVR with other different econometric models, to show the advantage of SVR in dealing with this kind of economical problem. The following of this paper is organized as follows. Details of SVM is introduced in Section 2. Section 3 shows the feature selection and lag time choosing process of our problem, as well as the result of different methods. We will conclude our work and conceive future work in Section 4.

2 Support Vector Machine

Basing on the research of convex quadratic programming, SVM algorithm takes a great development recently. Furthermore, taking advantage of kernel technique, we can apply SVM in many complicated real-life problem with small data set [2][3][12]. Mostly, it can be used to deal with classification problem and regression problem, which are called SVC and SVR respectively. Next, we will introduce these models.

2.1 Support Vector Classification Machine

Considering a typical binary classification problem, we can describe it as follows.

Suppose that there is a training set $T = \{(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \cdots, (\boldsymbol{x_l}, y_l)\} \in \mathscr{X} \times \mathscr{Y}$. Here, $\boldsymbol{x_i} \in \mathscr{X} = \mathbb{R}^n, y_i \in \mathscr{Y} = \{1, -1\}$. The algorithm tries to find a real function $g(x)$ on $\mathscr{X} = \mathbb{R}^n$.

Then, the decision function can be constructed as follows:

$$f(\boldsymbol{x}) = sgn(g(\boldsymbol{x})).$$

In this case, we can tell the class y of every single pattern $\boldsymbol{x}$. Here, $sgn(\cdot)$ can be defined as follows:

$$sgn(a) = \begin{cases} 1, & a \geqslant 0; \\ -1, & a < 0. \end{cases}$$

A two dimension case can be shown in Figure 1. In the figure, the symbols "+" and "-"

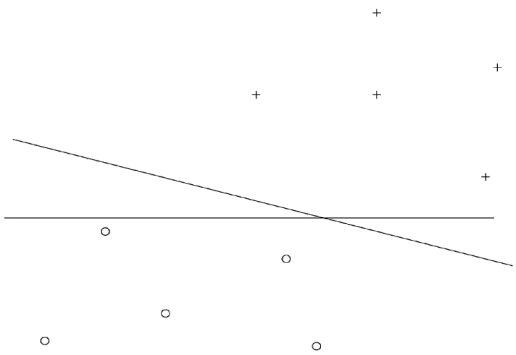

Figure 1: A linear separable case

represent two different classes respectively, i.e. "1" and "-1". Then, the lines represent different hyperplanes which are obtained by the algorithm. Each hyperplane separates the input space into two different parts. Then, the points distributing on different sides of the hyperplane belong to different classes.

Furthermore, we give the original problem and dual problem of soft margin SVC.

$$\begin{aligned} (Original\ Problem) \qquad & \min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad \tfrac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{l}\xi_i \\ s.t. \quad & y_i((\boldsymbol{w}\cdot\boldsymbol{x_i}) + b) + \xi_i \geqslant 1, i = 1, 2, \cdots, l, \\ & \xi_i \geqslant 0, i = 1, 2, \cdots, l. \end{aligned} \tag{1}$$

$$\begin{aligned} (Dual\ Problem) \qquad & \min_{\boldsymbol{\alpha}} \quad \tfrac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^{l}\alpha_j \\ & s.t. \qquad \sum_{i=1}^{l} y_i \alpha_i = 0, \\ & 0 \leqslant \alpha_i \leqslant C, i = 1, 2, \cdots, l. \end{aligned} \tag{2}$$

If we solve the original problem, the solution $(\boldsymbol{w^*}, b^*, \boldsymbol{\xi^*})$ will help to build the final decision hyperplane $\boldsymbol{w^*}\cdot\boldsymbol{x} + b^* = 0$. However, in fact, we solve the dual problem instead of the original problem. After getting the solution vector $\boldsymbol{\alpha^*}$, we calculate $\boldsymbol{w^*} = \sum_{i=1}^{l} y_i\alpha_i^*\boldsymbol{x_i}$ and $b^* = y_j - \sum_{i=1}^{l} y_i\alpha_i^*(\boldsymbol{x_i}\cdot\boldsymbol{x_j})$. Here, suppose that $\alpha_j^* > 0$. As a result, the final decision hyperplane can be expressed as follows:

$$\sum_{i=1}^{l} y_i\alpha_i^*(\boldsymbol{x_i}\cdot\boldsymbol{x}) + y_j - \sum_{i=1}^{l} y_i\alpha_i^*(\boldsymbol{x_i}\cdot\boldsymbol{x_j}) = 0 \tag{3}$$

2.2 Support Vector Regression

For the regression problem, suppose that $\boldsymbol{x_i} \in \mathscr{X} = \mathbb{R}^n, y_i \in \mathscr{Y} = \mathbb{R}$. Then, we give the original problem and dual problem of SVR as follows:

$$
\begin{aligned}
(\textit{Original Problem}) \quad & \min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\xi^*}} \quad \tfrac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\
& s.t. \quad (\boldsymbol{w}\cdot\boldsymbol{x_i}) + b - y_i \geqslant \varepsilon + \xi_i, i = 1,2,\cdots,l, \\
& \qquad y_i - (\boldsymbol{w}\cdot\boldsymbol{x_i}) - b \geqslant \varepsilon + \xi_i^*, i = 1,2,\cdots,l, \\
& \qquad \xi_i, \xi_i^* \geqslant 0, i = 1,2,\cdots,l.
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
(\textit{Dual Problem}) \quad & \min_{\alpha_T^{(*)}} \tfrac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i,x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) \\
& s.t. \quad \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0, \\
& \qquad 0 \leqslant \alpha_i, \alpha_i^* \leqslant C, i = 1,2,\cdots,l
\end{aligned} \tag{5}
$$

where $\alpha_T^{(*)} = (\alpha_1, \alpha_1^*, \cdots, \alpha_l, \alpha_l^*)$ and $K(x_i, x_j) = (\Phi(x_i), \Phi(x_j))$ is the kernel function.

After solving the dual problem above, we construct the decision function as follows:

$$
f(x) = (\overline{w}\cdot x) + \overline{b} = \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)K(x_i, x) + \overline{b},
$$

where $\overline{b}$ is computed as follows: either choose one $\overline{\alpha}_j \in (0, C)$, then

$$
\overline{b} = y_j - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)K(x_i, x_j) + \varepsilon;
$$

or choose one $\overline{\alpha}_i^* \in (0, C)$, then

$$
\overline{b} = y_j - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)K(x_i, x_j) - \varepsilon.
$$

3 Macro Assets Price Index

In this section, we apply SVR to macro assets price index construction. Actually, the purpose of constructing macro assets price index is to describe the overall macro assets price conditions. Noting that CPI is an important index of macro economy circumstance by measuring the prices of some certain goods and service. As a result, we choose CPI as the target. Then, regression between some financial indicators and nonfinancial indicators is executed. Finally, after comparing with other econometric methods, SVR is chosen for our regression problem. Here, we display all the tables in appendix A.

3.1 Flowchart of the Construction of Macro Index

In our problem, we have five important steps of constructing the final index. The first step is to select the original indicators (independent variable). Then, we need to select the time interval and collect the data. In the second step, preprocessing the data and filtering the features are

needed. Next, in the third step, because the data of financial indicators are often time series, we need to fix the lag time of each indicator by the accordance between the regression results and the signification in economy. In the fourth step, we compare the results of other different models and methods, to show the effectiveness of SVR in our problem. In the last step, construct the index by using the selected indicators with certain lag times.

The following Figure 2 shows the flowchart of the construction of macro index.

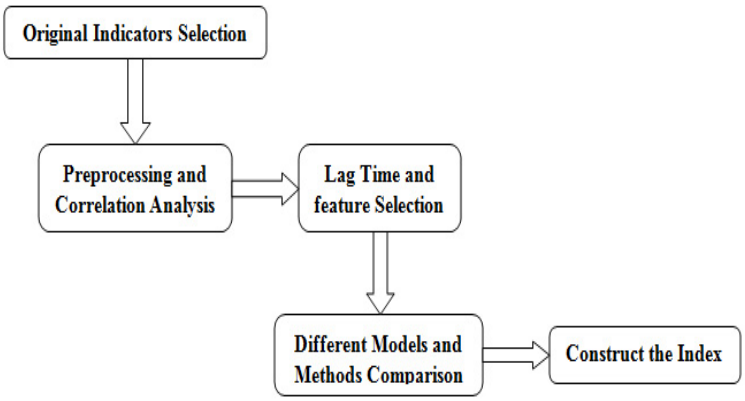

Figure 2: Flowchart of the construction of macro index

3.2 Original Indicators Selection

Based on the meaning of indicators in economy, we mainly choose indicators in three different but related fields, i.e. financial market, nonfinancial market and social investigation. Particularly, in the financial market, we choose China interbank offered rate (CHIBOR), Interbank collateral repo rate (IBCRR), RMB real effective exchange rate index (RMB-REER), RMB exchange rate against dollar (RMB-ERAD), Shanghai and Shenzhen 300 index (SHSE300), Shanghai Composite Index (SCI), Shanghai composite index of treasury bonds (SCITB), SSE fund index (SSEF), AU9995, AU9999, The dollar index (DI) and so on. In the nonfinancial market, we choose Purchase price index of industrial production (PPIIP), Producer price index of industrial (PPII), Producer price index of agricultural (PPIA), Price index of agricultural means of production (AMPI), Total corporate goods price index (CGPI), Retail price index (RPI), Price index for investment in fixed assets (FAPI) and so on. In social investigation, we choose Consumer confidence index (CCI), Entrepreneurs confidence index (ECI), Business climate index (BCI), Banker confidence index (BACI), Loan demand index (LDI) and so on.

Table 1 shows all the original indicators chosen in our method. The frequency of all the data, including CPI (year-on-year), is monthly. Some indicators with quarterly data have been interpolated into monthly data. The data period of the data is from January, 2006 to December, 2011 (72 months). Then, we do seasonal adjustment to the smoothed data.

3.3 Correlation Analysis

Firstly, in time series analysis, non-stationary data is often processed into stationary data. As a result, in our problem, we compare the effects between this two different kinds of data by

using SVR with linear kernel. The reason of the kernel choosing is we need to find the exact relationship between the indicators and CPI, which is clearer by using linear kernel than any other kernel. Another reason is that we can select indicators through the coefficients directly, according to the relationship between the indicators and CPI in economy. The SVR result shows that stationary data doesn't guarantee better result. To the contrary, the original data yields better result in this problem. Therefore, we keep the experiment data the same as the original one.

Next, we analyze the correlation between every indicator and CPI by calculating the correlation coefficients and maximum information coefficient (MIC) between them. Here, MIC is a kind of method to help finding the linear and nonlinear correlation between two variables, which was firstly proposed by Reshef et al. (2011) [11]. The value of MIC is between 0 and 1. The value 0 shows the weakest correlation between two variables. The value 1 shows the strongest correlation between two variables. We remove the indicator that shows weak correlation in both in correlation coefficients and MIC.

3.4 Lag Time and Feature Selection

After correlation analysis, we remove SCI, AU9999 and AU9995 for the following steps. Furthermore, existing data mining models cannot deal with high dimension well, so more features need to be cut. By virtue of traditional Econometrics method, we remove the indicators with weak prediction power or collinearity to each other. In other words, the indicators with strong prediction power to CPI and independence are kept as well.

On the other hand, taking advantage of lag time method and economic meaning, we remove the indicators that are not suitable for our experiment. In fact, the lag time will influence the relationship among the indicators in the following aspects.

Psychological Expecting Factor The main body of economic activity is human being. Psychological reasons make people have some time delay in reacting to the changes of economic condition and economic environment, namely, the lag of decision making. For example, the consumption amount of certain commodity is interrelated to not only the current price of this certain commodity, but also the price in the future people expecting.

Technique Factor In the operation of national economy, from production to currency, and then, to consumption, time lag is inevitable. This kind of situation is very easy to understand, because of the time every single economic link taking.

The Influence of Asset Price Transmission Channel in Monetary Policy Monetary policy is a set of measures that the government or central bank takes in order to influence or control the economic activity. In this way, the government can achieve certain economic goals, such restraining the inflation, maintaining the economic growth and increasing employment rate. The monetary policy determines money supply in some degree, which is the beginning of currency. The theoretical basis of asset price transmission mechanism are the q theory of investment and the wealth effect theory [13]. On one hand, Tobin's q theory defines a function q, which is the ratio between vendibility of company asset and capital replacement cost. The transmission mechanism of q theory is as follows:

$$\begin{aligned} &money\ supply \uparrow \rightarrow price\ of\ interest\ rate\ and\ stock \uparrow \\ &\qquad \rightarrow q \uparrow \rightarrow investment \uparrow \rightarrow output - input\ ratio \uparrow \end{aligned} \tag{6}$$

On the other hand, Modigliani's wealth effect theory says that consumer expenditure is determined by the consumer's life estate. The transmission mechanism of wealth effect theory is as follows:

$$\begin{aligned} &money\ supply \uparrow \rightarrow interest\ rate \uparrow \rightarrow the\ price\ of\ stock \uparrow \\ &\qquad \rightarrow wealth \uparrow \rightarrow consumption \uparrow \rightarrow output - input\ ratio \uparrow \end{aligned} \tag{7}$$

The theories introduced above indicate that the fluctuation financial indicators will effect inflation, and is ahead of the fluctuation of nonfinancial indicators. As a result, the choosing of lag time is important in the research of asset price.

3.5 The Choosing of Lag Time

In our experiment, considering the effect of time lag adequately, we need to fix the lag time of every single indicator, which is important to make sure our model can reflect economic phenomenon objectively. There are three different steps in our experiment by which we can fix the lag time of every selected indicator.

3.5.1 All the indicators with the same lag time

In the first round, we keep all the indicators with the same lag time, i.e. lag 1, lag 3. The following tables show the whole process.

As we can see in Table 2, the regression coefficients of marked indicators, IBCRR, AU9999 and AU9995, don't consist with the correlation analysis result, although they are stable with different parameters. The marked indicator DI is sensitive with data and parameter. Besides, the marked indicator SCI is not stable when the lag time changes. In addition, the correlation between gold products and China financial market is weak. At the same time, the correlation between SCI and HS300 is strong. As a result, in the round, we remove the indicators marked in bold, which are IBCRR, AU9999, AU9995, SCI, RMB-ERAD and DI.

In the second round, we still keep all the indicators with the same lag time, i.e. lag 1, lag 3 and lag 4.

In Table 3, the marked indicator RPI is not stable with diverse parameters of the same lag time. Meanwhile, RPI has a strong correlation with CGPI. The marked indicator BACI has a negative correlation with CPI, which conflicts with the real economic meaning. As a result, we remove RPI and BACI in our following experiment.

According to Table 4 and economic meaning, we remove eight different indicators, including IBCRR, RMB-ERAD, SCI, AU9999, AU9995, DI, RPI and BACI.

3.5.2 Financial and nonfinancial indicators with different lag times

As we all know, financial market is mostly ahead of economic market. As a result, the financial indicators will play a role of early warning to the development of economy, which leads us to choose financial and nonfinancial indicators with different lag times.

According to the former study, the financial indicator is usually ahead of the nonfinancial indicator with 4 to 6 lag time. Consequently, in this round, we choose 4 to 6 lag time for the financial indicators, 1 to 3 lag time for the nonfinancial indicators.

In this round, basing on the result of SVR on different lag times data, we can fix most of the indicators' lag time, such as HS300 (lag 4), SSEF (lag 4), CCI (lag 1) and PPIA (lag 3), excepting CHIBOR and SCITB. For CHIBOR and SCITB, the regression coefficient in this round is negative, which is not consistent with real economic analysis. As a result, we test

different lag time (lag 1, lag 2, lag 3) in the following fourth round. The result of this round is in Table 5

As we can see in Table 5, CHIBOR (Lag 3) and SCITB (Lag 3) are more effective than other lag times. At the same time, according to the regression coefficient of SVR, LDI, PPII and PPIIP have negative correlations with CPI, which doesn't agree with economic analysis. Therefore, we remove these three indicators in our following experiment.

Finally, the rest 12 indicators (with lag time) are used to construct our macro assets price index (MAPI). The result of SVR is displayed in Tabel 6.

3.6 Methods Comparison and Results

In this section, at first, we apply least square regression (LSR) and vector autoregression (VAR) impulse response to our data. Then, compare the results of different methods. At last, the overall result of SVR is displayed in

3.6.1 Methods results comparison

Least square regression First of all, feature selection and lag time choosing are inevitable, no matter what method is applied. Here, least square regression is used for feature selection and lag time choosing. As we can see, in Table 7, the coefficient is sensitive to different lag times and not stable to any lag time choice. As a result, LSR is not suitable to our data in this experiment.

Vector autoregression impulse response As we all know, the requirement of VAR impulse response is strict to the data. Particularly, same order stationary data is needed. According to the test result, some of the indicators don't accord with this requirement. There are only 9 indicators satisfy with one order stationary condition at last. However, there are more than half of the 9 indicators cannot converge at time 20, which means VAR impulse response method is not suitable to our data. Moreover, the result of VAR doesn't accord with economic meaning and with weak prediction effect. Figure 3 shows the result of VAR impulse response. In the figure, all the indicators are one order stationary sequence. Concretely, these indicators are CHIBOR, HS300, SSEF, CGPI, PPII, PPIIP, AMPI, PPIA and BCI, respectively.

Support vector regression Table 6 shows that we can obtain a more reasonable and reliable result by applying SVR to our data. Particularly, according to the weight coefficient in Table 6, there are some conclusion we can make, which are list in the following.
(1) **The sign of every indicator's coefficient agrees with the economic implication.** In particular, RMB-REER and CHIBOR have negative correlations to CPI, while the rest indicators all have a positive anticipation to the future inflation.
(2) **Generally speaking, the value of price indicators' coefficients agree with the study of most scholars.** For example, considering the current financial market in China, the conduction effect of interest rate, treasury bond and fund to economy is weaker than that of stock and exchange rate. Among these indicators, because of the largest proportion of stock market in asset market, stock has a major influence on the macro economy, which has been reflected by the relatively large value of its regression coefficient. In the nonfinancial market, because of the largest proportion of investment in the economy increase in China, especially the large proportion of fixed asset investment, the value of FAPI's coefficient is relatively larger.
(3) **The coefficients of social survey indicators, such as ECI and BCI, are larger**

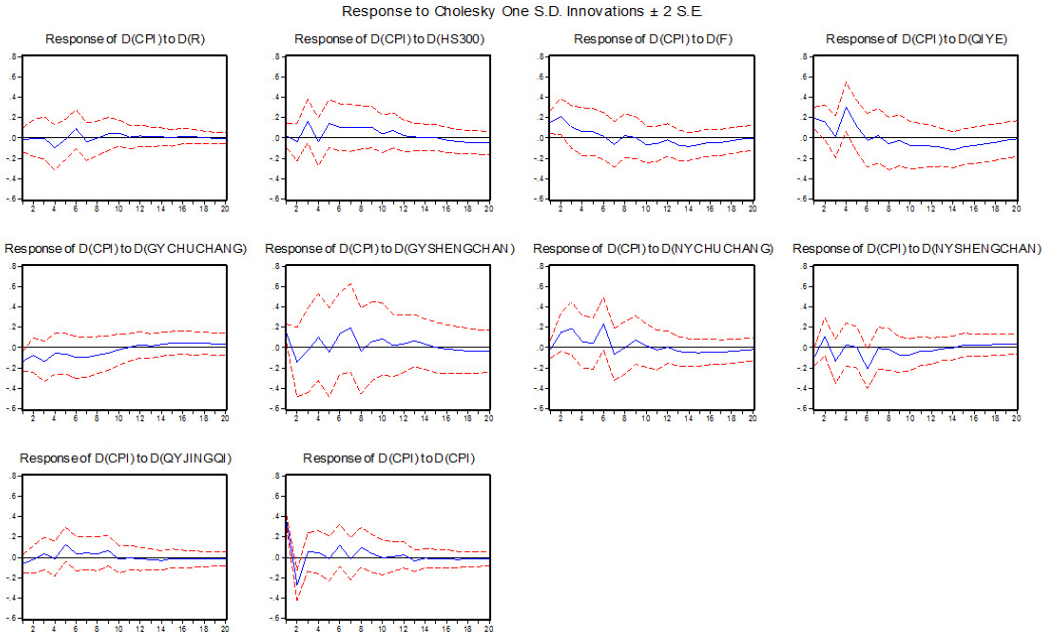

Figure 3: The result of VAR impulse response

than others. There are two possibilities of this situation. One is subjective anticipation has a comparatively large influence on China's economy. The other one is the interpolation has an impact on the regression result.

Macro index construction Here, we show the graph of our macro asset price index (MAPI) which is constructed by SVR. Figure 4 shows that our MAPI climaxes in the year of 2007, then, it begins to drop, which reaches the bottom in the end of 2008. Later, MAPI increases, with little fluctuation during 2009 to 2011, which can reflect the world's economy situation around the year of 2008. Meanwhile, MAPI has a similar trend with CPI, leading CPI with 4 to 6 months. Particularly, in the wave crest, MAPI leading CPI with about 4 months, in the trough of wave, MAPI leading CPI with about 6 months.

The leading characteristic of our index to CPI can provide a reference to economy policy decision makers and also help the investor to construct a intuitive anticipation to the trend of macro economy. Figure 4 shows the trends of our MAPI and CPI. As two vertical lines indicate, MAPI behaves a leading characteristic to CPI by 4 to 6 months. Furthermore, we display the trends of MAPI and macro prosperity consistent index (MPCI) in Figure 5, which shows MAPI has a similar trend with MPCI, leading with a conspicuous period.

4 Conclusion and Future Work

This paper proposes a brand new method of FCI construction, which is based on machine learning algorithm SVR. The work in this paper can be divided into two different part. The first one is feature selection. In this part, we choose 23 original indicators from financial market, nonfinancial market and social investigation. Considering the influence of lag time, we try to

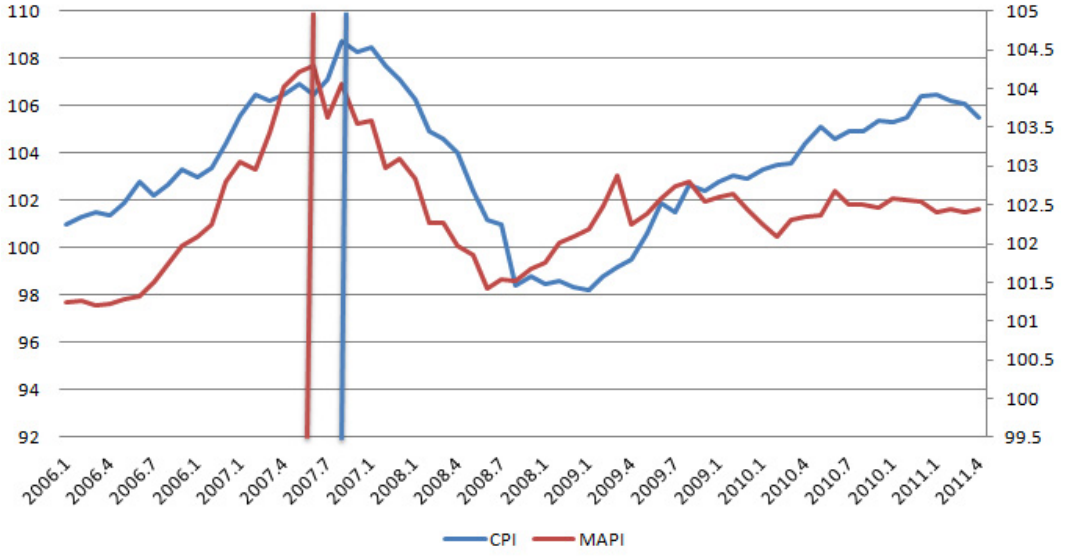

Figure 4: The trends of MAPI and CPI

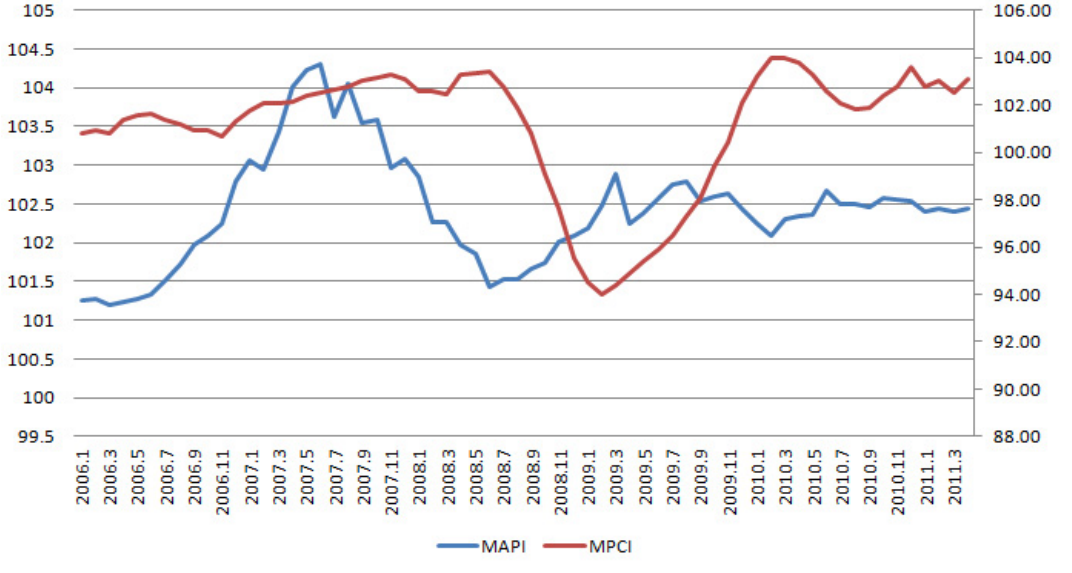

Figure 5: The trends of MAPI and MPCI

find the most significant lag time of each indicator, which can product a promising result and make sense in economy. After 5 rounds of wrapped feature selection, according to economical meaning of every single indicator, 12 indicators with certain lag time are involved in the final regression. Then, in the second part, taking advantage of SVR, we can obtain the weight vector of the indicators and build MAPI. Here, linear SVR is chosen, because we need to find the relationship between each indicator and CPI clearly. Intuitively, it can be reflected in the sign and value of each weight.

In addition, a comparison of different methods shows that SVR is more suitable for our problem than other econometric methods, such as LSR and VAR. Furthermore, our MAPI shows a good leading characteristic to CPI with 4 to 6 months. In the analysis of macro economy, MAPI has a good value for asset investment and macro economy policy making.

For the future work, we suggest to release our new index MAPI quarterly. As a result, weight vector of indicators should be adjust by SVR on new macro data. Also, different kernel functions can be used to our problem in order to find the best regression result.

4.1 Acknowledgments

This work has been partially supported by grants from National Natural Science Foundation of China (No.11271361, No.71331005), Major International (Regional) Joint Research Project (No. 71110107026).

References

[1] G. Dai and J. Zhang. Study on finance condition index and asset price channel of monetary policy in china. *Journal of Financial and Economics*, 7:52–62, 2009.

[2] N.Y. Deng and Y.J. Tian. *The New Approach in Data MiningSupport Vector Machines.* Science Press, Beijing, 2004.

[3] N.Y. Deng and Y.J. Tian. *Support Vector Machine: Theory, Algorithms and Extensions.* Science Press, Beijing, 2004.

[4] William English, Kostas Tsatsaronis, and Edda Zoli. Assessing the predictive power of measures of financial conditions for macroeconomic variables. *Bank for International Settlements, Investigating the relationship between the financial and real economy, BIS papers*, 22:228–252, 2005.

[5] Charles Freedman. The use of indicators and of the monetary conditions index in canada. *Frameworks for monetary stability: Policy issues and country experiences*, pages 458–476, 1994.

[6] Cline Gauthier, Christopher Graham, and Ying Liu. Financial conditions indices for canada. *Bank of Canada Working paper*, 2, 2004.

[7] Charles Goodhart and Boris Hofmann. Asset prices and the conduct of monetary policy. *Working paper of Conference on Asset Market and Monetary Policy*, 2001.

[8] Jan Hatzius, Peter Hooper, Frederic S Mishkin, Kermit L Schoenholtz, and Mark W Watson. Financial conditions indexes: A fresh look after the financial crisis. *NBER Working Paper, Issued in July*, 16150, 2010.

[9] Kyoung jae Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2):307–319, 2003.

[10] David G Mayes. Financial conditions indexes. *Bank of Finland Working Paper*, 9(17), 2001.

[11] David N. Reshef, Yakir A. Reshef, Hilary K. Finucane, Sharon R. Grossman, Gilean McVean, Peter J. Turnbaugh, Eric S. Lander, Michael Mitzenmacher, and Pardis C. Sabeti. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.

[12] Y Shi, Y Tian, G Kou, Y Peng, and J Li. *Optimization Based Data Mining: Theory and Applications.* Springer, 2011.

[13] J. Tobin. A general equilibrium approach to monetary theory. *Journal of money, credit and banking*, 1(1):15–29, 1969.

[14] Theodore B. Trafalis and Huseyin Ince. Support vector machine for regression and application to financial forecasting. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN)*, pages 348–353, 2000.

[15] V.N. Vapnik. *Statistical learning theory.* Wiley, New York, 1998.

[16] V.N. Vapnik. *The nature of statistical learning theory.* Springer, New York, 2000.

[17] Ye Wang, Bo Wang, and Xinyang Zhang. A new application of the support vector regression on the construction of financial conditions index to cpi prediction. *Procedia Computer Science*, 9:1263–1272, 2012.

A Tables

Table 1: The original indicators

Asset Type	Indicator Type	Original Indicators
Financial Market	Interest Rate	CHIBOR
		IBCRR
	Exchange Rate	RMB-REER
		RMB-ERAD
	Stock	SHSE300
		SCI
	Bond	SCITB
	Fund	SSEF
	Gold	AU9995
		AU9999
		DI
Nonfinancial Market	Industry	PPIIP
		PPII
	Agriculture	PPIA
		AMPI
		CGPI
	Retail	RPI
	Investment	FAPI
Social Investigation Index		CCI
		ECI
		BCI
		BACI
		LDI

Table 2: The first round indicator selection

Indicator	Lag 1 (C=2.1, e=0.024)	Lag 3 (C=3.68, e=0.045)
IBCRR	**0.1681**	**0.2731**
CHIBOR	-0.1800	-0.3919
RMB-ERAD	**0.2320**	**-0.0036**
RMB-REER	-0.0226	-0.2406
HS300	0.1753	0.1104
SCI	**-0.2812**	**0.0463**
SCITB	-0.2812	0.1764
SSEF	0.1897	-0.0992
AU9999	**0.1325**	**0.0358**
AU9995	**0.0676**	**0.0078**
DI	**-0.0153**	**0.0107**
RPI	0.0039	-0.0215
CGPI	0.1923	0.2910
PPII	-0.1875	0.0574
PPIIP	-0.0225	-0.4258
PPIA	0.3841	0.4068
AMPI	0.1010	0.0515
FAPI	0.2947	0.1657
BACI	-0.0794	-0.1108
LDI	0.0075	-0.0019
BCI	0.0395	0.1107
ECI	0.1092	0.2580
CCI	0.0361	0.0957
b	0.17	-0.06
MSE	0.0020	0.0057
R-squared	0.9698	0.9187

Table 3: The second round indicator selection

Indicator	Lag 1 (C=9.8, e=0.03)	Lag 3 (C=1.2, e=0.053)	Lag 4 (C=6.6, e=0.045)
CHIBOR	-0.0169	-0.0797	-0.0321
RMB-REER	-0.0935	-0.1942	-0.1399
HS300	0.0032	0.1252	0.1646
SCITB	0.0357	0.1479	0.0789
SSEF	0.0542	-0.0248	0.1264
RPI	**-0.0699**	**0.0324**	**0.0295**
CGPI	0.2379	0.1536	0.1479
PPII	-0.1841	0.0002	0.0307
PPIIP	0.0617	-0.1852	-0.065
PPIA	0.5665	0.3799	0.2584
AMPI	0.043	0.0131	-0.1596
FAPI	0.1992	0.0813	0.0078
BACI	**-0.0998**	**-0.1519**	**-0.1019**
LDI	0.0037	-0.0263	0.0114
BCI	0.0313	0.1758	0.1981
ECI	0.0717	0.2307	0.2588
CCI	0.0119	0.0977	0.0172
b	-0.02	-0.03	0.06
MSE	0.0029	0.0043	0.0055
R-squared	0.9569	0.9361	0.9196

Table 4: The third round indicator selection

Indicator	Lag 1 (C=2.2, e=0.04)	Lag 3 (C=3.9, e=0.045)	Lag 4 (C=1.18, e=0.009)
IBCRR			
CHIBOR	-0.0585	-0.0046	-0.478
RMB-ERAD			
RMB-REER	-0.0581	-0.1775	-0.1087
HS300	0.1505	0.1554	0.2287
SCI			
SCITB	-0.0855	-0.0003	0.0147
SSEF	0.1293	0.1079	0.0776
AU9999			
AU9995			
DI			
RPI			
CGPI	0.1172	0.1961	0.1444
PPII	0.0601	0.0073	0.0149
PPIIP	0.0491	-0.0991	-0.1464
PPIA	0.2653	0.3636	0.3098
AMPI	0.068	0.0172	-0.2099
FAPI	0.1069	-0.0151	0.1005
BACI			
LDI	-0.0104	0.045	0.0396
BCI	0.037	0.1657	0.1811
ECI	0.093	0.0698	0.2571
CCI	0.016	0.051	0.0325
b	0.05	-0.02	0.06
MSE	0.0036	0.006	0.0055
R-squared	0.9509	0.915	0.9199

Table 5: The fourth round indicator selection

Indicator	C=16, e=0.023 MSE=0.0053 R-squared=0.9226	C=24.9, e=0.031 MSE=0.0054 R-squared=0.9211
RMB-REER (Lag 4)	-0.0278	-0.0317
HS300 (Lag 4)	0.2484	0.2438
LDI (Lag 3)	-0.0209	-0.018
PPII (Lag 1)	-0.0294	-0.0274
PPIIP (Lag 1)	-0.097	-0.0955
FAPI (Lag 1)	0.0455	0.0431
PPIA (Lag 1)	0.0017	0.0071
AMPI (Lag 3)	0.3032	0.2962
ECI (Lag 3)	0.1883	0.1907
BCI (Lag 3)	0.179	0.1819
CGPI (Lag 1)	0.1125	0.1097
SSEF (Lag 4)	0.0559	0.0587
CCI (Lag 1)	0.0815	0.0759
CHIBOR (Lag 1)	**0.1316**	**0.1406**
CHIBOR (Lag 2)	**-0.023**	**-0.0158**
CHIBOR (Lag 3)	-0.0342	-0.0356
SCITB (Lag 1)	**-0.0192**	**-0.0253**
SCITB (Lag 2)	**-0.0133**	**-0.0193**
SCITB (Lag 3)	0.0055	-0.002

Table 6: The fifth round indicator selection

Indicator	C=17.96, e=0.063, MSE=00005 R-squared=0.9272
RMB-REER (Lag 4)	-0.0541
HS300 (Lag 4)	0.2341
FAPI (Lag 1)	0.0364
PPIA (Lag 1)	0.0206
AMPI (Lag 3)	0.2738
ECI (Lag 3)	0.1823
BCI (Lag 3)	0.1948
CGPI (Lag 1)	0.079
SSEF (Lag 4)	0.0411
CCI (Lag 1)	0.0356
CHIBOR (Lag 3)	-0.0132
SCITB (Lag 3)	0.0305

Table 7: The result of least square regression

Indicator	Coefficient	P-value
c	-90.7	0.000
RMB-REER (Lag 4)	-0.11	-0.008
HS300 (Lag 4)	0.002	0.000
LDI (Lag 3)	0.018	0.430
PPII (Lag 1)	0.774	0.000
PPIIP (Lag 1)	-0.6	0.000
FAPI (Lag 1)	0.204	0.004
PPIA (Lag 1)	0.053	0.210
AMPI (Lag 3)	0.062	0.220
ECI (Lag 3)	0.037	0.180
BCI (Lag 3)	0.049	0.110
CGPI (Lag 1)	0.025	0.850
SSEF (Lag 4)	0	0.000
CCI (Lag 1)	0.053	0.120
CHIBOR (Lag 3)	-0.51	0.000
SCITB (Lag 3)	0.27	0.000