An SVM Based Analysis of US Dollar Strength

Xiongfeng LI* and Xinyu FAN†

December 13, 2013

Abstract

The basic aim of this project is to provide a machine learning model to explain the strength of US Dollar. Support vector machine (SVM) is a promising method for the analysis of financial time series which we employ. Macroeconomic factors of fundamental analysis are selected as the features of the model. The learning results and out-of-sample test show that SVM provides a good performance to US Dollar Index analysis while the Gaussian kernel is the best through the cross validation process. Furthermore, feature selection is discussed, while CPI and PMI are the most significant features in assumption of Gaussian kernel.

Keywords: support vector machines, financial time series, US dollar strength

Introduction

A financial time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones Industrial Average. The distribution of financial time series is changing over the time. Modeling such dynamical and non-stationary time series is expected to be a challenging task.
Basically, the SVM maps the inputs into a higher dimensional feature space in which a linear classifier is constructed by minimizing an appropriate cost function. Using Mercer's theorem, the solution is obtained by solving a dual problem avoiding explicit knowledge of the high dimension and using only the related kernel function. This project applies SVM to financial time series analysis of US Dollars Index. In addition, this paper examines the features of both fundamental analysis and technical analysis.
There are some researches which show that SVM is a powerful tool for financial analysis: Kamruzzaman et al. (2003) investigated the effect of different kernel functions, namely, linear, polynomial, radial basis and spline on prediction error of SVM models for foreign currency exchange rates. Nakamori et al. (2005) use SVM method in predicting financial movement direction of NIKKEI 225 index[2]. Zhou, Ongsritrakul and Soonthornphisaj (2003) demonstrated the use of Support Vector Regression techniques for predicting the cost of gold by using factors that have an effect on gold to estimate its price.

Support Vector Machines

Recently, a support vector machine (SVM), a novel neural network algorithm, was developed by Vapnik and his colleagues[1]. SVM uses linear model to implement nonlinear class boundaries through some non- linear mapping the input vectors into the high-dimensional feature space. Therefore, SVM is known as the algorithm that finds a special kind of linear model, the maximum margin hyperplane, which gives the maximum separation between the decision classes. The training examples that are closest to the maximum margin hyperplane are called support vectors. All other training examples are irrelevant for defining the binary class boundaries[3]. The mathematical problem is constructed as follows:

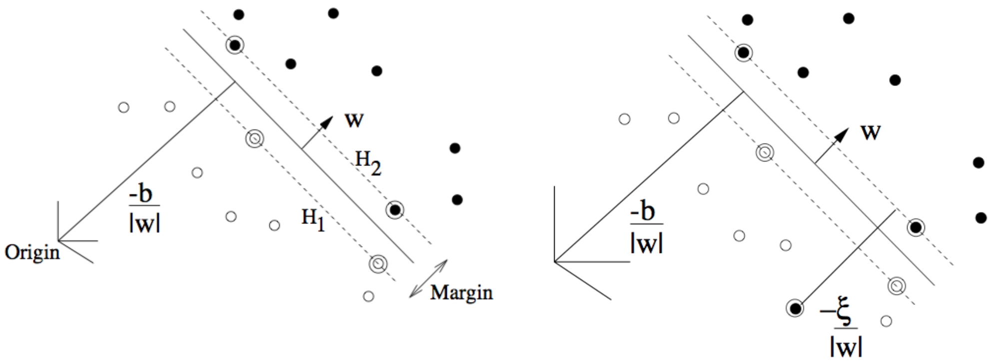
*ID: xiongfli, M.S. FinMath

†ID: xinyufan, M.S. FinMath

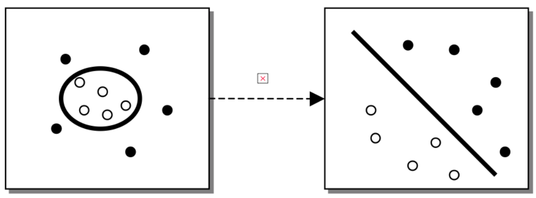
- To find the optimal margin classifier

$$
\begin{aligned}
min_{\gamma,w,b} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i \\
s.t. \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \qquad i = 1...m \\
& \xi_i \geq 0, \qquad i = 1...m
\end{aligned}
$$

- $\xi = 0$ in cases which are linearly separable (left figure below). While $\xi > 0$ in cases not linearly separable (right figure below), where examples are allowed to have functional margin less than 1.



- If the transformation is nonlinear and the dimensionality of the feature space is high enough, then input space may be transformed into a new feature space where the patterns are linearly separable with high probability[4]. This nonlinear transformation is performed in implicit way through so-called kernel functions.



(a) input space (b) feature space

Data

Five features deriving from fundamental analysis are selected out in the SVM model. All of them have significant effect on the price of US Dollar:

- US Dollar Index - USD Strength
  The USD Index measures the performance of the US Dollar against a basket of currencies: EUR, JPY, GBP, CAD, CHF and SEK. We label a monthly index increase as 1, while a decrease as 0 to train the following features.

- LIBOR - Interest Rate
  The biggest influence that drives the foreign-exchange market is interest rate. Generally speaking, a higher interest rate results in stronger currency and vice versa. A surprise of

interest rate driven by the central bank or the market can affect the foreign-exchange market in a great deal. There is also an equation known as Interest rate parity:
Forward Rate = Spot Rate · (1 + Interest Rate of Overseas country)/(1 + Interest Rate of Domestic country)

- CPI - Inflation rate
  Inflation rate have opposite effect on foreign-exchange rate, i.e., as a general rule, a country with a consistently lower inflation rate exhibits a rising currency value, as its purchasing power increases relative to other currencies. Consumer Price Index (CPI) - a measure of price changes in consumer goods and services such as gasoline, food, clothing and automobiles. This index is viewed as a common measure for inflation rate.

- GDP - Economics Growth
  No doubt, foreign investors seek for stable countries with strong economic performance in which to invest their capital. A country with such positive attributes will draw investment funds and hence result in stronger currency. And GDP is widely used as an index to measure a country's economic performance.

- PMI - Leading Economic Indicator
  Financial market players have their own view of the future, which drive their investment decisions in the market. So PMI, which is viewed as a leading indicator of the economics, has a similar role like GDP in foreign-exchange market. What is more, PMI is considered a leading indicator in the eyes of the Fed, who has dominant power in the market with their monetary policies.

- Unemployment Rate - Labor Market
  Labor market is considered by Fed in their policy making, too. Is it generally accepted that high employment is one of the goals of central bank when making monetary policy.

Figure 1 shows the scaled time series of our training data, the orange line is the scaled US dollar index measuring the performance of the US Dollar against a basket of currencies like EUR, JPY, GBP, etc. The time series are from Jan 1995 to Dec 2011 monthly.

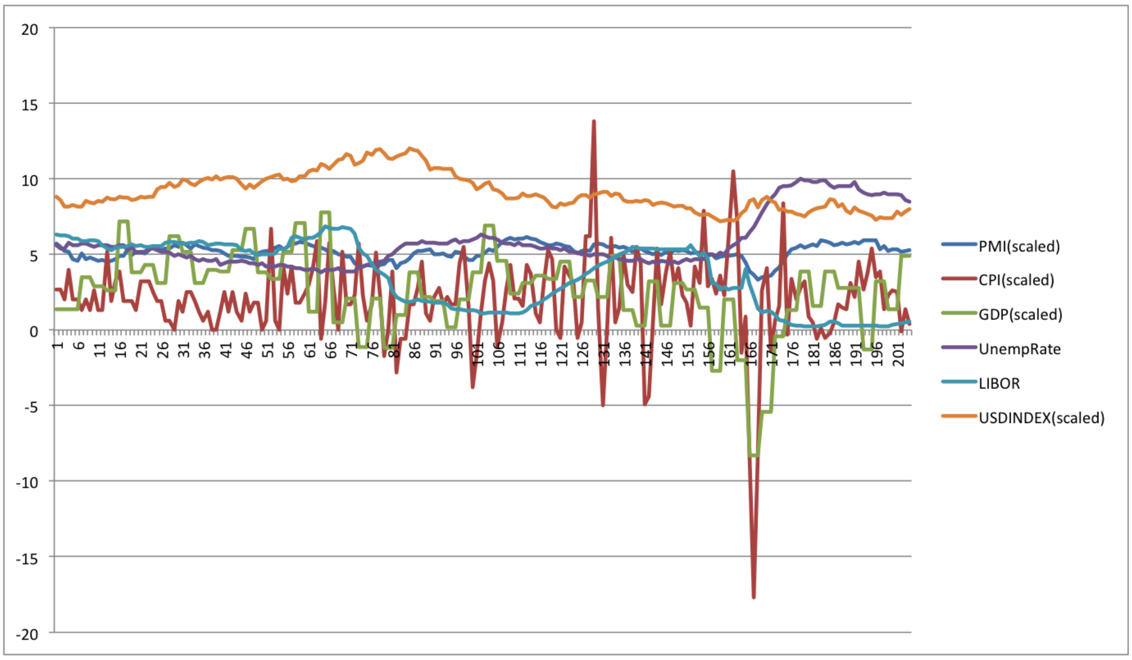


Figure 1: Time series of US dollar idx and macroeconomic features (204 data points)

Cross Validation

We use hold-out cross validation to randomly split all sample into training sample (87.5% of the data) and testing sample (12.5% of the data). Then we fit different kernel SVM to training and test their training error and generalization error as follows:

Training Error(%)	Mean	Standard Deviation	Max	Min
Gaussian	24.2989247	1.8356292	30.1075269	18.8172043
Polynomial	24.9435484	1.9580701	31.1827957	18.2795699
Linear	42.6586022	1.8295024	48.9247312	37.0967742
Quadratic	34.4698925	1.9147401	39.7849462	28.4946237

Table 1: Cross Validation Training Error

Gaussian and Polynomial have the least training error. The good performance of Gaussian is natural, and we want to point out that Polynomial kernel tends to be overfitting, so we need further research on its generalization error.

Generaliz.Error (%)	Mean	Standard Deviation	Max	Min
Gaussian	47.3666667	9.1629058	74.0740741	18.5185185
Polynomial	48.0259259	9.1249376	77.7777778	22.2222222
Linear	48.5185185	8.5179463	74.0740741	18.5185185
Quadratic	47.2555556	9.1106368	74.0740741	18.5185185

Table 2: Cross Validation Generalization Error

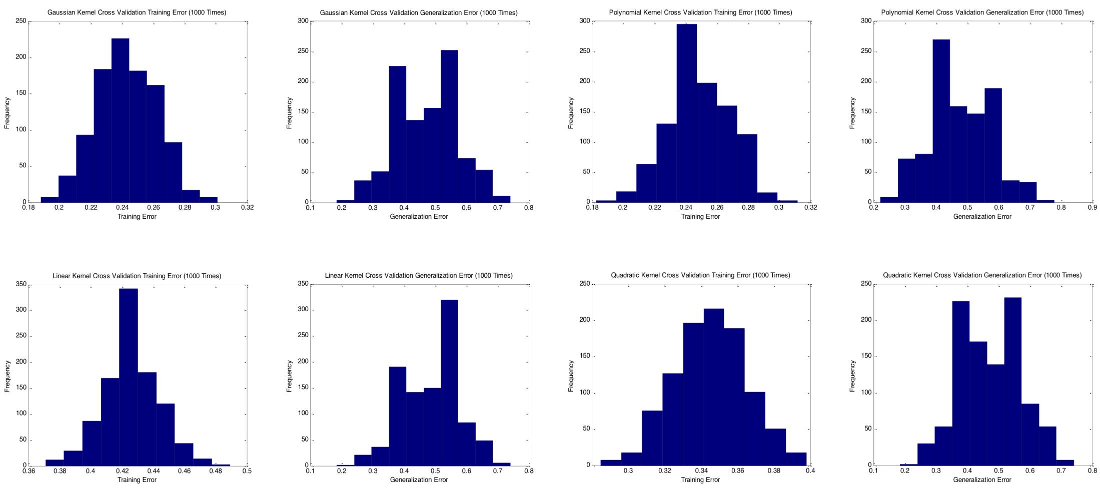


Figure 2: Training & Generalization Error in Gaussian, linear, quadratic and polynomial kernels

Generalization error of each kernel is not ideal, but we prove in next chapter that Gaussian out-of-sample test has good result. A explanation of this seemingly contradiction is that simple cross validation randomly split the sample into two groups, and this method is likely to destroy the property of time series. For instance, the training sample may have points in future, but testing sample may be constituted of many past points. Despite the defect mentioned above, cross validation is a good way to prove that Gaussian kernel has a relatively balanced and acceptable performance in both training and generalization error, which is exactly what we need.

Out-of-Sample Test

We use the latest 12 months to test the prediction ability of SVM model. Results of different kernels are as follows:

Out-of-Sample Test	Gaussian	Polynomial	Linear	Quadratic
Error	25%	33%	50%	50%

The result shows that Gaussian kernel has the best prediction ability (i.e. minimum generalization error). The error is 25%, which means our model is acceptable.

Feature Selection/Reduction Analysis

In this section, feature selection procedure is discussed. It is important and interesting to know which feature is more significant to help predict the USD strength and which one contribute more to the accuracy in the model. Through the component analysis, we found that different kernel functions gave different weight that each feature contribute to the training accuracy. Since we have justified the optimality of Gaussian kernel in previous section, we will discuss the component analysis in Gaussian kernel assumption. Table 3 below shows the results of the backward selection of the features with loss of training accuracy in each iteration with deleting each single feature.

					Loss of Accuracy (%)
Original Accuracy(%)	CPI	PMI	GDP	LIBOR	Unemployment Rate
74.1784	7.9812	6.1033	5.1643	2.3474	1.8779
72.3004	8.4502	11.2671	8.9197	6.5723	
65.7277	11.7371	5.1643	1.8779		
63.8498	7.9812	7.5117			
56.3380	N/A				

Table 3: Backward search with loss of accuracy for training data (213 data points)

As the table shows, the significance of features is in order of:
CPI > PMI > GDP > LIBOR > Unemployment Rate. When one feature is removed, unemployment rate gives the least loss of accuracy with 1.88% for training data. Therefore, 4 features with unemployment rate deleted provide 72.3% accuracy, which is already a very good proxy for the full set of features learning. Furthermore, CPI, PMI and GDP are the most significant features providing 65.73% of accuracy, although it loses 6.57% of accuracy when removing LIBOR's contribution. Interestingly, CPI, the most significant feature itself gives training accuracy of 56.3%.

Conclusion & Further Discussion

SVM approach for USD strength analysis has good performance when Gaussian kernel is employed, which is verified by cross validation procedure. While 5 selected macroeconomic features give 25% error in out-of-sample test. Moreover, CPI and PMI are found to be the most significant features through component analysis. Further discussion includes whether adding in other indigenous economic features and their technical analysis will yield better performance. Features of economic data in Eurozone and Japan should be analyzed as USD Index measure the strength against the currency in such economies, in which scenario, more accurate time filtration must be constructed.

References

[1] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[2] W., Huang, Y., Nakamori, S.Y., Wang, Forecasting Stock Market Movement Direction With Support Vector Machine, Computer & Operations Research, Vol. 32, pp. 2513 - 2522, 2005.

[3] K., Kim, Financial Time Series Forecasting Using Support Vector Machines, Neurocomputing., Vol. 55, pp. 307-319, 2003.

[4] S., Haykin, Neural Networks, Prentice Hall Inc., 1999.