

5-2012

in financial **Directional forecasting in financial time series using support vector machines: The USD/Euro exchange rate**

V. Plakandaras, T. Papadimitriou, P.Gogas

RESEARCH PAPERS IN ECONOMICS

Department of International Economic Relations and Development DEMOCRITUS UNIVERSITY OF THRACE Komotini 69100 Tel. +302531039000

Directional forecasting in financial time series using support vector machines: The USD/Euro exchange rate

V. Plakandaras, T. Papadimitriou, P.Gogas

Department of International Economics and Development

Democritus University of Thrace, University Campus Komotini 69100 Greece,

Abstract

In this paper, we present a novel machine learning based forecasting system of the EU/USD exchange rate directional changes. Specifically, we feed an overcomplete variable set to a Support Vector Machines (SVM) model and refine it through a Sensitivity Analysis process. The dataset spans from 1/1/1999 to 30/11/2011; the data of the last 7 months are reserved for out-of-sample testing. Results show that the proposed scheme outperforms various other machine learning methods treating similar scenarios.

1. Introduction

With more than \$4 trillion dollars traded daily in the foreign exchange market¹, exchange rate forecasting is of great importance to both traders and policy makers. Soon after the breakdown of the Bretton Woods system, the need to model and forecast the new highly volatile and complex phenomenon of floating currency rates (Trafalis et al, 2006), led to the introduction of several empirical techniques.

The first attempts to model the exchange rate market employed structural econometric models based on economic theory fundamentals. The goal was to predict future rates through the identification of the factors that determine exchange rates. These attempts suffered from mediocre forecasting abilities (B. Balassa, 1964; R. Dornbuch, 1976; J. Frankel, 1979; Meese & Rogoff, 1983) and in response, many participants of the foreign exchange market claim that the high volatility of the exchange rate market is a result of psychological trends discounting the predictive power of fundamentals (Fama, 1984).

Advances in statistical methods lead to the introduction of Autoregressive Integrated Moving Average (ARIMA) models proposed by Box and Jenkins (1973) that had better forecasting ability than structural models. The main drawback of this method is that it assumes a linear relationship between the variables involved in the estimation. A further development was the introduction of Structural Econometric Models Time Series Analysis (SEMTSA) by Arnold Zellner (1979) who combined both approaches. Nevertheless, the method suffered from high complexity and the results

 1 Bank of International Settlements, 2011

did not have straightforward economic interpretation. To overcome these drawbacks, C. Sims (1980) introduced the Vector Autoregressive (VAR) models adopting a more convenient interpretation of variable correlations, with the simplicity of a theory-free estimation framework.

Later, with the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) models by Engle (1982) and the generalization of this methodology by Bollerslev (1986) with the generalized ARCH (GARCH), a whole new group of empirical studies was introduced to model the conditional variance (K. Mitchell,2007; Chen S. et al, 2010; Hossein & Nasser, 2011).

The complexity and volatility of exchange rates intrigued researchers employing chaos theory as well. A number of studies focused on testing for deterministic behavior in various currencies, providing significant evidence in favor of a chaotic generating process in the exchange rates mechanism (B.Mandelbrot, 1963; De Grauwe & Grimaldi,2006; Gogas & Serletis, 1997; Bask, 2002; Das& Das,2007). Nonetheless, detection of a chaotic generating mechanism is only the first step towards forecasting. It needs to be followed by exact identification of the mechanism and necessarily produce only very short-term forecasts.

During the last decade, non parametric models such as Artificial Neural Networks (ANN) and Support Vector Machines have been used extensively. Due to their ability of dealing with non-linear systems through data mapping, machine learning has become a popular choice among researchers in various forecasting scenarios. In particular, models based on Neural Networks (Wang, 2004; Chang et al, 2009; Huang et al, 2004) and Support Vector Machines (Huang, 2010; Kamruzzaman et al, 2003; Trafalis, 2006; Brandl et al, 2009; Zhao et al, 2009) achieved accuracy at least at par with predictions from both structural econometric and naive models. In contrast to ANN estimation, the SVM solution derives from convex optimization making the optimal solution both global and unique.

There are only a few studies that deal with exchange rate directional forecasting using pure SVM classification techniques. Ullrich et al. (2007) implemented 6 kernels on 2439 daily observations yielding a structural model with out of sample forecasting accuracy of 56% on the exchange rate EUR/USD for next day's directional forecast, on a test data set of 350 observations. Zhang and Zhao (2009) studied various technical indicators over the EUR/USD exchange rate and achieve a 60% hit-rate in weekly directional forecast, with 56 observations used for the test sample. Brandl and Leopold-Wildburger (2009) introduced a structural SVM classification model with constraints imposed on the prediction results through Genetic Algorithms. The constraints provide a stable framework from classic economic theory, in a way that domain knowledge is introduced to the forecast model. The hit-rate for monthly forecast is 73%.

In this paper we aim at constructing a structural SVM model that can optimally classify input data in an attempt to forecast the directional change of the day-to-day EUR/USD exchange rate.

The selection of a directional forecast and not of the actual rate is due to the fact that the information about the future direction of the exchange rate can lead to a clear decision in a trading strategy. Under this machine learning framework, we construct a binary forecasting model with the use of an SVM classifier into two states: future rise or fall.

2. Support Vector Machines

The Support Vector Machines model is a supervised machine learning method used for two-class data classification. Roughly, the basic concept of an SVM is to select a small number of data points from our dataset, called Support Vectors (SV) that can define a hyperplane separating the two classes' data points. When the problem is not linearly-separable, then SVM is coupled with a non-linear Kernel mapping procedure, projecting the data points to a higher dimensional space, called feature space, where the classes are linearly separable.

The procedure has two steps: the training step and the testing step. In the training step, the largest part of the dataset is used for the estimation of the separating hyperplane; in the testing step, the generalization ability of the model is evaluated by investigating the model's performance in the small subset that was left aside in the first step. Typically, 80%-95% of the dataset is used for the training step and the rest 20%-5% for testing.

In the following we describe briefly the mathematical derivations of the SVM theory.

2.1 Linear separable case

We consider a dataset (vectors) $x_i \in R^2$ $(i = 1, 2, ..., n)$ belonging to two classes (output targets²) $y_i \in \{-1, +1\}$. If the two classes are linearly separable, then we define a separator

$$
f(\mathbf{x}_i) = \mathbf{w}^{\mathrm{T}} \mathbf{x}_i - b = 0 \tag{1}
$$

in such that $y_i f(x_i)$

where **w** is the weight vector and *b* is the bias.

The optimal hyper plane is selected as the decision boundary that classifies each data vector to the correct class and has the maximum distance from both class. This distance is often called "margin". In Figure 1, the SV's are represented with the pronounced contour, the margin lines (defining the distance of the hyperplane with

¹ ²In the SVM jargon

each class) are represented with the continuous lines and the hyper plane is represented with the dotted line.

Figure 1.Hyper plane selection and support vectors. The SV's are represented with the pronounced red contour, the margin lines are represented with the continuous lines and the hyper plane is represented with the dotted line.

The solution to the problem of finding the hyper plane can be dealt through the Lagrange relaxation procedure on the following equation:

$$
\min_{\mathbf{w},b} \max_{\mathbf{a}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N a_i \left[y_i (\mathbf{w}^T \mathbf{x}_i - b) - 1 \right] \right\}
$$
(2)

Where $\mathbf{a} = [a_1, ..., a_n]$ are the non negative Lagrange multipliers. Equation (2) is never used to estimated the solution. Instead we always solve the dual problem, defined as:

$$
\max_{\mathbf{a}} \left\{ \sum_{i=1}^{N} a_i - \sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k y_j y_k \mathbf{x}_j^{\mathrm{T}} \mathbf{x}_k \right\} \tag{3}
$$

Subject to $\sum_{i=1}^{N} a_i y_i = 0$ and $0 \le a_i$,

The solution of (3) gives the location of the hyper plane defined by:

$$
\widehat{\mathbf{w}} = \sum_{i=1}^{N} a_i y_i \mathbf{x}_i \tag{4}
$$

$$
\hat{b} = \hat{\mathbf{w}}^{\mathrm{T}} \mathbf{x}_i - y_i, i \in V \tag{5}
$$

Where $V = \{i : 0 < y_i\}$ is the set of the support vector indices.

2.2 Error Tolerant SVM

In order to allow a predefined level of error tolerance in the training procedure Cortes and Vapnik (1995) introduced non-negative slack variables $\xi_i \geq 0$, $\forall i$ and a parameter *C* describing the desired tolerance to classification errors. Equation (2) is now defined as:

$$
\min_{\mathbf{w},b,\xi} \max_{\mathbf{a}\mu} \left\{ \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{i=1}^N \xi_i - \sum_{j=1}^N a_j \left[y_j (\mathbf{w}^T \mathbf{x}_j - b) - 1 + \xi_j \right] - \sum_{k=1}^N \mu_k \xi_k \right\} \tag{4}
$$

where ζ ^{*i*} measures the distance of vector \mathbf{x}_i from the hyper plane when classified erroneously.

The hyper plane is defined as:

$$
\widehat{\mathbf{w}} = \sum_{i=1}^{N} a_i y_i \mathbf{x}_i \tag{5}
$$

$$
\hat{\mathbf{b}} = \hat{\mathbf{w}}^{\mathrm{T}} \mathbf{x}_i - y_i, i \in V \tag{6}
$$

Where $V = \{i : 0 < y_i < C\}$ is the set of the support vector indices.

2.3Kernel Methods

When the two class dataset cannot be separated by a linear separator (Figure 2), then the SVM classification is paired with kernel methods.

Figure 2.The Data Space. The non-separable two class scenario

The concept is quite simple: the dataset is projected though a kernel function into a richer space of higher dimensionality (called feature space) where the dataset is linearly separable (Figure 3)

Figure 3. The Feature Space. The two classes are linearly separable.

The solution to the dual problem with projection of eq. (4) now transforms to:

$$
\max_{\mathbf{a}} = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k y_j y_k \, K(\mathbf{x}_j, \mathbf{x}_k)
$$
(7)

Under the constraints $\sum_{i=1}^{N} a_i y_i = 0$ and $0 \le a_i \le C$, $\forall i$ where $K(\mathbf{x}_i, \mathbf{x}_k)$ is the kernel function.

As the SVM theory uses the structural risk minimization rule for the selection of the hyper parameters, it always seeks for a globally optimized solution avoiding model over-fitting.

3. Data Analysis

3.1 Data Collection and Kernel Selection

As already stated, our goal is to construct a structural prediction model. One of the most important steps in constructing our forecasting model is the correct model specification. In order to include the maximum input information, we gathered the majority of the input variables used in exchange rate theory and through sensitivity analysis we kept gradually just the variables that lead to an increase in prediction performance.

Based on economic theory and previous empirical studies, we selected 74 input variables (see the Appendix). These include macroeconomic variables for the USA, the EU, Japan and the UK, daily prices of commodities traded at the Chicago Mercantile Exchange, daily closing spot prices of precious and non-precious metals as traded at the London Metal Exchange, daily spot exchange rates for the EUR/USD and their cross rates with GBP and JPY, various stock market indices. We also introduce the Moving Averages (MA) of 3,5,10 and 30 days for the EUR/USD exchange rate, the interest rate on Treasury bonds with 6 months and 10 years maturity, the 1 week and 1 month Euribor rates and the Eonia overnight rate. In addition to that, we constructed a variable that counts the number of positive returns over the last 5 trading days for the exchange rate under investigation. Overall, the input data set contains all of the variables suggested by theory and empirical studies in the field of exchange rate forecasting. All the data series are collected from 1/1/1999 to 30/10/2011 yielding a set of 3280 observations.

The kernel used is the Gaussian Radial Basis Function (RBF) as it is considered superior in exchange rate forecasting (Kamruzzaman et al, 2003). The mathematical expression of the kernel is:

$$
K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2}, \gamma > 0
$$
\n(7)

Where γ is the width-length basis parameter of the Gaussian curves.

Before performing the tests, all numerical data where scaled to [-1, 1], in order to impose the same range on all variables. Finally, all the data set was randomly permutated in order to eliminate linearity from the data.

3.2 Sensitivity Analysis

The data set is separated as follows: 19/20 for training and 1/20 for test evaluation of the predicted results (163 observations). The hyper plane parameters C and γ were selected through a ten-fold cross-validation rule and grid search on the training set (Chang & Lin, 2003).

On a first step we explored the autoregressive models, using as many as 31 time lags in our tests. Results showed that the 11-lags autoregressive model outperformed the rest on the test sample, as presented on Figure 4.

Figure 4: Results of the autoregressive model. With the red bar we mark the results of the most accurate 11 lags autoregressive model

In Figure 5, we demonstrate the accuracy levels of the training sample for various combinations of parameters C,γ.

Figure 5: Best autoregressive model. The various contour colors represent different prediction accuracy levels for the various combinations of C, γ while the most accurate setting for the hyper plane parameters is marked with a green circle. The overall accuracy on the train sample is depicted on the headline of the graph.

On a second step, we introduced the rest of the input variables through a forward selection procedure. One by one every variable was added in the best autoregressive model and the optimum setting of the hyperplane parameters is selected. Then the new model is trained with the best combination of C, $γ$ and the accuracy of the model in out of sample forecasting is measured using the test sample. The overview of the proposed scheme is depicted in Figure 6.

Figure 6: Sensitivity analysis with forward introduction of parameters

Overall, we constructed 73 models each consisting of 12 input variables; the 11-lags autoregressive model as the base model and one at a time of the rest. When the first loop of 73 variables was completed, the variable with the best forecasting result was the index S&P500. The overall accuracy for training sample remained at 51.57% but the forecast accuracy on test sample set rose to 60%. The following diagram depicts the results for each variable.

Figure7: First loop results

Moving one step forward, we constructed 72 models by introducing one variable at a time to a 12 variable model consisting of the S & P 500 index and the 11-lags of the exchange rate. In other words the variable with the best out of sample forecasting result becomes a constant part of the model and the rest variables are reused. The results of the second loop are depicted in Figure 8.

Figure 8: Second loop results

From the comparison of the 72 models with 13 input variables, the one with the best out of sample forecasting accuracy performance was the one incorporating 11-lags of exchange rate observations, the price of index S & P 500 and the price of tin on moment t. The accuracy for the training sample was 52.12% and the hit-rate of correctly predicted directional movements for the test set rose to 61.34% (Figure 9).

Figure9: Best predictive model with 13 input variables consisting of 11-lags of the exchange rate observations, index S & P 500 and the daily price of tin.

Resuming the procedure, we constructed 71 models of 14 input variables consisting of the 11-lags of the exchange rate observations, index S & P 500 and the tin price on moment t and one by one the rest of the variables. In other words we try to measure the accuracy of the model with the best prediction accuracy so far, if we add to it a new variable. As in previous steps, all the variables are added one by one and the out of sample forecasting accuracy is measured. The outcome was models of lower predicting power in comparison to the best 13-variable model (less than 61.34% forecasting accuracy), as depicted in Figure 10.

Figure10: Third loop results

4. Conclusion

The construction of a structural SVM model with a forward input variable selection process, demonstrates considerable predictive power. Our approach produced predictions about next day's directional movement with a higher out-of-sample forecasting accuracy in comparison to previous efforts in the field of EUR/USD exchange rate forecasting with the need to collect data for only 3 variables. Moreover, it deals with daily forecasts which demonstrate wider fluctuations and thus have greater economic interest than weekly or monthly predictions.

Overall, structural SVM models seem promising and there is plenty of room for future research. Nevertheless, the main disadvantage of forward approaches is that they don't examine the models that can be produced with various combinations of the input variables, but are constrained in adding variables that improve the predicting performance of the initial model in an iterative procedure. To overcome this drawback, further research can focus on the implementation of different kernels and training techniques.

References

Balassa, B. (1964), *The purchasing power parity doctrine: a reappraisal*, Journal of Political Economy, Vol. 72, pp. 584-96.

Bask Mikael (1996) *Dimensions and Lyapunov exponents from exchange rate series*. Chaos, Solitons & Fractals 7, no. 12 (December): 2199-2214.

Belkin, M., Niyogi, P., & Sindhwani, V. (2006).*Manifold regularization: A geometric framework for learning from examples*. Journal of Machine Learning Research, vol. 7, pp.2399–2434.

BollerslevT.P. (1986), "*Generalized Autoregressive Conditional Heteroskedasticity*", Journal of Econometrics, vol. 31, pp. 307-327.

Box G.E.P. and Jenkins G.M. (1976) "*Time Series Analysis: Forecasting and Control"*, Holden-Day, San Francisco.

Brandl, B., & Leopold-Wildburger, U. (2009, *Increasing the Fitness of Fundamental,* Organizational Behavior Teaching Journal, 4(16), 779 - 798.

Chang Chih-Chung and Lin Chih-Jen (2011), *LIBSVM :a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.

Chang, J-F, C-W Chang, and Wen-Yan Tzeng (2009) *Forecasting Exchange Rates Using Integration of Particle Swarm Optimization and Neural Networks*. Fourth International Conference on Innovative Computing, Information and Control (ICICIC) (December): 660-663.

Chen, S., Härdle, W. K., &Jeong, K. (2010). *Forecasting Volatility with Support Vector Machine-Based GARCH Model*. Journal of Forecasting, vol.433,pp. 406-433.

Cortes, C. & Vapnik, V. (1995): *Support Vector Network. Machine Learning*, vol. 20, 273–297.

Das, a, and P Das (2007) *Chaotic analysis of the foreign exchange rates*. Applied Mathematics and Computation 185, no. 1 (February): 388-396.

Degrauwe, P, and M Grimaldi (2006) *Exchange rate puzzles: A tale of switching attractors*. European Economic Review 50, no. 1 (January): 1-33.

Dornbusch R.(1976) *Expectations and exchange rate dynamic*, Journal of Political Economy, 84(6), pp.1161–1176

Engle R.F. (1982) "*Autoregressive Conditional Heteroskedasticity with estimates of the variance of U.K. inflation*", Econometrica, vol. 50, pp. 987-1008.

Fama, E., (1984), *Forward and spot exchange rates*. Journal of Monetary Economics vol. 14, pp. 319–338.

Frankel J.A.(1979) *On the mark: A theory of floating exchange rates based on real interest differentials,* American Economic Review, 69(4),pp.610–622.

Hossain, A., & Nasser, M. (2011).*Comparison of the finite mixture of ARMA-GARCH, back propagation neural networks and support-vector machines in forecasting financial returns*. Journal of Applied Statistics, 38(3), pp. 533-551.

Huang Shian-Chang, Chuang Pei-Ju, Wu Cheng-Feng, and Lai Hiuen-Jiun,(2010) "*Chaos-based support vector regressions for exchange rate forecasting,*" Expert Systems with Applications, vol. 37 , pp. 8590-8598.

Huang W., Lai K.K., Nakamori Y. and Wang Shouyang (2004) "*Forecasting Foreign Exchange Rates with Artificial Neural Networks: A Review*", International Journal of Information Technology & Decision Making Vol. 3, No. 1 pp. 145–165.

Kamruzzaman, SarkerR.a., and Ahmad I., (2003) "*SVM based models for predicting foreign currency exchange rates,*" Third IEEE International Conference on Data Mining, IEEE Comput. Soc, 557-560.

Kilian L. & M.P. Taylor (2003), *Why is it so difficult to beat random walk forecast of exchange rates?*, Journal of International Economics 60, pp.85–107.

Macdonald, Ronald, and Jun Nagayasu (1998) *On the Japanese Yen – U .S. Dollar Exchange Rate: A Structural Econometric Model Based on Real Interest Differentials.* Journal of Economic Literature, Vol. 102, pp. 75-102.

Mandelbrot Benoit (1963), "*The Variation of Certain Speculative Prices"*, The Journal of Business, vol. 36, no. 4, pp. 394-419.

Meese, Richard & Kenneth Rogoff, 1983, *Empirical exchange rate models of the seventies: Do they tit out of sample?*, Journal of International Economics, pp. 3-24.

Mitchell, K., & Pearce, D. (2007).*Professional forecasts of interest rates and exchange rates: Evidence from the Wall Street Journal's panel of economists*. Journal of Macroeconomics, 29(4), 840-854. doi

Serletis Apostolos & Gogas Periklis (1997) *Chaos in East European black market exchange rates.* Research in Economics 51, no. 4 (December): 359-385.

Sims C.A. (2002) "*The role of models and probabilities in the monetary policy process ",* Brookings Panel on economic activity , Fall 2002.

T.B. Trafalis (1999), *Artificial neural networks applied to financial forecasting*, in: C.J.C. Dagli, Buczak, Ghosh, Embrechts, Ersoy (Eds.), Smart Engineering System Design: Neural Net- works, Fuzzy Logic, Evolutionary Programming, Data Mining and Complex Systems, ASME Press, pp. 1049–1054.

Taylor, L. (2004). *Exchange rate indeterminacy in portfolio balance, Mundell-Fleming and uncovered interest rate parity models*. Cambridge Journal of Economics, 28(2), 205-227.

Trafalis, T. &Ince, H., (2006).*A hybrid model for exchange rate prediction*. Decision Support Systems, 42(2), 1054-1062.

Ullrich C., Seese D. &Chalup S. (2007) *Foreign Exchange Trading with Support Vector Machines,* Advances in Data Analysis, part 8, pp. 539-546.

Wang T., Huang H., Tian S. &Xu J. (2010), *Feature selection for SVM via optimization of kernel polarization with Gaussian ARD kernels*, Expert Systems with Applications, vol. 37, pp. 6663–6668

Zellner A. (1979), *Statistical Analysis of Econometric Models*, Journal of the American Statistical Assosiation, vol. 74, No. 367, pp.628.

Zhang, Z., & Zhao, Q. (2009).*The Application of SVMs Method on Exchange Rates Fluctuation.* Discrete Dynamics in Nature and Society, 2009, 1-8.

APPENDIX

