

Predicting the Istanbul Stock Exchange Index Return using Technical Indicators: A Comparative Study

Şenol Emir^a

^a PhD, Beykent University, Computer Programming, Istanbul, Turkey

Abstract

The aim of this study to examine the performance of Support Vector Regression (SVR) which is a novel regression method based on Support Vector Machines (SVM) approach in predicting the Istanbul Stock Exchange (ISE) National 100 Index daily returns. For bechmarking, results given by SVR were compared to those given by classical Linear Regression (LR). Dataset contains 6 technical indicators which were selected as model inputs for 2005-2011 period. Grid search and cross validation is used for finding optimal model parameters and evaluating the models. Comparisons were made based on Root Mean Square (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Theil Inequality Coefficient (TIC) and Mean Mixed Error (MME) metrics. Results indicate that SVR outperforms the LR for all metrics.

Keywords: Support Vector Regression; Linear Regression; index return prediction, technical indicators; symmetric and asymmetric metrics

© 2013 Published by SSBFNET

1.Introduction

Financial markets are complex, nonlinear and dynamic systems. Financial prediction requires to process noisy, non-stationary, unstructured and uncertain data. Due to many factors including political events, general economic conditions and trader expectations, this is a quite difficult task (Huang and et al., 2005). In recent years a novel method called Support Vector Machines (SVM) has been widely used for predicting as well as statistical methods.

The SVM method was first developed by Vladimir N. Vapnik (Vapnik, 1995) based on the ideas from Statistical Learning Theory which uses Structural Risk Minimization (SRM) inductive principle instead of traditional Empirical Risk Minimization (ERM). Unlike ERM which focuses on minimizing the training error, SRM tries to minimize the generalization error upper bound. Overfitting risk is lower with SVM when compared with models such as Artificial Neural Networks (Wu and et al., 2010).

The method has been implemented in applications such as image recognition (Wei and et al., 2011), hand-writing recognition (Arora and et al., 2010), text categorization (Zaghloul and et al., 2009), bioinformatics (Ben-Hur and et al., 2008) successfully. Recently SVM has found a wide range of financial applications including stock selection (Huang, 2012), financial time series forecasting (Tay and Cao, 2001), evaluation of consumer loans (Li and et al., 2006), credit data fraud detection (Hejazi and Singh, 2012), stock trend prediction (Ni and et al., 2011) and so forth.

SVM was first designed to solve binary classification problem. Later SVM version for regression which is called Support Vector Regression (SVR) was proposed. Similar to Support Vector Classification (SVC), SVR depends on a small subset of training data.

Classical Linear Regression uses Least Squares approach. That is, regression function is a hyperplane that fits a given training set with with the minimum mean square error between this hyperplane and the data points. However SVR uses a different approach that aims to find a hyperplane which fits the data with a deviation less than a given quantity

^a Corresponding author. Tel.: +90 212 867 5458.

called epsilon (ϵ) for every training data point. SVR does not minimize errors less than ϵ , but only higher errors. By this way a regression function whose parameters are a linear combination of those training data points of which error is higher or equal to ϵ , can be constructed. This leads a unique, global and sparse solution (Ramon and Christodoulou, 2006, 14).

The remained of the paper is organized as follows. Second section focuses on literature review of stock market prediction and technical analysis. Third section describes Support Vector Regression method. Findings of the analysis are presented in fourth section. Fifth section discusses research findings and further research.

2. Literature Review

For predicting financial markets, technical inputs, fundamental inputs or both from one or more markets can be used. Fundamental inputs are economic indicators which are believed to influence the dependent variable. On the other hand technical indicators are calculated from the lagged values (Kaastra and Boyd, 1996).

Technical analysis is study of how securities prices behave and how to exploit that information to make money while avoiding losses. Main purpose is to predict the price of securities over some time interval in order to buy and sell to security to make a profit (Rockefeller, 2011: 9).

In this section literature review on stock market prediction and technical analysis is presented.

Dunis et al. (2012) examined the application of support vector machines to the task of forecasting the weekly change in the Madrid IBEX-35 stock index. They used Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) technical indicators as predictors.

Kara et al. (2011) compared performances of Artificial Neural Networks and Support Vector Machines in predicting the direction of movement in the daily Istanbul Stock Exchange (ISE) National 100 Index. Ten technical indicators (Simple Moving Average, Weighted Moving Average, Momentum, Stochastic K%, Stochastic D%, RSI, MACD, William's R%, A/D Oscillator, CCI) were used as inputs of the proposed models.

Kim (2003) applied SVM for predicting the stock price index. The research data used in study was technical indicators and the direction of change in the daily Korea Composite Stock Price Index (KOSPI). 12 technical indicators (%K, %D, Slow %D, Momentum, ROC, William's %R, A/D Oscillator, Disparity 5, Disparity 10, OSCP, CCI, RSI) were used as input variables.

Karymshakov and Abdykaparov (2012) examined performance of Artificial Neural Networks in forecasting stock market index movement. The forecasting was based on two samples of Istanbul Stock Exchange (ISE) data and each consisting of 150 observations. Daily high and low values of ISE, daily ISE 100 close value, stochastic oscillator, 5-day moving average, 1,2,3-period lag values of daily ISE 100, gold price and USD exchange rate were used as predictors.

Diler (2003) predicted the direction of Istanbul Stock Exchange (ISE) 100 Index by using Artificial Neural Networks for the 1990-2003 period. In analysis technical indicators (Moving Average, Weighted Moving Average, Momentum, Stochastic %K, RSI, and MACD) were used as input.

3. Methodology

Suppose a training data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \subseteq \mathcal{R}^m \times \mathcal{R}$ is given. A general linear regression function $f(x) = w \cdot x + b$ attempts to model the input-output relationship. Unlike classification problems where desired outputs y_i are discrete values, in regression tasks outputs are real valued (Kecman, 2001: 176).

For Support Vector Regression (SVR) an appropriate loss function is ϵ -insensitive loss (Equation 1)

$$|y - w \cdot x|_{\epsilon} = \begin{cases} 0 & \text{if } |y - w \cdot x| \leq \epsilon \\ |y - w \cdot x| - \epsilon & \text{otherwise} \end{cases} \quad (1)$$

Minimization of ε -insensitive loss can be formulated by using nonnegative slack variables ξ_i, ξ_i^* (Equation 2-3). These variables measure the deviation of training data points lying outside the ε -insensitive zone (Hamel, 2009: 199).

$$\xi_i = \begin{cases} 0 & \text{if } y_i - (w \cdot x_i) \leq \varepsilon, \\ |y_i - (w \cdot x_i)| - \varepsilon & \text{otherwise} \end{cases} \quad (2)$$

$$\xi_i^* = \begin{cases} 0 & \text{if } (w \cdot x_i) - y_i \leq \varepsilon, \\ |y_i - (w \cdot x_i)| - \varepsilon & \text{otherwise} \end{cases} \quad (3)$$

The problem of finding parameters w can be stated as

$$\min \frac{1}{2}(w \cdot w) + \frac{C}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4)$$

$$\text{Constraints} \begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n \end{cases} \quad (5)$$

Optimization formulated in (4) and (5) is a quadratic optimization problem which has linear constraints. C parameter in objective function controls the tradeoff between model complexity and training error (Cherkassky and Mulier, 2007: 441).

By constructing a Lagrangian function and then applying Karush-Kuhn-Tucker (KKT) optimality conditions primal optimization problem that is stated in (4) and (5) can be transformed into the dual form (Suykens and et al., 2002: 55).

In this dual form α_i, β_i coefficients are found by solving the quadratic optimization problem formulated in (6) and (7).

$$\max \mathcal{L}(\alpha_i, \beta_i) = -\varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) + \sum_{i=1}^n y_i (\alpha_i - \beta_i) - \frac{1}{2} \sum_{i,j} (\alpha_i - \beta_i)(\alpha_j - \beta_j)(x_i \cdot x_j) \quad (6)$$

$$\text{Subject to} \begin{cases} \sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i, \\ 0 \leq \alpha_i \leq C/n \\ 0 \leq \beta_i \leq C/n, i = 1, \dots, n \end{cases} \quad (7)$$

By solving optimization problem in (6) and (7) for a given training set $(x_i, y_i), i = 1, \dots, n$, the value of ε and C parameter, optimal values of α_i^* and β_i^* are obtained. Using these values regression function can be written as (8).

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \beta_i^*)(x_i \cdot x) + b \quad (8)$$

In regression function represented in (8) only a small fraction of training data points have a nonzero coefficients. These data points that determine the regression functions are called support vectors. They are the samples that lay at or outside ε -insensitive zone.

The bias term (b) can be found by using one of the support vector pairs (x_s, y_s) as shown in (9).

$$b^* = y_s - \sum_{i=1}^n (\alpha_i^* - \beta_i^*)(x_i \cdot x_s) \quad (9)$$

Linear regression formulation can be extended to nonlinear case by using kernels. A kernel (K) is function that achieves the inner product in input space instead of high dimensional feature space (Steinwart and Christmann, 2008: 18). Any inner product kernel that satisfies Mercer conditions can be used in SVR to create nonlinear regression functions (Schölkopf and Smola, 2001: 110). The most popular kernels are linear, polynomial, sigmoid, radial basis function (RBF) kernels.

Nonlinear regression function can be stated as in (10).

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \beta_i^*) K(x_i, x) + b \tag{10}$$

where $\alpha_i^*, \beta_i^* \in [0,1], i = 1, \dots, n$

In SVR prediction quality depends on setting of proper model parameters as well as parameters like ϵ and C and kernel parameters.

4. Findings

In the analysis technical indicators (Table 1) were used as predictors and index returns as target variable. Dataset contains the daily observations of these variables for the 2005-2011 period. Technical indicators were obtained from MetaStock Software (<http://www.equis.com>). Closing prices of ISE-100 Index were collected from ISE official web site. Training set were composed of observations from 04.01.2005 to 02.08.2010 (1397 observations) and test set includes 03.08.2010-30.12.2011 period (349 observations).

Table 1. Technical indicators used in analysis and what they measure.

Technical Indicator	Measures (Achelis, 2001).
MO (Momentum)	The amount that a security's price has changed over a given time span.
CCI (Commodity Channel Index)	Variation of a price from its statistical mean.
MFI (Money Flow Index)	The strength of money flowing in and out of a security.
RSI (Relative Strength Index)	The internal strength of a single security.
STOCH (Stochastic)	Where a security's price closed relative to its price range over a given time period.
WILLR (William's %R)	Overbought / oversold levels.

Return of the index (r_t) was calculated by using closing prices of current and previous day as follow.

$$r_t = \left(\ln \frac{P_t}{P_{t-1}} \right) * 100 \tag{11}$$

P_t : Price at time t , P_{t-1} : Price at time $t - 1$.

For comparing performance of SVR and LR, both symmetrical (Table 2) and non symmetrical performance criteria was used.

Table 2. Symmetrical performance criteria.

Criteria	Formula
RMSE (Root Mean Square Error)	$\sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{r}_t - r_t)^2}$
MAE (Mean Absolute Error)	$\frac{1}{N} \sum_{t=1}^N \hat{r}_t - r_t $
MAPE (Mean Absolute Percentage Error)	$\frac{1}{N} \sum_{t=1}^N \left \frac{\hat{r}_t - r_t}{r_t} \right $
TIC (Theil Inequality Coefficient)	$\frac{RMSE}{\sqrt{\frac{1}{N} \sum_{t=1}^N \hat{r}_t^2 + \frac{1}{N} \sum_{t=1}^N r_t^2}}$

\hat{r}_t : predicted index return, r_t : actual index return, t : prediction period, N : number of observations in test set.

Non symmetric performance criteria that was utilized in the analyses is MME. MME (Mean Mixed Error) is based on giving different weights when predicted index return is under (U) or Over (O) actual index return. Shortly, U and O can be shown as

$$U = \{t: \hat{r}_t - r_t < 0\} \text{ and } O = \{t: \hat{r}_t - r_t > 0\}$$

If penalizing the predicted returns that are under actual returns is needed

$$MME(U) = \frac{1}{N} \left[\sum_{t \in O} |\hat{r}_t - r_t| + \sum_{t \in U} |\hat{r}_t - r_t|^{2^A} \right] \tag{12}$$

criteria is used. On the contrary to penalize predicted returns that are over actual returns

$$MME(O) = \frac{1}{N} \left[\sum_{t \in U} |\hat{r}_t - r_t| + \sum_{t \in O} |\hat{r}_t - r_t|^{2^A} \right] \tag{13}$$

criteria is preferred. A is a parameter that has value of -1 if $|\hat{r}_t - r_t| < 1$, and otherwise having the value of 1.

For obtaining succesfull prediction results from SVR models, determining optimal parameters is crucial. For this objective a grid search was done on predefined search ranges. For epsilon(ϵ) 0.0001, 0.001, 0.002, 0.005, 0.01, 0.05 values were tested. For each of these epsilon values, optimal C and Gamma kernel parameters were search within 0-20 and 0-2 intervals respectively. The highest prediction accuracy is given by best combination of these three (namely epsilon, C and Gamma) parameters.

Only RBF kernel is used as it is the most often preferred kernel in financial applications in the literature. To decrease overfitting risk 10-fold cross validation was applied.

After grid search the lowest prediction error that is obtained from the SVR model which has a parameter combination of C=10, epsiolon=0.05 and Gamma=0.5.

Results of the Linear Regression Model is presented on (Table 3). All the predictions have a acceptable significance levels. Regression Function given by LR is as follow.

$$r_t = -0.06404 * MO_t + -0.0025 * CCI_t + -0.03768 * MFI_t + 0.044256 * RSI_t + -0.01895 * STOCH_t + 0.059886 * WILLR_t + 9.860439 \tag{14}$$

Table 3. Results of Linear Regression.

	Coeff.	Std.Err	t-stat	p-value
MO	-0.06404	0.011548	-5.54513	0.0000
CCI	-0.0025	0.001293	-1.93054	0.0636
MFI	-0.03768	0.005012	-7.51871	0.0000
RSI	0.044256	0.008402	5.267332	0.0000
STOCH	-0.01895	0.003136	-6.04308	0.0000
WILLR	0.059886	0.004395	13.62469	0.0000
(Intercept)	9.860439	1.095733	8.998939	0.0000

(Table 4) shows the results of SVR and LR models based on 6 performance criteria. SVR has provided better results than LR for all metrics.

Table 4. Analysis results for SVR and LR.

Model	RMSE	MAE	MAPE	TIC	MME(O)	MME(U)
SVR	1.12738	0.86080	2.60304	0.381186	1.15421	1.21414
LR	1.34070	1.01747	2.684342	0.517769	1.47541	1.51829

All the analyses were done by using RapidMiner 5.0 (<http://rapid-i.com>) which is open source data mining software. For SVR implementation, MySVM library in RapidMiner was utilized.

5. Conclusion

In this study, Support Vector Regression method was applied to predict Istanbul Stock Exchange (ISE) National 100 Index returns by using technical indicators. For constructing the best model, optimal SVR parameters were found by using grid search. SVR model gave better results than Linear Regression based on all the metrics. It can be concluded that SVR can be a useful tool for predicting stock market returns.

For further research, hybrid models can be utilized to improve prediction accuracy. In addition to technical indicators, fundamental inputs can be used in the model. Tuning model parameters by more advanced techniques is another way of building more successful SVR models.

References

- Achelis, S. B. (2001). *Technical Analysis from A to Z*. New York: McGraw Hill.
- Arora, S., Bhattacharjee, D., Nasipuri, M., Malik, L., Kundu, M., & Basu, D. K. (2010). Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition. *IJCSI International Journal of Computer Science Issues*, 7(3), 1-10.
- Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support Vector Machines and Kernels for Computational Biology. *PLoS Computational Biology*, 4(10), 1-10.
- Cherkassky, V., & Mulier, F. M. (2007). *Learning From Data: Concepts, Theory, and Methods*. (2nd ed.). New Jersey: Wiley-IEEE Press.
- Diler, A. İ. (2003). İMKB Ulusal 100 Endeksinin Yönünün Yapay Sinir Ağları Hata iye Yayma Yöntemi ile Tahmin Edilmesi, *İMKB Dergisi*, 7(25-26), 65-81.
- Dunis, C. L., Rosillo, R., Fuente, D. de la, & Pino, R. (2012). Forecasting IBEX-35 Moves Using Support Vector Machines. *Neural Computing and Applications*, 23(1), 229-236.
- Hamel, L. H. (2009). *Knowledge Discovery with Support Vector Machines*. New Jersey: Wiley-Interscience.
- Hejazi, M., & Singh, Y. P. (2012). Credit Data Fraud Detection Using Kernel Methods with Support Vector Machine. *Journal of Advanced Computer Science and Technology Research*, 2(1), 35-49.
- Huang, C. F. (2012). A Hybrid Stock Selection Model Using Genetic Algorithms and Support Vector Regression. *Applied Soft Computing*, 12(2), 807–818.
- Huang, W., Nakamorii Y., & Wang, S. Y.(2005). Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*, 32(10), 2513–2522.
- Kaastra, I., & Boyd, M. (1996). Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing*, 10(3), 215–236.
- Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38(5), 5311–5319.
- Karymshakov, K., Abdykaparov, Y. (2012). Forecasting stock index movement with artificial neural networks: The case of İstanbul stock exchange. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(2), 231–242.

- Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. Cambridge, MA, USA: A Bradford Book.
- Kim, K. J. (2003). Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, 55(1–2), 307–319.
- Li, S. T., Shiue, W., & Huang, M. H. (2006). The Evaluation of Consumer Loans Using Support Vector Machines. *Expert Systems with Applications*, 30(4), 772–782.
- Ni, L. P., Ni, Z. W., & Gao, Y. Z. (2011). Stock Trend Prediction Based on Fractal Feature Selection and Support Vector Machine. *Expert Systems with Applications* 38(5), 5569–5576.
- Ramon, M. M. & Christodoulou, C. (2006). *Support Vector Machines for Antenna Array Processing and Electromagnetics*. Morgan & Claypool Publishers.
- Rockefeller, B. (2011). *Technical Analysis for Dummies*. (2nd. ed.). Indiana: Wiley.
- Schölkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge Massachusetts: The MIT Press.
- Steinwart, I., & Christmann A. (2008). *Support Vector Machines*, Springer.
- Suykens, J. A. K., Gestel, T. Van, De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least Squares Support Vector Machines*. Singapore: World Scientific Publishing Company.
- Tay, F. E. H, & Cao, L. (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, 29(4), 309–317.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.
- Wei, J., Jian-qi, Z., & Xiang, Z. (2011). Face Recognition Method Based on Support Vector Machine and Particle Swarm Optimization. *Expert Systems with Applications*, 38(4), 4390–4393.
- Wu, C. H., Ken, Y., & Huang, T. (2010). Patent Classification System Using a New Hybrid Genetic Algorithm Support Vector Machine. *Applied Soft Computing*, 10(4), 1164–1177.
- Zaghloul, W., Lee, S. M., & Trimi, S. (2009). Text Classification: Neural Networks Vs Support Vector Machines. *Industrial Management & Data Systems*, 109(5), 708–717.