European Journal of Operational Research 247 (2015) 831–846

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations

Georgios Sermpinis [a,*], Charalampos Stasinakis [a], Konstantinos Theofilatos [b], Andreas Karathanasopoulos [c]

[a] University of Glasgow Business School, University of Glasgow, Adam Smith Building, Glasgow G12 8QQ, UK
[b] Pattern Recognition Laboratory, Department of Computer Engineering & Informatics, University of Patras, 26500 Patras, Greece
[c] Andreas Karathanasopoulos Royal Docklands Business School, University of East London, UK

ARTICLE INFO

Article history:
Received 19 May 2014
Accepted 19 June 2015
Available online 27 June 2015

Keywords:
Genetic algorithms
Support vector regression
Feature subset
Forecast combinations
Exchange rates

ABSTRACT

The motivation of this paper is to introduce a hybrid Rolling Genetic Algorithm-Support Vector Regression (RG-SVR) model for optimal parameter selection and feature subset combination. The algorithm is applied to the task of forecasting and trading the EUR/USD, EUR/GBP and EUR/JPY exchange rates. The proposed methodology genetically searches over a feature space (pool of individual forecasts) and then combines the optimal feature subsets (SVR forecast combinations) for each exchange rate. This is achieved by applying a fitness function specialized for financial purposes and adopting a sliding window approach. The individual forecasts are derived from several linear and non-linear models. RG-SVR is benchmarked against genetically and non-genetically optimized SVRs and SVMs models that are dominating the relevant literature, along with the robust ARBF-PSO neural network. The statistical and trading performance of all models is investigated during the period of 1999–2012. As it turns out, RG-SVR presents the best performance in terms of statistical accuracy and trading efficiency for all the exchange rates under study. This superiority confirms the success of the implemented fitness function and training procedure, while it validates the benefits of the proposed algorithm.

1. Introduction

Forecasting financial time series appears to be a challenging task for the scientific community because of its non-linear and non-stationary structural nature. On one hand, traditional statistical methods fail to capture this complexity, while on the other hand non-linear techniques present promising empirical evidence. However, their practical limitations and the expertise required to optimize their parameters are creating skepticism on their utility.

This study introduces a hybrid Rolling Genetic Algorithm-Support Vector Regression (RG-SVR) algorithm for optimal parameter selection and features subset combination when applied to the task of forecasting and trading the EUR/USD, EUR/GBP and EUR/JPY exchange rates. The proposed model genetically searches over a feature space (pool of individual forecasts) and then combines the optimal feature subsets for each exchange rate. A novel fitness function specialized for financial purposes is used to simultaneously minimize the error of the obtained forecasts, increase the profitability of the final forecast combination and reduce the complexity of the algorithm. This is crucial in financial applications, where statistical accuracy is not always synonymous with the financial profitability of the deriving forecasts. The reduced complexity of the algorithm decreases the computational cost of the proposed methodology and makes it ideal for trading applications, where time efficiency is important. At the same time, it acts as a protection against overfitting and impeded generalization abilities. The model employs a sliding window training approach and is capable of capturing the time-varying relationship that dominates the financial trading series.

RG-SVR is benchmarked against seven models. Their statistical and trading performance is compared during the period of 1999–2012. The rationale behind the selection of the benchmarks is twofold. Firstly, Support Vector Machine (SVM) and Support Vector Regression (SVR) architectures with genetically and non-genetically optimized parameters are dominant in the relevant literature. The six most popular and promising variants are identified and included in the study (see Section 4.3). Secondly, it would be unfair to

* Corresponding author. Tel.: +447914053349.
E-mail addresses: georgios.sermpinis@glasgow.ac.uk, sermpinis@hotmail.com (G. Sermpinis), charalampos.stasinakis@glasgow.ac.uk (C. Stasinakis), theofilk@ceid.upatras.gr (K. Theofilatos), a.karathanasopoulos@uel.ac.uk (A. Karathanasopoulos).

http://dx.doi.org/10.1016/j.ejor.2015.06.052
0377-2217/

compare the hybrid RG-SVR only with hybrids of the same methodology class, especially when other proposed methodologies have shown statistical and trading superiority in a similar task. One such example is the hybrid Neural Network (NN) that combines Adaptive Radial Basis Functions with Particle Swarm Optimization (ARBF-PSO), as introduced recently by Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013). The authors apply ARBF-PSO to the task of forecasting and trading the same exchange rates that the present study is investigating. Their results show that ARBF-PSO is outperforming several linear and non-linear models, while its structural complexity is not high. Consequently, ARBF-PSO's selection as a benchmark to this application is required and justified.

To the best of our knowledge, the proposed RG-SVR methodology has not been presented in the literature. Similar hybrid applications exist but they are either limited in classification problems (Dunis, Likothanassis, Karathanasopoulos, Sermpinis & Theofilatos, 2013; Huang & Wang, 2006; Min, Lee, & Han, 2006; Wu, Tzeng, Goo, & Fang, 2007) or the Genetic Algorithm (GA) does not extend to optimal feature subset selection (Chen & Wang, 2007; Pai, Lin, Hong, & Chen, 2006; Yuan, 2012). In addition to this fact, the RG-SVR hybrid is the first GA hybrid SVR algorithm that deploys v-SVR models, minimizes the number of support vectors and applies a specialized loss function and a sliding window training approach. A detail comparison of the algorithm and the previous two algorithms is presented in the next section. The genetically optimized SVM model (GA-SVM) proposed by Huang and Wang (2006), Min et al. (2006), Wu et al. (2007) and Dunis et al. (2013) and the genetically optimized ε-SVR model (GA-ε-SVR) proposed by Chen and Wang (2007) , Pai et al. (2006) and Yuan (2012) will act as benchmarks to the RG-SVR algorithm. Compared to non-adaptive algorithms presented in the literature, the proposed model does not require from the practitioner to follow any time-consuming optimization approach (such as cross validation or grid search) and is free from the data snooping bias (all parameters of RG-SVR are optimized in a single optimization procedure).

From the results of this analysis, it emerges that RG-SVR presents the best performance in terms of statistical accuracy and trading efficiency for all exchange rates under study. RG-SVR's trading performance and forecasting superiority not only confirms the success of the implemented fitness function, but also validates that applying GAs in this hybrid model to optimize the SVR parameters is more efficient compared to the optimization approaches (cross validation and grid search algorithms), that dominate the relevant literature.

The rest of the paper is organized as follows. Section 2 is a literature review of previous relevant research on SVMs and SVRs in forecasting. A detailed description of the study's dataset, the EUR/USD, EUR/GBP and EUR/JPY European Central Bank (ECB) fixing series, is presented in Section 3. Section 4 includes the complete description of the hybrid RG-SVR model, while the statistical and trading performance of the implemented models is presented in Sections 5 and 6 respectively. The concluding remarks are provided in Section 7. The essential theoretical background for the complete understanding of the proposed methodology is given in Appendices A–D, along with the technical characteristics of the models used in this study.

2. Literature review

SVMs were originally developed for solving classification problems in pattern recognition frameworks. The introduction of Vapnik's (1995) insensitive loss function has extended their use in non-linear regression estimation problems. SVRs' main advantage is that they provide global and unique solutions and do not suffer from multiple local minima, while they present a remarkable ability of balancing model accuracy and model complexity (Kwon & Moon, 2007; Suykens, De Brabanter, Lukas, & Vandewalle, 2002).

The literature of SVM and SVR applications is voluminous, especially when they are applied in financial tasks. This study aims to delve deeper into their hybrid structures that are already very popular (Lo, 2000). Lee, Lin, and Wahba (2004) propose the multi-category SVM as an extension of the traditional binary SVM and apply it in two different case studies with promising results. They note that their proposed methodology can be a useful addition to the class of nonparametric multi-category classification methods. Liu and Shen (2006) advance the previous mentioned approach by presenting the multi-category ψ-learning methodology. The main advantage of their method is that the convex SVM loss function is replaced by a non-convex ψ-loss function, which leads to smaller number of support vectors and sparser solution. Martens, Baesens, van Gestel, and Vanthienen (2007) introduce two extraction techniques for SVMs and prove their utility in a series of tests. Hsu, Hsieh, Chih, and Hsu (2009) integrate SVR in a two-stage architecture for stock price prediction and present empirical evidence that shows that its forecasting performance can be significantly enhanced compared to a single SVR model. Lu, Lee, and Chiu (2009) and Yeh, Huang, and Lee (2011) propose also hybrid SVR methodologies for forecasting the TAIEX index and conclude that that they perform better than simple SVR approaches and other autoregressive models. Wu and Liu (2007) introduce the robust truncated hinge loss SVM and claim that their method can overcome drawbacks of traditional SVM models, such as the outliers' sensitivity in the training sample and the large number of support vectors. Huang, Chuang, Wu, and Lai (2010) forecast the EUR/USD, GBP/USD, NZD/USD, AUD/USD, JPY/USD and RUB/USD exchange rates with a hybrid chaos-based SVR algorithm. In their application, they confirm the forecasting superiority of their proposed technique compared to chaos-based NNs and several traditional non-linear models. Lin and Pai (2010) introduce a fuzzy SVR model for forecasting indices of business cycles, Kim and Sohn (2010) forecast the credit score of small and medium enterprises with SVM, while Wu and Akbarov (2011) apply successfully weighted SVRs to the task of forecasting warranty claims. Moreover, Jiang and He (2012) propose a hybrid SVR that incorporates the Grey relational grade weighting function. When applied to financial time series forecasting, the local Grey SVR outperforms locally weighted counterparts in terms of computational speed and prediction accuracy. A hybrid architecture for computer products' sales forecasting is also introduced by Lu (2014) based on SVR and multivariate adaptive regression splines.

Most recently, Yao, Crook, and Andreeva (2015) use SVRs in the credit risk modeling framework. Specifically, the authors evaluate the predictive ability of SVR over recovery rates of defaulted corporate instruments between 1985 and 2012. The results show the superiority of the SVR techniques in forecasting these rates compared to other commonly used methods, such as linear regression, fractional response regression and the two-stage methodology. Finally, Geng, Bose, and Chen (2015) present a forecast competition of methodologies, such as NNs, SVM, decision trees and majority voting classifiers, to the task of predicting financial distress of listed Chinese companies. The empirical evidence of that study shows that NNs outperform the SVM, but they acknowledge that this is contradicting previous literature.

Similar applications to the proposed hybrid approach of this study can be found in the literature. For example Min et al. (2006) and Wu et al. (2007) use hybrid GA-SVM models in order to forecast the bankruptcy risk. In both applications, the GAs optimize the parameters of the SVM and select the financial ratios that most affect bankruptcy. Dunis et al. (2013) developed a GA-SVM algorithm and applied it to the task of trading the daily and weekly returns of the FTSE 100 and ASE 20 indices. This approach deals with financial forecasting as a classification problem and has limited applicability. In financial forecasting, though, it is crucial to obtain forecasts that predict not only the sign but also the size of the examined financial indices.

Pai et al. (2006) apply epsilon SVR with genetically optimized parameters (GA-ε-SVR) in forecasting exchange rates, while