A Combination Forecasting Model Using Machine Learning and Kalman Filter for Statistical Arbitrage

Jarley P. Nóbrega
Centro de Informática
Universidade Federal de Pernambuco
Recife (PE), Brazil
jpn@cin.ufpe.br

Adriano L. I. Oliveira
Centro de Informática
Universidade Federal de Pernambuco
Recife (PE), Brazil
alio@cin.ufpe.br

Abstract—In this paper we evaluate the combination of Extreme Learning Machine (ELM) and Support Vector Regression (SVR) with a Kalman filter regression model for financial time series forecasting. We also compare the forecast performance with a set of linear regression combination methods. The application of the traditional Kalman Filter for the statistical arbitrage strategy improves the statistical performance of ELM and SVR individual forecasts. The accuracy of the models is statistically tested and an investigation is performed to confirm the impact of the forecasts combination in terms of annualized returns and volatility

Keywords— Statistical Arbitrage; Pair Trading; Extreme Learning Machine; Support Vector Regression; Kalman Filter; Forecast Combinations

I. Introduction

There has been growing interest in financial time series forecasting in recent years and accurate predictive models are taking a central place for decision making. In this context, neural networks have become one of the most useful tools for applications in financial time series analysis and forecasting [1].

However, it is known that neural networks have the issue of heavy computation, since all parameters need to be tuned during the training phase. Such iterative approach will take a long time for tuning and the learning process can be inefficient [2].

To overcome the weakness of traditional neural networks, new classes of learning methods were introduced. For instance, Extreme Learning Machine (ELM) was introduced by Huang et al. in [3]. One advantage of ELM over traditional neural networks is that it is not necessary to adjust parameters iteratively.

Support Vector Regression (SVR) is another class of learning algorithms that has also been receiving increasing attention to solve nonlinear estimation problems, including financial time series forecasting [4]. According to the model described by Vapnik et al. [5], it presents a global minimum solution for regression problems and faster training speed when compared with traditional neural networks architectures.

Financial data can also be described in terms of a state space model for representing dynamic systems with unobserved variables in order to process complex nonlinear and non-stationary signals with strong component correlations [7]. In particular, Kalman Filter is largely applied to build forecasting models as a noisy observation of some mean-reverting state process [8].

In addition to the forecasting methods mentioned before, some new approaches based on the combination of them have been successfully applied in order to outperform the accuracy of individual forecasts. Experiments involving the combination of time-varying financial forecasts with neural networks and a Kalman Filter regression model were conducted with encouraging results [9].

In particular, the results reported by [6] present evidences that the combination of individual forecasts using a Kalman Filter regression model increase the forecast accuracy when applied for financial time series. However, in this work it is not presented a statistical test in order to confirm the differences of accuracy for each model. Moreover, the short time period used in those experiments do not assert that the improvements in statistical performance have a correlation with the annualized returns for a given trading strategy.

In this paper, we propose to extend the method presented in [6] in order to investigate the statistical and trading performance of the combination of ELM and SVR with the Kalman Filter regression model. We also compare the proposed method with three linear combination methods: the Bayesian Average [10], Granger and Ramanathan Regression (GRR) [11] and Least Absolute Shrinkage and Selector Operator (LASSO) [12]. Using the same approach as defined in [6], the statistical performance of the models described in this paper was estimated using the statistical arbitrage trading strategy [13]. We also evaluated the trading risk for the series described in this paper, using the Sharpe ratio as a risk measure [14].

The novelty of this research is the extension of the framework proposed in [6]: for dimensionality reduction purpose we apply a feature selection procedure in the datasets in order to find the best attributes for the models. We also

expand the period of the training and testing experiments to include different market conditions. As the original method, we put together two different classes of learning methods and combine them using a state space model. We also compare the forecast performance with a set of linear regression combination methods. At the end, the accuracy of the models is statistically tested and an investigation is performed to confirm the impact of the forecasts combination in terms of annualized returns and volatility.

The remainder of this paper is organized as follows. Section II discusses the fundamentals for the proposed method. In Section III, we present the proposed method. In Section IV, we present the description of the sample data, the statistical and trading performance, respectively. In Section V, we briefly discuss the results of the experiments. Section VI concludes this work.

II. FUNDAMENTALS

A. Statistical Arbitrage

In the world of finance, the term statistical arbitrage encompasses a variety of strategies that attempt to profit from pricing discrepancies that appear in a group of assets. This process identifies pairs of securities whose prices tended to move together. In Fig. 1 is presented an example of pair with correlated price movement. In this example, the time series for AMBV3 and AMBV4 presents similar movement for the period of February, 2009 to November, 2013. This strategy was designed to exploit short-term deviations from a long-run equilibrium pricing relationship between two assets, based on cointegration, correlation and other nonparametric decision rules [13].
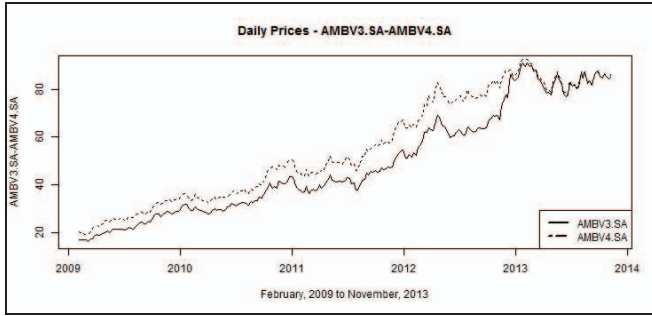

Fig. 1.   AMBV3 and AMBV4 daily price series.

The detection of price anomalies is based upon the identification of a linear combination of assets, whose time series is mean-reverting and has finite variability. This model will allow us to make predictions for the relative difference of a pair of financial assets, named *pair spread*. A history and discussion about statistical arbitrage and pairs trading can be found in [15].

B. Extreme Learning Machine

Unlike the traditional learning algorithms for neural networks, the main characteristic of ELM is learning without iterative training, as proposed by Huang et al [3]. Let the

training set be $\left\{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m, i = 1, ..., N\right\}$, where x_i is an $n \times 1$ input vector and t_i is a $m \times 1$ target vector. The training process is briefly described as follows.

Step 1: Randomly assign values to the inputs weights and the hidden neuron biases.

Step 2: The output weights are analytically determined through the generalized inverse operation of the hidden layer matrices, according to the following equation:

$$\sum_{i=1}^{L} \beta_i G(a_i, b_i, x_j) = t_j, \quad j = 1, ..., N \qquad (1)$$

where a_i is the input weights, b_i is the hidden layer biases, β_i is the output weight that connects the ith hidden node and output node, and G is the activation function. L is the number of hidden neurons. N is the number of distinct input or output data. This is equivalent to $H\beta = T$, where

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_L) \\ \vdots & \cdots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \qquad (2)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N \times m}$$

H is the hidden layer output matrix. The activation function $G(x)$ should be assigned before training is carried out. For this paper, we used the sigmoid function. It is given by

$$G(a_i, b_i, x_i) = \left(1 + e^{-(a_i \cdot x_i + b_i)}\right)^{-1}, \quad i = 1, ..., N \qquad (3)$$

Step 3: Calculate the output weight by $\beta = H^+ T$, where H^+ is the Moore-Penrose generalized inverse of H.

C. Support Vector Regression

Support Vector Machines (SVM) use an implicit mapping Φ of the input data into a high-dimensional feature space defined by a kernel function, i.e., a function returning the inner product $\langle \Phi(x_i) \Phi(x) \rangle$ between the images of two data points x_i, x in the feature space [5]. If a projection $\Phi: X \to H$ is adopted, the dot product $\langle \Phi(x_i), \Phi(x) \rangle$ can be represented by the following kernel function k

$$k(x_i, x) = \langle \Phi(x_i) \Phi(x) \rangle \qquad (4)$$

If applied for classification problems, SVM separate the different classes of data by a hyper-plane $\langle w, \Phi(x) \rangle + b = 0$, corresponding to the decision function

$$f(x) = sign(\langle w, \Phi(x) \rangle + b) \qquad (5)$$

1295

where $w = \sum_i \alpha_i \Phi(x_i)$, α_i are the coefficients and b is a constant. The optimal hyper-plane is the one with the maximal margin of separation between the two classes, with the optimization problem solved by a constrained quadratic method for a subset of training patterns. These patterns are called support vectors.

For the regression task, let the training data $\{(x_1, y_1), ..., (x_n, y_n)\} \subset \chi \times \Re$, where χ denotes the space of input patterns. We have to find a function $f(x)$ that has at most ε deviation from targets y_i for all training data and is as flat as possible. The formulation of support vector regression stated in [5] implies in the minimization of the following expression:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \quad (6)$$

subject to

$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i$$

$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

where the constant $C > 0$ defines the trade-off between the flatness of f and the level of tolerance for ε. ξ_i, ξ_i^* are two positive slack variables which can be used to measure the deviation from the boundaries of the ε-sensitive zone. The construction of a Lagrange function for Eq. (6) generates the following regression:

$$f(x) = \left(\alpha_i - \alpha_i^*\right)\langle x_i, x \rangle + b \quad (7)$$

where α_i, α_i^* are Lagrange multipliers. The expansion of Eq. (7) for the nonlinear case can be written as

$$f(x) = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)K(x_i, x) + b \quad (8)$$

For this paper, we are adopting the RBF kernel

$$K(x_i, x) = \exp\left(-\gamma\|x_i - x\|^2\right), \ \gamma > 0 \quad (9)$$

where γ is the parameter of the kernel.

D. Kalman Filter Regression

The Kalman Filter can be described as a recursive method to estimate the state of a dynamic system from a series of incomplete and noisy measurements [16]. The state representation of the dynamics of the time-varying regression coefficients is given by the following system of equations:

$$P_{Y_t} = P_{X_t}\beta_t + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \quad (10)$$

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim N\left(0, \sigma_\eta^2\right) \quad (11)$$

where P_{Y_t} is the dependent variable, β_t is a time-varying regression coefficient, and P_{X_t} is the independent variable at time t, respectively. ε_t and η_t are independent uncorrelated error terms with standard variances σ_ε^2 and σ_η^2, respectively. Eq. (10) is also named the measurement equation and Eq. (11) is named the state equation, which defines the regression coefficient as a simple random walk. When we apply the Kalman Filter, the β_t coefficients are estimated by maximizing a likelihood function with a numerical algorithm based on σ_η^2. The coefficients are estimated at time t based on the new observations and the states estimates are propagated in time $t+1$.

For this paper, the regression model is defined by the following equations:

$$f_t = \alpha_{ELM_t} f_{ELM_t} + \alpha_{SVR_t} f_{SVR_t} + \varepsilon_t, \ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \quad (12)$$

$$\alpha_{ELM_t} = \alpha_{ELM_{t-1}} + \eta_{ELM_t}, \ \eta_{ELM_t} \sim N\left(0, \sigma_{\eta_{ELM}}^2\right) \quad (13)$$

$$\alpha_{SVR_t} = \alpha_{SVR_{t-1}} + \eta_{SVR_t}, \ \eta_{SVR_t} \sim N\left(0, \sigma_{\eta_{SVR}}^2\right) \quad (14)$$

where, α_{ELM_t} and α_{SVR_t} are time-varying regression coefficients at time t for ELM and SVR. Eq. (12) is the measurement equation and Eq. (13) and Eq. (14) describe the states equations. Here, η_{ELM_t} and η_{SVR_t} are the error terms with variances $\sigma_{\eta_{ELM}}^2$ and $\sigma_{\eta_{SVR}}^2$, respectively.

III. Proposed Method

As a preprocessing step for training the ELM and SVR models, the proposed method reduces the dimensions of the training datasets by performing a single feature selection which evaluates a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them [17]. We choose the subsets that are high correlated with the current pair spread by calculating a correlation matrix for each attribute subset. In order to explore the attributes space we applied a Particle Swarm Optimization (PSO) [18] algorithm to find the best subsets. For this paper, we defined empirically the PSO parameters, which are the social, inertia and individual weights. The number of particles in the swarm is 20 and the probability of mutation is 0.01. The number of iterations for the PSO search is 200. The full training set for all pairs were used to find the best attributes.

After the feature selection step, we train the ELM and SVR models by adjusting their parameters. For ELM, we adjusted iteratively the size of the hidden layer nodes considering the range of [50,100] to obtain the best network setup. To optimize the values of the SVR parameters, γ and C, we used the Caret R package [19] in the in-sample dataset, according to Eq. (6) and Eq. (9).

Next, we applied the Eq. (12) to generate the pair spread forecast by combining the ELM and SVR individual forecasts using the Kalman Filter regression model. In order to evaluate the statistical performance of the model, we applied this

procedure for traditional classes of learning algorithms and regression combination models. We compared our approach with the combinations of ELM and SVR forecasts using regression models based on BMA, GRR, and LASSO for comparison purposes. All evaluation was made in both in-sample and out-of-sample datasets.

Finally, the last step of the proposed method is to evaluate the econometric performance of the model and compare it with the BMA, GRR and LASSO. In general, pairs trading are a bet on the mean reversion property of the linear combination of cointegrated time series. As presented in [20], this property can be defined by two quantities: the cointegration coefficient and the equilibrium value. Moreover, the equilibrium adjustment can be estimated by modeling the mean-reverting process as an Ornstein-Uhlenbeck (O-U) model [21].

The trading rule implies in opening a long or short position for the pair depending the following conditions:

- Enter a long position, buying the first part of the pair and selling the second one when $Z_t < \hat{Z}_{t+h}^{L,\alpha}$ and unwind the operation when Z_t converges to its mean.

- Enter a short position, selling the first part of the pair and buying the second one when $Z_t > \hat{Z}_{t+h}^{H,\alpha}$ and unwind the operation when Z_t converges to its mean.

where Z_t is the value of the forecast of the pair spread at time t, $\hat{Z}_{t+h}^{L,\alpha}$ and $\hat{Z}_{t+h}^{H,\alpha}$ denote the *(1-α)* low and high confidence bound on the forecasted value of the spread h minutes ahead, and h is the estimate of mean reversion speed.

In this paper, the cost for opening and closing positions for a selected pair was estimated by market benchmarks, including the brokerage and fee values. Another source of cost for trading transaction is the *slippage*, the difference between ask and bid values, which leads to a small loss when we buy or sell an asset. The slippage values for each pair are calculated at the first level of the book orders. As a single risk management policy, we established the maximum percentage level of loss for an opened trade position. If the current net operation exceeds 1% of loss, the position is immediately closed.

The evaluation of trading performance was performed in terms of the annualized return, the annualized volatility and the Sharpe ratio [14]. Annualized return describes the average amount of money earned by an investment each year over a given time period. The annualized volatility is determined by the standard deviation of the return and the Sharpe ratio measures the excess return or risk premium per unit deviation in a trading strategy. For this paper, the Sharpe ratio characterizes how well the return of the statistical arbitrage strategy compensates the investor for the risk taken.

In Fig. 2 is presented a diagram with the summary of the proposed method.
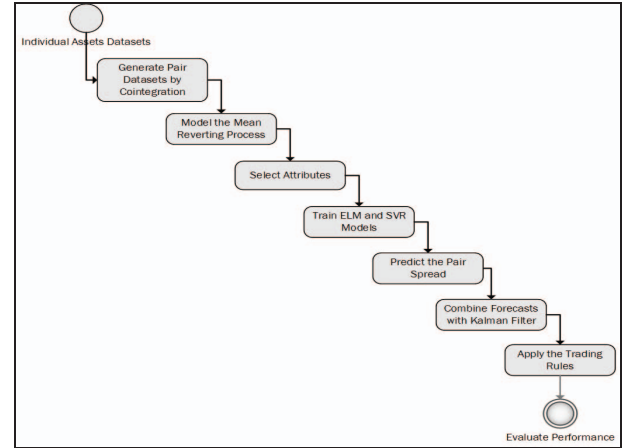

Fig. 2.   Kalman Filter Combination Approach.

IV.   RESULTS

A. Sample Data

We selected a set of tradable financial assets that are listed on BM&FBovespa Exchange[1]. For this task, we captured all asset quotes between the period of February, 2009 and November, 2013. The combination of assets results in twelve pair spreads generated from the cointegration approach. The pairs and their underlying assets are described in Table I.

The cointegration approach was defined by Engle and Granger [20] as a statistical feature whereby two time series that are integrated of order 1, *I(1)*, can be linearly combined to produce one stationary time series, denoted as *I(0)*. Let $X_{1,t}, X_{2,t},..., X_{k,t}$ a sequence of time series *I(1)*. If exist non-zero real numbers $\beta_1, \beta_2,..., \beta_k$ and $\beta_1 X_{1,t}, \beta_2 X_{2,t},..., \beta_k X_{k,t}$ is a time series *I(0)*, then $X_{1,t}, X_{2,t},..., X_{k,t}$ is a sequence of cointegrated time series. In Table II we present the datasets used in all experiments. Since all pairs in this study were generated by cointegration, their stationary property is confirmed at the 1% confidence level. The Jarque-Bera statistics confirms that the spread series are non-normal at the 99% confidence interval. By applying the Phillips-Ouliaris test for cointegration [23], we confirm at the 1% confidence level that all pairs have at least one cointegration vector. The inputs of the training dataset are presented in Table III.

TABLE I.   Selected Cointegrated Pairs

Pair	Description
AMBV3-AMBV4	Companhia de Bebidas Das Americas ON/PN
BBDC3-BBDC4	Banco Bradesco ON/PN
BOVA11-PIBB11	Bovespa IShares / PIBB IShares
BRAP3-BRAP4	Bradesco Holding ON/PN
CMIG3-CMIG4	Companhia Energetica MG ON/PN
CPLE3-CPLE6	Companhia Paulista de Energia ON/PNA
ELET3-ELET6	Eletrobras ON/PNA
GOAU4-GGBR4	Gerdau Metalurgica ON/PN
LAME3-LAME4	Lojas Americanas ON/PN
PETR3-PETR4	Petrobras ON/PN
USIM3-USIM5	Usiminas ON/PN
VALE3-VALE5	Vale ON/PN

---

[1] http://www.bmfbovespa.com.br

TABLE II.   Training, Validation and Testing datasets

	Trading Days	Start Date	End Date
Total dataset	1209	02/02/2009	22/11/2013
Training dataset	605	02/02/2009	01/07/2011
Validation dataset	302	02/07/2011	10/09/2012
Testing dataset	302	11/09/2012	22/11/2013

TABLE III.   Dataset attributes

Number	Attribute	Lag[(*)]
1-10	Pair Spread	1-10
11-20	Spread Volatility	1-10
21-30	Spread Mean	1-10
31-40	Mean Reversion Speed Estimation	1-10

[(*)] Lag 1 implies in forecasting tomorrow close price
using today's close price.

B. Statistical Performance

In order to evaluate the statistical performance of the ELM and SVR forecasts, as well the performance of the combination among them, we have executed 200 repetitions for each individual experiment. We compute the RMSE statistics to support our analysis in terms of the accuracy of the models. In Fig. 3 and 4 we present a summary of the statistical performance for the in-sample and out-of-sample datasets. When comparing the individual results, the SVR forecast outperforms the ELM accuracy in both in-sample and out-of-sample datasets for most of cases, unlike what was presented in [6]. When we apply the combinations techniques, the overall performance is quite similar to all of them, except for the Kalman Filter which outperforms all models studied here, similarly to what was highlighted in the previous study.
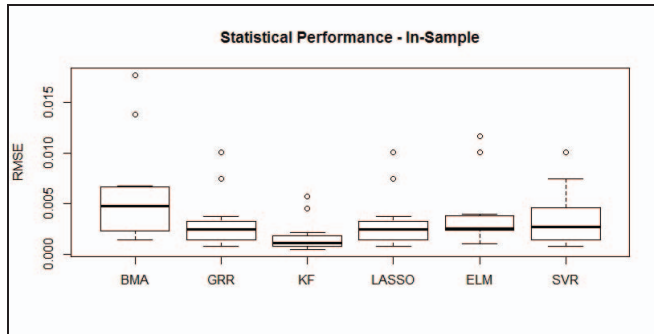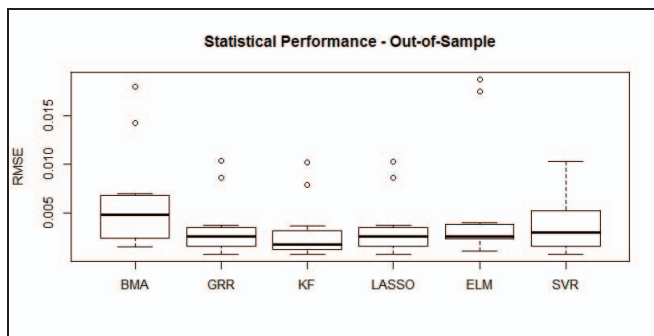

Fig. 3.   RMSE for In-Sample datasets.

Fig. 4.   RMSE for Out-of-Sample datasets.

C. Econometric Performance

In Tables IV and V we present a summary of the trading performance for ELM and SVR models. The econometric results for the combination models are presented in Tables VI and VII. The criteria for performance here is the highest average of annualized return with low average of volatility and high Sharpe ratio. We note that all pairs presented positives annualized returns and corresponding positive annualized Sharpe ratios. The best model in terms of annualized returns was the combination of ELM and SVR forecasts using the Kalman filter regression model. This confirms the findings reported by [6], even when the period of training and testing of the models increases. For the in-sample dataset the Kalman filter model achieved 6.98% of annualized return when combining the ELM and SVR forecasts with the Kalman Filter model. The trading performance for the same model in the out-of-sample dataset achieved 26.13% of annualized return and a Sharpe ratio of 5.29. We conclude that the iterative nature of Kalman Filter model gives it some advantage when compared to the remaining models, since it computes forecasts propagating the states estimates when new observations are feeding the model.

The individual ELM and SVR forecasting models only outperforms the trading returns for the BMA model, both in-sample and out-of-sample, which confirms our research hypothesis. Just as [6], we have not found an evidence of the superior performance of Kalman Filter method for all criteria.

TABLE IV.   ECONOMETRIC PERFORMANCE – ELM AND SVR – IN-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
ELM	-2.31%	1.80	**3.05%**	5.83%
SVR	**-2.18%**	1.73	3.20%	5.39%

TABLE V.   ECONOMETRIC PERFORMANCE ELM AND SVR – OUT-OF-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
ELM	-2.80%	3.83	5.64%	20.18%
SVR	-2.72%	4.31	5.29%	21.32%

TABLE VI.   ECONOMETRIC PERFORMANCE – COMBINATION MODELS– IN-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
BMA	-2.24%	1.38	**3.05%**	4.94%
GRR	-2.34%	**2.07**	3.20%	6.37%
KALMAN	-2.30%	1.97	3.57%	**6.89%**
LASSO	-2.43%	2.06	3.39%	6.30%

TABLE VII.   ECONOMETRIC PERFORMANCE – COMBINATION MODELS – OUT-OF-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
BMA	-2.97%	3.83	**5.12%**	19.33%
GRR	**-2.53%**	4.76	5.49%	23.69%
KALMAN	-2.64%	**5.29**	5.17%	**26.13%**
LASSO	-2.64%	4.76	5.49%	23.79%

V.   EVALUATION

In order to confirm the results obtained from the statistical evaluation, we performed a set of tests for the mean of forecast errors. We aim to compare the performance of the Kalman filter with the remaining models. It was performed 60 Student's t-tests with confidence level of 95%. The null

hypothesis of all tests is that there is no difference between the residuals of Kalman model and the remaining models. The test statistics rejected the null hypothesis for 11 of 12 selected pairs. The Exception was the VALE3-VALE5 pair where the null hypothesis cannot be rejected for GRR, LASSO and SVR models. When compared with the results reported in [6], these results presents a statistical evidence that the forescast errors decreases when we apply the combination of ELM and SVR forecasts using the Kalman filter regression model. However, it was unable to verify the same behavior when we apply the tests for the mean of returns. Only the pair AMBV3-AMBV4 had the null hypothesis rejected for the equality of means. These results are consistent with the work shown in [6] due to profitability of statistical arbitrage strategies is dependent on trading rules.

The results of the experiments indicate that the combination of forecasts using a regression model increases the annualized return for both in-sample and out-of-sample datasets. Moreover, the level of volatility of the combined models is less than those of the ELM and SVR individual forecasts. The Kalman filter regression model also achieved the best Sharpe ratio for the out-of-sample dataset, which implies in a reduced risk level when applying the model in a real trading environment.

VI. CONCLUSIONS

In this paper we have investigated the statistical and economic performance for pairs trading strategy using ELM and SVR models, and their forecast combination through Kalman filter regression models. We first selected a set of pairs, which had their cointegration and stationary properties confirmed by ADF and Phillips-Ouliaris tests. Next, we modeled the pair spreads dynamics for each pair using the Ornstein-Uhlenbeck process in order to estimate the speed of mean reversion. As a preprocessing step, we applied a feature selection method by calculating a correlation matrix for the pairs spread attributes. In order to explore the attributes space we applied a Particle Swarm Optimization algorithm to find the best subsets. The statistical arbitrage model feeds both ELM and SVR models that aimed to produce forecasts for the price spread direction. We also combined the ELM and SVR forecasts with three linear regression models in order to compare the overall performance.

The results have shown that Kalman Filter outperforms all models studied in terms of statistical performance which confirm the findings reported in [6] even for a large test period. We confirmed the statistical difference of the Kalman filter accuracy by applying t-tests with the remaining models. On the other hand, the statistical tests for the returns of each model indicate that there is no significant difference between them. However, when applying the Kalman filter model for combining the ELM and SVR forecasts the overall level of profitability increases and the volatility level decreases, which is an evidence of lower risk to run the model in a real trading environment.

REFERENCES

[1] M. Deutsch, C.W.J. Granger, T. Teräsvirta, "The combination of forecasts using changing weights," International Journal of Forecasting 10 (1) (1994) 47–57

[2] G.-B. Huang, "Learning capability and storage capacity of two hiddenlayer feedforward networks," IEEE Trans. Neural Networks, Vol. 14, no. 2, pp. 274-281, 2003.

[3] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, "Extreme learning machine: Theory and applications", Neurocomputing, Vol.70, no.1-3, pp. 489–501, Dec. 2006.

[4] C.-J. Lu, T.-S Lee, C.-C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," Decision Support Systems, 47(2), 2009,115–125. doi:10.1016/j.dss.2009.02.001.

[5] V. Vapnik, S. Golowich, A. Smola, "Support vector machine for function approximation, regression estimation, and signal processing," Advances in Neural Information Processing Systems, vol. 9, p. 281–287, 1996.

[6] J. P. Nóbrega, A. L. I. Oliveira, "Improving the Statistical Arbitrage Strategy in Intraday Trading by Combining Extreme Learning Machine and Support Vector Regression with Linear Regression Models", IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI), 182- 188, 2013.

[7] S.L. Goh, D.P. Mandic, "An augmented extended Kalman Filter algorithm for complex-valued recurrent neural networks," Neural Computation 19 (4) (2007) 1039–1055.

[8] G. Sermpinis, C. Dunis, J. Laws, C. Stasinakis, "Forecasting and Trading the EUR/USD Exchange Rate with Stochastic Neural Network Combination and Time-varying Leverage," Decision Support Systems, ISSN 0167-9236 (2013)

[9] N. Terui, H.K. Van Dijk, "Combined forecasts from linear and nonlinear time series models," International Journal of Forecasting 18 (3) (2002) 421–438.

[10] S.T. Buckland, K.P. Burnham, N.H. Augustin, "Model selection: an integral part of inference", Biometrics 53 (2) (1997) 603–618.

[11] C.W.J. Granger, R. Ramanathan, "Improved Methods of Combining Forecasts," Journal of Forecasting 3 (2), 197–204 (1984)

[12] H. Wang, G. Li, G. Jiang, "Robust Regression Shrinkage and Consistent Variable Selection through the LAD–Lasso," Journal of Business and Economic Statistics 25 (3), 347–355 (2007)

[13] G. Vidyamurthy, "Pairs Trading, Quantitative Methods and Analysis," John Wiley & Sons, (2004)

[14] W.F. Sharpe, "The Sharpe Ratio," The Journal of Portfolio Management 21 (1),2010, 49–58, doi:10.3905/jpm.1994.409501

[15] A. Pole, "Statistical Arbitrage: Algorithmic Trading Insights and Techniques," Wiley Finance, 2007.

[16] A.C. Harvey, "Forecasting, Structural Time Series Models and the Kalman Filter," Cambridge University Press, Cambridge, U.K., 1990.

[17] M. A. Hall, "Correlation-based Feature Subset Selection for Machine Learning", Hamilton, New Zealand, 1998.

[18] A. S. Moraglio, C. D. Chio, R. Poli, "Geometric Particle Swarm Optimisation", In: Proceedings of the 10th European Conference on Genetic Programming, Berlin, Heidelberg, 125-136, 2007.

[19] "Caret R Package", June, 2013 [online], Available: http://cran.r-project.org/web/packages/caret/caret.pdf

[20] R.F. Engle, C.W.J. Granger, "Cointegration and error-correction: representation, estimation and testing," Econometrica, 55, 1987, 251-276.

[21] E. Bibbona, G. Panfilo and P. Tavella: "The Ornstein-Uhlenbeck process as a model of a low pass filtered whitenoise," Metrologia, 45, 2008, S117-S126.