

Accepted Manuscript

Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices

Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura



PII: S2405-9188(18)30006-0

DOI: [10.1016/j.jfds.2018.04.003](https://doi.org/10.1016/j.jfds.2018.04.003)

Reference: JFDS 44

To appear in: *The Journal of Finance and Data Science*

Received Date: 22 January 2018

Revised Date: 14 March 2018

Accepted Date: 20 April 2018

Please cite this article as: Henrique BM, Sobreiro VA, Kimura H, Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices, *The Journal of Finance and Data Science* (2018), doi: 10.1016/j.jfds.2018.04.003.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices[☆]

Bruno Miranda Henrique^a, Vinicius Amorim Sobreiro^{a,*}, Herbert Kimura^a

^aUniversity of Brasília, Department of Economics, Campus Darcy Ribeiro, Brasília, Federal District, 70910-900, Brazil.

Abstract

The purpose of predictive stock price systems is to provide abnormal returns for financial market operators and serve as a basis for risk management tools. Although the Efficient Market Hypothesis (EMH) states that it is not possible to anticipate market movements consistently, the use of computationally intensive systems that employ machine learning algorithms is increasingly common in the development of stock trading mechanisms. Several studies, using daily stock prices, have presented predictive system applications trained on fixed periods without considering new model updates. In this context, this study uses a machine learning technique called Support Vector Regression (SVR) to predict stock prices for large and small capitalisations and in three different markets, employing prices with both daily and up-to-the-minute frequencies. Prediction errors are measured, and the model is compared to the random walk model proposed by the EMH. The results suggest that the SVR has predictive power, especially when using a strategy of updating the model periodically. There are also indicative results of increased predictions precision during lower volatility periods.

Keywords: Prediction, Stock Market, Machine Learning, Support Vector Regression, High Frequency Trading.

[☆]This document was a collaborative effort.

*Corresponding author

Email addresses: brunomhenrique@hotmail.com (Bruno Miranda Henrique), sobreiro@unb.br (Vinicius Amorim Sobreiro), herbert.kimura@gmail.com (Herbert Kimura)

1. Introduction

Stock price prediction mechanisms are fundamental to the formation of investment strategies and the development of risk management models (Dash and Dash, 2016, p. 43). The Efficient Market Hypothesis (EMH), however, states that it is not possible to consistently obtain risk-adjusted returns above the profitability of the market as a whole (Malkiel and Fama, 1970). Computational advances have led to several machine learning algorithms used to anticipate market movements consistently and thus estimate future asset values such as company stock prices (Gerlein et al., 2016, pp. 193–194). Models based on the Support Vector Machine (SVM) are among the most widely used techniques.

Information is a valuable resource when building predictive models in the pursuit of profitable financial market transaction systems. Given the peculiarities of financial time series, various challenges must be faced when developing price forecasting systems (Araújo et al., 2015, p. 4081). From a theoretical point of view, under the EMH, relevant information would be widely available to all market participants and immediately reflected in price, according to Malkiel and Fama (1970, p. 383). Malkiel and Fama (1970)'s EMH claims that it is impossible, consistently and over the long term, to achieve above-market returns adjusted to the level of risk assumed. As summarised by Malkiel (2003), the EMH has been questioned since its introduction, especially with the development of predictive systems, as shown in studies based on SVM and other algorithms (for example, Ballings et al. (2015); Nayak et al. (2015); and Qu and Zhang (2016)) that can generate profit in the long term. Malkiel and Fama (1970, pp. 386–387), however, argue that the market follows a random walk and that attempts to predict its movements in a consistent manner will be vain.

Computational advances have led to the introduction of machine learning techniques for predictive systems in financial markets. In a review of articles on predictive systems, Hsu et al. (2016, p. 215) observed that it is common to use financial series to measure the efficiency of predictive algorithms and classifiers in machine learning. Classifiers are systems that can learn, through training, to recognise patterns and thus assign a class to new data. As an example, machine learning algorithms can be used to predict insolvency, as observed by Zhou et al. (2012) and Li et al. (2012). In such cases, the aim is to classify companies with the highest probability of insolvency, according to an automatic classifier algorithm. Other examples are credit risk measurement, as in Li et al. (2006), and asset price forecasting, as proposed by Kao et al. (2013) and Xiao et al. (2013).

In addition to developing transactional strategies, progress in computational information systems has enabled rapid electronic transactions to take place in financial markets. Based on high-frequency trading algorithms, the submission and execution of purchase, sale or cancellation orders can be performed in seconds and microseconds, as Goldstein et al. (2014, pp. 182–183) note. The intensive use of rapid computational systems by some market participants may increase profitability, but the effects on normal

35 market functioning are questionable, as not all participants have access to this type of technology (Goldstein
36 et al., 2014, pp. 182–183).

37 To analyse the EMH, as Hsu et al. (2016) have sought to do, this study tests stock predictability in Brazil,
38 the United States and China, based on prediction error analysis. The study does not seek to identify trading
39 strategies that can lead to extraordinary gains but rather to evaluate prediction errors by comparing a
40 machine learning model with a base model that follows a random walk. The choice of countries is due to the
41 desire to evaluate results of machine learning techniques in both developed and developing markets. In
42 regard to stock price frequency, daily data are traditional in prediction studies, such as those of Kumar et al.
43 (2016), Zbikowski (2015) and Patel et al. (2015b). However, other studies, such as those of Qu and Zhang
44 (2016) and Manahov et al. (2014), use up-to-the-minute data. Our study shows results for both daily and
45 up-to-the-minute prices.

46 Blue chip and small cap stocks are selected, as higher and lower capitalisations in each country, re-
47 spectively. One hypothesis tested involves the argument that a more accurate prediction can be obtained
48 using a frequency greater than daily. Moreover, aiming to capture changing market conditions more quickly,
49 results are evaluated by comparing periodic updating of models with an absence of updating in terms of
50 prediction performance in each case. The selected prediction method is a regression method based on SVM,
51 as used by Qu and Zhang (2016), Patel et al. (2015b) and Choudhury et al. (2014), called Support Vector
52 Regression (SVR). Finally, it should be noted that three kernel functions are tested for SVR to identify the
53 most suitable kernel function for this type of stock price prediction. A random walk model is used as a
54 reference to evaluate predictions of returns.

55 In addition to minimising risks to stock market investors, strategies based on price prediction may
56 provide evidence against the EMH. Predictive studies such as that presented here contribute to the building
57 of profitable strategies, especially risk-adjusted ones, as greater predictability can affect an investment
58 portfolio's exposure level. Thus, greater accuracy in price forecasting may imply potential risk-adjusted
59 profits for investors. Other studies, such as those of Yeh et al. (2011), Choudhury et al. (2014) and Patel et al.
60 (2015b), select just one frequency for prices used in predictions. Our study contrasts different frequencies,
61 using the same model. Specifically, predictions are made using both daily prices and up-to-the- minute
62 prices. Another important contribution of this study is that it contrasts the results of a training model based
63 on a fixed period with those with dynamically updated training periods, thus comparing the predictive
64 performances of these two strategies.

65 Regarding the up-to-the-minutes trading frequency, costs of obtaining and processing large prices
66 database are usually high. However, for illustrating the robustness of a price prediction method, longer
67 periods of testing are always preferable. The same goes for the variety of the selected securities. In this
68 context, this article brings results from SVR price predictions for 2 years of the 1-minute historical stock
69 prices for Brazilian small and large capitalization companies. Finally, an analysis of SVR predictions and

70 basics statistics of each stock is introduced, indicating the existence of a relationship between the predictions
71 precision and volatility in prices.

72 The paper is organised as follows. The next section presents a brief theoretical framework, explaining the
73 methods and data frequencies used in previous studies. Sec. 3 describes the method used to predict stock
74 prices, the performance measures considered and the data used. Sec. 4 presents and discusses the results.
75 Finally, Sec. 5 presents conclusions and limitations of the study and suggests areas of future research.

76 2. Brief literature review

77 Describing the EMH, [Malkiel and Fama \(1970\)](#) state that, on balance, prices reflect all relevant information
78 available when pricing an asset. This hypothesis arises from empirical observations of changes in price
79 time series that are very similar to a random walk process. According to these authors, even a system in
80 which a number of buy and sell orders are generated in the short term is not profitable, due to transaction
81 costs and commissions [Malkiel and Fama \(1970, p. 396\)](#). Despite the evidence obtained for market efficiency,
82 [Malkiel and Fama \(1970, pp. 413–416\)](#) further encourage a search for more data confirming or disproving
83 their hypothesis. Since then, academic papers have sought to show that stock market prices are, to some
84 extent, predictable. [Malkiel \(2003, p. 80\)](#) concludes that not all market participants are rational and that there
85 are irregular price formations, leading to exploitable return patterns over short time periods. [Timmermann
86 and Granger \(2004, p. 20\)](#) consider the possibility that a profitable predictive system may exist but only up to
87 the point of its discovery. In this case, the performance of that system would deteriorate when more market
88 participants begin to use it.

89 The development of consistently profitable systems may constitute evidence against the EMH, as
90 suggested by [Hsu et al. \(2016, pp. 217–218\)](#). Such systems may benefit from computationally intensive
91 techniques, such as those that exploit machine learning algorithms. [Hsu et al. \(2016, p. 229\)](#) show that
92 machine learning algorithms commonly use financial time series to evaluate their predictive capabilities. In
93 this context, [Ballings et al. \(2015\)](#) and [Gerlein et al. \(2016\)](#) have obtained good results in predictions when
94 applying classifiers such as SVM, k-Nearest Neighbours (KNN), neural networks and decision trees.

95 [Zbikowski \(2015\)](#), for example, utilised SVM and a predictive variable selection method within Technical
96 Analysis (TA) indicators. In an attempt to develop an optimal market transactions strategy, [Choudhury et al.
97 \(2014\)](#) used k- means to predict market volatility and SVR to predict prices in the Indian stock market. The
98 authors analysed daily data to estimate prices for two days. [Ballings et al. \(2015\)](#) studied the direction of the
99 stock market and, in so doing, evaluated classifiers. In [Ballings et al. \(2015\)](#)'s study, annual data from more
100 than 5,000 European companies were used in classifiers such as logistic regression, neural networks, KNN
101 and SVM. Classifiers are used to determine the direction of the respective company stocks in the following
102 year. The authors compared the results of those classifiers to ensemble approaches, such as Random Forests

103 (RF), AdaBoost and kernel factory. The so-called ensemble techniques involve multiple classifiers, usually
104 of the same type or algorithm, resulting in independent classifications but with some decision method
105 for determining a single final classification. Applying these techniques, [Ballings et al. \(2015\)](#) calculated
106 their predictive variables, considering companies' balance sheets and financial statements. Based on the
107 prediction of the direction that a particular stock would take during the year, the authors showed how a
108 profitable strategy could be built. Also seeking to predict the direction of stock prices, [Kumar et al. \(2016\)](#)
109 and [Kim \(2003\)](#) applied SVM-based systems to (TA) indicators.

110 Criticising classification approaches based only on the direction of the stock market, [Barak and Modarres](#)
111 [\(2015\)](#) attempted to predict risks and returns of Tehran stocks. The authors proposed a predictor variable
112 selection method and applied decision tree and neural network mechanisms. The predictive variables
113 used by [Barak and Modarres \(2015\)](#) were constructed from financial statements published by companies.
114 [Zbikowski \(2015\)](#) used TA indicator values as predictive variables in his short-term trends prediction model.
115 [Zbikowski \(2015\)](#) proposed a modified SVM, with variable selection determined by Fisher scores, a method
116 used to rank definitions into classes according to certain predictor variables. The TA variables used by
117 [Zbikowski \(2015\)](#) included On Balance Volume (OBV), the Relative Strength Index (RSI) and the Williams
118 oscillator. The results obtained through the SVM approach exceeded those based on a buy-and-hold strategy
119 in [Zbikowski \(2015\)](#)'s study. In turn, [Yeh et al. \(2011\)](#) and [Lu et al. \(2009\)](#) applied SVR-based systems to
120 predict the TAIEX and Nikkei 225 indices, both using daily data.

121 [Gerlein et al. \(2016\)](#) proposed using computationally simpler classifiers for intraday strategies in the
122 Foreign Exchange (FOREX) market. Decision tree algorithms and lazy models were applied to TA variables
123 for strategies on USDJPY, GBPUSD and EURUSD prices. [Patel et al. \(2015a\)](#) also used TA variables as inputs
124 in their model but in a different manner from other authors. Instead of using the indicator values directly, as
125 [Gerlein et al. \(2016\)](#) and [Zbikowski \(2015\)](#) did, [Patel et al. \(2015a\)](#) used the trend indication given by the
126 indicators. A TA indicator can identify the market trend as bullish or bearish. Thus, [Patel et al. \(2015a\)](#)'s
127 model uses this information as a predictive variable in algorithms such as SVM, random trees and neural
128 networks to predict trend rather than price.

129 Some authors have used hybrid machine learning algorithms to increase predictive performance. [Xiao](#)
130 [et al. \(2013\)](#), for example, proposed integrating various neural networks with SVM to predict the daily values
131 of stock indices. Using TA indicators as input variables, [Nayak et al. \(2015\)](#) proposed a hybrid SVM and
132 KNN system for predicting index values. The results were better than those of traditional neural networks.
133 In turn, [Patel et al. \(2015b\)](#) evaluated the efficiency of the daily closing price predictions performed by SVM,
134 RF and neural network hybrids, comparing the results with the isolated use of these same algorithms. The
135 authors concluded that hybrid uses of these algorithms offered better results. Finally, [Dash and Dash \(2016\)](#)
136 investigated an approach to neural networks with TA compared to traditional machine learning algorithms
137 in stock transaction decisions.

138 The development of computing has not only enabled the development of complex prediction algorithms
 139 but also quantitative analysis of high-frequency data, as in [Brownlees and Gallo \(2006\)](#)'s study. This data
 140 type requires a different treatment from that of traditional lower frequency data. [Brownlees and Gallo \(2006\)](#)
 141 suggest ways of treating high frequency data with regard to outliers and temporal irregularities. Once
 142 treated, high frequency data can be used to develop very fast market transaction strategies, known as High
 143 Frequency Trading (HFT). This type of transaction has become common in today's financial markets, and its
 144 effects have been studied by authors such as [Lee \(2013\)](#) and [Goldstein et al. \(2014\)](#).

145 Given the advances in the use of HFT discussed by [Araújo et al. \(2015, p. 4082\)](#), there is plenty of room
 146 for the development of models in this area. For example, [Araújo et al. \(2015\)](#) developed a new mathematical
 147 model to forecast price changes, measured in seconds, of companies listed on the BM&F Bovespa. Meanwhile,
 148 ([Manahov et al., 2014](#)) investigated the profitability of a genetic learning algorithm for the FOREX market,
 149 applied to price changes measured in minutes. SVR is applied to up-to-the-minute prices in the Chinese
 150 market in [Qu and Zhang \(2016\)](#)'s study. Other studies have applied prediction techniques to even higher
 151 frequency data. [Brownlees and Gallo \(2006\)](#), for example, suggest the use of HFT on immediate price
 152 variations, that is, on ticks that vary according to each purchase or sale transaction performed.

153 The articles used in the construction of the following prediction models are shown in Tab. 1. The main
 154 prediction methods for each article and frequency of empirical data are also shown in the table. Note that
 155 most of the studies analysed apply a daily data model, highlighting the need to produce research with
 156 higher data frequencies. Therefore, as stated earlier, this study applies SVR prediction models not only to
 157 daily data but also to data with an up-to-the-minute frequency.

Please insert Tab. 1 here.

159 3. Method

160 The prediction of closing stock prices in this study is performed by SVR. The results are compared to
 161 returns obtained by the random walk model, assuming zero average returns on stocks and a given variance.
 162 The Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used to evaluate
 163 the adequacy of the models' price predictions. These measures have also been used by [Nayak et al. \(2015\)](#),
 164 [Patel et al. \(2015b\)](#), [Araújo et al. \(2015\)](#), [Manahov et al. \(2014\)](#) and [Choudhury et al. \(2014\)](#). Based on these
 165 authors' studies, the MAPE and RMSE can be calculated according to Eqs 1 and 2, respectively.

$$\text{MAPE} = \frac{1}{T} \sum_{i=1}^T \left| \frac{d_i - \hat{d}_i}{d_i} \right| \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{i=1}^T (d_i - \hat{d}_i)^2} \quad (2)$$

166 The following notations were adopted for the variables used in this study:

- 167 • N : total samples;
- 168 • T : total test samples;
- 169 • d : real sample value;
- 170 • \hat{d} : value estimated by the model;
- 171 • P : period (minutes or days);
- 172 • Cl : closing price for the period;
- 173 • Hi : maximum price in the period;
- 174 • Lo : minimum price in the period;
- 175 • Up : number of price rises;
- 176 • Dw : number of price reductions;
- 177 • r : return.

178 3.1. Technical analysis indicators

179 The predictor variables commonly used in the literature for SVR and SVM models are TA indicators
 180 [Kumar et al. \(2016\)](#); [Dash and Dash \(2016\)](#); [Gerlein et al. \(2016\)](#); [Hsu et al. \(2016\)](#); [Patel et al. \(2015b\)](#); [Huang](#)
 181 [and Tsai \(2009\)](#); [Kim \(2003\)](#). According to [Nayak et al. \(2015, p. 672\)](#), a TA indicator is composed of data
 182 derived from the application of a certain formula to the past prices of a stock. This paper considers the
 183 values of these indicators, detailed below, as predictive variables in the SVR model.

184 The simplest selected indicators are the a moving averages, which are easy to understand and calculate.
 185 The Simple Moving Average (SMA) is the arithmetic mean of T past prices Cl_i , according to [Nayak et al.](#)
 186 [\(2015, p. 672\)](#) and as described by Eq. 3.

$$\text{SMA} = \frac{1}{T} \sum_{i=1}^T Cl_i \quad (3)$$

187 One moving average variation, known as the Weighted Moving Average (WMA), used in this study,
 188 assigns higher weightings to more recent prices. A WMA of P periods is given by Eq. 4, as shown in [Patel](#)

189 et al. (2015b, p. 2164). This study defines Cl_i as either daily closing or up-to-the-minute prices, depending on
 190 the frequency used in the analysis.

$$\text{WMA} = \frac{P Cl_i + (P - 1) Cl_{i-1} + \dots + Cl_{i-P}}{P + (P - 1) + \dots + 1} \quad (4)$$

191 The Relative Strength Index (RSI) is a comparison indicator between losses and recent gains and deter-
 192 mines an overbought or oversold market. At time t , it has the form of Eq. 5 (Patel et al., 2015b, p. 2164). In
 193 this study, the RSI is calculated for the stock's closing prices in the selected period.

$$\text{RSI} = 100 - \frac{100}{1 + \frac{\sum_{i=0}^{P-1} Up_{t-i}/n}{\sum_{i=0}^{P-1} Dw_{t-i}/n}} \quad (5)$$

194 The Accumulation/Distribution Oscillator (ADO) is an indicator of momentum, i.e., the strength of a
 195 price trend Hsu et al. (2016, p. 221), and is given by Equation 6.

$$\text{ADO} = \frac{Hi_t - Cl_{t-1}}{Hi_t - Lo_t} \quad (6)$$

196 The final TA indicator considered in this study is the Average True Range (ATR). This indicator seeks to
 197 measure bull and bear price trends, using the True Range (TR), defined in Eq. 7. According to Nayak et al.
 198 (2015, p. 672), the ATR represents a mean, given by Eq. 8, in which the first value of the period is given by
 199 $1/P \sum_{i=1}^P TR_i$.

$$\text{TR} = \max[Hi_t - Lo_t, |Hi_t - Cl_{t-1}|, |Lo_t - Cl_{t-1}|] \quad (7)$$

$$\text{ATR}_t = \frac{\text{ATR}_{t-1}(P - 1) + \text{TR}_t}{P} \quad (8)$$

200 3.2. Support Vector Regression

201 A classification method based on SVM maps the independent variables of N samples available into
 202 a space of more dimensions and is typically used to classify observations between groups. This method,
 203 developed by Vapnik (1995), uses $\{(\mathbf{x}_k, y_k)\}_{k=1}^N$ training observations to build a linear model using non-linear
 204 classification thresholds, mapping variables on a greater number of dimensions. Separation between classes
 205 is achieved using an optimal hyperplane, calculated based on N observations, where \mathbf{x} is the independent
 206 variable vector, and y is classification $y_k \in \{-1, 1\}$ for each sample. Thus, the classification hyperplane is
 207 given by Eq. 9, which satisfies the conditions of Eqs 10 and 11.

$$\mathbf{w}^T \phi(\mathbf{x}_k) + b = 0 \quad (9)$$

$$\mathbf{w}^T \phi(\mathbf{x}_k) + b \geq 1 \text{ para } y_k = 1 \quad (10)$$

$$\mathbf{w}^T \phi(\mathbf{x}_k) + b \leq -1 \text{ para } y_k = -1 \quad (11)$$

208 In the context of price prediction, the goal is not necessarily classification into groups but estimation of
 209 real values. This study therefore uses SVR, which is employed to obtain a regression model used to predict
 210 asset prices. The SVR model in this paper includes the following notations:

- 211 • \mathbf{x} : vector with predictor variables;
- 212 • y : sample classification;
- 213 • \mathbf{w} : weight vector;
- 214 • b : constant;
- 215 • C, c : model parameters.

216 The function $\phi(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_k}$ is the said mapping of independent variables to a space with a greater
 217 number of dimensions, in which it is possible to linearly separate the samples according to each class. Eq. 9
 218 and its conditions in Eqs. 10 and 11 are summarised in Eq. 12.

$$y_k[\mathbf{w}^T \phi(\mathbf{x}_k) + b] \geq 1 \quad (12)$$

219 The classification condition originally proposed by Vapnik (1995) is $y(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \phi(\mathbf{x}) + b)$. However,
 220 even in the new space mapped by the function $\phi(\cdot)$, there may not be a perfect separation of N samples into
 221 two classes $\{-1, 1\}$. Thus, we define a variable $\xi \geq 0$ as a tolerance margin in the classification thresholds,
 222 making the classifier more flexible in accepting possible errors. With this flexibilisation, the hyperplane
 223 condition in Equation 12 becomes Eq. 13, and the problem of finding the optimal hyperplane becomes a
 224 convex optimisation problem given by Eq. 14. In this equation, C is the adjustment parameter for the edge
 225 of the hyperplane with the smallest possible misclassification, under the conditions of Eq. 13.

$$y_k[\mathbf{w}^T \phi(\mathbf{x}_k) + b] \geq 1 - \xi_k \quad (13)$$

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{k=1}^N \xi_k \quad (14)$$

226 The above formulated optimisation can be converted into Wolfe (1961)'s dual, given by Eq. 15, in which
 227 $\mathbf{y}^T \boldsymbol{\alpha} = 0$ e $0 \leq \alpha_i \leq C, i = 1, \dots, N$. In this model, $\mathbf{e} = [1, \dots, N]^T$ represents a vector of unit values, and

228 Q is a l by l matrix in which $Q_{ij} \equiv y_i y_j K(x_i, x_j)$. The function $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is called the kernel
 229 function.

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - \mathbf{e}^T \alpha \quad (15)$$

230 The solution of Wolfe (1961)'s dual function turns the classification decision function into the form of Eq.
 231 16.

$$\text{sgn}(w^T \phi(x) + b) = \text{sgn} \left(\sum_{i=1}^N y_i \alpha_i K(x_i, x) + b \right) \quad (16)$$

232 As already indicated, SVR uses principles similar to SVM, but the response variable is a continuous
 233 value $y \in \mathbb{R}$. However, as shown by Huang and Tsai (2009, p. 1530) and Patel et al. (2015b, p. 2164), instead
 234 of seeking the hyperplane in Eq. 13, SVR seeks the linear regression function, given by Eq. 17. To achieve
 235 this, a threshold error ϵ is defined to be minimised in the expression in Equation 18. This expression is called
 236 the ϵ -insensitivity loss error function. The SVR regression process therefore seeks to minimise ϵ in Eq. 18
 237 and $\|\mathbf{w}\|^2$ in the expression of R , defined in Eq. 19.

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + b \quad (17)$$

$$|y - f(\mathbf{x}, \mathbf{w})|_{\epsilon} = \begin{cases} 0, & \text{case } |y - f(\mathbf{x}, \mathbf{w})| \leq \epsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \epsilon, & \text{otherwise} \end{cases} \quad (18)$$

$$R = \frac{1}{2} \|\mathbf{w}\|^2 + c \left(\sum_{i=1}^N |y_i - f(\mathbf{x}_i, \mathbf{w})|_{\epsilon} \right) \quad (19)$$

238 Tolerance variables are again introduced, defining ζ as the value in excess of ϵ and ζ^* to limit the value
 239 to the regression target. Thus, the minimisation of Eq. 19 becomes Equation 20, under the conditions of Eqs.
 240 21 and 22 for ζ_i and $\zeta_i^* \geq 0$ and $i = 1, 2, \dots, N$.

$$R = \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{i=1}^N (\zeta_i + \zeta_i^*) \quad (20)$$

$$(\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \epsilon + \zeta_i \quad (21)$$

$$y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \epsilon + \zeta_i^* \quad (22)$$

241 This paper considers common forms of kernel functions from the literature Huang and Tsai (2009); Nayak

242 et al. (2015); Patel et al. (2015b); Ballings et al. (2015), given explicitly by the linear, radial and polynomial
 243 functions in Eqs 23, 24 and 25, respectively.

$$K(x_i, x_j) = x_i^T x_j \quad (23)$$

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \text{ para } \gamma > 0 \quad (24)$$

$$K(x_i, x_j) = (x_i^T x_j + 1)^d \quad (25)$$

244 Note that the shape of the kernel function directly influences the values obtained by the SVR regression.
 245 Similarly, the constant c in Eq. 19 and the parameters γ and d in Eqs. 24 and 25 should be optimised. For this
 246 purpose, a training data set is divided into two new sets: the first is used to choose the optimal parameters,
 247 and the second is used to validate the smallest error possible, given these choices. This process, called k-fold
 248 cross validation, selects the parameters c , γ and d , according to the lowest RMSE.

249 To illustrate the general idea of price prediction, Figure 1 visually demonstrates the difference between
 250 returns prices observed in the market and those predicted by the SVR model. The data used in this
 251 illustration are daily and are not used in the following experiments. As seen in Figure 1, the SVR model
 252 implies considerable prediction error, and the literature seeks to minimise these errors, as in this study.

 Please insert Fig. 1 here.

254 3.3. Random walk

255 According to Malkiel and Fama (1970, p. 383), prices in efficient markets should reflect all information
 256 available at a particular time. Any price changes are difficult to predict and should be independent of
 257 previous values. Following a basic approach to efficiency analysis, returns must be independent and
 258 identically distributed. It should be noted that this study only investigates price unpredictability, using a
 259 random walk model. Thus, if the price of a stock market asset at time t follows a random walk, as described
 260 in Araújo et al. (2015, pp. 4083–4084), the price behaviour must be based on Equation 26, with return r at
 261 time t , and follow a normal distribution, with zero mean and a variance of σ^2 ($r \sim \mathcal{N}(0, \sigma^2)$), i.e., white
 262 noise. With this return distribution, the market would be unpredictable and efficient enough not to allow
 263 the formation of profitable transactional strategies on the long-term.

$$Cl_t = Cl_{t-1} + r_t \quad (26)$$

264 This study uses random walk predictions as a reference model for comparison with SVR predictions,
 265 as set out by Lu et al. (2009). According to Eq. 26, the model predicts that the next stock price will be the
 266 current price plus a random value taken from the distribution $\mathcal{N}(0, \sigma^2)$.

267 3.4. Data

268 This study considers Brazilian, American and Chinese stocks, with three blue chip and three small cap
 269 stocks for each country, totaling 18 assets. The stocks were chosen to obtain a distribution of companies of
 270 different sizes in different markets of both developed and developing nature. The stocks selected in this
 271 study are shown in Tab. 2. The time period selected for daily prices comprises 15 years. It should be noted,
 272 as Tab. 2 shows, some of the stocks do not have all those historical prices available, specially some small
 273 capitalization companies. However, most stocks selected have more than 10 years of historical daily prices,
 274 including bull and bear periods.

275 Due to processing costs and data availability, 1-minute historical prices were constrained as detailed in
 276 Tab. 3. The data period for 1-minute prices for all stocks is from March 1 to May 26, 2017, i.e., three months
 277 of up-to-the-minute prices, a window of analysis selected following Manahov et al. (2014) high frequency
 278 study. This period is used to compare up-to-the-minutes predictions across markets and stocks regarding
 279 their capitalisation, in both fixed trained SVR models and dynamically updated ones. A longer period of
 280 historical 1-minute prices is also considered for six stocks, limited to one market because of processing
 281 costs limitations. Brazilian stocks, therefore, are selected for this longer period of testing, as a means to
 282 evaluate SVR predictions stability over the long run in high frequency trading. As shown by Tab. 3, 2 years
 283 of 1-minute prices are used for that case. Data were obtained from Reuters[©], Yahoo!Finance and BM&F
 284 Bovespa. Minutes without information were considered to have the same prices as the previous minute.

285 Please insert Tab. 2 here.

286 To compare the SVR prediction results with regard to different price frequencies, daily and up-to-the-
 287 minute data are used for the periods described above. Daily prices are widely used in academic studies Tay
 288 and Cao (2001); Kim (2003); Huang and Tsai (2009); Lu et al. (2009); Yeh et al. (2011); Choudhury et al. (2014);
 289 Patel et al. (2015b). The results obtained with daily prices can be compared to the use of higher frequency,
 290 up-to-the-minute prices. It is worth mentioning that valid prices in this study include only those obtained
 291 during the respective sessions in each market considered. Thus, the data were limited in advance to the
 292 official opening and closing times of each market. It should be noted that high frequency data, for example,
 293 those expressed in milliseconds, are not within the scope of this study, due to the method employed.

294 Two strategies are considered when using SVR. The first is to separate the data into training observations,
 295 for the optimisation of the SVR model, and test observations, with prediction errors calculated on the closing

296 prices, as adopted by Gerlein et al. (2016), Manahov et al. (2014), Nayak et al. (2015) and Patel et al. (2015b).
 297 For daily prices, each period in Tab. 2 is split in a training set, with approximately 70% contiguous days, and
 298 a test set, with the remaining 30% days. For the 1-minute historical prices, Tab. 3 shows the divisions into
 299 training and testing periods for all studied stocks, following the same strategy of separating 70% of data for
 300 training the models and 30% for testing them.

301 Please insert Tab. 3 here.

302 The second SVR optimisation strategy involves updating the model as new information becomes
 303 available in the market, as suggested by Lessmann et al. (2011, p. 2122) and Hsu et al. (2016, p. 223). It is a
 304 dynamic optimisation that uses periodically updated training data, known as a sliding or moving window.
 305 It is expected that this procedure will capture new market conditions as soon as possible.

306 Regarding the up-to-the-minute prices, although three months seems a short interval, it should be
 307 noted that period comprises more than 33000 data points, making the task of obtaining and processing the
 308 prices for all 18 selected securities a challenging effort. However, a longer period of analysis is desirable,
 309 mainly because three months may not include all possible market conditions for real tests. In this context,
 310 as stated before, this article brings yet another SVR evaluation for price prediction, using 2 whole years
 311 of up-to-the-minutes prices. Brazilian stocks are selected for this simulation over the long run, with data
 312 gathered directly from BM&F Bovespa. Such a long period, in terms of 1-minute prices, contains short-term
 313 bull and bear markets for this timeframe, being suitable for evaluating the models' predictions stability.

314 4. Analysis and results

315 Before applying SVR to the prices described in this paper, prices and the TA indicators highlighted above
 316 – namely, SMA, WMA, RSI, ADO and ATR – were calculated and normalized. A calculation period of $P = 10$
 317 was fixed for all indicators. The SVR prediction model transactions for the up-to-the-minute data prices
 318 thus start just 10 minutes after the beginning of each trading session. The analysis of results are organised
 319 into two sections, one dedicated to the use of a fixed training period as described before and other section
 320 dedicated to constantly updating the model in a moving training window. Both these sections consider the
 321 15 years historical daily prices, the 3 months period for up-to-the-minutes prices and the results of predicting
 322 prices using 2 years of 1-minute Brazilian stock data. Finally, a dedicated section introduces a correlation
 323 study between average returns and volatility and the SVR prediction precision.

324 4.1. Fixed training

325 Once the TA indicators were calculated and normalized, the SVR models were optimised for the fixed
 326 training prices described in Tab. 3, for the 1-minute prices, and the 70% first contiguous available daily

327 prices shown in Tab. 2. The optimum parameters for each kernel function, according to Eqs. 19, 24 and
 328 25, are given in Tab. 4, in the case of daily prices, in Tab. 5, in the case of up-to-the-minute prices for the
 329 3-months period and finally in Tab. 6 for the 2-years 1-minute prices period. Each table also shows the RMSE
 330 associated with optimisation, using training data.

331 Applying the parameters in Tab. 4, 5 and 6, the SVR models with each kernel function were run on
 332 daily and up-to-the-minute price test sets, respectively, according to the periods shown in Tab. 2 and 3.
 333 The resulting error in RMSE and MAPE for each stock is given in Tab. 7 for daily prices, in Tab. 8 for
 334 up-to-the-minute prices during the 3-months data period and in Tab. ?? for the 2-years period of 1-minute
 335 prices. In most cases, almost all of the RMSE errors are greater in the test data. This behavior is expected, as
 336 the models are optimised for the training data. However, there are exceptions, especially with the use of the
 337 linear kernel for all periods and price timeframes considered.

338 Comparison of Tab. 4 and 7 reveals that the daily test data SVR had smaller errors, compared to those
 339 RMSE errors on the optimization phase on the daily training data for the following stocks: PETR4, VALE5,
 340 DIRR3, BAC, ANGI, HL, 601318, 1970 and 2030. When using daily prices, the use of radial kernel resulted in
 341 smaller errors for the test data set only in the case of BAC stock. Usage of polynomial kernels did not, for
 342 any stock, result in smaller errors in the test data than in the training data.

343 Up-to-the-minute test data frequencies, for the 3-months period data, are compared in Tab. 5 and 8. In
 344 this case, the application of SVR with a linear kernel to the test data produces smaller errors than the same
 345 SVR applied to training data for the following stocks: BBAS3, PETR4, VALE 5, ALPA4, DIRR3, LEVE3, XOM,
 346 600028, 601318, 601668, 1432, 1970, and 2030. Interestingly, the application of the linear kernel to the test
 347 data resulted in a smaller error than its application to the training data for all Brazilian and Chinese sample
 348 stocks. When using the longer period of training/testing of 2-years of 1-minute prices, that behavior is not
 349 observed for BBAS3, DIRR3 and LEVE3, which can be noted by comparing Tab. 6 and 9.

350 The results in Tab. 7, 8 and 9 are indicative of the SVR's superior predictive power when using a linear
 351 kernel compared with radial and polynomial kernels, for this study's selected stocks. To measure the
 352 significance of these results, the errors produced were compared with the errors produced by a random
 353 walk model. To that end, random predictions were generated according to the model given by Eq. 26. To
 354 model white noise in that equation, the distribution variance was estimated according to price frequency.
 355 The variance in daily prices was therefore estimated using the closing prices of the most recent 7 days.
 356 Similarly, the 90 most recent minutes were used to estimate the variance in the distribution of returns for
 357 up-to-the-minute prices. Based on this procedure, the RMSE and MAPE results from obtaining the random
 358 walk model's predictions of daily closing and up-to-the-minute prices are as shown in Tab. 10.

359 Comparison of errors produced by the SVR and random walk models for the fixed training period with
 360 respect to daily prices, shown in Tab. ?? and 10, respectively, reveals that the SVR model with a linear kernel
 361 has superior predictive power to the random walk model only for the following stocks: DIRR3, BAC, HL,

1970, and 2030. This means that the RMSE and MAPE error measures are smaller for the linear kernel SVR model for these stocks than those obtained for daily price predictions generated by a random walk model. Such stocks are exemplars and have common characteristics, most being classified as small caps in the three studied countries, as shown in Tab. 2, exception being the BAC stock. For the SVR with radial and polynomial kernels, no stocks had the prices predicted better than with the random model in the case of daily prices. Although by few stocks, these results also indicate superior predictive power using a simple linear kernel.

The errors of the SVR models with a fixed training period with regard to up-to-the-minute prices, shown in Tab. 8 and 9, were also compared with those of the random walk model, shown in Tab. 10. The fixed training SVR results for up-to-the-minute prices reveal that this model does not have superior predictive power to a random prediction model for almost any of the selected stocks. Except for the application of the linear kernel to the Chinese stock, 1432, all the errors produced in the SVR predictions, measured by RMSE, were greater than the errors produced by the random walk model for the up-to-the-minute closing prices with fixed training and test periods. In the case of a fixed training period, therefore, the SVR model's predictive power is only evidenced for daily prices, while the model is ineffective for higher frequency, up-to-the-minute periods.

4.2. Moving training window

Having recorded the closing price prediction errors for the SVR models with fixed training and test sets, we turn to an examination of the strategy of constantly updating the models. In this study, the frequency selected for updating the models was the extreme case of an update made whenever a new price becomes available. Thus, for daily prices, the model was updated every day, and the next day's closing price served as a test observation. The daily prices of the 7 most recent days were selected for training, leaving the closing price on the 8th day for the prediction test. Similarly, for up-to-the-minute prices, the model was updated every minute, and the next minute's closing price served as a test observation. The previous 90 minutes' prices were used for training, leaving the 91st closing price for the prediction test. The SVR parameters, i.e., c , γ and d , were fixed as the optimum values obtained in Tab. 4 for daily prices, Tab. 5 for up-to-the-minute prices in the 3-months historical data period and Tab. 6 for the 2-years period.

The results obtained when periodically updating the model are recorded in Tab. 11 (for daily prices) and 12 (for up-to-the-minute prices in the 3-months period). The Brazilian cases selected for the 2-years 1-minute price prediction study using the moving training window SVR strategy are reported in Tab. 13. These results were measured in RMSE and MAPE and compared with the results obtained by the random walk model. Therefore, comparing the results in Tab. 10 and 11, smaller errors were observed in the SVR model predictions, using linear and radial kernels for virtually all selected stocks, regardless of country of origin or capitalisation, for daily prices. The exception was LEVE3 stock, which presented odd results,

396 possibly due to data errors. Moreover, on average, errors measured by MAPE were reduced when the linear
 397 kernel was used.

Please insert Tab. 4 here.

Please insert Tab. 5 here.

Please insert Tab. 6 here.

Please insert Tab. 7 here.

Please insert Tab. 8 here.

Please insert Tab. 9 here.

Please insert Tab. 10 here.

405 Tab. 12 and 10 allow for a comparison of the SVR model with the random walk model for up-to-the-
 406 minute prices during the 3-months period proposed. In this case, attention is drawn to the errors obtained
 407 using the linear kernel for the US stocks, GOOGL, XOM and ANGI. In these cases, the SVR model had
 408 no predictive power for the up-to-the-minute prices selected for this study possibly due to data errors.
 409 However, for the other stocks, the linear kernel returned smaller errors than the random walk model. Use of
 410 the radial kernel returned smaller errors than the random model for all stocks, regardless of the country of
 411 origin or capitalisation, as shown in Tab. 2. Finally, the use of SVR with a polynomial kernel had predictive
 412 power only for the American small cap stock, PZZA.

413 Results of 1-minute price prediction in a training moving window fashion for Brazilian stock for a 2-years
 414 period are shown in Tab. 13. They are used as indicative SVR prediction power over the long run in the
 415 up-to-the-minutes prices timeframe. Comparing those results with the predictions obtained with the random
 416 walk model of Tab. 10, SVR updated regularly confirms its predictive power observed in the 3-months cases
 417 for almost all stocks, specially using the linear kernel. LEVE3 stands out as having results worse than using
 418 the random walk model.

419 4.3. SVR prediction models and stocks volatility

420 After registering SVR prediction results and comparing them with the random walk model-generated
421 predictions in the previous paragraphs, this section verifies possible relationships between stocks basic
422 statistics and the predictions themselves. Specifically, the daily SVR predictions using the moving training
423 window are split yearly for evaluation against prices statistics. Two measures are examined for each of the
424 15-years historical prices per stock: average returns and volatility, calculated as the standard deviation of
425 closing prices. Then we tabulate the correlations between yearly volatility and RMSE values obtained by
426 SVR prediction models. Correlations are also calculated between the RMSE values and average daily return.
427 Results are given in Tab. 14.

428 Examining the correlation values from average returns and RMSE errors, i.e. the left-most columns of
429 Tab. 14, it is not possible to draw any conclusions about trends and relationships. Most of those correlation
430 values are close to zero, exception being BAC and 1970. The American blue ship stock presented a relatively
431 negative correlation between average return and SVR predicted price. That can indicate SVR predictions,
432 considering constantly updated models, may be more precise during periods with larger daily returns for
433 the BAC stock only. In the case of Chinese small cap 1970, we refrain from any comment, since the historical
434 period considered for that specific stock, as shown by Tab. 2, is too short in aggregate years data.

435 Although the average returns do not seem related to the SVR predictions, there are indications of a
436 strong relationship between the prediction errors and volatility. That can be observed for almost all the
437 cases in the right- most columns of Tab. 14. Specially for the SVR predictions using linear kernel, correlation
438 values are close to one for virtually all stocks, exception being LEVE3, which seems to contain too many
439 historical data inconsistencies. A closer look in our data for that stock reveals heavy occurrences of missing
440 data. However, apart from that case, SVR predictions, considering constantly updated models using linear
441 and radial kernels, seem to be more precise during periods with lower volatility in prices.

442 5. Conclusion

443 Developing predictive price models for the stock market is challenging, but it is an important task
444 when building profitable financial market transaction strategies. Computationally intensive methods, using
445 past prices, are developed to facilitate better management of market risk for investors and speculators. Of
446 the machine learning techniques available, this study uses SVR and measures its performance on various
447 Brazilian, American and Chinese stocks with different characteristics, for example, small cap or blue chip.
448 The predictive variables are calculated using TA indicators on asset prices. The results show the magnitude
449 of the mean squared errors for the three common kernels in the literature, using specific algorithm training
450 strategies with different price frequencies of days and minutes. The results are contrasted with those of a
451 random walk-based model.

452 This study shows that using a fixed training set on daily prices, it is possible to obtain smaller prediction
453 errors in the test set than in the training set when using a linear kernel. Moreover, this kernel was more
454 adequate for price predictions than the radial and polynomial kernels in the case of daily prices and fixed
455 training models and outperformed the random model for some stocks classified as blue chips and small
456 caps in the three studied countries. However, increasing the price frequency to minutes reduced the model's
457 predictive power using a fixed training period. In particular, SVR obtained inferior predictive results relative
458 to a random walk model for almost all stocks studied in up-to-the-minute prices, using fixed training,
459 regardless of the adopted kernel function.

460 The periodically updated models provided important evidence. In these cases, the use of linear and
461 radial kernels resulted in smaller errors than the random walk model for almost all daily stock prices. The
462 only exception was a stock with a high missing data rate. Constant model updating was also beneficial in
463 the up-to-the-minute price frequency, and SVR models with linear and radial kernels achieved better results
464 than the random walk model when this strategy was used. To emphasize the stability of the predictions
465 over the long run, we processed a 2-years up-to-the-minute prices period for the selected Brazilian stocks,
466 confirming better results with a constantly updated model. The analyses presented in this study suggest
467 that periodically updating the SVR model reduces the mean square error compared to using a rigid model
468 without periodic updating. This result contrasts with that of [Hsu et al. \(2016\)](#), who did not achieve better
469 performance when using a sliding window on the training data.

470 An important contribution of this study is a comparison of price prediction results of the presented
471 SVR models with those of the random walk model, according to which markets are unpredictable in the
472 long term. In this respect, the results presented here show that some SVR models, with periodic or fixed
473 updates, may achieve better than random predictive performance, especially with the use of the linear
474 kernel. Another result which prompts further investigation is the indication of a strong relationship between
475 SVR price prediction and volatility, considering a moving training window.

476 Importantly, despite the evidence of asset price predictability presented here, this article does not propose
477 transactional strategies applicable to the stock market. The results therefore do not directly refute the EMH.
478 Given that the focus of the analysis is not the identification of purchasing or sales strategies that allow for
479 extraordinary gains, the study does not address issues such as transaction costs or portfolio risk levels.

480 As the focus of the study is the analysis of asset price prediction errors, it is possible to build risk
481 management models using SVR-based estimates. Exposure limits may be obtained by evaluating model
482 errors. This study therefore provides a basis for the construction of systems that, while not directly evaluating
483 the EMH, make possible the study of market efficiency and risk analysis. This study obtained results using
484 SVR that were better than those of a null mean return random model.

485 Despite comparing daily rates with the use of high frequency up-to-the-minute trading, this paper
486 considers only a predictive algorithm based on machine learning. Furthermore, the SVR model allows for

487 testing of many kernel functions, while this study is limited to only the three most common in the literature.
 488 It should be noted that, for a more robust simulation of high frequency stock market strategies, it would be
 489 necessary to include transaction costs, communications network delays, differentiation between the market
 490 price and actual value of a purchase or sale transaction (slippage) and transaction liquidity.

491 This article has the limitation of not considering data quality assurance methods. Some results presented
 492 here suffer from poor data inputs and future studies should consider data treatment before usage. The
 493 present research approach equates missing minute prices to the previous values, not considering contiguous
 494 missing minutes or interpolating values. Some of the selected stocks illustrate the influence over the results
 495 of long streaks of missing data as well as outliers.

496 Another limitation of these research results is the length of the periods of historical prices considered,
 497 specially the 3-months up-to-the-minutes prices data. Although the 3-months period seem short compared
 498 to the 15 years daily historical prices data, it should be noted it contains nearly 33000 data points per
 499 stock, compared to the approximately 3700 data points for the daily period selected, posing a challenging
 500 processing task. For future reference, all processing of the 15 years daily data, 3-months 1-minute data
 501 and the 2-years 1-minute Brazilian stock data took about 15 hours per stock in a powerful machine, with
 502 24 processors type Intel[®] Xeon[®] CPU E5-2650 v4 @ 2,20 GHz with 226GB of RAM, running Linux 3.10.0,
 503 distribution CentOS 7.3. The implementations of SVR and related functions used in this research are scripted
 504 in the R[®] statistical language, version 3.4.1, using functions from *e1071* and *caret* packages. To reduce
 505 computation time, we took advantage of the parallel capabilities of the computer environment, allocating
 506 each stock simulations to an exclusive processor. Therefore, careful considerations are necessary for any real
 507 trading implementation attempts.

508 Future studies may include a larger number of test stocks and markets other than those selected
 509 here. Other predictive models could also be compared, including classifiers of the directions of asset
 510 prices. Independent variables may include other TA indicators, trend predictors or past prices. In addition,
 511 fundamental analysis indicators, such as company size, liquidity, indebtedness, profitability and activity
 512 measures, could be included. The inclusion of such data could improve the machine learning mechanism.
 513 It is also recommended that other model updating periodicities be tested, especially those with higher
 514 frequency than up-to-the-minute prices.

515 Please insert Tab. 11 here.

516 Please insert Tab. 12 here.

517

Please insert Tab. 13 here.

518

Please insert Tab. 14 here.

519 **Bibliography**

- 520 Araújo, R., Oliveira, A. L., Meira, S., 2015. [A hybrid model for high-frequency stock market forecasting](#). *Expert Systems with*
521 *Applications* 42 (8), 4081–4096.
- 522 Ballings, M., Van den Poel, D., Hespels, N., Gryp, R., 2015. [Evaluating multiple classifiers for stock price direction prediction](#). *Expert*
523 *Systems with Applications* 42 (20), 7046–7056.
- 524 Barak, S., Modarres, M., 2015. [Developing an approach to evaluate stocks by forecasting effective features with data mining methods](#).
525 *Expert Systems with Applications* 42 (3), 1325–1339.
- 526 Brownlees, C., Gallo, G., 2006. [Financial econometric analysis at ultra-high frequency: Data handling concerns](#). *Computational Statistics*
527 *& Data Analysis* 51 (4), 2232–2245.
- 528 Choudhury, S., Ghosh, S., Bhattacharya, A., Fernandes, K. J., Tiwari, M. K., 2014. [A real time clustering and SVM based price-volatility](#)
529 [prediction for optimal trading strategy](#). *Neurocomputing* 131 (1), 419–426.
- 530 Dash, R., Dash, P. K., 2016. [A hybrid stock trading framework integrating technical analysis with machine learning techniques](#). *The*
531 *Journal of Finance and Data Science* 2 (1), 42–57.
- 532 Gerlein, E. A., McGinnity, M., Belatreche, A., Coleman, S., 2016. [Evaluating machine learning classification for financial trading: An](#)
533 [empirical approach](#). *Expert Systems with Applications* 54 (1), 193–207.
- 534 Goldstein, M. A., Kumar, P., Graves, F. C., 2014. [Computerized and High-Frequency Trading](#). *Financial Review* 49 (2), 177–202.
- 535 Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., Johnson, J. E., 2016. [Bridging the divide in financial market forecasting: machine learners](#)
536 [vs. financial economists](#). *Expert Systems with Applications* 61 (1), 215–234.
- 537 Huang, C.-L., Tsai, C.-Y., 2009. [A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting](#). *Expert Systems*
538 *with Applications* 36 (2), 1529–1539.
- 539 Kao, L.-J., Chiu, C.-C., Lu, C.-J., Yang, J.-L., 2013. [Integration of nonlinear independent component analysis and support vector](#)
540 [regression for stock price forecasting](#). *Neurocomputing* 99 (1), 534–542.
- 541 Kim, K., 2003. [Financial time series forecasting using support vector machines](#). *Neurocomputing* 55 (1-2), 307–319.
- 542 Kumar, D., Meghwani, S. S., Thakur, M., 2016. [Proximal support vector machine based hybrid prediction models for trend forecasting](#)
543 [in financial markets](#). *Journal of Computational Science* 17 (1), 1–13.
- 544 Lee, E. J., 2013. [High frequency trading in the Korean index futures market](#). *Journal of Futures Markets* 35 (1), 31–51.
- 545 Lessmann, S., Sung, M.-C., Johnson, J. E. V., 2011. [Towards a methodology for measuring the true degree of efficiency in a speculative](#)
546 [market](#). *J Oper Res Soc* 62 (12), 2120–2132.
- 547 Li, H., Sun, J., Li, J.-C., Yan, X.-Y., 2012. [Forecasting business failure using two-stage ensemble of multivariate discriminant analysis and](#)
548 [logistic regression](#). *Expert Systems* 30 (5), 385–397.
- 549 Li, S., Shiue, W., Huang, M., 2006. [The evaluation of consumer loans using support vector machines](#). *Expert Systems with Applications*
550 30 (4), 772–782.
- 551 Lu, C.-J., Lee, T.-S., Chiu, C.-C., 2009. [Financial time series forecasting using independent component analysis and support vector](#)
552 [regression](#). *Decision Support Systems* 47 (2), 115–125.
- 553 Malkiel, B. G., 2003. [The Efficient Market Hypothesis and Its Critics](#). *Journal of Economic Perspectives* 17 (1), 59–82.

- 554 Malkiel, B. G., Fama, E. F., 1970. [Efficient Capital Markets: a Review of Theory and Empirical Work](#). *The Journal of Finance* 25 (2),
555 383–417.
- 556 Manahov, V., Hudson, R., Gebka, B., 2014. [Does high frequency trading affect technical analysis and market efficiency? And if so, how?](#)
557 *Journal of International Financial Markets, Institutions and Money* 28 (1), 131–157.
- 558 Nayak, R. K., Mishra, D., Rath, A. K., 2015. [A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark](#)
559 [indices](#). *Applied Soft Computing* 35 (1), 670–680.
- 560 Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015a. [Predicting stock and stock price index movement using Trend Deterministic Data](#)
561 [Preparation and machine learning techniques](#). *Expert Systems with Applications* 42 (1), 259–268.
- 562 Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015b. [Predicting stock market index using fusion of machine learning techniques](#). *Expert*
563 [Systems with Applications 42 \(4\), 2162–2172.](#)
- 564 Qu, H., Zhang, Y., 2016. [A New Kernel of Support Vector Regression for Forecasting High-Frequency Stock Returns](#). *Mathematical*
565 [Problems in Engineering 2016 \(1\), 1–9.](#)
- 566 Tay, F. E., Cao, L., 2001. [Application of support vector machines in financial time series forecasting](#). *Omega* 29 (4), 309–317.
- 567 Timmermann, A., Granger, C. W., 2004. [Efficient market hypothesis and forecasting](#). *International Journal of Forecasting* 20 (1), 15–27.
- 568 Vapnik, V. N., 1995. [The Nature of Statistical Learning Theory](#). Springer New York.
- 569 Wolfe, P., oct 1961. [A duality theorem for non-linear programming](#). *Quarterly of Applied Mathematics* 19 (3), 239–244.
- 570 Xiao, Y., Xiao, J., Lu, F., Wang, S., 2013. [Ensemble ANNs-PSO-GA Approach for Day-ahead Stock E-exchange Prices Forecasting](#).
571 *International Journal of Computational Intelligence Systems* 6 (1), 96–114.
- 572 Yeh, C.-Y., Huang, C.-W., Lee, S.-J., 2011. [A multiple-kernel support vector regression approach for stock market price forecasting](#).
573 *Expert Systems with Applications* 38 (3), 2177–2186.
- 574 Zbikowski, K., 2015. [Using Volume Weighted Support Vector Machines with walk forward testing and feature selection for the purpose](#)
575 [of creating stock trading strategy](#). *Expert Systems with Applications* 42 (4), 1797–1805.
- 576 Zhou, L., Lai, K. K., Yen, J., 2012. [Bankruptcy prediction using SVM models with a new approach to combine features selection and](#)
577 [parameter optimisation](#). *International Journal of Systems Science* 45 (3), 241–253.



Figure 1: Real returns and those predicted by the SVR model for daily prices. The continuous curve represents returns observed in the market, and the dashed curve represents returns predicted by the model.

Reference	Method	Data frequency
Tay and Cao (2001).	SVM.	Daily data.
Kim (2003).	SVM.	Daily data.
Lu et al. (2009).	SVR.	Daily data.
Yeh et al. (2011).	SVR.	Daily data.
Choudhury et al. (2014).	SVR.	Intraday data.
Araújo et al. (2015).	Based on neural networks.	Data in seconds.
Ballings et al. (2015).	SVM, KNN, neural networks.	Annual data.
Nayak et al. (2015).	SVM, KNN.	Daily data.
Patel et al. (2015b).	SVR, RF, neural networks.	Daily data.
Patel et al. (2015a).	SVR, RF, neural networks.	Daily data.
Zbikowski (2015).	SVM.	Daily data.
Dash and Dash (2016).	Neural networks.	Daily data.
Gerlein et al. (2016).	Multiple classifiers.	Daily data.
Hsu et al. (2016).	SVM, neural networks.	Intraday data.
Qu and Zhang (2016).	SVR.	Intraday data.
Kumar et al. (2016).	SVM.	Daily data.

Table 1: Prediction methods used by this study's references and data frequency.

Code	Company name	Country	Classification	Start date	End date
BBAS3.	<i>Banco do Brasil.</i>	Brazil.	Blueship.	01/05/2002	31/05/2017
PETR4.	<i>Petrobrás.</i>	Brasil.	Blueship.	02/05/2002	31/05/2017
VALE5.	<i>Vale do Rio Doce.</i>	Brazil.	Blueship.	01/05/2002	30/05/2017
ALPA4.	<i>Alpargatas.</i>	Brazil.	Small cap.	02/01/2007	29/05/2017
DIRR3.	<i>Direcional Engenharia.</i>	Brazil.	Small cap.	19/11/2009	26/05/2017
LEVE3.	<i>Metal Leve.</i>	Brazil.	Small cap.	01/05/2002	31/05/2017
BAC.	<i>Bank of America.</i>	USA.	Blueship.	01/05/2002	30/05/2017
GOOGL.	<i>Google.</i>	USA.	Blueship.	19/08/2004	30/05/2017
XOM.	<i>Exxon Mobil.</i>	USA.	Blueship.	01/05/2002	30/05/2017
ANGL.	<i>Angie's List.</i>	USA.	Small cap.	17/11/2011	26/05/2017
HL.	<i>Hecla Mining.</i>	USA.	Small cap.	01/05/2002	30/05/2017
PZZA.	<i>Papa Johns.</i>	USA.	Small cap.	01/05/2002	30/05/2017
600028.	<i>China Petroleum.</i>	China.	Blueship.	01/05/2002	31/05/2017
601318.	<i>Ping an Insurance.</i>	China.	Blueship.	08/05/2007	31/05/2017
601668.	<i>China State Construction Eng.</i>	China.	Blueship.	29/07/2009	26/05/2017
1432.	<i>China Shengmu Organic Milk.</i>	China.	Small cap.	15/07/2014	31/05/2017
1970.	<i>IMAX China Holding.</i>	China.	Small cap.	08/10/2015	31/05/2017
2030.	<i>Cabbeen Fashion.</i>	China.	Small cap.	28/10/2013	29/05/2017

Table 2: Selected stocks and historical periods considered in this article for daily prices.

Set	Period	Minutes	Days
3-months cases (all stocks):			
Train.	1/3/2017 to 2/5/2017	23555	39
Test.	5/5/2017 to 26/5/2017	10006	16
2-years cases (Brazilian stocks):			
Train.	10/2/2016 to 4/7/2017	123972	350
Test.	5/7/2017 to 8/2/2018	53132	150

Table 3: Fixed training and testing periods for the 1-minute prices.

Stock	Linear kernel		Radial kernel			Polynomial kernel		
	RMSE	c	RMSE	c	γ	RMSE	c	d
BBAS3	0,05745	1,00	0,05603	1,00	0,37132	0,05605	1,00	2,00
PETR4	0,07704	1,00	0,07856	1,00	0,35509	0,07343	1,00	2,00
VALE5	0,07368	1,00	0,07488	1,00	0,48782	0,06912	1,00	2,00
ALPA4	0,05421	1,00	0,05792	1,00	0,53674	0,05179	1,00	2,00
DIRR3	0,09526	1,00	0,09024	1,00	0,26466	0,09122	1,00	3,00
LEVE3	0,06949	1,00	0,10048	1,00	5,82473	0,07177	1,00	1,00
BAC	0,07108	1,00	0,07053	1,00	0,67116	0,06965	1,00	3,00
GOOGL	0,02864	1,00	0,02810	1,00	0,28154	0,02756	1,00	2,00
XOM	0,05099	1,00	0,05163	1,00	0,41009	0,04896	1,00	3,00
ANGI	0,13171	1,00	0,12078	1,00	0,45076	0,11700	1,00	3,00
HL	0,13238	1,00	0,12264	1,00	0,28910	0,12342	1,00	3,00
PZZA	0,01458	1,00	0,01452	1,00	0,47304	0,01370	0,25	3,00
600028	0,09759	1,00	0,09259	1,00	0,62462	0,08162	1,00	2,00
601318	17,3848	1,00	0,08261	1,00	0,63188	0,08073	1,00	3,00
601668	0,04796	1,00	0,05288	1,00	0,54471	0,04295	0,50	3,00
1432	0,18971	1,00	0,15882	1,00	0,35996	0,17126	1,00	3,00
1970	0,25036	1,00	0,22665	1,00	0,43222	0,21895	0,50	3,00
2030	0,10831	1,00	0,11371	1,00	1,07655	0,09542	0,50	3,00

Table 4: Optimal parameters and RMSE errors for each stock and kernel with daily prices.
Note: c : required parameter for all kernels; γ : Radial kernel parameter; d : Polynomial kernel parameter.

Stock	Linear kernel		Radial kernel			Polynomial kernel		
	RMSE	c	RMSE	c	γ	RMSE	c	d
BBAS3	0,05232	1,0	0,02098	1,0	0,37626	0,01644	1,00	2,0
PETR4	0,05704	1,0	0,03171	1,0	0,43427	0,02073	0,50	3,0
VALE5	0,05935	1,0	0,03744	1,0	0,42226	0,02168	0,50	3,0
ALPA4	0,05435	1,0	0,02998	1,0	0,51963	0,02147	0,50	3,0
DIRR3	0,05152	1,0	0,02748	1,0	0,42272	0,01988	1,00	3,0
LEVE3	0,05453	1,0	0,02376	1,0	0,33712	0,01893	0,25	3,0
BAC	0,03005	1,0	0,02766	0,5	1,36658	0,01473	0,25	2,0
GOOGL	0,00970	1,0	0,01770	1,0	5,80610	0,01017	1,00	2,0
XOM	0,00666	1,0	0,00715	1,0	2,37655	0,00375	1,00	3,0
ANGI	0,00953	1,0	0,02128	1,0	3,84291	0,00594	0,50	2,0
HL	0,03343	1,0	0,03927	1,0	1,43961	0,02188	0,25	1,0
PZZA	0,01143	1,0	0,01683	1,0	1,42533	0,01143	1,00	1,0
600028	0,08102	1,0	0,02319	1,0	0,40770	0,01705	1,00	2,0
601318	0,04970	1,0	0,01157	1,0	0,68792	0,00695	0,50	3,0
601668	0,06899	1,0	0,04563	1,0	0,48394	0,02585	1,00	2,0
1432	0,04145	1,0	0,02251	1,0	0,75254	0,01464	0,25	3,0
1970	0,07106	1,0	0,03555	1,0	0,42316	0,02500	1,00	3,0
2030	0,05315	1,0	0,04332	1,0	0,56395	0,02638	0,25	2,0

Table 5: Optimal parameters and RMSE errors for each stock and kernel with up-to-the-minute prices (3-months cases).

Note: c : required parameter for all kernels; γ : Radial kernel parameter; d : Polynomial kernel parameter.

Stock	Linear kernel		Radial kernel			Polynomial kernel		
	RMSE	c	RMSE	c	γ	RMSE	c	d
BBAS3	0,02521	1,00	0,03139	1,00	0,39859	0,02019	1,00	2,00
PETR4	0,02808	1,00	0,03247	1,00	0,41870	0,02278	1,00	2,00
VALE5	0,02279	1,00	0,03458	1,00	0,49505	0,02063	1,00	2,00
ALPA4	0,01361	1,00	0,01749	1,00	0,47601	0,01135	2,00	1,00
DIRR3	0,04192	1,00	0,04371	1,00	0,33309	0,03758	1,00	3,00
LEVE3	0,03084	1,00	0,03370	1,00	0,26088	0,02863	1,00	2,00

Table 6: Optimal parameters and RMSE errors for each stock and kernel with up-to-the-minute prices (2-years cases).

Note: c : required parameter for all kernels; γ : Radial kernel parameter; d : Polynomial kernel parameter.

Stock	Linear kernel		Radial kernel		Polynomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,16427	0,20141	1,13716	1,06158	19,1481	32,3253
PETR4	0,06727	0,40963	0,57590	1,00977	2,18556	7,60178
VALE5	0,05786	0,57993	0,86932	2,96455	3,46965	29,2500
ALPA4	0,08676	0,31555	0,40642	0,54263	1,07965	2,92223
DIRR3	0,08743	0,06276	1,15706	0,64556	1,52443	0,67216
LEVE3	0,14270	0,10383	1,61960	1,25777	0,12765	0,09086
BAC	0,04395	0,12505	0,06929	0,26302	0,08167	0,22308
GOOGL	0,06248	0,03752	1,97066	1,30933	1,18256	0,72193
XOM	0,07517	0,06040	1,08478	0,87865	6,87381	4,13189
ANGI	0,10194	0,42575	0,53587	1,59145	0,78755	1,11141
HL	0,08209	0,29055	0,32289	0,52430	1,33855	1,53013
PZZA	0,07511	0,09601	1,97670	1,49839	30,7060	10,3271
600028	0,17402	3,40886	0,16198	3,27387	2,20295	17,3848
601318	0,31341	0,31624	0,63316	0,45609	3,23618	2,20723
601668	0,16790	0,11452	1,49302	0,94396	69,1139	29,1619
1432	0,19716	0,72379	0,30898	0,93350	0,82217	3,02214
1970	0,19135	0,27265	0,46560	0,52747	0,33288	0,36864
2030	0,07756	0,14436	1,74701	2,57181	3,24215	4,18583

Table 7: SVR results, using the optimal parameter in Tab. 4 on the daily price test data.

Stock	Linear kernel		Radial kernel		Polynomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,04280	0,10205	1,17903	1,11416	1,22156	5,78860
PETR4	0,04412	0,71739	0,20380	0,75658	2,41968	3,48905
VALE5	0,02790	0,15494	0,33634	0,36753	1,008094	8,495684
ALPA4	0,04380	0,25884	0,67884	2,11752	230,1143	2004,279
DIRR3	0,04566	0,03532	0,64196	0,28137	4,52389	1,81663
LEVE3	0,02787	0,10794	0,69095	1,73337	3,52646	4,06868
BAC	0,04049	0,56743	0,02367	0,20404	0,43350	3,00627
GOOGL	0,02330	0,02525	1,03981	1,21231	0,47182	0,29781
XOM	0,00575	0,04100	0,01217	0,07729	0,03574	0,20269
ANGI	0,03928	0,02649	2,00272	1,34247	4,05219	1,29621
HL	0,04158	0,12863	0,07357	0,17985	0,02613	0,07963
PZZA	0,01866	0,06666	0,20137	0,36305	0,01872	0,06700
600028	0,02321	0,07113	0,52052	0,65335	6,36125	9,70272
601318	0,03883	0,04743	1,58233	1,00422	35,7510	11,4795
601668	0,03725	0,44597	0,03686	0,30912	0,39956	2,63612
1432	0,01575	0,00806	1,27500	0,76444	0,94059	0,52970
1970	0,03319	0,22744	0,09702	0,36440	0,33793	1,79328
2030	0,03708	0,21539	0,50188	0,55486	1,52496	4,67396

Table 8: SVR results, using the optimal parameter in Tab. 5 on the up-to-the-minute price test data (3-month cases).

Stock	Linear kernel		Radial kernel		Polinomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,02364	0,10063	0,38569	0,18316	0,38382	0,63054
PETR4	0,04348	0,18961	0,30784	0,28624	0,72348	1,04578
VALE5	0,03699	0,04743	0,09120	0,09199	0,53417	0,52307
ALPA4	0,03879	0,02882	1,81449	1,08270	4,89478	2,76474
DIRR3	0,03397	0,20610	0,03820	0,19452	0,38056	0,83559
LEVE3	0,03067	0,09856	0,55004	0,23597	1,17608	1,91663

Table 9: SVR results, using the optimal parameter in Tab. 6 on the up-to-the-minute price test data (2-years cases).

Stock	Daily prices		Up-to-the-minute prices	
	RMSE	MAPE	RMSE	MAPE
BBAS3	0,06982	0,31553	0,01214	0,13480
PETR4	0,06299	0,41335	0,01511	0,57371
VALE5	0,05713	0,36126	0,01454	0,12852
ALPA4	0,05624	0,76789	0,01976	0,18754
DIRR3	0,09596	0,74089	0,02238	0,13788
LEVE3	0,07263	0,10570	0,01681	0,16462
BAC	0,04598	0,45303	0,00832	0,09031
GOOGL	0,03656	0,10691	0,00484	0,01745
XOM	0,05310	0,40824	0,00444	0,11338
ANGI	0,09856	0,47464	0,00712	0,01068
HL	0,11783	0,77834	0,01443	0,08547
PZZA	0,02499	0,13904	0,00866	0,12760
600028	0,06476	0,71341	0,02195	0,23547
601318	0,08782	0,33966	0,00773	0,04590
601668	0,07430	0,29177	0,01767	0,10858
1432	0,17248	0,69774	0,01621	0,02047
1970	0,25802	0,57245	0,02284	0,21137
2030	0,08892	0,26102	0,02759	0,15310

Table 10: Random walk model results for daily closing and up-to-the-minute prices.

Stock	Linear kernel		Radial kernel		Polynomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,05687	0,25892	0,05894	0,26003	0,09175	0,44025
PETR4	0,05270	0,36937	0,05351	0,31293	0,20236	0,61956
VALE5	0,04403	0,24956	0,04783	0,22266	0,07257	0,32720
ALPA4	0,04652	0,74441	0,04960	0,46956	0,07247	0,96709
DIRR3	0,07727	0,52189	0,07677	0,52011	0,18988	0,97750
LEVE3	319,044	886,815	0,05974	0,10358	70,3598	185,093
BAC	0,03658	0,32727	0,03993	0,38629	0,11997	0,60985
GOOGL	0,02980	0,09312	0,02987	0,11268	0,05404	0,18005
XOM	0,04224	0,36616	0,04199	0,33754	0,10866	0,63326
ANGI	0,08268	0,35844	0,08508	0,37845	0,50205	1,59174
HL	0,09456	0,69319	0,09481	0,71560	0,23611	1,66588
PZZA	0,02067	0,10521	0,02126	0,08679	0,04162	0,21016
600028	0,05120	0,54468	0,05669	0,52282	0,09285	0,74604
601318	0,06868	0,23194	0,07748	0,26111	0,18146	0,76056
601668	0,05779	0,16590	0,06226	0,21833	0,14252	0,44786
1432	0,14203	0,56543	0,13364	0,54647	0,66641	1,30121
1970	0,22315	0,68089	0,22564	0,65871	0,49861	1,53855
2030	0,06776	0,17963	0,07659	0,23437	0,22052	0,52482

Table 11: Results of SVR models updated periodically for daily prices.

Stock	Linear kernel		Radial kernel		Polinomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,00787	0,08743	0,01083	0,11790	0,02364	0,25295
PETR4	0,00970	0,40789	0,01369	0,59740	0,01806	0,69828
VALE5	0,00926	0,08162	0,01317	0,12653	0,01756	0,15890
ALPA4	0,01187	0,10699	0,01687	0,16579	0,03964	0,26407
DIRR3	0,01300	0,08051	0,01751	0,11830	0,02477	0,15939
LEVE3	0,00973	0,09519	0,01309	0,11936	0,01870	0,16123
BAC	1656531	7910363	0,00916	0,09557	2,2e+25	1,5e+26
GOOGL	617306,3	722994,3	0,00497	0,01878	3,7e+22	4,3e+22
XOM	8163009	83056969	0,00431	0,06861	1,8e+37	2,5e+38
ANGI	4853022	12952911	0,00708	0,01052	5,8e+24	1,6e+25
HL	0,00773	0,04365	0,01388	0,07773	43769951	190231018
PZZA	0,00419	0,06079	0,00746	0,10127	0,00419	0,06071
600028	0,01321	0,14705	0,01586	0,17251	0,10413	0,60028
601318	0,00475	0,03128	0,00753	0,05383	0,01444	0,07176
601668	0,01127	0,07303	0,01610	0,10458	0,02678	0,15542
1432	1,92099	4,00243	0,01150	0,01531	830235	1,7e+12
1970	0,01208	0,10058	0,01729	0,14582	0,06873	0,49109
2030	0,01259	0,07017	0,01811	0,10148	0,04341	0,21062

Table 12: Results of SVR models updated periodically for up-to-the-minute prices (3-months cases).

Stock	Linear kernel		Radial kernel		Polinomial kernel	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BBAS3	0,00369	0,03717	0,00542	0,04813	0,01119	0,19015
PETR4	0,00416	0,02895	0,00598	0,04254	0,01224	0,09986
VALE5	0,00292	0,00797	0,00456	0,01167	0,00904	0,02352
ALPA4	0,00619	0,04543	0,00931	0,07025	0,01824	0,13474
DIRR3	0,01983	0,13016	0,02756	0,17858	0,06632	0,32251
LEVE3	0,01728	0,08113	0,02326	0,11044	0,04923	0,23288

Table 13: Results of SVR models updated periodically for up-to-the-minute prices (2-years cases).

Stock	Average return			Volatility		
	Linear	Radial	Polynomial	Linear	Radial	Polynomial
BBAS3	0,167	0,084	0,244	0,996	0,971	0,963
PETR4	-0,360	-0,390	0,076	0,973	0,988	-0,347
VALE5	-0,146	-0,186	-0,238	0,999	0,996	0,957
ALPA4	0,379	0,368	0,361	0,989	0,983	0,916
DIRR3	0,049	-0,153	0,117	0,834	0,671	0,431
LEVE3	0,395	0,084	0,399	-0,055	-0,261	-0,052
BAC	-0,807	-0,800	-0,561	0,998	0,996	0,780
GOOGL	0,130	0,322	0,225	0,992	0,957	0,850
XOM	-0,201	-0,164	-0,229	0,991	0,938	0,930
ANGI	-0,241	-0,146	-0,315	0,897	0,929	0,277
HL	-0,208	-0,271	-0,019	0,988	0,987	0,919
PZZA	0,238	0,192	0,394	0,998	0,993	0,932
600028	-0,001	-0,054	-0,038	0,994	0,982	0,599
601318	0,055	0,167	0,192	0,988	0,965	0,944
601668	-0,024	-0,002	0,078	0,982	0,996	0,924
1432	-0,612	-0,857	-0,533	0,953	0,928	0,959
1970	0,983	0,985	0,952	0,966	0,964	0,991
2030	0,411	0,282	0,621	0,975	0,968	0,581

Table 14: Correlation values between RMSE errors resulting from SVR daily prediction using the moving training window strategy for each kernel and two stock prices statistics: average return and volatility (calculated as the standard deviation of closing prices).