Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists.

Ming-Wei Hsu[1], Stefan Lessmann,[1,2]
Ming-Chien Sung[1], Tiejun Ma[1], Johnnie E.V. Johnson[1],

Abstract— Financial time series forecasting is a popular application of machine learning methods. Previous studies report that advanced forecasting methods predict price changes in financial markets with high accuracy and that profit can be made trading on these predictions. However, financial economists point to the informational efficiency of financial markets, which questions price predictability and opportunities for profitable trading. The objective of the paper is to resolve this contradiction. To this end, we undertake an extensive forecasting simulation, based on data from thirty-four financial indices over six years. These simulations confirm that the best machine learning methods produce more accurate forecasts than the best econometric methods. We also examine the methodological factors that impact the predictive accuracy of machine learning forecasting experiments. The results suggest that the predictability of a financial market and the feasibility of profitable model-based trading are significantly influenced by the maturity of the market, the forecasting method employed, the horizon for which it generates predictions and the methodology used to assess the model and simulate model-based trading. We also find evidence against the informational value of indicators from the field of technical analysis. Overall, we confirm that advanced forecasting methods can be used to predict price changes in some financial markets and we discuss whether these results question the prevailing view in the financial economics literature that financial markets are efficient.

Keywords: Financial time series forecasting, market efficiency, machine learning

[1] Centre for Risk Research, Southampton Business School, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom , email: {mh5e11; ms9; tiejun.ma; jej}@soton.ac.uk.
[2] School of Business and Economics, Humboldt-University of Berlin, D-10099 Berlin, Germany, email: stefan.lessmann@hu-berlin.de.

1. Introduction

Financial markets facilitate international trade, aggregate, discount and convey information concerning the future prospects of organizations and economies, and are enablers of economic growth. Given their importance, financial markets have attracted much research, including the modeling of market prices. Applications of such prediction models include the management of financial risks (W.-S. Chen & Du, 2009) and the support of investment decisions (Chang, Liu, Lin, Fan, & Ng, 2009; de Oliveira et al., 2013; C.-J. Huang et al., 2008).

It is common practice to demonstrate the merit of a novel machine learning (ML) model for time series forecasting using financial market data such as stock or commodity prices, currencies, or financial indices. We survey such experiments in our literature review and show that many studies report high levels of forecasting accuracy. Some studies also identify opportunities to trade profitably on model predictions (Chang et al., 2009; de Oliveira et al., 2013; Doeksen, Abraham, Thomas, & Paprzycki, 2005; C.-J. Huang et al., 2008; Huck, 2010; Kara, Acar Boyacioglu, & Baykan, 2011; Schumaker & Chen, 2009). However, few of these studies discuss the implications for market efficiency (see Table 1).

The success of financial time series forecasting demonstrated by ML experiments is surprising given the theory and evidence from the financial economics literature. More specifically, the efficient market hypothesis (EMH) suggests that current stock prices discount available information and that it is not possible to obtain systematic returns by exploiting any predictability of prices (Malkiel, 2003). This theory is widely supported by financial economists (Fama, 1970, 1991). For example, Jensen (1978, p. 96) claims that “there is no other proposition in economics which has more solid empirical evidence supporting it.”

Our research is motivated by the need to develop an understanding of the reasons for the disagreement between the EMH and empirical evidence in the ML literature. Given the prominence of the EMH, for example in the design and regulation of financial markets, such understanding is important. In addition, the paper helps fill a research gap, as cross-fertilization between the ML and financial economics literatures is limited. For example, the ML community relies on advanced, data-driven forecasting methods in the form of, for example, support vector machines (SVM) or artificial neural networks (ANN). Such methods are rarely considered by financial economists who prefer econometric, often linear methods (Campbell & Thompson, 2008; Fama & French, 1993, 2012). Contradictory findings may thus result from different modeling cultures. In addition, few ML studies discuss their empirical findings in the light of the EMH. The feasibility of predicting a financial market with accuracy above 80 percent seems to be taken for granted in the ML literature, although such figures suggest significant inefficiency, with important implications for the use of financial markets as effective allocative mechanisms. In summary, the objectives of this study are to: (i) examine the magnitude of disagreement between previous findings of the ML and financial economics literature, (ii) investigate the experimental factors in ML studies which may help explain this discrepancy, and (iii) explore the implications for studies examining financial market efficiency.

To achieve our objectives, we compare the predictive accuracy of the most widely used ML and econometric methods and find that the best ML methods outperform the best econometric methods. In doing so we extend previous ML studies in financial time series forecasting (de Oliveira et al., 2013; Kara et al., 2011) by performing an independent evaluation of ML methods (i.e., we do not propose a new method, which may bias an evaluation (Hand, 2006)). This enables us to examine the degree to which previous findings in the ML literature (i.e., high forecasting accuracy) generalize to novel experimental conditions. For example, unlike most previous studies, which examine forecast accuracy and resulting trading profitability in *one* financial market, either within a day or across trading days, we examine forecast accuracy and trading profitability across a *large number* of financial markets. We then compare the performance of forecasts of market prices *within a day and across trading days*, which allows us to make three contributions: First, we assess the degree to which factors associated with forecasting methodology (e.g., forecasting method and horizon, etc.) affect predictive accuracy and trading profitability. Second, our results enable us to provide guidance on how to organize benchmarking experiments in financial time series forecasting and to identify the origins of disagreement between the ML and the financial economics literatures. Third, we examine the implications that results from ML-based forecasting studies should have on the financial economists' prevailing view of widespread market efficiency. In particular, we attempt to build a bridge between the ML and the financial economics domains. This is important because advances in either domain are rarely reflected in the other. Thus, our study is a first step toward unlocking the potential for collaborative gains.

The remainder of the paper is organized as follows: In the following section, we review related work in the financial and ML literatures. Next, we develop six hypotheses concerning how different methodological factors affect predictive accuracy. We then elaborate on our experimental design, before presenting empirical results and discussing their implications. We conclude with a summary of our main findings.

2. Related Work

2.1 The Efficient Market Hypothesis

It is widely believed that financial markets are efficient in aggregating diverse sources of information concerning an asset's future prospects. In particular, the EMH states that asset prices reflect all relevant information and it is impossible to generate excessive returns through 'more informed' (e.g., model-guided) investment decisions (Fama, 1970). Milder versions of the EMH relax the assumption of prices incorporating all relevant information: semi-strong and weak form efficient markets being ones where prices reflect all publicly available information and historic price information, respectively.

The EMH is based on the belief that market participants make rational decisions and that any mis-pricing will quickly be eliminated by those seeking to gain from these anomalies. However, persistent evidence of irrationality has been identified, such as overreaction (Bondt & Thaler, 1985) and the disposition effect (Dhar & Zhu, 2006; Grinblatt & Keloharju, 2001; Shefrin & Statman, 1985). Furthermore, persistent market anomalies, such as the January effect (Keim, 1983) and the weekend effect (French, 1980), have been observed and these are difficult to explain in terms of the EMH. In addition, Lo et al. (2000) show that technical analysis can be used to predict stock movements.

Proponents of the EMH argue that market efficiency is a simplification that may not always hold true, but will on most occasions and for most investors. Equally, they argue that anomalies may appear, but will disappear when they become known to the market. In addition, the ability to predict certain market prices does not imply that these can be exploited to earn excess profit (Malkiel, 2003). Summing up the literature, Fama (1998) indicates that although some studies appear to identify market inefficiency, the EMH cannot be rejected unless three elements are present: endurance - the inefficiency should survive in the long term; homogeneity - the inefficiency should be apparent in the same form across different markets; robustness - the methodology used to demonstrate the inefficiency should be sufficiently robust to provide confidence that the inefficiency exists. It is in this light that we examine the literature which employs ML for financial market prediction.

2.2 Price Prediction in Financial Markets using Machine Learning

Despite financial economists' widespread belief in the veracity of the EMH, a large body of ML literature examines the predictability of financial market prices and the profitability of model-based trading (see Table 1). For example, financial market forecasting is a popular application domain to develop new modeling methodologies and to demonstrate their potential. The prevailing approach is to develop dynamic regression models, which predict future market prices on the basis of past movements in those prices and other price time series (e.g., stocks, indices, currencies, etc.).

A researcher faces many degrees of freedom in organizing a forecasting study. We believe that the specific choices made may have a sizeable effect on the observed results. Consequently, previous findings related to price predictability are best appraised in the context of the specific experimental conditions under which they were obtained. To that end, we review previous studies and their experimental design in Table 1. The common denominator is that all these studies employ some advanced prediction method to forecast price developments in some financial market. In particular, we concentrate on studies that consider support vector machines (SVM) and/or artificial neural networks (ANN). Due to their ability to recognize patterns in nonlinear, dynamic time series data (Chang et al., 2009; Lee, 2009; Żbikowski, 2015), these methods are especially popular in financial market forecasting. Additional factors, other than the prediction method, considered in Table 1 are the financial instrument that is being forecast (individual stocks and indices), the geographic market, the forecasting horizon, whether the study uses a static or dynamic approach to simulate model-based trading, and whether it uses technical indicators among the covariates. We further motivate these factors when developing our hypotheses in Section 3.

[Table 1 about here]

The information in Table 1 suggests that previous studies predominantly seek to forecast financial indices in Asian markets and the Taiwanese market in particular, although US markets also occur frequently. Only six out of 28 studies consider other markets (e.g., Australian, German). Overall, ANN is the most popular prediction method and only four studies consider both ANN and SVM. One of the interesting findings is that very few studies evaluate prediction models in a dynamic fashion. Rather, the prevailing approach, used in 25 out of 28 previous studies, is to split a financial time series into a training and a hold-out test set. We refer to this approach as ‘static’ because it uses the same prediction model throughout the whole testing period, without updating. A dynamic approach such as sliding-window cross-validation performs model training and evaluation multiple times using smaller chunks of sequentially ordered data (Lessmann, Sung, & Johnson, 2011).

The merit of technical analysis is hotly debated in the financial economics literature (Fama, 1970; Lesmond, Schill, & Zhou, 2004; Lo et al., 2000). Advocates of technical analysis stress the predictive ability of technical indicators, whereas opponents refute their value. Nonetheless, technical indicators are widely used in previous ML studies, with 21 out of 28 studies using these indicators, possibly together with other covariates (e.g. financial news).

Table 1 also summarizes previous studies according to whether they: (a) predict the direction of price movements (e.g., rise or fall); (b) predict actual price changes; (c) examine whether investing according to model predictions would produce a profit; (d) discuss their findings in the light of the EMH.

Predicting price direction/size of changes involves forecasting a discrete target variable (i.e., classification)/continuous target variable (i.e., regression). In the former case, Table 1 also reports the directional accuracy, which can be as high as 90 percent (Patel et al., 2015). Clearly, a model that forecasts the direction of price/market movements correctly in 90 out of 100 cases, say days, would facilitate enormous profits, contradicting the view that financial markets are efficient. The directional accuracies shown in Table 1 therefore illustrate the disagreement between (several) ML forecasting studies and studies in the financial economics literature. If the results of the ML studies are reproducible across markets in the long term, the excess returns will be strong evidence against the EMH and most asset pricing models. However, this implication is discussed in only 3 of 28 studies (see Table 1). Our objective is to identify the factors that explain this disagreement.

3. Hypothesis Development

To shed light on the origin of differences in findings concerning market efficiency in the finance and ML literature, we develop a series of hypotheses. These examine the extent to which experimental factors in ML studies affect forecast accuracy and trading profit derived from forecasts and whether the best performing ML technique outperforms the best performing econometric modelling procedures.

The financial economics literature suggests that informational efficiency differs between established and emerging financial markets (Griffin, Kelly, & Nardari, 2010; Ojah & Karemera, 1999); established markets being more efficient and, thus, more difficult to predict. One argument to support this view is regulation. For example, in a highly regulated market, governmental institutions create and enforce detailed rules concerning the release of new information that might affect a company's stock prices. In general, these rules enforce greater and wider disclosure of relevant information. Consequently, capitalizing on private information becomes illegal and less likely.

The relevance of regulation for our study is that selecting a particular financial market for analysis might predetermine the level of predictive accuracy. Specifically, a selection bias toward easier to predict/less efficient markets could explain the high forecasting accuracy reported in the ML literature. However, this is a valid explanation only if predictability and market efficiency do indeed differ with market maturity. To clarify this, we test:

H1a: Predictive accuracy is higher in emerging cf. mature markets, and

H1b: Model-based trading gives higher profits in emerging cf. mature markets.

Another explanation for observing high accuracy/profits in forecasting studies could be the methods used for model assessment. Most ML stock price forecasting studies use a static approach to evaluate model performance (see Table 1), where predictions are generated by a model that is trained with a fixed set of samples. This enables a large number of samples to be included in the training set, which increases the opportunity to recognize price patterns. The static approach is based on the assumption that stock markets exhibit long-term memory (Lo, 1991). However, it disregards the latest price information, which never enters the (single) training set. Sliding-window cross-validation, on the other hand, evaluates a model from multiple origins of a financial time series (i.e., trading periods) and thus uses recent samples for model training. We compare the static approach and sliding-window cross-validation to examine the effect of model assessment methods on model performance.

A static model evaluation paradigm might overestimate the performance of a prediction model. For example, by selecting a single evaluation period (i.e., the test set) the static approach suffers the risk of picking a *lucky sample*. Consider for example the years 2008/2009. Global stock prices all over the world fell significantly in response to the subprime credit crises. In such a period, a naïve forecasting model, which simply predicts price decreases, would display high accuracy, although it does not embody any predictive insight. The risk of over-estimating the value of a model is far less using sliding-window cross-validation, since this repeats the evaluation of a prediction model multiple times. Consequently, we test the following hypotheses:

H2a: Predictive accuracy is higher under a static model evaluation approach, and

H2b: Model-based trading gives higher profits under a static model evaluation regime

The predictive performance of any forecasting model depends largely on the degree to which the covariates are correlated with the target variable. Many financial market forecasting models are auto-regressive models, where the covariates refer to past realizations of the target variable. For example, to predict the closing price of stock X on day *t*, a model might use the closing prices of X at day *t*-1, *t*-2, etc. Indicators from technical analysis are another type of covariate. They apply additional transformations of prices with the goal of creating more informative variables. The value of technical indicators remains an open question. For example, Lo et al. (2000) find some indicators with practical value. Similarly, Brock et al. (1992) conclude that trading strategies that rely on technical indicators can produce excess returns. However, other studies argue that technical indicators have little predictive value and do not facilitate profitable trading, (e.g., Fama, 1970; Lesmond et al., 2004).

Many ML studies employ technical indicators but several studies do not, relying rather on raw prices or simple price differences (e.g., closing price – opening price) (see Table 1). Informative covariates play a key role in predictive modeling. The use or omission of technical indicators may, therefore, affect the degree of predictive accuracy considerably. More specifically, the financial economics literature may under-estimate the degree of inefficiency in financial markets because the predictive value of technical indicators in that literature has been underestimated. ML studies commonly use advanced data-driven models that discern nonlinear relationships and complex variable interactions automatically, but these models are not common in financial economics studies. It may be that ML studies, therefore, extract full information value from the technical indicators. To clarify the influence of technical indicators on predictive accuracy and market efficiency, we test:

H3a: Predictive accuracy is higher if a model incorporates technical indicators, and

H3b: Model-based trading gives higher profits if a model incorporates technical indicators

We also examine the degree to which the selection of a specific forecasting horizon affects predictive accuracy. In general, one expects accuracy to decrease when forecasting further ahead (Hyndman & Athanasopoulos, 2014). This view is supported by the EMH, which predicts that time may be needed for stock prices to fully reflect all available information (Fama, 1970); possibly explaining the rising interest in high-frequency trading (Chordia, Goyal, Lehmann, & Saar, 2013; Menkveld, 2013).

The information reported in Table 1 indicates that previous ML studies in financial time series forecasting predominantly consider forecast horizons of one day or more; possibly because financial data of such granularity is freely available online, whereas high-frequency, intraday data is not available for free. This could indicate that previous forecasting studies, albeit evidencing market inefficiency, have actually underestimated the degree of inefficiency, because they do not consider short forecasting horizons (e.g., intraday forecasting). To examine this proposition we test the following hypotheses:

H4a: Predictive accuracy is higher for shorter forecast horizons, and

H4b: Model-based trading gives higher profits if forecasting shorter periods into the future

A number of studies have suggested that SVM outperform ANN in terms of predictive accuracy (e.g., W.-H. Chen, Shih, & Wu, 2006; W. Huang et al., 2005; K. Kim, 2003; Ou & Wang, 2009; Tay & Cao, 2001). Consequently, to examine the effect of the ML forecasting model on predictive accuracy for financial time series and on the assessment of market efficiency, we test the following hypotheses:

H5a: Accuracy is higher for predictions based on SVM (cf. ANN), and

H5b: Model-based trading gives higher profits when using SVM (cf. ANN) for prediction

The financial economics and ML literature examining forecasting accuracy largely rely on econometric and ML methods, respectively. To examine whether the adoption of econometric *vs.* ML methods effects predictive accuracy, we perform a comparison of the commonly employed ML methods (ANN and SVM) with the commonly employed econometric models: autoregressive model (AR), generalized autoregressive conditional heteroscedasticity model (GARCH) and autoregressive integrated moving average model (ARIMA) (Charles, Darné, & Kim, 2011). According to our experiments, reported in online Appendix A, AR outperforms ARIMA and GARCH. We also find that SVM outperforms ANN. In order to assess the effect of the forecasting model (econometric vs. ML) on predictive accuracy for financial time series and on assessment of market efficiency, we test the following hypotheses:

H6a: Accuracy is higher for predictions based on the best performing ML cf. econometric method (i.e. SVM vs. AR), and

H6b: Model-based trading gives higher profits when based on predictions from the best performing ML cf. econometric method (i.e. SVM vs. AR)

The proposed set of hypotheses, enable us to examine the extent to which various factors, such as emerging market data, static evaluation setting, technical indicators, shorter forecasting horizon and selected models (e.g. SVM vs ANN and econometric vs. ML), influence the accuracy of financial time series forecasting. Integrating such insight with information how ML studies are typically conducted (i.e., Table 1) enables us to explain apparent inconsistencies between the results of the ML and financial economics literatures.

4. Experimental Design

We investigate the influence of five experimental factors (market maturity, model simulation methodology, covariate composition, forecast horizon and prediction method (SVM vs. ANN; best ML (SVM) vs. best econometric (AR)): see Table 2) on the predictability of price movements in financial markets and the profitability of model-based trading. We test hypotheses based on these experimental factors using a data set of 34 financial time series. Whilst this is not a large sample, it facilitates a reasonably inclusive study, since most major markets in the world are examined. In addition, since most ML studies use only a single financial time series, our study is relatively large and we feel, therefore, that the conclusions can be relied upon.

More specifically, we examine all combinations of factors in a full-factorial setup. To illustrate our procedure, consider one experiment with specific choices for each of the experimental factors. For example, the experimental setup could be such that we use *SVM* to forecast the *FTSE 100* using a forecasting horizon *of one day, excluding technical indicators* from the covariates and measuring forecast accuracy using a *static evaluation approach*. This experiment produces a set of predictions for all periods (i.e., days) of the test set, from which we can calculate forecast accuracy and the profitability of model-based trading. Having completed the first experiment, we change one factor (e.g. incorporate technical indicators among the covariates), and repeat the forecasting simulation. We continue this process for all combinations of factor levels (see Table 2) and all financial time series. This enables us to compare predictability across different experiments and to examine the impact of the experimental factors. The design is similar to that of Gerlein, McGinnity, Belatreche and Coleman's (2016) study in which repeated experiments are performed to clarify the influence of several factors (e.g., the size of sliding window and the number of covariates), on the prediction performance of simple ML classifiers. In the following subsections, we discuss the factors and motivate our choices of individual factor levels.

[Table 2 about here]

4.1. Data, Variables, and Forecasting Horizon

We obtain our data from TickWrite Data Inc.[3] It comprises time series of financial indices from 34 markets for both established and emerging markets over a 6-year period (2008-14). An exception is the Brazilian market for which we have data for 4 years (2010-14). We included as many markets as possible to cover both mature and less mature markets, since one of the aims of the study is to examine the influence of market maturity. It was important for comparability that the same period was employed for each market included in the study. This restricted the sample period because the availability of intraday data is limited and, for many markets, is only available from 2008 onwards. This forced us to choose the data period 2008 to 2014, where intraday data is available for most markets.

Table 3 summarizes the data set. We have focused on predicting national stock indices because these indices are used in the majority of previous ML studies that predict direction of price changes (e.g. Bodyanskiy & Popov, 2006; de Oliveira, Nobre, & Zárate, 2013; C.-J. Huang, Yang, & Chuang, 2008; W. Huang, Nakamori, & Wang, 2005; Pan, Tilakaratne, & Yearwood, 2005; Qian & Rasheed, 2007).

[3] http://www.tickdata.com/

[Table 3 about here]

4.1.1 Market maturity

To test our first hypotheses (H1a and H1b) related to market maturity, we use the World Bank income level[4] to categorize financial markets (fourth column in Table 3): 26 markets in our sample stem from high income (mature) economies and 8 from middle income (less mature) economies. This approach to categorization is consistent with previous experiments (Choong, Baharumshah, Yusop, & Habibullah, 2010; Claessens, Klingebiel, & Schmukler, 2006). However, idiosyncrasies of particular classification systems may introduce bias. To test the robustness of our findings in relation to market maturity, we also employ market maturity classifications provided by the Morgan Stanley Capital Index (MSCI) (Zunino et al., 2009) and the International Monetary Fund (IMF) (J. H. Kim & Shamsuddin, 2008).

4.1.2. Model simulation methodology

We examine hypotheses related to the model simulation methodology employed. In particular, we compare results from the most common approach adopted, namely a 'static model evaluation' (e.g., de Oliveira et al., 2013), with those from a 'dynamic model evaluation' (e.g., Lessmann et. al., 2011). The full details of these approaches are provided in 4.2.2. These models are developed using a fixed number of data points (e.g. days or hours) or 'sliding window size). In comparing results for the static and dynamic approaches, we examine the effect of alternative sliding window sizes of 25, 50 and 100.

4.1.3 Model covariates

To test hypotheses related to the effect of technical indicators (H3a and H3b), we consider the seven technical indicators that have been included in at least one study in our literature survey (these, together with the papers in which they appear are shown in Table 4). We develop sixteen covariates by varying parameters of these technical indicators, using popularvariations found in the literature (e.g., Bollen et al., 2011). In addition, we create a second set of covariates using the characteristic points of a financial time series, namely, the opening, closing, highest and lowest price and the change of price (from opening to closing) observed in a predefined interval (e.g. 1 day). To test H3a and H3b, we compare the predictive ability of forecasting models incorporating covariates based on the reference values (open, high, low, close and change) to models that incorporate both the reference values *and* the sixteen technical indicator-based covariates. We deliberately employ simple covariates as a

[4] http://data.worldbank.org/about/country-and-lending-groups

benchmark against which to examine the predictive value of the technical indicators because some previous research studies question the value of their informational content (Lesmond et al., 2004). Clearly, these suspicions will be confirmed if technical indicators cannot improve the accuracy of predictions of a model based solely on a simple benchmark.

[Table 4 about here]

4.1.4 Forecast horizons

To test hypotheses concerning the influence of different forecast horizons on market predictability and model-based trading profitability (H4a and H4b), we compare forecast horizons of one day and one hour. A horizon of one day is selected as it is the predominant setting in previous research (see Table 1). We selected an alternative horizon of less than one day because a comparison of intraday forecasting models and models that forecast one day or more ahead has, to our best knowledge, not been undertaken. We select one hour for the intraday setting to gain the benefits of a short forecast horizon (which increases the difference between the intraday setting and the setting where models predict one day into the future) whilst reducing the high computational costs associated with performing a large number of comparisons of computationally complex forecasting models for even shorter horizons. To compare experiments with forecast horizons of one day and one hour, we need to align the time periods of forecast model training, validation and testing. To illustrate this, consider a test set of five trading days. In the daily setting, a model produces five individual predictions of price change direction (one for each day), which we aggregate to estimate the accuracy of the (daily) model. To compare this model on an equal footing to to an alternative model that predicts one hour ahead (i.e., in the hourly settings), the latter should predict the same test set of five days. To achieve this, we let the hourly model produce one prediction for every hour for each of the five days in the test set. For example, if a stock exchange opens from 9am to 5pm (i.e, 8 hours), we create 8*5=40 predictions, estimate the accuracy of the hourly model on the basis of these 40 predictions, and compare this accuracy with that of the daily model.

4.1.5 Forecasting methods

A large number of ML prediction methods are available (Hastie et al., 2009). However, we focus on the two methods predominantly used in the ML literature on financial time series forecasting, SVM and ANN (Ballings, Van den Poel, Hespeels, & Gryp, 2015). One of the explanations for their popularity for time series forecasting (cf. other advanced techniques such as bagged or boosted decision trees), is that they are better-suited to handle continuous covariates (Lessmann, Baesens, Seow, & Thomas, 2015), which occur frequently in financial time series forecasting (e.g., prices, price differences or technical indicators).

In its basic form, a SVM can be characterized as a regularized linear classifier, which estimates the target label, y, of an observation $\boldsymbol{x}$ by means of a linear function. Let the vector $\boldsymbol{\beta}$ and the scalar β_0 denote the coefficients and the intercept of such a linear function. Then, to develop a SVM, one solves the following mathematical program:

$$\min_{(\beta_0,\boldsymbol{\beta})} \mathcal{L} = \|(\beta_0, \boldsymbol{\beta})\|_2 + \lambda \sum_{i=1}^{n} \max(1 - y_i(\boldsymbol{\beta} \cdot \boldsymbol{x} + \beta_0), 0). \qquad (1)$$

To prevent an SVM from overfitting the training data, the first term on the right-hand side of (1) penalizes model complexity through minimizing the magnitude of the model coefficients (Hastie et al., 2009). The second term measures the degree to which the model fits the training data accurately and is called the Hinge loss (Hastie et al., 2009). The scalar λ is a meta-parameter of SVM that allows the user to control the trade-off between high model fit and low model complexity (Smola & Schölkopf, 2004).

In our study, the scalar y refers to the change in direction (up or down) of a financial index in a future period (i.e., next day or next hour) and $\boldsymbol{x}$ represents a vector of covariates which we use to predict y (previous values of y, technical indicators, etc.). We estimate the vector of model parameters, $\boldsymbol{\beta}$, during model training. We then compute predictions, $\hat{y}$, as follows:

$$\hat{y} = sign(\boldsymbol{\beta} \cdot \boldsymbol{x} + \beta_0) \qquad (2)$$

SVMs are able to implicitly project the input data into a nonlinear feature space of higher dimension. Creating a linear model in the transformed space is equivalent to creating a nonlinear model in the original input space (Cristianini & Shawe-Taylor, 2000). The nonlinear transformation of the data is (computationally) feasible because the dual program of (1) incorporates the input data only in the form of scalar products. Using a kernel function, SVMs are able to compute the scalar product in the feature space directly (i.e., without actually transforming the data). We use the radial-basis-kernel-function, which is the standard kernel in SVM applications (Keerthi & Lin, 2003).

ANNs are another nonlinear prediction method. In general, a feed-forward neural network consists of input, hidden and output layers, where each layer has multiple information processing units called neurons. The neurons of one layer are fully-connected to the neurons of the next layer. The neurons in the input layer are simply the original covariates. We design our ANNs such that the output has only one neuron, which is the standard setup for regression and binary classification problems (Hastie et al., 2009). In the former case, the output neuron models the value of the continuous target variable and in the latter case the posterior probability of the discrete target variable. The number of neurons in the hidden layer, Z, is a meta-parameter. Let $\boldsymbol{a}_z$ be a vector of weights that connect the input neurons to the z^{th} hidden neuron, b_z a threshold attached to hidden neuron z, and let g^h be a nonlinear function. We can then write the output of the hidden layer as follows:

$$G(\boldsymbol{a}_z, b_z, \boldsymbol{x}) = \begin{bmatrix} g^h\left(b_1 + \sum_{j=1}^{m} a_{1j}x_j\right) \\ \vdots \\ g^h\left(b_z + \sum_{j=1}^{m} a_{zj}x_j\right) \end{bmatrix}_{Z\times 1}. \tag{3}$$

In a similar way, the result of the ANN, $f(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}_Z, b_Z)$, is:

$$f(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}_Z, b_Z) = g^o\left(\sum_{z=1}^{Z} \beta_z\, G(\boldsymbol{a}_z, b_z, \boldsymbol{x})\right), \tag{4}$$

where $\boldsymbol{\beta}$ denotes the weight vector that connects the hidden and the output layer, and g^o is the function that transforms the result of the output neuron. We follow Chang et al. (2009) and choose g^o and g^h to be logistic functions. This allows us to interpret the output of the neural network as an estimate of the posterior probability of an upward/downward price change.

To determine the parameters of the ANN ($\boldsymbol{a}_z, b_z,$ and $\boldsymbol{\beta}$), one minimizes a suitable loss-function using gradient-based methods. Specifically, we create an ANN model through solving (5) using a quasi-Newton algorithm, where λ is once more a regularization parameter to penalize model complexity and prevent overfitting.

$$\min_{\boldsymbol{\beta}} \mathcal{L} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\beta}))^2 + \lambda \|\boldsymbol{\beta}\|_2. \tag{5}$$

ANN and SVM contain two meta-parameters. First, the regularization term λ appears in both models. The second meta-parameter for SVM is a parameter of the radial-basis kernel function, (denoted by γ). For ANN, the second meta-parameter determines the number of neurons in the hidden layer. We employ grid-search to identify suitable values for the meta-parameters of ANN and SVM (Cherkassky & Ma, 2004). In particular, based on recommendations from the literature (Berry & Linoff, 1997; Xu & Chen, 2008), we define candidate values for each meta-parameter and empirically assess all possible combinations of meta-parameter settings. We select the combination with maximal forecast accuracy on a separate validation sample to compute predictions for an out-of-sample test set. Table 5 reports the candidate settings for meta-parameter values.

[Table 5 about here]

4.2. Performance Measurement

4.2.1. Indicators of predictive accuracy

As indicated above, our target variable is ‘direction change’:1 if the index prices increase from one period to the next (e.g., from day t to day t+1), and 0 otherwise. We assess the predictive accuracy of a forecasting model in terms of the percentage of correctly classified observations (hit rate).

Since the main aim of the paper is to examine the influence of methodological factors on prediction performance, we do not attempt to optimize accuracy. Previous ML studies generally focus on one specific market. However, to achieve our objective, we attempt to forecast *all* markets with a consistent, clearly documented methodology. This facilitates replicability across indices, which we consider an integral part of our research. Consequently, we deliberately do not try to maximize predictive accuracy for an individual index.

We estimate the profitability of trading on a forecasting model’s predictions based on a simple trading strategy, which we call ‘follow prediction’. In particular, we assume that a trader, via futures contracts, buys into the market if the predicted ‘direction change’ equals 1, and sells otherwise. We calculate the return on investment (ROI) in period t (roi_t), as follows:

$$roi_t = 1 + \frac{ABS|CI_t - CI_{t+1}|}{CI_t} * PC_t, \tag{6}$$

where CI_t is the closing price of the financial index at period t and PC_t is 1 if the model-predicted ‘direction change’ in period t+1 is correct, and -1 otherwise. We employ this ROI formula as we use the information available in period t to predict the ‘direction change’ in period t+1. A period is either one day or one hour, depending on the experimental setting. Consequently, we obtain a profit/loss if the prediction is correct/wrong. The overall ROI over all test periods t = 1, 2, …,T is then calculated as follows:

$$ROI = \prod roi_t\,, t \in T \tag{7}$$

4.2.2. Model evaluation and simulation of model-based training

To test the accuracy of forecasting models, we distinguish two approaches for out-of-sample model evaluation. The standard approach in the literature is to split a financial time series into three non-overlapping sets for model training, validation, and testing (see Table 1). The training set is used to estimate the parameters of the forecasting model (e.g., $\boldsymbol{\beta}$ in (1)). The validation set is used to tune meta-parameters (e.g., the regularization parameter λ in (1)) by means of empirical experimentation. A fully-specified model with fixed meta-parameters is then developed on the union of the training and validation sets (to maximize data utilization) and applied to generate predictions for the observations in the test set. To measure forecast accuracy, the predictions are compared to the actual values of the target variable in the test set. This allows us to compare models in terms of their predictive accuracy on hold-out data. We call this approach a static evaluation, because the same model is used to forecast all observations in the test set (de Oliveira et al., 2013; Kara et al., 2011).

To implement the static approach we split the data points in a financial time series chronologically: 50, 25 and 25 percent for training, validation and testing, respectively. Our data set, which runs from February 2008 to February 2014, contains 1,500 trading days for each of the 34 financial time series. Consequently, by way of example, when developing forecasting models to predict one day ahead, we use the last 25% of trading days (375) for hold-out testing. The models that we eventually compare (e.g., SVM vs. ANN) are trained and validated on the preceding 75% (1,125) of trading days. In particular, we use the first 50% (750) of trading days to train the forecasting models with alternative meta-parameter settings (see Table 5) and assess their accuracy on the following 375 trading days (our validation sample). The validation sample predictions reveal the best meta-parameter setting for a given time-series. We use this setting to estimate a final model using the full 1,125 days sub-sample and use the resulting model to predict the hold-out test set.

We contrast the static model evaluation approach to a dynamic sliding-window cross-validation (Lessmann et al., 2011). Here, we use the same data partitioning as in the static setting. For example, when developing forecasting models to predict one day into the future we again use the first 50% (750) days of a time series to estimate one forecasting model per meta-parameter setting (e.g., 12 settings for ANN). Every meta-parameter setting provides a candidate model. However, we use these candidate models to develop predictions for a single data point, namely for day 751, which is the first day in the validation partition. Next, we shift the training and validation window one period (day) forward. That is, using days 2 to 751 for model training, we estimate one forecast model per meta-parameter setting and let the resulting models predict day 752. We repeat the estimation of models and prediction of one day ahead until we have developed predictions for all days in the validation period (i.e., days 751 to 1,125). We then calculate the accuracy of candidate forecasting models on the validation sample and select the meta-parameters that give maximal accuracy.

This dynamic approach differs from the static approach in that every data point of the validation set is predicted with a different model. In other words, we update forecasting models to incorporate recent information (e.g., we train forecasting models that predict the price 'direction change' for trading day 752 on a data set that includes trading day 751).

After completing model selection in the dynamic approach we collect predictions for the test set. To that end, we proceed as in the model selection step. For example, when comparing SVM and ANN models used to predict one day ahead, for each day of the test set we estimate one SVM and ANN model. Considering, for example, the first day in the test set (day 1,126), we use the preceding days to train the forecasting models and let them predict trading day 1,126. We then shift the evaluation window one period forward and repeat. The main difference to the model selection stage is that we do not consider all meta-parameter settings. Instead, we estimate one SVM and one ANN model, for which we use the best meta-parameter values as identified in the preceding model selection step. In summary, each data point of the test set is predicted with a different forecasting model, but all models use the same settings for meta-parameters.

If the forecast horizon is not one day but one hour, we proceed in exactly the same way as described. However, the size of the initial data sample is much larger because we now have one data point per opening hour of the stock exchange.

Sliding-window cross-validation offers some degrees of freedom. In particular, when predicting a particular data point (e.g., day), one can either use all previous data points for model training or work with a fixed training window (Lessmann et al., 2011). We opt for the latter approach and use a window size of 50. Using a fixed window size (cf. all available trading days prior to a day *t*) has the advantage that it decreases computational costs. More importantly, assuming that stock returns show no long-term memory (Cheung & Lai, 1995; Lo, 1991), using very old data for training a forecasting model might actually harm its accuracy. However, we acknowledge that the choice of the window size might affect the results of sliding-window cross-validation. Therefore, to secure reliability, we empirically analyze alternative window sizes of 25 and 100 data points.

5. EMPIRICAL RESULTS

5.1. Tests of the Hypotheses of Experimental Factors on Market Predictability

The experimental design includes five main factors: market maturity, model simulation methodology, covariate composition, forecast horizon and prediction method (SVM vs. ANN; best ML (SVM) vs. best econometric (AR))(see Table 2). Each factor has two alternative settings (e.g., SVM and ANN for the factor prediction method) and these are compared in multiple simulations. That is, we obtain 272 individual prediction results for SVM and ANN

(34 financial markets × 2 forecast horizons × 2 covariate compositions × 2 model simulation methodologies), enabling us to show the performance difference of the two prediction methods. We compare prediction results for other factors in a similar way. We provide an overview of the empirical results in Figure 1, which reports the distribution across the experimental factors for predictive accuracy and ROI.

A number of conclusions emerge from the results presented in Figure 1. First, a model-based trading approach produces higher ROI in high (cf. middle) income markets. More specifically, the majority of prediction models produce a ROI greater than one in high income markets; suggesting that profitable trading on model predictions is possible and questioning market efficiency in the corresponding markets. In middle income markets, a larger number of prediction models do not produce a profit, although the median ROI is still greater than one.

We also examine the performance of the forecasting models at the individual market level. We obtain a distribution of prediction performances for each market because we develop multiple models to examine the other experimental factors: forecast horizon, prediction model, etc. These results are displayed in Figures 2 and 3.

[Figure 1, Figure 2, Figure 3 about here]

Figures 2 and 3 evidence substantial variation in predictability across markets. This is true for the comparison across high and medium income markets but also for comparisons between markets within these groups (e.g., the Finnish and German markets belong to the high income group but display large differences in their spread in predictive accuracy). Results at an individual market level also suggest that profitable model-based trading is feasible based on the covariates we employ (see Figure 3) since the median ROI is considerably greater than one for most markets. Examples of exceptions include markets in Austria, Denmark, Estonia, Hungary and Germany.

In summary, the main finding from the results displayed in Figures 2 and 3 is that the selection of the financial market exerts a major influence on the observed level of predictability and thus market efficiency. Our analysis of the literature reveals that previous ML studies predominantly rely on data from a single market. The wide variation that we observe in market efficiency between markets suggests that this focus on a single market risks over- or under-estimating the potential for prediction across all financial markets.

The results presented in Figure 1 indicate that a sliding-window methodology typically indicates lower levels of accuracy and ROI than a static methodology. This suggests that the cross-validation method employed affects prediction performance and should be considered in future comparisons. The static approach, which prevails in prior research (see Table 1), benefits from using a larger number of samples for model training compared to the sliding-window cross-validation. The dynamic approach benefits from including the most recent samples. Therefore, our results indicate that including recent price information does not improve stock market prediction more than incorporating a larger number of samples. This supports the view that financial markets exhibit long-term memory (Lo, 1991).

The results presented in Figure 1 suggest that technical indicators for financial market modeling do not offer much advantage over simple covariates, as the accuracy and ROI distributions of the two settings are similar. This is surprising since we deliberately select simple covariates in the form of reference prices (open, high, low, close) as a benchmark for the technical indicators. For example, we do not consider price differences or moving averages, which have been used in prior research (see Table 1). Therefore, our results provide strong evidence against using technical indicators for financial time series forecasting.

The results displayed in Figure 1 suggest that predictive accuracy is lower when forecasting price movements one hour (cf. one day) ahead but the ROI distribution does not display much difference between the two settings. However, it is noteworthy that both daily and hourly forecasting horizons produce, on average, a positive return (i.e., ROI > 1).

Our results support previous findings that SVM perform better than ANN in financial market modeling (K. Kim, 2003; Ou & Wang, 2009; Tay & Cao, 2001). The median accuracy and the median ROI are higher for SVM compared to ANN. In addition, the distribution of ROI displays less spread for SVM, implying greater stability of SVM-based prediction models. The lower quartile of the ROI distribution is higher for SVM (cf. ANN) and the ROI distribution of ANN shows an outlier value where ROI drops below 50%, whereas SVM do not suffer from such large losses. SVM can thus be considered more robust.

To formally test our hypotheses, we examine the statistical significance of observed mean differences across factor levels. We achieve this using regression analysis. In particular, we estimate the following regression models to explain predictive accuracy and ROI:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{8}$$

$$\text{ROI} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{9}$$

where *MI, ST, T, D* and *SVM* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates (open, high, low, close and change), a daily forecast horizon and SVM, respectively, and taking the value 0 otherwise (i.e. for high income markets, a dynamic model simulation, when only using basic price covariates, an hourly forecast horizon and ANN).

To formally test our hypotheses related to the relative performance of, what we show to be, the best ML and econometric models (i.e. SVM and AR, respectively), we estimate the following regression models to explain predictive accuracy and ROI:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \quad (10)$$

$$\text{ROI} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \quad (11)$$

where *MI*, *ST*, *T*, *D* and *SVM* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates, a daily forecast horizon and the best performing ML model (SVM) prediction, respectively, and 0 otherwise (the reference model is AR).

If we expect a regression model to go through the origin, we can remove the intercept term from the model without introducing bias in estimating coefficients. In the case of prediction accuracy, *going through the origin* means the average accuracy is 0%. However, since the accuracy of randomly guessing the price direction is around 50%, the accuracy of our models are unlikely to be around 0%. Similarly, the regression line for ROI is also unlikely to go through the origin, unless our models generate huge negative returns and lose all the original capital. Consequently, we expect that the regression lines for both accuracy and ROI will not go through the origin. Hence, we include the intercept terms to avoid potential bias in estimating the parameters of the regression models (Brooks, 2014). Tables 6 and 7 summarize the results of the regression analyses for accuracy and ROI, respectively, related to the comparison of the two commonly employed ML methods (SVM and ANN). Tables 8 and 9 summarize the results of the regression analyses for accuracy and ROI, respectively, related to the comparison of the best performing ML and econometric methods (SVM and AR).

[Tables 6, 7, 8 and 9 about here]

The F-statistics and their corresponding p-values displayed in Tables 6-9 confirm the statistical significance of the four regressions. The adjusted R^2 values suggest that the independent variables explain about 20 (21) percent and 17 (23) percent of the observed variation in predictive accuracy and ROI, respectively (The numbers in brackets relating to the regressions incorporating the best performing ML and econometric methods (SVM and AR). These R^2 values may appear rather low. However, it is important to remember that the development of prices in financial markets is driven by a multitude of factors, many of which are not considered in this study. As a consequence, the performance of the prediction models exhibit substantial unexplained variation. However, it is noteworthy that R^2 is higher in the regression of predictive accuracy. This emphasizes the subtle difference between predicting market movements using a ML model and trading profitably on model-predictions. Our results show that the degree to which the latter is possible depends even more on external factors other than the prediction method (and on other factors considered in this study). As we discuss below (see Section 6), this has important implications for conclusions regarding market efficiency.

We first examine results related to experiments simply incorporating ML forecasting models (SVM and ANN).The results presented in Table 6 and Table 7 demonstrate that the effect of all experimental factors (other than covariate composition) on predictive accuracy and ROI is statistically significant. This confirms the relevance of the chosen factors and that the potential of sophisticated technical indicators to help predict financial market movements (over that achievable via simple reference-price based covariates) might be limited. Note that we also run experiments in the markets in which data are available for longer periods (8 ~ 30 years). The corresponding results, which are available in online Appendix C, agree with those presented in the main part of the paper.

In interpreting the other results of the regression, it is important to remember the reference model that forms the basis of the comparison. Given our coding of the independent variables, the reference model is an ANN model that predicts price movements in high income markets using simple reference prices as covariates one hour ahead and is evaluated using sliding-window cross-validation. Considering the regression coefficient of the intercept in the accuracy regression model (Table 6), such a model produces a directional accuracy of 51 percent. Assuming that the distribution of hourly price movements is roughly balanced, this is only a little better than random. However, we observe in the ROI regression (Table 7) that the seemingly small improvement over a random model is enough to produce a sizeable profit, with a ROI of about 18 percent. The regression coefficient of market maturity (-0.1496) indicates that profitability erodes in middle income markets. Similarly, ROI decreases with forecast horizon. Keeping everything else constant, predicting one day into the future reduces ROI approximately 4 percentage points (regression coefficient of -0.0442). However, predictive accuracy increases in a setting with a daily forecast horizon (regression coefficient 0.0174 in Table 6). The results also confirm that a static model simulation methodology (cf. sliding-window cross-validation) and using SVM instead of ANN both significantly increase accuracy and ROI. It is worth noting that the significance of the intercept terms in both the regressions reported in Tables 6 and 7 supports our decision to include these terms.

To complement the analysis of the five experimental factors, Table 6 and Table 7 also include measures of effect size. In particular, Cohen's d captures the mean difference between groups in standard deviation units. As a rule of thumb, values of 0.2, 0.5, and 0.8 indicate small, medium, and large effect sizes, respectively (Cohen, 1969). The Cohen's d results displayed in Tables 6 and 7, indicate that market maturity, prediction model (SVM vs ANN), and model simulation methodology have a medium effect on predictive accuracy and ROI. Effect sizes for the factor covariate composition are small, which supports the lower levels of significance that we observe for this factor. Finally, the forecast horizon has a medium to large effect on predictive accuracy but only a small effect on ROI. For completeness, we include an additional measure of effect size in Table 6 and Table 7, the amount of variation that is explained by an individual factor – commonly referred to as partial η^2 in analysis of variance. Overall, the results are in line with those related to the Cohen's d.

We now turn to our analysis of the regressions incorporating the best performing ML and best performing econometric models. As discussed above, the results displayed in Tables 6 and 7, show that SVM outperforms ANN. This result is line with that reported in the literature (W.-H. Chen et al., 2006; W. Huang et al., 2005; K. Kim, 2003; Ou & Wang, 2009; Tay & Cao, 2001). Our experiments also show that AR outperforms ARIMA and GARCH (see online Appendix A). To formally test hypotheses 6a and 6b, we compare SVM and AR by examining the statistical significance of observed mean differences across factor levels in our estimations based on equations (10) and (11).

The results presented in Table 8 and Table 9 demonstrate that SVM significantly outperforms AR in terms of accuracy and ROI, providing support for Hypotheses 6a and 6b. It is important to note that the effects of all experimental factors (with the exception of covariate composition) on predictive accuracy and ROI are statistically significant. The significance and direction of the effect of the experimental factors on accuracy and ROI are in line with the results discussed above for the regressions based on the two most commonly employed ML techniques (i.e. equations (8) and (9)). This confirms that most of the experimental factors have a significant effect on the predictive accuracy and ROI achievable using both ML and econometric methods.

5.2. Sensitivity Analysis

To confirm the robustness of our conclusions, we perform additional experiments related to different sliding window sizes and different systems for classifying market maturity.

5.2.1. Analysis of Sliding Window Size

In sliding-window cross-validation, the size of the training window is a potentially important parameter that might affect the observed level of predictive accuracy. Thus, to investigate whether the above conclusions are robust toward alternative settings, we perform experiments with sliding window size 25 and 100. We estimate the following regression models to explain predictive accuracy and ROI:

$$\text{Accuracy} = \alpha + \beta_{\mathrm{MI}} MI + \beta_{\mathrm{SW25}} SW25 + \beta_{\mathrm{SW50}} SW50 + \beta_{\mathrm{SW100}} SW100 + \beta_{\mathrm{T}} T + \beta_{\mathrm{D}} D + \beta_{\mathrm{SVM}} SVM + \varepsilon, \tag{12}$$

$$\text{ROI} = \alpha + \beta_{\mathrm{MI}} MI + \beta_{\mathrm{SW25}} SW25 + \beta_{\mathrm{SW50}} SW50 + \beta_{\mathrm{SW100}} SW100 + \beta_{\mathrm{T}} T + \beta_{\mathrm{D}} D + \beta_{\mathrm{SVM}} SVM + \varepsilon, \tag{13}$$

where *MI, SW25, SW50, SW100, T, D* and *SVM* are dummy variables, taking the value 1 for middle income markets, when the sliding window size is 25, 50 and 100, when using technical indicators alongside basic price covariates, a daily forecast horizon and SVM prediction, respectively, and 0 otherwise. Recall that the data to estimate these equations also includes the results of (static) forecasting models that are not subject to a sliding window evaluation. Therefore, using three dummies to encode the three settings for sliding window size is appropriate and does not create linear dependence among covariates.

The results of regressions (12) and (13) are shown in Table 10. For both accuracy and ROI, all of the coefficients of *SW25*, *SW50* and *SW 100* are negative, indicating that accuracy and ROI are significantly higher under a static model evaluation compared to using a dynamic sliding-window cross-validation. For both accuracy and ROI, the coefficients of sliding window size 100 are very close to that of sliding window size 50, whereas the coefficient of sliding window size 25 is much lower than the others. Hence, increasing the size of the sliding window appears to improve prediction performance, but the marginal benefit appears to decrease. We take the small differences between the results for window sizes of 50 and 100 as evidence that the choice of 50 is appropriate and does not affect the conclusions of the comparison of alternative model evaluation regimes (sliding-window vs. static).

5.2.2. Analysis of Different Market Maturity Classification Methods

There are several stock market maturity classification methods. For example, Claessens et al. (2006) and Choong et al. (2010) classify stock market development based on the income level data of World Bank (as used in our main analysis), Zunino et al. (2009) use the Morgan Stanley Capital Index (MSCI), and Kim and Shamsuddin (2008) adopt the International Monetary Fund (IMF) classification.

To examine whether the adoption of different classification methods could lead to a different conclusion regarding the impact of market maturity, we estimate the following regression models to explain predictive accuracy and ROI:

$$\text{Accuracy} = \alpha + \beta_{\mathrm{IMFE}} IMFE + \beta_{\mathrm{ST}} ST + \beta_{\mathrm{T}} T + \beta_{\mathrm{D}} D + \beta_{\mathrm{SVM}} SVM + \varepsilon, \tag{14}$$

$$\text{ROI} = \alpha + \beta_{\text{IMFE}} IMFE + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{15}$$

$$\text{Accuracy} = \alpha + \beta_{\text{MSCIE}} MSCIE + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{16}$$

$$\text{ROI} = \alpha + \beta_{\text{MSCIE}} MSCIE + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{17}$$

where *IMFE*, *MSCIE*, *ST*, *T*, *D* and *SVM* are dummy variables, taking the value 1 for IMF emerging markets, MSCI emerging or frontier markets, a static model simulation, when using technical indicators alongside basic price covariates (open, high, low, close and change), a daily forecast horizon and SVM prediction, respectively, and 0 otherwise. Table 11 provides details of the manner in which different markets are classified using the World Bank, the IMF and the MSCI classification systems.

Table 12 and Table 13 summarize the results of the regression analyses for accuracy and ROI, respectively. All the estimated coefficients of IMFE and MSCIE are negative and statistically significant. That is, predictive accuracy and the profitability of model-based trading are higher in the markets with greater (cf. lower) maturity. This confirms the conclusion we drew based on the World Bank classification system.

6. Discussion

We first discuss our results relating to the factors that influence the prediction accuracy and the profitability of trading on the forecasts based on the most commonly employed ML techniques (SVM and ANN). These results demonstrate that the selection of the data source (i.e., the financial market being forecast) significantly and substantially influences the observed level of prediction accuracy and the profitability of trading on the forecasts. The impact on the profitability of trading on the forecasts is particularly large (Cohen’s d equal to 0.6264). However, we cannot accept hypotheses *H1a* and *H1b* because the direction of the influence differs from expectations. Specifically, we find that predictive accuracy and the profitability of model-based trading are higher in financial markets with high (cf. medium) income levels. This is surprising since regulations and rules related to information disclosure should be further developed in such markets. On the other hand, the markets in the high- and medium income groups differ in many ways. It may be that the middle-income markets we consider are sufficiently regulated to rule out the effect of information disclosure policies on market predictability. An analysis of the predictability of low-income markets might give further insights into which other factors might govern the relationship between market income-level and predictability, and thus explain the result observed. Unfortunately, intraday time series data of sufficient length for markets with low income levels is not available from our data provider (TickWrite Data Inc.[5]). Consequently, a focused comparison of high *vs*. low income markets with forecasting horizons of one day or above might be a fruitful avenue for future research. An important implication of our findings is that the financial market selected for a forecasting study has a significant effect on the resulting accuracy of the forecasts. This implies that results which have been observed for one market may not be generalized to a different market. Drawing conclusions from findings derived from a single market, as is undertaken in the majority of existing ML studies, may mask the true degree of informational efficiency in financial markets. Rather, researchers are well-advised to test novel forecasting methods on multiple financial time series. This will offer a clearer picture of the relative advantages of competing methods and enhance the generalizability of empirical studies.

Our results offer support for hypotheses *H2a* and *H2b*, that predictive accuracy and the profitability from model-based trading are higher when evaluating forecasting models using a static train/test set approach (cf. a dynamic approach). The results suggest that the selection of the cross validation method is influential on prediction performance. This suggests that there are considerable risks in drawing conclusions regarding market efficiency based on the results of a single cross validation method.

The results suggest that the prediction performance of ML techniques are more sensitive to the size of samples than to the use of more recent price information. ML techniques make predictions by recognizing patterns. One potential price pattern is long-term memory, which describes stock price behavior with autoregressive models (Greene & Fielitz, 1977; Lo, 1991; Mandelbrot, 1971). Our results suggest that the size of the training sample is a key ingredient of prediction performance. In our experimental settings, there are over 700 samples in the training set for a static approach. The width of a sliding window is fixed at 50, so the training set is always the latest 50 samples before the predicted sample. Hence, the higher accuracy we observe from the static approach implies that the size of the training set is at least as important as recent price information in predicting stock prices. On the other hand, doubling the size of the training window to 100 observations, as we have done in the sensitivity analysis, has not increased predictive accuracy substantially over the 50 observation window size. This suggests that the trade-off between the recency and size of data in financial time series forecasting experiments is indeed complex and would benefit from future research. However, from a practical point of view, ignoring the possibility and necessity to update a prediction model, the static approach appears naïve and is very unlikely to be adopted. For example, consider an experiment where stock prices are forecast one day into the future and we have a

[5] http://www.tickdata.com/

financial time series of four years. Assuming there are roughly 350 trading days per year, this gives 1400 data points. A 70:30 train/test set split will reserve 420 of these data points (i.e., trading days) as a hold-out test set. Such an evaluation implies that a trader uses her stock prediction model for more than a year without updating. We argue that this is not realistic. More importantly, a static model methodology is vulnerable to selecting a test set that does not represent the overall population well (i.e., lucky sample effect).

Our results lead us to reject *H3a* and *H3b*. In particular, our results suggest that the performance of a forecasting model that includes technical indicators is not significantly better than one that uses basic reference prices as covariates. We find that even advanced nonlinear prediction methods such as SVM and ANN, which are able to discern complex relationships among covariates and the target variable, are unable to distill predictive information from technical indicators beyond that contained in basic price covariates. Consequently, our results further support previous criticism of technical indicators (Fama, 1970; Lesmond et al., 2004) and cast doubt on their value for predictive modeling.

Our results also suggest that the length of the forecast horizon has a significant effect on predictive accuracy and the profitability from model-based trading. However, the forecast horizon appears to have a different effect on the predictability (cf. profitability) of model-based trading. In particular, we find evidence in favor of *H4b*, in that the profitability of model-based trading is higher for an hourly (cf. daily) forecast horizon. On the other hand, predictive accuracy is significantly higher if forecasting price movements a day (cf. hour) into the future. Furthermore, the size of the forecast horizon effect on the profitability of model-based trading is small (Cohen's d equal to 0.1795), whereas the forecast horizon effect on predictive accuracy is actually the largest effect observed in the study (Cohen's d equal to 0.7105). We suggest that two effects maybe important to explain these results. First, a market that is more volatile is ceteris paribus harder to predict. Moreover, the (weak form) EMH suggests that stock prices can take some time to reflect all information. Consequently, volatility is higher in short-term forecasting, which, in turn, suggests lower predictive accuracy. However, in terms of the profitability of model-based trading, another effect comes into play. Second, forecasting shorter horizons creates more opportunities to trade on model predictions. We can only trade one daily forecast in a day, but multiple hourly forecasts. Furthermore, hourly forecasts and trading allow an investor to exploit the variation of prices within a day. In summary, the volatility argument predicts less accuracy from model-based trading for a forecasting horizon of one hour (cf. one day). However, the feasibility of higher trading frequency and the opportunity to capitalize on intraday price movements suggests that

ROI might be higher in the intraday setting. Consequently, the positive effect on profitability for higher frequency trading may be larger than the negative influence of volatility. Given that predictive accuracy is not directly linked to trading frequency and intraday price movements, it is certainly plausible that we observe higher accuracy for a daily forecast horizon. There, higher intraday volatility is the only relevant effect and this effect has a negative influence on accuracy.

It is important to note that the results observed in the forecast horizon comparison have implications for market efficiency. In particular, the review of previous forecasting studies indicates a bias toward forecast horizons of one day or above in the ML literature. Many such studies report very high accuracies (see Table 1), which, in the light of the EMH, is surprising. However, the results observed in this study demonstrate that predictive accuracy and trading profit are not perfectly correlated in financial time series forecasting. Market efficiency is only questioned if empirical evidence suggests that trading on model-based predictions produces excessive returns. Hence, observing high accuracy contradicts the EMH only if the corresponding prediction models facilitate profitable trading. We observe that this is not necessarily the case. More specifically, we find a forecast horizon of one day, the setting predominantly considered in previous ML research, to be associated with higher accuracy. However, we demonstrate that this forecast horizon is associated with less model-based trading profit than a shorter horizon of one hour. Consequently, the disagreement between the financial economics and the ML literatures concerning market efficiency might be less than the results of published studies suggest.

Our results also support *H5a* and *H5b* that SVM predict price movements more accurately than ANN and also produce higher trading profits. Furthermore, Cohen's d suggests that the effect is large (0.4154 for accuracy and 0.4842 for ROI). This finding is interesting given the recent history of the application of SVM and ANN. The former gained much popularity in early 2000, leading to a partial replacement of standard feedforward ANN. However, with the advent of extreme learning machines (G.-B. Huang, Chen, & Siew, 2006) and, more recently, deep neural networks (Schmidhuber, 2015), attention has shifted back to neural prediction models. It seems likely that such advanced types of neural networks will be applied in the domain of financial time series forecasting. An implication of our results is that studies which aim to investigate the potential of these and advanced prediction methods for financial time series forecasting should routinely compare them with SVM, since SVM predict price movements with high accuracy. In other words, SVM represents an important benchmark for assessing the marginal utility of new methods.

Our results also support *H6a* and *H6b* that the best ML model (which we show to be SVM) predicts price movements more accurately than the best econometric method (which we show to be AR) and also produces higher trading profits. Furthermore, Cohen's d suggests that the effect is sizeable (0.51 for accuracy and 0.66 for ROI). These results are in line with the literature that ML techniques outperform econometric methods (Donaldson & Kamstra, 1999; Pai & Lin, 2005). Our results, therefore, show that ML techniques, such as SVM and ANN, are useful techniques for detecting market anomalies. The conventional approach in the financial economics literature (Fama, 1970; Fama & French, 1993) is to use autocorrelation and linear regression models to examine the relation between explanatory factors and stock prices. (e.g., Keim, 1983; French, 1980). ML methods work in a different way. They are trained to recognize patterns in a data-driven manner and do not require human intervention. We find that such an approach facilitates profitable model-based trading, even when using models with fairly naive covariates. Both ANN and SVM achieve an ROI greater than one in most of our settings and the reference setting in particular (see Figures 1 and 3). Clearly, studies that scrutinize the degree to which financial markets are efficient must ensure that a modeling method is employed that fully exploits all predictive information contained in the covariates. Our results indicate that ML methods are well suited for this task and, thus, deserve a place in the financial economists' toolbox.

7. CONCLUSIONS

The EMH predicts that excess returns cannot be earned in a systematic way, by, for example, model-based trading. However, many ML-based financial time series forecasting studies seem to find ways to anticipate market developments with surprisingly high accuracy. The direction of price movements in these studies is often predicted with 80 percent accuracy and above. Some studies also report that their models facilitate profitable trading (Bitvai & Cohn, 2014). We set out to clarify the origins of the apparent contradiction between the ML and EMH literatures. To that end, we perform an extensive forecasting benchmark in which we use two established ML methods to predict price movements in most major stock markets and we compare, what we show to be, the best ML and econometric models, for predicting these movements. This study, to our best knowledge, is the first to compare intraday and daily ML and econometric prediction models across most major markets.

We find that the maturity of a financial market, the prediction method, the horizon for which it generates forecasts, and the methodology to simulate model-based trading all have a significant effect on market predictability and the feasibility of profitable model-based trading. Consequently, decisions that forecasting studies have taken with respect to these factors can help explain the results observed. This is not true for our last experimental factor, covariate composition, since we find that popular technical indicators are no more predictive than basic reference prices.

Overall, we do not find overwhelming evidence which contradicts the EMH, since the results cannot be said to pass Fama's (1998) tests of endurance, homogeneity and robustness; the EMH acknowledging that stock prices are partly predictable in the short run in some markets. Most of the predictive accuracies we observe are well below 60 percent. However, the level of accuracy we observe is substantially lower than that commonly published in previous research and it is acknowledged that we might have increased the levels of accuracy by including more sophisticated covariates. Consequently, we would suggest that inefficiency may exist in some markets.

However, the EMH is a theory related to general market behavior and acknowledges that anomalies can occur at certain times and in certain situations. In this context, the importance of our results lie in the insight they provide into the methodological issues which might explain the difference between the prevailing view in support of the EMH in the financial economics literature and the high accuracy in predicting financial market prices achieved in ML studies. Clearly, our results suggest that the econometric models generally employed in the financial economics literature may have led to overly pessimistic views of the degree to which financial price series can be predicted. It is clear from our results that practitioners interested in predicting financial time series are well-advised to consider ML techniques in their arsenal of methods. In fact, to aid this process, we show in Table B.1 in online Appendix B, the experimental settings that have given the best results for each individual market.

Furthermore, our analyses help to provide a realistic estimate of the potential and limitations of ML techniques for financial index forecasting. In particular, they suggest that the following factors may have given a false sense of the degree to which the results of ML studies contradict the EMH: First, there may be a bias toward the specific markets studied, with an emphasis on those that might be easier to predict. Second, there may have been a focus on less suitable model evaluation methodologies (i.e., a static approach as opposed to sliding-window cross-validation), which are prone to give optimistic estimates of model accuracy. Third, the focus on forecast horizons of one day and above in prior research may have led to a false impression of the predictability of financial markets. In summary, our study provides some evidence that it would be unwise to draw conclusions regarding the degree to which financial markets in general are inefficient from the accuracy figures reported in some previous ML financial time series forecasting studies.

The main implication of the EMH is that, in an informational efficient market, it is not possible to obtain systematic, excessive returns from trading the predictions of a forecasting model. The results of our study indicate an imperfect link between the predictive accuracy of a forecasting model and the profitability of trading on the model's forecasts. For example, we show that predicting the direction of price movements one day into the future produces higher accuracy compared to predicting prices one hour into the future, but produces a lower return from model-based trading. Consequently, through concentrating on forecast horizons of one day and above, the predictive accuracy that previous studies observe tends to be higher than what would be observed if examining shorter, intraday, forecast horizons. However, our results suggest that even then this does not necessarily imply that the corresponding models would facilitate profitable trading. For example, we find a forecast horizon of one day to be associated with relatively lower ROI and the returns observed have not taken account of transaction costs (which might well make our results consistent with the implications of the EMH (Malkiel, 2003)).

Our study has focused on predicting national stock indices because these indices are used in the majority of previous ML studies that predict direction of price changes. However, to confirm the robustness of our findings, it would be useful to extend the analyses to prediction of stock prices associated with a variety of industry sectors. It would also be useful to incorporate data from a wider range of middle income markets and to employ higher frequency data, once sufficient data becomes available. It should also be noted that due to data availability issues outlined in section 4.1, our main analyses were conducted using data from 2008-14, a period which covered two financial crises, the US subprime mortgage crisis and the European sovereign debt crisis. Clearly, these unexpected events may have influenced our results. However, additional analyses, reported in online Appendix C, for the 13 stock indices for which we were able to obtain a longer time series of data, produced similar conclusions to the earlier analyses. This gives some comfort that our results are robust to unexpected events such as financial crises. It would be valuable if future studies which employ longer time series of data could confirm this view.

Our results also have implications for the financial economics literature. In particular, we provide empirical evidence for the ability of advanced ML techniques in the form of SVM and ANN to detect market anomalies across many major financial markets. ML methods are rarely employed by financial economists. However, these techniques can capture complex nonlinear interactions in a financial data set and approximate their relationship to a target variable. This is probably the reason we find evidence that the best ML method (SVM) outperforms the best econometric method (AR) when predicting financial prices. Consequently, our results suggest that ML methods offer the prospect of studying informational efficiency and, thus, providing financial economists with new insights concerning the manner in which financial markets employ information.

References

Armano, G., Marchesi, M., & Murru, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, *170*, 3–33.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Systems with Applications*, *36*, 5932–5941.

Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, *21*(1), 167–183.

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, *42*(20), 7046–7056.

Berry, M. J., & Linoff, G. (1997). *Data mining techniques: for marketing, sales, and customer support*. USA: John Wiley & Sons, Inc.

Bitvai, Z., & Cohn, T. (2014). Day trading profit maximization with multi-task learning and technical analysis. *Machine Learning*, 1–23.

Bodyanskiy, Y., & Popov, S. (2006). Neural network approach to forecasting of quasiperiodic financial time series. *European Journal of Operational Research*, *175*(3), 1357–1366.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1–8.

Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, *40*(3), 793–805.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, *47*(5), 1731–1764.

Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge university press.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531.

Chang, P.-C., Liu, C.-H., Lin, J.-L., Fan, C.-Y., & Ng, C. S. (2009). A neural network with a case based dynamic window for stock trading prediction. *Expert Systems with Applications*, *36*, 6889–6898.

Charles, A., Darné, O., & Kim, J. H. (2011). Small sample properties of alternative tests for martingale difference hypothesis. *Economics Letters*, *110*(2), 151–154.

Chen, A.-S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index. *Computers & Operations Research*, *30*, 901–923.

Chen, K.-Y., & Ho, C.-H. (2005). An improved support vector regression modeling for Taiwan Stock Exchange market weighted index forecasting (Vol. 3, p. nil15-1638). Presented at the Neural Networks and Brain, 2005. ICNN&B'05. International Conference on, IEEE.

Chen, W.-H., Shih, J.-Y., & Wu, S. (2006). Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *International Journal of Electronic Finance*, *1*(1), 49–67.

Chen, W.-S., & Du, Y.-K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, *36*, 4075–4086.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*(1), 113–126.

Cheung, Y.-W., & Lai, K. S. (1995). A search for long memory in international stock market returns. *Journal of International Money and Finance*, *14*(4), 597–615.

Choong, C.-K., Baharumshah, A. Z., Yusop, Z., & Habibullah, M. S. (2010). Private capital flows, stock market and economic growth in developed and developing countries: A comparative analysis. *Japan and the World Economy*, *22*(2), 107–117.

Chordia, T., Goyal, A., Lehmann, B. N., & Saar, G. (2013). High-frequency trading. *High-Frequency Trading*, *16*(4), 637–645. http://doi.org/10.1016/j.finmar.2013.06.004

Claessens, S., Klingebiel, D., & Schmukler, S. L. (2006). Stock market development and internationalization: Do economic fundamentals spur both similarly? *Journal of Empirical Finance*, *13*(3), 316–350.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic press.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Dai, W., Shao, Y. E., & Lu, C.-J. (2012). Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Computing and Applications*, *23*(6), 1551–1561.

de Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to Prediction of Stock Price and Improvement of the Directional Prediction Index-

Case Study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, *40*(18), 7596–7606.

Dhar, R., & Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, *52*(5), 726–740.

Doeksen, B., Abraham, A., Thomas, J., & Paprzycki, M. (2005). Real stock trading using soft computing models (Vol. 2, pp. 162–167). Presented at the Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on, IEEE.

Donaldson, R. G., & Kamstra, M. (1999). Neural network forecast combining with interaction effects. *Journal of the Franklin Institute*, *336*(2), 227–236.

Fama, E. F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, *46*(5), 1575–1617.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, *105*(3), 457–472.

French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, *8*(1), 55–69.

Gerlein, E. A., McGinnity, M., Belatreche, A., & Coleman, S. (2016). Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, *54*, 193–207.

Greene, M. T., & Fielitz, B. D. (1977). Long-term dependence in common stock returns. *Journal of Financial Economics*, *4*(3), 339–349.

Griffin, J. M., Kelly, P. J., & Nardari, F. (2010). Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. *Review of Financial Studies*, *23*, 3225–3277.

Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? *The Journal of Finance*, *56*, 589–616.

Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*, *23*, 800–808.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, *21*(1), 1–14.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications*, *33*, 171–180.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning.* New York: Springer.

Huang, C.-J., Yang, D.-X., & Chuang, Y.-T. (2008). Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, *34*, 2870–2878.

Huang, G.-B., Chen, L., & Siew, C.-K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Neural Networks, IEEE Transactions on*, *17*(4), 879–892.

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, *32*, 2513–2522.

Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, *207*(3), 1702–1716.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. Australia: OTexts.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, *6*(2), 95–101.

Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*, 5311–5319.

Keerthi, S. S., & Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, *15*(7), 1667–1689.

Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, *12*(1), 13–32.

Kim, J. H., & Shamsuddin, A. (2008). Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance*, *15*(3), 518–532.

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1), 307–319.

Kuo, R. J., Chen, C., & Hwang, Y. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, *118*, 21–45.

Lee, M.-C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, *36*(8), 10896–10904.

Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, *71*(2), 349–380.

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136.

Lessmann, S., Sung, M., & Johnson, J. E. (2011). Towards a methodology for measuring the true degree of efficiency in a speculative market. *Journal of the Operational Research Society*, *62*, 2120–2132.

Lo, A. W. (1991). Long-Term Memory in Stock Market Prices. *Econometrica*, *59*(5), 1279–1313.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, *55*(4), 1705–1770.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *The Journal of Economic Perspectives*, *17*(1), 59–82.

Mandelbrot, B. B. (1971). When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models. *The Review of Economics and Statistics*, *53*(3), 225–236.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *High-Frequency Trading*, *16*(4), 712–740.

Mittermayer, M.-A. (2004). Forecasting Intraday stock price trends with text mining techniques. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, 10 pp.

Ojah, K., & Karemera, D. (1999). Random walks and market efficiency tests of Latin American emerging equity markets: a revisit. *The Financial Review*, *34*(2), 57.

Ormoneit, D., & Neuneier, R. (1996). Experiments in predicting the German stock index DAX with density estimating neural networks. *Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on*, 66–71.

Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science*, *3*(12), 28.

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497–505.

Pan, H., Tilakaratne, C., & Yearwood, J. (2005). Predicting Australian stock market index using neural networks exploiting dynamical swings and intermarket influences. *Journal of Research and Practice in Information Technology*, *37*, 43–56.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268.

Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, *26*, 25–33. http://doi.org/10.1007/s10489-006-0001-7

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, *27*, 12.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, *40*(3), 777–790.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222.

Steiner, M., & Wittkemper, H.-G. (1997). Portfolio optimization with a neural network implementation of the coherent market hypothesis. *European Journal of Operational Research*, *100*(1), 27–40.

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, *29*(4), 309–317.

Tsai, C.-F., & Hsiao, Y.-C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, *50*, 258–269.

Tseng, C.-H., Cheng, S.-T., Wang, Y.-H., & Peng, J.-T. (2008). Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices. *Physica A: Statistical Mechanics and Its Applications*, *387*, 3192–3200.

Vanstone, B., Finnie, G., & Tan, C. (2005). Evaluating the application of neural networks and fundamental analysis in the Australian stockmarket. In *Information Technology papers*. Calgary, AB, Canada.

Xu, S., & Chen, L. (2008). A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. Presented at the 5th International Conference on Information Technology and Applications (ICITA 2008), Cairns, Queensland, Australia.

Żbikowski, K. (2015). Using Volume Weighted Support Vector Machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications*, *42*(4), 1797–1805.

Zhang, D., Jiang, Q., & Li, X. (2004). Application of neural networks in financial data mining. *International Journal of Computational Intelligence*, *1*, 116–119.

Zhang, Y., & Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, *36*, 8849–8854.

Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G., & Rosso, O. A. (2009). Forbidden patterns, permutation entropy and stock market inefficiency. *Physica A: Statistical Mechanics and Its Applications*, *388*(14), 2854–2864.