ScienceDirect

Applied Soft Computing
Volume 71, October 2018, Pages 685-697

Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models

Bin Weng [a], Waldyn Martinez [b], Yao-Te Tsai [c], Chen Li [d], Lin Lu [e], James R. Barth [f], Fadel M. Megahed [g]

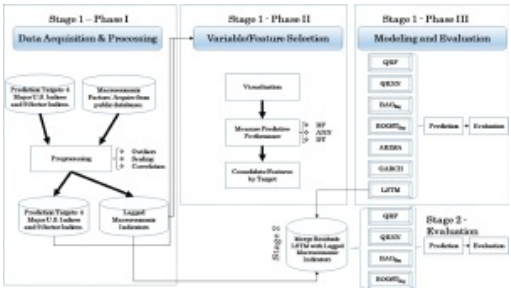
Show more

https://doi.org/10.1016/j.asoc.2018.07.024
Get rights and content

Abstract

This paper proposes a two-stage approach that can be used to investigate whether the information hidden in macroeconomic variables (alone) can be used to accurately predict the one-month ahead price for major U.S stock and sector indices. Stage 1 is constructed to evaluate the hypothesis that the price for different indices is driven by different economic indicators. It consists of three phases. In phase I, the data is automatically acquired using freely available APIs (application programming interfaces) and prepared for analysis. Phase II reduces the set of potential predictors without the loss of information through several variable selection methods. The third phase employs four ensemble models and three time-series models for prediction. The prediction performance of the seven models are compared using the Mean Absolute Percent Error (and two additional metrics). If the hypothesis were to be true, one expects that the performance of the ensemble models to outperform the time-series models since the information in the economy is more important than the information in previous prices. In Stage 2, a hybrid approach of the recurring neural network used for time-series prediction (i.e., the LSTM) and the ensemble models is constructed to examine the secondary hypothesis that the residuals from the time-series models are not random and can be explained by the macroeconomic indicators. To test the two hypotheses, the monthly closing prices for 13 U.S. stock and sector indices and the corresponding values for 23 macroeconomic indicators were collected from 01/1992–10/2016. Based on the case study, the four ensembles prediction performance were superior to that of the three time-series models. The MAPE of the best model for a given index was < 1.87%. The Stage 2 results also show that the three evaluation metrics (RMSE, MAPE and MAE) can be typically improved by 25–50% by incorporating the information hidden in the macroeconomic indicators (through the ensemble approach). Thus, this paper shows that, for the analysis period and the indices studied, the macro-economic indicators are leading predictors of the price of 13 U.S. sector indices.

Graphical abstract



Download : Download high-res image (174KB)

Download : Download full-size image

Introduction

The prediction of stock prices has continued to fascinate both academia and business. The question driving early stock market research was: "To what extent can the past history of a common stock's price be used to make meaningful predictions concerning the future price of the stock?" [1]. The *Efficient Market Hypothesis* (EMH) [1] and the *Random Walk Theory* [2] provided a theoretical foundation for tackling this question. These models posited that stock prices cannot be forecasted since they are driven by new information and not present/past prices. Thus, prices will follow a random walk and cannot be predicted accurately.

There has been an increasing number of studies [3], [4], [5], [6], [7] that provide evidence contrary to what is suggested by the EMH and random walk theory. These studies show that the stock market can be predicted to some degree and therefore, questioning the EMH's underlying assumptions. Moreover, many practitioners refer to two main examples, which demonstrate that the stock market can be accurately predicted: (a) Warren Buffet's ability to consistently beat the S&P index [8], [9]; and (b) the successful prediction of the 2008 Stock Market crash based on the "housing bubble", which was popularized by the *New York Times Bestseller* book (turned movie): "The Big Short: Inside the Doomsday Machine" [10].

The literature on stock market prediction can be divided, based on the type of prediction models used, into statistical time-series models and machine learning techniques [11]. Based on the review of [12], autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroscedasticity (GARCH) are the most commonly used time-series approaches for stock prediction. Despite their widespread use, limitations of these models include: (a) needing the model to be prespecified [13]; (b) an increased effect of estimation error as the models become more complex [13]; and (c) sub-par predictive ability when compared to machine learning methods [14], [15], [16]. On the other hand, the machine learning (ML) techniques can be categorized into (a) non-voting approaches, which include artificial neural networks (ANNs) [17], support vector machines (SVM) [18], [19], [20], and classification & regression trees (CART) [20], [21]; and (b) voting/ensembles [22], [23], [24] and hybrid methods [25], [26], [27]. The reader is referred to [14], [15] for detailed reviews of ML stock market prediction methods.

Based on the above discussion and the reviews of [12], [14], [15], [17], there are four important observations to be made. First, most (if not all) of the stock market prediction papers used some form of the previous price as a predictor/feature. In our estimation, this can be explained by the following logic: if the market can be predicted, then the previous price (or a feature based on it, e.g., through a technical indicator) should explain some of the variation in prices/returns. Second, only a small subset of ML papers considered using macroeconomic predictors [28], [29], [30], [13], [19] (see Table 1 for a summary of their contributions). In our estimation, this can be explained by the following: (a) majority of stock market prediction papers focus

on short-term predictions, and (b) macroeconomic indicators are released, at best, monthly. Thus, any paper focusing on the short-term prediction cannot utilize an invariant predictor. The third observation relates to the papers discussed in Table 1. These papers typically showed that macroeconomic indicators can be strong predictors of price (when machine learning models are used). However, these papers generally: (a) had a single index, and (b) utilized both macroeconomic indicators and past prices as predictors so it is not clear how generalizable the results are (to other indices and whether macroeconomic indicators alone can be predictors of future prices). Fourth, the use of ensemble and hybrid-based approaches improves the prediction results through voting/averaging, which is an expected result based on the data mining literature. Based on these observations, this paper will investigate the utility of macroeconomic variables (including those highlighted in [10]) in predicting the one-month ahead price for major U.S stock and sector indices. *The overarching hypothesis is that the price for different indices can be predicted fairly accurately by different economic indicators.* Such effects will be quantified/validated using a novel soft computing approach.

In this paper, the main research questions are: (a) to examine whether macroeconomic factors can predict the 1-month ahead price of four major U.S stock indices (*the Dow Jones Industrial Average Index,* $DJI, *the NYSE Composite Index,* $NYA, *the NASDAQ Composite Index,* $IXIC, and the *S&P 500 Index,* $GSPC) and the nine U.S. sector indices; and (b) if the answer to question (a) is "yes", then which factors are the most predictive to each index. To examine these research questions, a two-stage experimental-based framework is proposed.

The first stage is comprised of two main phases. In phase I, an automated data acquisition procedure is developed to capture the monthly values for the different macroeconomic factors (i.e., the independent variables) and closing price for different stock indices (i.e., the response variables). In phase II, four ensemble models and three time-series models are used for predicting the closing price of the different indices. The ensemble models chosen for the analysis are: (i) quantile regression random forest (QRF), (ii) quantile regression neural network ensemble (QRNN), (iii) bagging regression ensemble (BAG_{Reg}), and (iv) boosting regression ensemble ($BOOST_{Reg}$). These have been chosen since they are the most commonly used ensembles for continuous predictions. The performance of these ensembles are then compared with ARIMA and GARCH models, as well as a deep long-term memory recurrent neural network (LSTM) for time-series forecasting [32] (see recent applications to stock predictions in [33], [34], [35]). If the overarching hypothesis in this paper is true (i.e., medium term index prices are driven by macro-economic factors), then one would expect that the performance of the ensemble methods would outperform the time-series methods since the information affecting the medium-term price is in the economy (and not contained completely in past prices).

To validate the insights gained from the first stage, a hybrid approach of the LSTM and the ensemble models will be constructed and utilized in the second stage. In the hybrid approach, the LSTM model (chosen given its nonparametric nature) is used to predict the closing price of the different indices (i.e., the same approach from stage 1). The residuals from this model are then used as a target for prediction (i.e., the dependent variable) for the four ensemble models, then the predictions from the LSTM model are adjusted by adding the corrections predicted by the ensembles. Just as the first stage, the macroeconomic indicators are used to predict the 1-month ahead residuals from the LSTM model. Thus, the hybrid approach is used to test the following secondary hypothesis: *the errors/residuals from the time-series models are not completely random and can be explained by the macroeconomic indicators.*

The remainder of this paper is organized as follows. Section 2 presents the macroeconomic factors that are used in this study and provides a justification for their selection. In Section 3, the proposed two-stage approach is detailed. The results of the experimental study for Stages 1 and 2 are shown in Sections 4 and 5,

respectively. Finally, the main contributions of the paper, its practical implications, and suggestions for future work are highlighted in Section 6.

Section snippets

Justification for the macroeconomic indicators used

Researchers list several different macroeconomic factors that could potentially have an impact on stock market movements including oil prices [36], [37], [38], housing prices [39], interest rates [40], foreign markets [41], and inflation [21]. Ref. [42] explored the effects of important macroeconomic variables on stock market returns. From the results, they concluded that industrial production, risk premium change, yield curve twist, and inflation all have significant effects on the variability …

Two-stage approach to demonstrate the utility of macroeconomic indicators in predicting monthly stock prices

Fig. 1 presents the process to build up the model. In Stage 1, the data from several different online resources are first collected. The data acquisition phase is divided into two main steps, where the dependent indices' monthly closing prices and the independent macroeconomic predictors (used in the ensemble models) are obtained. Then, in phase II the variables are selected using three machine learning models and consolidated into one final set of features. Phase III compares the seven…

Stage 1: Results and discussion

In this section, the Stage 1 results for the proposed method are presented. First, the phase I results are highlighted, where irrelevant and redundant features that do not contribute to, or have a minimal contribution to, the predictive models are identified. Then, the results of the seven prediction models (four ensembles and three time-series models are presented). The performance of these models are compared using three metrics as mentioned in Section 3.1.3. To facilitate the replication of…

STAGE 2: Experimental results and discussion

In this section, the evidence supporting the secondary hypothesis: "*the errors/residuals from the time-series models are not entirely random and can be explained by the macroeconomic indicators*" is evaluated. As explained in Section 3, the results for the proposed hybrid Deep LSTM-Ensemble formulation are presented here. Recall that the residuals from the Deep LSTM model e_t are used as Target for the four ensembles analyzed. The predicted residuals $\hat{e}_t$ are then used to correct the errors in the …

An overview of the impacts and contributions of this paper

The overarching goal behind this paper was to investigate if macroeconomic indicators are drivers for the monthly prices of the main stock and sector indexes in the U.S. To investigate this hypothesis, a two-stage approach was proposed. The first stage was comprised of three phases. In phase I, the data from 01/1992 to 10/2016 was acquired, covering the monthly values of 13 major indexes and 23 potentially relevant macroeconomic indicators. Phase II involved the use of variable selection…

References (80)

J. Bollen *et al.*
Twitter mood predicts the stock market
J. Comput. Sci. (2011)

J.-J. Wang *et al.*
Stock index forecasting based on a hybrid model
Omega (2012)

D. Enke *et al.*
The use of data mining and neural networks for forecasting stock market returns
Expert Syst. Appl. (2005)

G.S. Atsalakis *et al.*
Surveying stock market forecasting techniques – Part II: Soft computing methods
Expert Syst. Appl. (2009)

K.-j. Kim
Financial time series forecasting using support vector machines
Neurocomputing (2003)

W. Huang *et al.*
Forecasting stock market movement direction with support vector machine
Comput. Oper. Res. (2005)

C.-F. Tsai *et al.*
Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches
Decis. Support Syst. (2010)

C.-F. Tsai *et al.*
Predicting stock returns by classifier ensembles
Appl. Soft Comput. (2011)

N.F. Da Silva *et al.*
Tweet sentiment analysis with classifier ensembles
Decis. Support Syst. (2014)

L.-J. Kao *et al.*
A hybrid approach by integrating wavelet-based feature extraction with mars and SVR for stock index forecasting
Decis. Support Syst. (2013)

View more references

Cited by (26)

Stock market predictor using prescriptive analytics

2023, Materials Today: Proceedings

Show abstract

Intelligent candlestick forecast system for financial time-series analysis using metaheuristics-optimized multi-output machine learning

2022, Applied Soft Computing

Show abstract

Structural rule-based modeling with granular computing

2022, Applied Soft Computing

> *Citation Excerpt :*
>
> …Subsequently, monitoring and studying these systems' dynamic behaviors becomes difficult and significantly important for industry safety and social economic development [1]. To accomplish system behavior analysis, a feasible way is to study system's output behaviors or variance of a specific variable in applications, e.g., industrial and engineering analysis, social science and economic study [2,3]. It is known that conventional analysis was realized by physical models based on system's mechanism or statistical models from data [4]….

Show abstract

AComNN: Attention enhanced Compound Neural Network for financial time-series forecasting with cross-regional features

2021, Applied Soft Computing

> *Citation Excerpt :*
>
> …The detailed label, feature generation procedures and their configuration in our experiment are shown in the later parts. Although lots of literature have proposed various determinant indicators that are highly related to stock index price movement of HSI, S&P 500, DJIA, NASDAQ composite index, and FTSE 100 [13,16,19,26–30], there is no unified criterion on related feature selection. In this work, our raw data are retrieved from regions of the U.K., the U.S., and China, including Open, High, Low, Close price, and Volume (called as OHLCV variables) of FTSE 100, S&P 500, and HSI as well as regional indicators such as macro-economic indicators, commodity indicators, currency indicators….

Show abstract

A two-stage machine learning framework to predict heart transplantation survival probabilities over time with a monotonic probability constraint

2020, Decision Support Systems

> *Citation Excerpt :*
>
> …In the context of ML methodologies, our framework represents an extension to the use of "hybrid" methodologies. Existing "hybrid" methods typically incorporate two or more approaches for the purposes of improving prediction accuracy [e.g., see 48]. However, our proposed framework introduces the idea of using a hybrid approach to constrain the predictions from the initial ML modeling stage, which can influence the development of other "hybrid" approaches where different criteria for calibration are to be enforced….

Show abstract

Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange

2020, Applied Soft Computing Journal

Show abstract

View all citing articles on Scopus

Recommended articles (6)

Research article

A novel optimal replication allocation strategy for particle swarm optimization algorithms applied to simulation optimization problem

Applied Soft Computing, Volume 71, 2018, pp. 591-607

Show abstract

Research article

Shape optimisation method based on the surrogate models in the parallel asynchronous environment

Applied Soft Computing, Volume 71, 2018, pp. 1189-1203

Show abstract

Research article

Enhancing robustness of the inverted PBI scalarizing method in MOEA/D

Applied Soft Computing, Volume 71, 2018, pp. 1117-1132

Show abstract

Research article

A 2-dimension uncertain linguistic DEMATEL method for identifying critical success factors in emergency management

Applied Soft Computing, Volume 71, 2018, pp. 386-395

Show abstract

Research article

Green Swarm: Greener routes with bio-inspired techniques

Applied Soft Computing, Volume 71, 2018, pp. 952-963

Show abstract

Research article

A novel phase I fuzzy profile monitoring approach based on fuzzy change point analysis

Applied Soft Computing, Volume 71, 2018, pp. 488-504

Show abstract

View full text