Knowledge-Based Systems 50 (2013) 151–158

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Posterior probability model for stock return prediction based on analyst's recommendation behavior

Jiangjiao Duan [a,*], Hongzhong Liu [a], Jianping Zeng [b]

[a] School of Economics, Fudan University, 200433 Shanghai, PR China
[b] School of Computer Science, Fudan University, 200433 Shanghai, PR China

ARTICLE INFO

Article history:
Received 16 December 2012
Received in revised form 5 June 2013
Accepted 13 June 2013
Available online 25 June 2013

Keywords:
Stock recommendations
Return prediction
Bayesian inference
Posterior return distribution
Behavioral finance

ABSTRACT

Existing studies on stock return forecasting mainly formulate the issue in a numeric analysis framework. Various kinds of time series models and optimization methods are applied. In this paper, we explore a new prediction approach based on the behavior of analysts' recommendations. By combining each recommendation and stock return, a posterior probability model associated with an analyst's recommendation is built based on Bayesian inference. It provides an estimation of stock return distribution for next several days after recommendation, and thus serves as a novel predictor from point of view of behavioral finance. Based on the empirical studies on China stock market, we demonstrate the superior forecasting performance over traditional methods. The model's maximum accuracy can be reached between 84.3% and 94.2%. The average accuracy falls between 58.6% and 60.3%, while it is just from 43.5% to 56.2% or lower by traditional prediction methods. We also find that most of the analysts can produce recommendations which fitness lies between 0.5 and 0.6 at the successive recommendation time. The finding is in accordance with early conclusion which indicates that stock analysts tend to maintain their reputation when they issue recommendations. The consistency also confirms the effectiveness of the proposed method.

1. Introduction

Forecasting stock price and return have attracted researchers' attention for several decades [1–5]. Accurate prediction might provide a sound foundation for government administrators to maintain the order of stock markets. On the other hand, it also helps investors to obtain more profits, and to avoid investment risk.

By considering the returns as a time series, most of the forecasting methods utilize statistical models to reveal the hidden properties in the sequence, such as dependency and data distribution. Several kinds of statistical models, such as AR process [6], auto-regression moving average (ARMA), Markov process and the artificial neural network model [7], generalized auto-regression conditional heteroscedasticity (GARCH) model [8,9], hidden Markov model (HMM) [2], hierarchical hidden Markov chain model [10], and so on, have been employed to describe the statistical characteristics.

The predictability of intraday stock market returns has been studied by using both linear and nonlinear time series models, including simple autoregressive, Markov switching and support vector machine models. The empirical results indicated that nonlinear models outperformed linear models [11]. The co-integration concept and the related vector error correction model (VECM) were employed to model daily high and daily low of stock price [13]. A fusion model combining the HMM, artificial neural network (ANN) and a genetic algorithm (GA) was proposed to achieve better forecast performance [12]. Model-based prediction methods are helpful in obtaining desired accurate results when the market trading is stable.

However, parameters optimization is a critical problem in utilizing the time series based prediction models. Several methods have been devised for this problem. Artificial fish swarm algorithm (AFSA) and a *K*-means clustering algorithm were proposed to optimize basis function neural network (RBF) which was employed to forecast the stock indices of the Shanghai Stock Exchange [14]. A GA-based algorithm was used to build an associative classifier that can discover trading rules from the numerical indicators [15]. An integrated approach based on genetic fuzzy systems (GFS) and artificial neural networks (ANN) for constructing a stock price forecasting expert system was proposed to achieve best results using minimum required input data [3]. In order to provide flexible methods to adjust the model parameters for further enhancement on the prediction models, a new architecture entitled as the Multi-level and Interactive Stock Market Investment System (MISMIS) which combined different areas – financial economics, prediction techniques, and dynamical systems theory to handle financial data

* Corresponding author.
E-mail address: jjduan@fudan.edu.cn (J. Duan).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.06.007

was developed for investors to build their financial models to forecast stock price [16]. In addition, over fitting is one of the main problems in model optimization methods since they try to maximize objective function with limited number of return time series.

In this paper, we apply behavioral finance theory to establish a new prediction approach. As we know, the stock market is a complex system that is characterized by dynamics, uncertainty, and chaos. The rate of price fluctuations depends on many factors, such as interest rate, temporary disturbance and merger of large financial corporations [17]. Behavioral finance theory emphasizes that the uncertainty and chaos in stock markets are largely due to information asymmetric. Financial analysts who can effectively process all kinds of information related to stock market are generally regarded as an important information source for stock investors [18]. They play an important role in promoting market efficiency by reducing the information asymmetry between company management and external market participants.

Hence, researches on the behavior of security analysts' forecasting and recommendation have attracted a lot of attention recently. Ref. [19] proposed a rational learning-based explanation for the predictability in financial analysts' earnings forecasting errors. Using simulations and real data, they showed that the predictability evidence was more consistent with rational learning than with irrationality. Some literatures provided evidences to clarify analyst's ability in predicting stock price or influencing price [20–22]. A portfolio of the stocks with the most (least) favorable consensus analyst recommendations was found to provide an average annual abnormal gross return of 4.13 (−4.91) percent, after controlling for market risk, size, book-to-market, and price momentum effects [20]. The persistence in stock picking ability of security analysts was verified [23]. It was indicated that the performance of recommendations-based investment strategies could be significantly enhanced by transacting quickly following recommendation changes [21]. It was also demonstrated the usefulness of forecast-derived information in characterizing analysts' recommending behavior, and their incremental informativeness even in the presence of information obtained from past recommendations [22]. Barniv et al. used large international samples and found that the relation between stock recommendations and future returns was negative in high investor participation countries but positive in low investor participation countries [24].

Bayesian rule was employed to model analyst's predictive ability recently since it can provide a reasonable approach to update investor's brief. Chen et al. developed a predictive model of investor learning in Bayesian manner about analysts' ability to forecast quarter profit [25]. The model showed that when investors rely on an analyst's record of history forecasting performance to update their beliefs about the analyst's true predictive ability, they tend to increase weights on the analyst's past performance and decrease weights on prior perceived ability.

We propose to model stock return based on the behavior of analysts' recommendations and Bayesian rule. Different from the approach in [25], we do not require a common prior which is an unconditionally normally distribution about the analyst's true predictive ability. Hence, the model can be more appropriate in actual applications. The performance of the model is evaluated in terms of prediction accuracy which is commonly used in other similar researches. In addition, stability and persistence are also proposed to evaluate the model so that we can gain more intuition about the new model.

This study is different from prior work in respect to the prediction method and the evaluation measurements.

Firstly, we address the problem of stock return prediction by utilizing analysts' recommendations issued at arbitrary date interval. The new method is based on the theory of behavioral finance which emphasizes on the relation between participant's activities and the movement of stock price.

Secondly, we establish the posterior return distribution based on Bayesian inference. It provides an estimation of return distribution for the next few days after recommendation according to Bayesian rule. Thus it can serve as a novel predictor from point of view of behavioral finance. Hence, the prediction method is basically different from time series analysis methods, such as ARMA, Markov process, and so on.

Thirdly, the performance of the predictive model is extensively studied. The accuracy of model prediction and its variance are investigated by varying ***TL*** (Training Length) and ***DRR*** (Date Range of Recommendation). The ability in persistence of prediction is also investigated. The result is in accordance with early conclusion which indicates that stock analysts try to maintain their reputation when they publish recommendations. This finding also confirms the effectiveness of the proposed method.

The remainder of this article is organized as follows. Data preparation for prediction modeling is described in the next section. The model structure and model training process are described in Section 3. Then, the evaluation and results are provided in Section 4, and the final section reports conclusion and future work.

2. Data preparation for prediction modeling

Although there are many researches on the prediction of daily returns in different stock markets including US, France, Germany, UK, Cyprus and Athens [7], we pay more attention to China stock market. China stock market is well-known for being full of much uncertainty, which leads to unsatisfactory performance in return prediction based on statistical time series analysis methods. Hence we select China stock market as the data source to evaluate the proposed model with aim at providing a new method for return prediction in other similar stock markets.

In order to evaluate the model, we need two data collections which describe stock price and analysts' recommendations, respectively. The data collections, that is, recommendations and stock returns, are obtained by the following ways.

Since the model is actually inferred by investors based on the observations of recommendations and stock return, the recommendations are required to be publicly available. Hence, we get the recommendation records from public Internet web sites, instead of some special databases which are not free to be accessed. A well-known Chinese website named sina-finance (http://finance.sina.com.cn) provides a large number of messages about several kinds of financial products, such as funds and stocks. All the China stocks, including those in Shanghai market and Shenzhen market are indexed by the website. News and notices about each stock company are also published in time on the website. Most important of all, a large number of stock recommendations are also publicly accessible on the website. Thus, it provides us a free and suitable way to get the recommendation information.

An automatic computer program which simulates the working process of Web browser in dealing with web pages is designed to launch HTTP (Hypertext Transfer Protocol) type requests to get recommendation web pages from sina-finance. Then the recommendation records are extracted from the pages using a parser-based algorithm [26]. As a result, a dataset which contains stock recommendations published from January 1, 2009 to December 31, 2010 is constructed. Each record in the dataset includes four fields which are as follows:

(1) analyst name
(2) stock code
(3) recommendation date