

Chapter III

Automatically Identifying Predictor Variables for Stock Return Prediction

Da Shi

Peking University, China

Shaohua Tan

Peking University, China

Shuzhi Sam Ge

National University of Singapore, Singapore

ABSTRACT

Real-world financial systems are often nonlinear, do not follow any regular probability distribution, and comprise a large amount of financial variables. Not surprisingly, it is hard to know which variables are relevant to the prediction of the stock return based on data collected from such a system. In this chapter, we address this problem by developing a technique consisting of a top-down part using an artificial Higher Order Neural Network (HONN) model and a bottom-up part based on a Bayesian Network (BN) model to automatically identify predictor variables for the stock return prediction from a large financial variable set. Our study provides an operational guidance for using HONN and BN in selecting predictor variables from a large amount of financial variables to support the prediction of the stock return, including the prediction of future stock return value and future stock return movement trends.

INTRODUCTION

The stock return prediction, including both the future stock return value prediction and the future stock return movement trends prediction,

has gained unprecedented popularity in financial market forecasting research in recent years (Keim & Stambaugh, 1986; Fama & French, 1989; Basu, 1977; Banz, 1980; Jegadeesh, 1990; Fama & French, 1992; Jegadeesh & Titman, 1993; Lettau

& Ludvigson, 2001; Avramov & Chordia, 2006a; Avramov & Chordia, 2006b). Because any current stock market is not “efficient,” researchers believe that appropriate techniques can be developed for the prediction of the stock return for a certain period of time to allow investors to benefit from the market inefficiency. Actually, some previous works have proved this point of view to a certain extent (Fama & French, 1989; Fama & French, 1992; Avramov & Chordia, 2006b, Ludvigson & Ng, 2007). In general, stock return prediction can be divided into two steps:

1. Identifying those predictor variables which can explain the stock return closely
2. Setting up a linear or nonlinear model which expresses qualitative or quantitative relationships between those predictor variables and the stock return. The stock return is then predicted by computing these models.

Obviously, the first step is the foundation of the prediction. However, there has not been a systematic technique developed in the past to effectively implement this step. This chapter focuses on developing an effective technique for this purpose.

There exist a large number of financial variables for a stock market (typically, over 100 variables or more), but not all of them are directly relevant to the stock return. Researchers always want to identify, among this large set of variables, those underlying predictor variables with a prominent influence on the stock return to support their further prediction. However, in the past two decades, because there have not been effective tools to fulfill this task, researchers have to select predictor variables manually according to their domain knowledge and experience or simply forced to use all the available financial variables when they want to predict the stock return (Fama & French, 1989; Fama & French, 1992; Kandel & Stambaugh, 1996; Lettau &

Ludvigson, 2001; Avramov & Chordia, 2006a; Avramov & Chordia, 2006b).

Although the domain knowledge and experience may provide some help in selecting predictor variables, relying on them alone often causes the following two problems which prevent them from obtaining quality predictive results:

1. Because different researchers may have different domain knowledge and experiences, selecting predictor variables manually may introduce researchers’ subjective biases, even some wrong information into the prediction procedure.
2. Another problem of manual selection is that in many cases, the domain knowledge or experience may not at all be sufficient to determine whether some financial variables will influence the stock return or not. A trial and error approach is often resorted to in order to test out each of these variables and their combinations to ascertain the relevance, leading to too large a test problem to handle computationally.

Our main objective in this study is to develop a constructive technique that can effectively select predictor variables for stock return prediction from a large number of financial variables (typically, over 100 variables or more) automatically to overcome the above disadvantages caused by manual selection. The technique consists of a top-down part and a bottom-up part. In the top-down part, we define an information credit for each candidate financial variable based on an artificial Higher Order Neural Network (HONN) model. A heuristic selection procedure will go through all the candidate variables, compute the information credit value for each of them and select the k variables with the highest value. Simultaneously, in the bottom-up part, a Bayesian Network (BN) is used to build up all the relationships that exist among all these variables using a directed acyclic graph. Using the generated BN graph, a validation

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/automatically-identifying-predictor-variables-stock/5277

Related Content

Analysis of Precipitation Variability using Memory Based Artificial Neural Networks

Shyama Debbarma, Parthasarathi Choudhury, Parthajit Roy and Ram Kumar (2022). *Research Anthology on Artificial Neural Network Applications* (pp. 955-970).

www.irma-international.org/chapter/analysis-of-precipitation-variability-using-memory-based-artificial-neural-networks/288994

Comprehensive Modelling of ANN

Meghna Babubhai Patel, Jagruti N. Patel and Upasana M. Bhilota (2022). *Research Anthology on Artificial Neural Network Applications* (pp. 31-40).

www.irma-international.org/chapter/comprehensive-modelling-of-ann/288949

Neuroglial Behaviour in Computer Science

Ana B. Porto and Alejandro Pazos (2006). *Artificial Neural Networks in Real-Life Applications* (pp. 1-21).

www.irma-international.org/chapter/neuroglial-behaviour-computer-science/5361

Tripartite and Quadpartite Size Ramsey Numbers for All Pairs of Connected Graphs on Four Vertices

Chula J. Jayawardene (2020). *Handbook of Research on Advanced Applications of Graph Theory in Modern Society* (pp. 251-266).

www.irma-international.org/chapter/tripartite-and-quadpartite-size-ramsey-numbers-for-all-pairs-of-connected-graphs-on-four-vertices/235539

A Recurrent Probabilistic Neural Network for EMG Pattern Recognition

Toshio Tsuji, Nan Bu and Osamu Fukuda (2006). *Neural Networks in Healthcare: Potential and Challenges* (pp. 130-153).

www.irma-international.org/chapter/recurrent-probabilistic-neural-network-emg/27276