PERGAMON

Expert Systems with Applications 24 (2003) 115–122

Expert Systems with Applications

www.elsevier.com/locate/eswa

Impact of momentum bias on forecasting through knowledge discovery techniques in the foreign exchange market

Se-Hak Chun[a,*], Steven H. Kim[b]

[a]*Hallym University, 1 Okehon-Dong, Chunchon, Kangwon-do 200-702, South Korea*
[b]*Sookmyung Women's University, 53-12 Chungpa-dong 2 Ka, Yongsan-ku, 140-742 Seoul, South Korea*

Abstract

To an increasing extent since the late 1980s, software learning methods including neural networks (NN) and case based reasoning (CBR) have been used for prediction in financial markets and other areas. In the past, the prediction of foreign exchange rates has focused on isolated techniques, as exemplified by the use of time series models including regression models or smoothing methods to identify cycles and trends. At best, however, the use of isolated methods can only represent fragmented models of the causative agents, which underlie business cycles. Experience with artificial intelligence applications since the early 1980s points toward a multistrategy approach to discovery and prediction.

This paper investigates the impact of momentum bias on forecasting financial markets through knowledge discovery techniques. Different modes of bias are used as input into learning systems using implicit knowledge representation (NNs) and CBR. The concepts are examined in the context of predicting movements in the Japanese *yen*.

Keywords: Forecasting; Foreign exchange; Knowledge discovery; Neural network; Case based reasoning

1. Introduction

In the past, the prediction of foreign exchange rates has focused primarily on isolated techniques, as exemplified by the use of time series models including regression models or smoothing methods to identify cycles and trends. At best, however, the use of isolated methods can only represent fragmented models of the causative agents, which underlie business cycles. Experience with artificial intelligence applications since the early 1980s points toward a multi-strategy approach to discovery and prediction. In particular, statistical methods such as factor analysis may be used for exploratory analysis to determine the most salient characteristics behind foreign exchange rate behavior. The results of such analysis may be used as input into a learning system using implicit knowledge representation (neural networks, NN) and case based reasoning (CBR).

The rest of this paper is organized into five sections. Section 2 describes research background. In Section 2, the review of chaotic analysis and knowledge discovery techniques is presented. In Section 3, we present the case study. The case study intends to investigate the effect of bias on the predictive performance of learning methods in forecasting a foreign exchange rate. Section 4 reports the results of the case study. Finally, the concluding remarks are presented in Section 5.

2. Research background

Increasing evidence over the past decade indicates that financial markets exhibit chaotic behavior. The level of chaos in a data stream can be characterized by a number of methods. Two of the most popular parameters are the correlation dimension and the Lyapunov exponent.

2.1. Lyapunov characteristic exponents and correlation dimension estimates

The Lyapunov exponent characterizes the dynamics of a complex process. Each dimension of the process is associated with a Lyapunov exponent. A positive exponent indicates the sensitivity of initial conditions; that is, how much a forecast diverges based on approximately similar starting conditions. From a slightly different perspective, a Lyapunov exponent indicates the loss of predictive ability as one looks forward in time. On the other hand, a negative exponent indicates the degree to which points converge toward one another. For instance, a point attractor is characterized by negative values for each exponent (Ott,

* Corresponding author. Tel.: +82-33-240-1327; fax: +82-33-256-3424.
E-mail address: shchun@hallym.ac.kr (S.H. Chun).

0957-4174/03/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0957-4174(02)00089-1

Sauer, & Yorke, 1994; Pesaran & Potter, 1993; Peters, 1991).

The correlation dimension is an estimate of the fractal dimension. A chaotic system can be modeled by a number of coupled non-linear first-order differential equations. The minimum number of differential equations is equal to the integer that embeds the fractal dimension. The dimension of the phase space that spans the minimal number of differential equations is called the embedding dimension (Embrechts, 1994).

In addition, the level of chaos in a time series data can be characterized by a number of methods. One of the methods widely used by the physicists to test for chaos in time series data is the estimation of correlation dimension (Cecen and Erkal, 1996). The correlation dimension is an estimate of the fractal dimension and is used to differentiate between deterministic chaos and stochastic systems. It measures the correlation integral $C(\varepsilon)$, the probability that two point chosen at random will be within a certain distance of each other, and tests how the probability changes as the distance is increased (Peters, 1991).

For a given time series $\{Y_t : t = 1, \ldots, T\}$ of D-dimensional vectors, the correlation integral is formally defined as

$$C(\varepsilon) = \lim_{T \to \infty} \frac{2}{T(T-1)} \sum_{i<j} I_\varepsilon(Y_i, Y_j)$$

where $I_\varepsilon(x, y)$ is an indicator function which is equal to one if $\|x - y\| < \varepsilon$, and zero otherwise; where $\|x - y\|$ is the norm as measured by the Euclidean distance (Wasserman, 1989).

Grassberger and Procaccia (1983) defined the correlation dimension of the time series $\{Y_t\}$ as follows

$$D^m = \lim_{\varepsilon \to 0} [\log C_m(\varepsilon, T)/\log \varepsilon]$$

where m is embedding dimension.

The Hurst exponent H is a measure of the bias in random motion. For Brownian motion, the value of H is 0.50. For a persistent, or trend-reinforcing series, $0.50 < H \leq 1.00$. On the other hand, $0 \leq H < 0.50$ for an antipersistent, or mean-reverting system.

The calculation of the Hurst exponent involves a preliminary step known as rescaled (R/S) range analysis. The analysis developed to determine long-memory effects and fractional Brownian motion. R/S analysis measures how the distance covered by a particle increases as we look at longer and longer time scales (Peters, 1996).

2.2. Data mining and knowledge discovery techniques

Principal components analysis. The goal of principal components analysis is to take p variables $X_1, X_2, \ldots, X_p$ and find combinations of these to produce indices $Z_1, Z_2, \ldots, Z_p$ which are uncorrelated. The indices correspond to an orthogonal set of vectors, which simplifies the representation and interpretation of observations. Moreover, the indices are ordered so that Z_1 exhibits the greatest amount of variation, Z_2 displays the second largest amount of variation, and so on. The Z_i are called the principal components (Dunteman, 1989).

In many cases, the first several principal components collectively account for the bulk of the variation. In that case, the leading indices may be regarded as a sparse set of variables, which largely model the underlying situation; the remaining variables may be viewed as secondary and often ignored without much loss of modeling accuracy.

Factor analysis. In a way, factor analysis has similar aims to principal components analysis. The goal is still to describe a set of p variables $X_1, X_2, \ldots, X_p$ in terms of an orthogonal set of indices or factors. However, factor analysis has one further ambition: each factor should ideally represent an underlying dimension with physical relevance to an application, rather than just an arbitrary amalgamation of variables as in principal components analysis.

Backpropagation neural network. The NN methodology has been applied extensively to solve practical problems following the publication of the backpropagation algorithm for the multilayer perceptron (Rumelhart, Hinton, & Williams, 1986). The algorithm was developed for the perceptron model, a simple structure to simulate a neuron (Rosenblatt, 1962). Today the backpropagation network (BPN) is the most widely used neural algorithm in science, engineering, finance and other fields.

The general structure of a multilayer perceptron plus the backpropagation algorithm is shown in Fig. 1. The data entering an input node is multiplied by a set of weights. All such weighted inputs are summed at each node of next layer. The summed value enters an activation function, which depends on the learning algorithm. The output of the activation function then becomes the raw input for a node in next layer. This process is called *feed-forward*.

The output of nodes in the last layer may differ from the target value because the weights are initialized randomly. The error between the target value and the calculated value can be adjusted by varying the weights. The weights are adjusted by a delta rule derived from a cost function, which

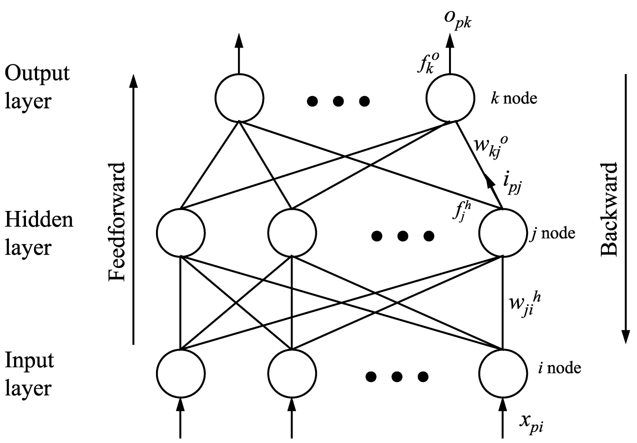

Fig. 1. General structure of the multilayer perceptron using the backpropagation algorithm.