Stock Tracking: A New Multi-Dimensional Stock Forecasting Approach

Huifeng Shen, Congfu Xu, Xuemei Han, Yunhe Pan
College of Computer Science, Zhejiang University
Zheda Road 38, Hangzhou 310027 Zhejiang
P.R.China
useraddcn@yahoo.com.cn, xucongfu@cs.zju.edu.cn
zjuwendy@yahoo.com.cn, panyh@zju.edu.cn

Abstract – This paper proposes a new approach—Stock Tracking that can forecast multi-dimensional stock information simultaneously. In this approach, a vector is used to represent stock information in point and a stock model is constructed to describe the stock fluctuation. A finite stock trend set is defined for simplifying the stock model. When the stock keeps the same trend, the model can be degenerated to a linear one and each trend has its own unique model parameters. The approach devises a new way to compute the transition matrix of the stock model and employs it for forecasting stock prices, volumes and indices etc. As for the stock trend changing, the discrete Markov process is adopted for stock forecasting. The experiments demonstrate the effectiveness of this approach. Furthermore, our approach can be used to solve those multi-dimensional financial forecasting problems where the state and observation space are the same Hilbert Space, the trend set is a finite set, and each state corresponds to one observation.

Keywords: stock tracking, multi-dimensional forecasting, stock model, target tracking, Markov processes.

1 Introduction

Stock forecasting is one of the hottest topics since prediction of future prices, indices, volumes etc. must be incorporated into the decision-making process. Owing to its importance, a well-established school of concepts and techniques, known as the fundamental analysis [1] and technical analysis [2], has been devised in recent decades. Since making the research of fundamental analysis is beyond our ability, this paper only considers the technical analysis methods which assume the future value is solely on the past value. Efforts to forecast stock information in this area have used statistical models and the Box-Jenkins time series approach [3], neural networks [4, 5], genetic programming [6, 7], fuzzy stochastic prediction method [8] and Kalman filter [9], etc.

In this paper we propose a new forecasting approach—Stock Tracking which can do multi-dimensional stock forecasting. As stock markets are dynamic and nonlinear systems in nature and target tracking is also a typical time-variant problem, it is feasible to introduce some methods of target tracking to solve the problem of multi-dimensional stock forecasting. Stock Tracking use a vector, which is composed of various prices, indices and volumes, etc., to represent the stock information at a given moment. The classical discrete time target tracking model [10] is applied to describe the stock fluctuation. The fluctuation is classified as a few finite trends for simplifying the stock model and each trend has its own model parameters. This paper devises a new way to compute the transition matrix—the most important parameter of the stock model. After that the model can be applied to forecast the stock information of interest when the stock keeps the same trend. As for the stock trend is uncertain, the discrete Markov process is adopted.

The rest of this paper is organized as follows. In Section 2, the stock model is formulated for stock forecasting and the way to compute the transition matrix is presented. Section 3 explains how the Stock Tracking approach is used for real stock forecasting. Section 4 presents experiment results of the approach. The last section offers conclusions and directions of future work.

2 Stock model

In this paper, we suppose that the stock market is only weak form EMH(Efficient Market Hypothesis) [11] that the price contains enough history information. This assumption makes it possible to infer the future price only from historical data.

Generally speaking, every moving object of interest can be considered as a target, not exception the stock. In this paper, we treat the stock forecasting as a special target tracking. We say it is special because all the forecasting data can be obtained finally which implies the state space and observer space are the same.

2.1 Problem formulation

Before forecasting, a stock model must be constructed. In this paper the stock fluctuation is treated as a discrete-time system. In order to construct the stock model, Stock Tracking approach defines stock state to denote the stock information.

Definition 1 (stock state) A vector $x_k$ reflects the stock information in point at time $k$ and its composition are mainly:

- various prices, such as open price, close price, low price, high price, average price, and 5-day average price etc.
• various indices, such as open index, close index and other history average volume etc.
• the volume and various history average volume etc.

In this paper, the classical target tracking model [10] is re-explained as the stock model to depict the stock fluctuation. In this new model, the stock data are gotten at fixed time interval and the stock information of \( k+1 \) can be calculated by the formula

\[
x_{k+1} = F(x_k) + \Gamma(w_k) + u_k,
\]

where
- \( x_k \) is a \( n \)-dimensional stock state vector;
- \( F \) is a known transition function;
- \( w_k \) is a \( r \)-dimensional input vector, which contains the factors outside the stock market, such as interest rates, foreign exchange rates, bond prices and the economic policies of the government, etc. at time \( k \) which will influence the stock price more or less;
- \( \Gamma \) is a known input function, which can be used to denote the effects of the government policy, such as monetary policy, fiscal policy;
- \( u_k \) is a state model noise, assumed to be normally distributed with mean zero and known variance

\[
E(u_ku_k') = Q_k \delta_{kj}.
\]

The accepted observer function is modeled as

\[
z_k = H(x_k) + v_k,
\]

where
- \( z_k \) is a \( m \)-dimensional measurement vector;
- \( H \) is a known observer function;
- \( v_k \) is normally distributed with mean zero and known variance

\[
E(v_kv_k') = R_k \delta_{kj}.
\]

In the classical target tracking model, the state space and observer space are almost not the same and the observer space is a subspace of the state space generally. But in our stock model the two spaces are the same and all the estimations can be verified in future. So the observer function of the stock model is very simple and \( H \) is a unit matrix. And Eq. (2) can be simplified as

\[
z_k = x_k + v_k.
\]

In addition, three aspects of the stock model are worthy of notice:

- all the elements of vector have their economic meanings.
- \( x_k \) contains enough information of the stock considered from time period 1 to \( k \) by contain a few of history average value so that the state \( x_{k+1} \) can be directly inferred from \( x_k \).
- since \( v_k \) is very small in most cases, \( x_k \) approximates to \( z_k \) for most situation and \( z_k \) is replaced with \( x_k \) in the real forecasting.

### 2.2 Stock forecasting

The Eq. (1) only has the theoretical sense and can not be used for real forecasting. As the stock market is volatility, to give an accurate stock model is an immense challenge. This paper defines a infinite trend set to embrace the most of the stock movements and each trend has its own parameter. **Stock trend set** and **stock trend function** are defined to simplify the stock model for real forecasting.

**Definition 2 (Stock Trend Set)**

\[
S = \{1, 2, \ldots, n\}
\]

is a finite set, which has \( n \) elements and each element corresponds to a certain stock fluctuation when considering the change of price and volume, etc.

\[
S^* = S \cup \{0\}
\]

"0" is used to represent that the trend is uncertain.

**Definition 3 (Stock Trend Function \( \mathcal{S} \))**

\[
\mathcal{S}(x_k) = j, \quad j \in S^*,
\]

where \( x_k \) is a stock state at time \( k \), \( j \) is a stock trend, \( S^* \) is a stock trend set. It tells which trend the stock state \( x \) will be during time \( k \) to \( k+1 \).

The **stock trend function** can also be applied to measurements and

\[
\mathcal{S}(z_k) = \mathcal{S}(x_k).
\]

By defining the above two definitions, the stock fluctuation can be classified as a series of trends. Based on the below observations, the input function can be ignored in most cases and the stock model can be degenerated to a linear equations when the stock keeps the same trend:

1. predictions over a relatively short time are easier to do reliably;
2. attempts to profit on short-term moves need this reliability to compensate for risks, taxes, and transaction costs etc.;
3. the effects of external factors are transitory and will be assimilated in a short time.

So the stock model (1) can be degenerated to a series of linear equations when the stock keeps the same \( j \)th trend,

\[
x_{k+1} = F_j x_k + u_k, \quad j \in S
\]
where \(x_t\) and \(u_t\) are the same as Eq. (1), and \(F_j\) is the \((n \times n)\) transition matrix of \(j\)th trend.

For convenience, this paper uses \(\hat{x}_k\) to denote the estimation of stock state \(x_k\), \(\hat{\xi}_{ij}\) to denote the stock estimation of \(x_k\) given data up to time \(j\), and \(Z^k\) to denote the set of validate measurements from time period 1 to \(k\)

\[
Z^k \triangleq \bigcup_{i=1}^{k} [z_i]
\]

When the stock trend does not change, the optimal estimation of \(\hat{x}_{ij}\) is formulated as

\[
\hat{x}_{ij} = E(\hat{x}_k|Z^i), \quad \forall i \in \{j, j + 1, \ldots, k - 1\}
\]

\[
\forall j, k, l \in S, \quad \forall i \in \{j, j + 1, \ldots, k - 1\}, \quad \forall i \in S.
\] (5)

But when the stock trend is not uncertain, Stokk Tracking uses discrete Markov process \([12, 13]\) to depict the jumps between the different trends. Let

\[
P = [P_{ij}], \quad i, j \in S
\]

be the (one step) transition probability matrix, where

\[
P_{ij} = P(\mathcal{S}(x_{n+1}) = j | \mathcal{S}(x_n) = i) \quad \forall n \geq 0, \quad i, j \in S.
\]

Then the optimal estimation of \(\hat{x}_{jk}\) can be formulated as

\[
\hat{x}_{jk} = \sum_{i=1}^{n} [P^{i-k}]_{il} E(\hat{x}_k|Z^i),
\]

\[
\mathcal{S}(x_l) = i \quad \& \quad \mathcal{S}(x_k) = l; \quad i, j \in S.
\] (6)

where \([P^{i-k}]_{il}\) is the \(lth\) row and \(ith\) column element of \(P^{i-k}\) and \(P^{i-k} = P \times P \times \cdots \times P\) \((j - k)\) times. In this paper, we only discuss one step transition.

### 2.3 Transition matrix computation

The popular stock models are various ARIMA(p,d,q) processes \([14]\), but its form is too complex. The average value such as five-day average price is used very common in technical analysis especial for those stock of the mature industry. Our approach decrease number of model parameter estimated by introducing these average history value. Kalman filter was used in \([9]\) for stock forecasting and an impressive tool is adopted to construct the transition matrix from ARMA process. But the vector constructed by this method can only contain the homogeneous factors from time \(t+1\) to \(t+k\). The defect of this method is that the information the vector contains is too little and its elements are correlated. In Stokk Tracking more heterogeneous factors, such as volumes and indices etc., can be taken into account for stock forecasting.

As the appropriate transition matrix \(F\), which is the most important parameter of our stock model, is not always available, a fitting method is proposed to compute it for a given stock trend based on the below facts: that the state and observer space are the same, and that the stock model is approximately linear when the stock trend does not change.

All the historical measurements of the same stock trend can be grouped as such pair

\[ (z_k, z_{k+1}) \quad \text{st.} \quad \mathcal{S}(z_k) = \mathcal{S}(z_{k+1}) = j, j \in S \]

By substituting Eq. (3) for \(x_k\), Eq. (4) becomes

\[
z_{k+1} - v_{k+1} = F_j(z_k - v_k) + u_k.
\] (7)

All s pairs of the \(j\)th trend can be grouped as

\[
\begin{align*}
z_k + 1 - v_k + 1 & = F_j(z_k - v_k) \quad + u_k, \\
z_k + 2 - v_k + 1 & = F_j(z_k - v_k) \quad + u_k, \\
\cdots & \\
z_k + s - v_k + 1 & = F_j(z_k - v_k) \quad + u_k.
\end{align*}
\]

The above simultaneous equations are rewritten to one equation as below:

\[
[z_{k+1}, \cdots, z_{k+s}] - [v_{k+1}, \cdots, v_{k+s}] = F_j \left( [z_k, \cdots, z_k] \right) [u_k, \cdots, u_k].
\] (8)

So the computation of the transition matrix \(F_j\) of stock trend \(j\) is translated into an estimation problem, which is formulated as below:

\[
\hat{F}_j = E(F_j|Z^j),
\]

where \(\hat{F}_j\) is the optimal estimation of \(F_j\) and \(Z^j\) is a set of measurements of \(j\)th trend defined as below:

\[
Z^j \triangleq \bigcup_{i=1}^{k} [z_i], \quad \forall i \quad \text{st.} \quad \mathcal{S}(z_i) = j, j \in S.
\]

When the noises are eliminated or discarded with measurements smoothing operation of data training phase mentioned in Section 3.1, the Eq. (8) can be simplified to

\[
[z_{k+1}, \cdots, z_{k+s}] = F_j \cdot [z_k, \cdots, z_k]
\] (9)

at last. \(F_j\) can be calculated by working out the above overdetermined equation and all \(F_j\) comprise the set \(\{F\}\).

### 3 Stocktracking process

In section 2, the stock model is constructed and a new way to compute the transition matrix is devised. In this section, we present how this model is used for real stock forecasting. The forecasting process is composed of two phase: data training and stock forecasting.

#### 3.1 Data training

Before the stock model can be applied to real forecasting, some historical data are used to train the stock model parameters. The training process includes three operations:

- **measurements smoothing**—that removes the measurements’ noise.
- **measurements classifying**—that the measurements are first segmented and then the segments are classified as a few trends.
The recursive stock forecasting algorithm can be depicted as Algorithm 1. Before forecasting, $x_k$ is known and $\mathcal{F}(x_k)$ can be known or not. If $\mathcal{F}(x_k)$ is known, the Eq. (4) is used to estimate $\tilde{x}_{k+1}$, otherwise Eq. (6) is adopted. When the real data $x_{k+1}$ arrives in the checking step, some checking functions, such as decision trees are used to determine $\mathcal{F}(x_{k+1})$ for next forecasting. Since time series processes of stock market are non-ergodic and non-stationary in nature, the model parameters trained by too old data will decrease the forecasting precision. $F$ and $P$ can be recomputed online via the algorithms such as Gauss–Newton [18], which can perform online Recursive Parameter Estimation (RPE) in the updating step.

Finally the Stock Tracking approach is summarized as below. Both of the input and output of this approach are vectors. Before the approach is applied to stock forecasting, some stock measurements are used to train the stock model parameters in batch mode. Then the stock measurements arrive sequentially the stock forecasting process is executed iteratively and the model parameters is updated on-line.

### 4 Experiment Results

We chosen Visual C++ as the programing language and Numerical Recipes as the numerical library. The the data segmenting, a simple trends classifying and the computa- tion of $F$ of data training phase have been coded. As for $P$ we used enumeration method, which means it can not be used for multi-step forecasting currently. In stock forecasting phase, a very simple check function that only considers the close price was adapted and the online recomputation of $F$ and $P$ was not implemented.

The indices and stocks of Shanghai Stock Exchange, PRC have been studied to validate the approach. The experiment chose the index 1000A1—SSE Fund Index as object investigated for index and volume forecasting, the stock 600016—China Minsheng Banking Corp.Ltd and 600059—ZheJiang GuYueLongShan ShaoXing Wine Co.Ltd as the objects investigated for index and volume forecasting. In addition, the stocks chosen are both the compositions of SSE Fund Index. The data chosen are the history from Jan 4, 2000 to Nov 30, 2004. The mixed Sliding windows and Bottom up method was used for segmenting. And the close index or price is chosen as the indicator for segmenting. Since the the number of segments is small, only five trends were classified. The first 60% segments are used to train the model, the later 40% segments are used to verify the approach. We have also experimented many combinations of the elements of vector.

### Algorithm 1: One step recursive tracking algorithm

**Data:** $x_k, \mathcal{F}(x_k), P, \{F_i, i \in S\}, x_{k+1}$

**Result:** $\hat{x}_{k+1}, \mathcal{F}(x_{k+1})$

if $\mathcal{F}(x_k) = 0$ then

$\hat{x}_k = F \mathcal{F}(x_k) \times x_k$

else

$\hat{x}_k = \sum_{i=1}^b [P]_i \cdot F_i \cdot x_k, \quad j = \mathcal{F}(x_k)$

\[ \mathcal{F}(x_{k+1}) = \text{check}(x_k, x_{k+1}, \ldots) \]

**updating** $(F)$ and $P$ with $x_{k+1}$
4.1 State vector construction and model training

The experiment chose low price, high price, close price, average price, 5-day average price, 15-day average price, 30-day average price, close volume and the corresponding average volume, close index and the corresponding average indices as candidates to construct the vector \( x_k \). The criterion of selecting is plain that it must generate more precise results, which may imply each element of vector is uncorrelated in nature. The open price was not considered since it is equal to about the close price the day before. As the amount can be acquired by performing multiplication on the close volume and the average price, it is not considered.

After many experiments, only various prices were picked to construct vector \( x_k \) for price forecasting. But when forecasting the volume, all the prices and volumes were considered. Low index, high index, close index and various average indices were used to forecast SSE Fund Index. We also made attempt at other possible combing and find they will decrease the accuracy of forecasting result.

The mixed *Sliding windows* and *Bottom up* method was adopted to segment the stock data. The least-squares fit was selected for *Sliding windows* and the close price or index was picked as indicator. Fig. 1 and Fig. 2 show the distributions of segment slope and days of segment of stock 600016 respectively. Five trends were constructed and each trend took about 20 percents segments considering the close price and volume. The five stock trends is showed as below. After classifying trend, \( \{F\} \) and \( \Pi \) were computed.

4.2 Forecasting results

Table. 1 and Table. 2 list some real price and the responding forecasting price of stock 600016 and each row of the table is an instance of vector \( x_k \). Fig. 3, 4 and 5 show some results of price predicting of stock 600016 of one forecasting. Fig. 6, 7 show volume and index forecasting results of stock 600059 and index 1000A1 respectively. The blue solid lines in above figures are the real values and the red dot lines are the forecasting values.
### Table 1: Some of the real price of stock 600016

<table>
<thead>
<tr>
<th>Date</th>
<th>Low</th>
<th>High</th>
<th>Close</th>
<th>Average</th>
<th>5-day</th>
<th>15-day</th>
<th>30-day</th>
<th>60-day</th>
</tr>
</thead>
</table>

### Table 2: Some of the forecasting price of stock 600016

<table>
<thead>
<tr>
<th>Date</th>
<th>Low</th>
<th>High</th>
<th>Close</th>
<th>Average</th>
<th>5-day</th>
<th>15-day</th>
<th>30-day</th>
<th>60-day</th>
</tr>
</thead>
</table>

Fig. 5: The high price forecasting of stock 600016

Fig. 6: The close volume forecasting of stock 600059
umes forecasting are acceptable in most cases. The results of one-step forecasting results are satisfying. The results of information, one step jumping comes into existence and the vectors we have constructed contain enough historical approach for real stock forecasting. Since the fluctuation of volume is much more violent than the index and price. Neither increasing the ratio of training to 80 percent nor decreasing it to 50 percent does not affect the forecasting results notably. As the number of segment is too small, the attempt to increase number of trends is not made.

The above experiment results show the effectiveness of Stcok Tracking approach for real stock forecasting. Since the vectors we have constructed contain enough historical information, one step jumping comes into existence and one-step forecasting results are satisfying. The results of price and index forecasting are exciting and the results volumes forecasting are acceptable in most cases.

5 Conclusions and future works

This works is a attempt to use the theory of information fusion for economic application—stock forecasting. Such efforts are made in [19, 20], which present the multiresolution filter application on the typical macroeconomic and microeconomic system. The Stcok Tracking approach treats the fluctuating stock as a moving target, and use an recursive method to predict its value. This approach is very simple and can be implemented easily. Its forecasting result is exciting.

Using vectors to represent stock information makes it convenient to deal with multi-dimensional data and increases the flexibility of this approach greatly. It also make it facile to analyze the interrelationship between various factors. such as the relation between prices and volumes of the same stock, the relation between two or more stocks and the relation between the price of a single stock and market indices. This approach can also be applied to other financial data forecasting, such as foreign exchange rate, time-bargain, and the selling of some merchandise etc. Our approach can be used to solve the problems of the system where the state and observer space are the same Hilbert Space, the number of trend set are finite and each state corresponds to an observation.

In future work, we intend to consider several related questions, for example, doing some comparison with more stock forecasting methods especially the ARIMA(p,d,q) model, extending this approach to do real time forecasting, analyzing the correlation of the elements which construct the vector.

Acknowledgements

This paper is supported by National Natural Science Foundation of China (No.60402010) and Advanced Research Project of China Defense Ministry(No.413150804), partially supported by the Aerospace Research Foundation (No. 2003-HT-ZIDX-13) and Zhejiang Provincial Natural Science Foundation of China (No. M603169).

We are grateful to the anonymous referees for their insightful comments and suggestions, which clarified the presentation.

References


