TITLE:
A trend-shift model for global factor analysis of investment products

AUTHOR(S):
Kirihata, Makoto; Ma, Qiang

CITATION:

ISSUE DATE:
2019-11

URL:
http://hdl.handle.net/2433/245892

RIGHT:
© 2019 The Institute of Electronics, Information and Communication Engineers
A Trend-Shift Model for Global Factor Analysis of Investment Products*

Makoto KIRIHATA, Nonmember and Qiang MA, Member

SUMMARY Recently, more and more people start investing. Understanding the factors affecting financial products is important for making investment decisions. However, it is difficult to understand factors for novices because various factors affect each other. Various techniques have been studied, but conventional factor analysis methods focus on revealing the impact of factors over a certain period locally, and it is not easy to predict net asset values. As a reasonable solution for the prediction of net asset values, in this paper, we propose a trend shift model for the global analysis of factors by introducing trend change points as shift interference variables into state space models. In addition, to realize the trend shift model efficiently, we propose an effective trend detection method, TP-TBSM (two-phase TBSM), by extending TBSM (trend-based segmentation method). Comparing with TBSM, TP-TBSM could detect trends flexibly by reducing the dependence on parameters. We conduct experiments with eleven investment trust products and reveal the usefulness and effectiveness of the proposed model and method.

key words: factor analysis, state space model, trend detection

1. Introduction

Recently, the Japanese government introduced the NISA (NIPPON Individual savings account) system, which encourages people to shift from savings to investments. Approximately 70% of the balance in NISA accounts is invested in investment trusts. Investment trust products are very popular and many people begin investing with investment trusts, because trust products do not require thorough knowledge of investments unlike stocks and bonds. However, there are too many similar trust products, which make determining appropriate ones for investments difficult. Revealing the factors that can be used to distinguish trust products is a considerable solution to support decisions on trust investments [3], [8].

In order to support investment by considering various factors that affect the NAV (net asset value) of investment trust products, research on factor analysis has been conducted. For example, methods for quantitatively analyzing factors affecting investment trust products have been proposed. They analyze investment trust products by using text data such as monthly reports and numeric data such as NAVs of investment trusts. However, they attempt to analyze factors to explain the current situation, and they cannot be applied for predictions. In addition, some researchers report that introducing the notation of trends into a state space model is useful to improve the performance of factor analysis. However, to the best of our knowledge, there is scant work on effectively detecting trends and analyzing factors from the global viewpoint (i.e., analyzing factors from a long-term perspective including multiple trends), which could help predict NAVs.

In this paper, we propose a trend shift model for the global analysis of factors by introducing trend change points as shift interference variables into state space models. In addition, to realize the trend shift model efficiently, we propose an effective trend detection method, TP-TBSM (two-phase TBSM), by extending TBSM (trend-based segmentation method).

The major contributions of this paper can be summarized as follows.

- We enable factor analysis across trends using a trend shift model (Sect. 3.1) and improve the accuracy of prediction. (Sect. 4)
- We enable to detect flexible trends while reducing the dependence on parameters using TP-TBSM (Sect. 3.2). The experimental results demonstrate that TP-TBSM is superior to conventional methods (Sect. 4).

2. Related Work

2.1 Financial Analysis with Text Data

In order to obtain information that cannot be attained using only numerical data, many studies have analyzed text data. These studies have demonstrated outstanding results in forecasting field and market understanding [1]–[3].

Johan Bollen et al. [1] proposed a method to predict the stock price by detecting the mood on Twitter. They achieved an accuracy of 86.7% in predicting the daily fluctuations in the closing values of the DJIA, and reduced the mean average percentage error more than 6%. Mahajan et al. [2] attempted to extract topics on the background of financial news using Latent Dirichlet Allocation, and discovered the topic that highly affected stock price by estimating the correlation between them. They also predicted a rise and fall in the market using extracted topics, and the average accuracy was 60%. Awano et al.[3] attempted to extract factors...
using the sentence structure of a monthly report on investment trust products, and developed a visualization system to support understanding of investment trust products.

These studies demonstrate that incorporating text data analysis could improve the market analysis. In this study, we use factors extracted from a monthly report of investment trust products by using the existing methods [8].

2.2 Financial Analysis with Time Series Data

Various time series analysis methods are used to study financial products and market analysis. Among them, the state space model is often used because it can flexibly build a model tailored to the purpose by incorporating various factors [4]–[8].

Bräuning et al. [4] used the state space model to analyze the effects of various factors on macroeconomic variables, and proposed a method to predict future values of the macroeconomic changes of the United States.

Marcellino et al. [5] proposed a kind of State-Space model called mixed-frequency Dynamic Factor Model. They used this model to analyze GDP data to clarify the dynamics of business cycles and evaluate the uncertainty of news content for monthly hard, soft and financial indicators. Holston et al. [6] analyzed the natural rate of interest in the United States, Canada, the euro area, and the state space model. They have found that drastic drop in GDP growth rates and decline in natural rates coincided in the four economies over the past 25 years.

Ando et al. [7] proposed a method to analyze point of sales data, which is important in marketing, using the state space model. Onishi et al. [8] quantitatively analyzed factors affecting NAV using the state space model. They extracted macro factors and micro factors from monthly reports and news, and used them in combination with numerical data such as NAV to determine the degree of influence of each factor. They concluded that considering trends could improve the accuracy.

Many other studies focused on the analysis of trends. Suzuki et al. [9] improved the accuracy of long-term prediction with non-linear prediction methods by handling trend change points. The shortcut prediction method proposed in [7] yields good results in predicting trend change points.

Chang et al. [10] proposed a method called intelligent piecewise linear representation (IPLR) for maximizing trading profit. IPLR detects a trend change point and uses it to convert time series data into a trading signal such as buying or selling. Using optimal parameters to maximize the profit learned in the neural network, it achieves better profit than rule-based transactions. Jheng-Long et al. [11] predicted buying and selling timings by using a method called TBSM together with support vector regression.

These studies show that consideration of trends and the state space model are useful for factor analysis. However, the existing trend detection methods require the specification of appropriate parameters, which is a difficult task.

3. Methodology

In this section, we first introduce a trend shift model for the global analysis of factors. Subsequently, we describe our TP-TBSM method, which detects trends automatically to realize the trend shift model efficiently.

3.1 Trend Shift Model

Generally, time series data such as stock prices are non-stationary time series whose mean and variance fluctuate with time. Therefore, it is necessary to deal with trends for analysis of such time series data. Onishi et al. [8] handled trends by delimiting data at the trend change point and constructing a state space model within it. However, as the analysis has been completed in each trend, it is not useful for future prediction. In this study, we propose a state space model incorporating the detected trend change points as slope shift interference variables. Hereafter, this model will be referred to as a trend shift model.

Assuming that the time of the i-th trend change point is \( \tau_i \), the slope shift interference variable can be defined as follows:

\[
Z_{i,t} = \begin{cases} 
0 & t \leq \tau_i \\
1 & t > \tau_i 
\end{cases}
\]

(1)

where \( Z_{i,t} \) is a variable whose value increases with time changing from \( \tau_i \). By obtaining the regression coefficient of this variable, the slope of the trend can be estimated.

By extending the state space model proposed in [8], the trend shift model incorporating the slope shift interference variable is described as follows.

\[
y_i = \mu_i + \sum_j a_{ij} Z_{i,t} + \sum_k \beta_{kj} x_{kj} + \sum_m \lambda_{mj} w_{mj} + \epsilon_i
\]

(2)

\[
\epsilon_i \sim \text{NID}(0, \sigma^2_i)
\]

(3)

\[
\mu_{i+1} = \mu_i + \xi_i
\]

(4)

\[
\xi_i \sim \text{NID}(0, \sigma^2_i)
\]

(5)

\[
a_{i+1} = a_{ij}, \quad \beta_{k+1} = \beta_{kj}, \quad \lambda_{m+1} = \lambda_{mj}
\]

(6)

where \( y_i \) is the logarithm value of NAV at time \( t \). \( \mu_i \) represents irregular variations. \( x_{kj} \) denotes the logarithmic value of a macro variable factor \( k \), such as the exchange rate. \( w_{mj} \) denotes a macro interference factor \( m \), such as policy announcement; it is 0 until the event occurs, and becomes 1 after the event occurs. The parameters \( \sigma^2_i, \sigma^2_i, \beta, \lambda \) are learned by using maximum likelihood estimations. The regression coefficients \( \beta \) and \( \lambda \) quantitatively represent the degree of influence of each factor.

3.2 TP-TBSM

We propose TP-TBSM, a method to detect trends effectively to realize the trend shift model by extending TBSM [11].

TBSM segments time series data into three kinds of
KIRIHATA and MA: A TREND-SHIFT MODEL FOR GLOBAL FACTOR ANALYSIS OF INVESTMENT PRODUCTS

2207

(a) Detect change points

(b) Detect stagnating trend

Fig. 1 Explanation of each step of TBSM (d: Distance from straight line, X_thld: Parameter of the length of trend, Y_thld: Parameter of the magnitude of variation)

trends i.e., rising, falling, and stagnating using three parameters and the point farthest from a linear function. An example is shown in Fig. 1. In the second trend Fig. 1 (a), the point where the distance from the straight line representing the trend becomes the maximum is determined. If the distance d exceeds the parameter δd, this point is set as a change point. If the variation is small around the change point, it is segmented into three trends (Fig. 1 (b)). This judgment is made based on whether the point is included in the rectangle of X_thld and Y_thld. The second trend in (a) is segmented into three trends in (b).

It is difficult to determine appropriate parameters according to time series data. Therefore, we propose TP-TBSM, which relaxes the dependency on parameters. We introduce the concept of trend error, and recursively detect trends by reducing the trend error.

A trend error is an average value of distance between each data point and a trend line (which can be represented by a linear function). The trend line is a straight line connecting the start and end points of the trend. The trend error is a measure showing the distance of the points from the trend line. The trend error is calculated as follows.

\[
TE(y(t), t_s, t_e) = \frac{\sum_{t=t_s}^{t_e} e(t)}{t_e - t_s} \\
e(t) = |f(t) - y(t)|
\]

(7)
(8)

where \(t_s\) and \(t_e\) are the start and end points of a trend respectively, \(y(t)\) is a value of time series data, and \(f(t)\) is a linear function representing a trend line.

In this study, a trend is considered good if \(TE(y(t), t_s, t_e)\) is small. \(e(t)\) is the distance between the real point \(y(t)\) and the corresponding point \(f(t)\) on the trend line.

As shown in Algorithm 1, the proposed method detects trends by alternately repeating two phases: evaluation and segmentation. The evaluation phase is shown in Algorithm 2, and the segmentation phase is shown in Algorithm 3. After describing these two phases, the algorithm of TP-TBSM will be explained. The symbols commonly used in the algorithms are listed in Table 1.

In the evaluation phase, we determine trends, which should be further segmented by considering their trend errors.

### Table 1 Symbols in TP-TBSM

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y(t)</td>
<td>Time series data</td>
</tr>
<tr>
<td>((t_s, t_e))</td>
<td>Trend represented by a combination of points (t_s) and (t_e)</td>
</tr>
<tr>
<td>f(t)</td>
<td>Linear function representing a trend line</td>
</tr>
<tr>
<td>e(t)</td>
<td>Distance between (y(t)) and (f(t))</td>
</tr>
<tr>
<td>C</td>
<td>Set of trend change points</td>
</tr>
<tr>
<td>(c_i)</td>
<td>The (i)-th element of (C)</td>
</tr>
<tr>
<td>(E)</td>
<td>Set of trends whose trend error is large</td>
</tr>
<tr>
<td>(\delta_e)</td>
<td>The size of the minimum trend. Needs to be set</td>
</tr>
<tr>
<td>(\delta_d)</td>
<td>The magnitude of (e(t)). Calculated by the algorithms</td>
</tr>
<tr>
<td>(\delta_e)</td>
<td>The magnitude of trend error. Calculated by the algorithms</td>
</tr>
</tbody>
</table>

**Fig. 2** Trend Error \(e(t)\) of trend \((t_s, t_e)\) defined by Eq. (7), (8).

**Step 1** Calculate the trend error for each trend and set the parameter \(\delta_e\) as their average value (Line:2-5).

**Step 2** A trend whose trend error is larger than \(\delta_e\) is subject to segmentation (Line:6-10).

In the segmentation phase, we segment trends, as follows.

**Step 1** Determine the point whose distance to the trend line is the maximum. Such a point is a candidate for a trend change point. We are considering the interval \([start + \delta_e, end - \delta_e]\) to ensure that the length of the trend is greater than or equal to the parameter \(\delta_e\) to avoid segments that are too short. (Line 1)

**Step 2** Determine whether to segment by using the parameter \(\delta_d\). (Line 2)

**Step 3** Check whether there is a stagnating trend around the trend change point. A stagnating trend indicates that the value variation in the trend is small.
Algorithm 1 TP-TBSM

Input: $y(t), \delta_t$

Output: $C$

1. $C = \{1, n\}$
2. $E = (1, n)$
3. repeat
   4. if not first iteration then
      5. $E = \text{Evaluation}(C)$
   6. end if
   7. $C_{odd} = C$
   8. for $(t_s, t_e) \in E$ do
      9. if $t_e - t_s < 2\delta_t$ then
         10. Go to the next trend, because the trend length is short
      11. else
         12. $d_{max} = \max e(t)$ in the interval $[t_s + \delta_t, t_e - \delta_t]$
         13. $\delta_d = d_{max}$
         14. $C = C \cup \text{Segmentation}(y(t), \delta_t, \delta_d, t_s, t_e)$
      15. end if
   16. end for
17. until $C_{odd} = C$
18. return $C$

Algorithm 2 Evaluation

Input: $E$

Output: $E$

1. $E = \emptyset$
2. for $i = 1 : p$ do
   3. $e_{list}[i] = TE(y(t), c_i, c_{i+1})$ // $e_{list}$: List of length $p$
4. end for
5. $\delta_x = \text{Average}(e_{list})$
6. for $i = 1 : p$ do
   7. if $e_{list}[i] > \delta_x$ then
      8. $E = E \cup (c_i, c_{i+1})$
   9. end if
10. end for
11. return $E$

Algorithm 3 Segmentation

Input: $y(t), \delta_t, \delta_d, (t_s, t_e)$

Output: $C$

1. $d_{max} = \max e(t)$ in the interval $[t_s + \delta_t, t_e - \delta_t]$. Let $t_d$ be that time
2. if $(d_{max} \geq \delta_d)$ then
3. $p = 0$ // $p$: Number of points included in $H$
4. for $t_i = (t_d - \delta_d) : (t_d + \delta_d)$ do
5. if $\left|y(t_d) - y(t_i)\right| < \frac{\delta_d}{2}$ then
6. $H[p] = i, p = p + 1$ // $H$: Point list for a stagnating trend
7. end if
8. end for
9. if $(H[p] - H[1] > \delta_d)$ and $(p > \frac{H[p] - H[1]}{2})$ then
10. $c_a = \text{Segmentation}(y(t), \delta_t, \delta_d, t_s, H[1])$
11. $c_b = \{H[1], H[k]\}$
12. $c_c = \text{Segmentation}(y(t), \delta_t, \delta_d, H[k], t_e)$
13. return $\{c_a, c_b, c_c\}$
14. else
15. $c_a = \text{Segmentation}(y(t), \delta_t, \delta_d, t_s, t_d)$
16. $c_c = \text{Segmentation}(y(t), \delta_t, \delta_d, t_s, t_e)$
17. return $\{c_a, c_c\}$
18. end if
19. end if
20. return $\{t_s, t_e\}$

(1) As preparation for the checking, we construct a list $H$ consisting of points whose values are close to that of the candidate trend change point. (Line 3-8)

(2) If $H$ is sufficiently long, and more than half of the points in $H$ have a value close to that of the candidate trend change point, we conclude that a stagnating trend exists, and thereafter divide the current trend into three sub-trends including a stagnating trend. (Line 9-13)

(3) If no stagnating trend exists, we simply segment the current trend into two sub-trends using the (candidate) trend change point. (Line 15-17)

The TP-TBSM algorithm is shown in Algorithm 1.

Step 1 The start and end points of the time series data are considered as the initial trend change points, and the trend line connecting these points is considered as the initial trend. (Line 1-2)

Step 2 An evaluation phase is performed. A trend with large trend error is selected and placed in the set $E$. (Line 4-6)

Step 3 The length of the trends in $E$ is examined. If the trend length is shorter than $2\delta_t$, we do not perform further segmentation for this trend to avoid trends shorter than $\delta_t$. (Line 8-10)

Step 4 If segmentation is possible, $\delta_d$ for segmentation is determined, and the segmentation phase is performed. The parameter $\delta_d$ is set to the maximum distance to the trend line. (Line 11-16)

Step 5 Steps 2-4 are repeated until the result does not change. (Line 17)

Figure 3 shows an example of detecting trends by using TP-TBSM. In Fig. 3 (a), each trend is evaluated using trend error. The trend error of the second trend is large. In Fig. 3 (b), the point where $e(t)$ becomes maximum is detected as the trend change point. In Fig. 3 (c), it is verified whether
there is a stagnation trend. In Fig. 3 (d), segmentation is performed. In the next step, the fourth trend is considered to have large trend error. This process is repeated to detect trends.

4. Experiments

First, we evaluate the usefulness of the trend shift model by comparing the trend shift with the basic state space models. Second, we construct trend shift models with different trend detection methods to evaluate our TP-TBSM method.

Since trusts products are more suitable for factor analysis (fundamental analysis) than stocks and foreign exchanges, we used the trusts data in our experimental evaluation.

4.1 Outline of the Experiment

We used the data set collected by Onishi et al. [8] consisting of 13 investment trust products from January 4, 2016 to October 31, 2016.

The data for the last 20 days are used for testing m predictions, and the other data are used for learning. The 20 days will be about a month’s worth of data excluding days with no NAV data such as Saturdays and Sundays. The parameter \( \delta_t \) used to detect trends using TP-TBSM was also set as 20 days. We used the macro and micro factors extracted using the existing method [8].

As the state space model assumes that the standardized prediction error is independent and normal, we analyzed 13 trust products with each model and used only 11 products for further analysis. These 11 products satisfied the Ljung–Box test and the Shapiro–Wilk test with the significance level 5%.

4.2 Evaluation Measures

4.2.1 Average Error of Prediction

State space models are rarely used for prediction and are often used for factor analysis. Therefore, the focus is often on how much data can be reproduced in training data. However, in investment trust products, accuracy of prediction is also important. The prediction in this experiment means the future daily value generated by the learned model. By comparing the average error between generated values and ac-
tual values, it is possible to evaluate how much the price movement mechanism can be modeled. Therefore, in this study, the average error of prediction is used for the evaluation of the model. However, as the regression components are included in the model, it is necessary to use the observed data with respect to them, and hence, this prediction is closer to completion than pure prediction.

4.2.2 AIC (Akaike Information Criterion)

In addition to the prediction error, the Akaike information criterion (AIC) is used for the model evaluation. Let \( L \) be the maximized log-likelihood, \( r \) be the number of unknown parameters, \( q \) be the number of initial points in a diffuse initial state, and \( n \) be the number of points; the AIC in time series is expressed as follows.

\[
AIC = -2L + 2(q + r) \frac{n}{n - q - r - 1} \tag{9}
\]

AIC is penalized by the number of parameters that must be estimated for maximum log likelihood. As the likelihood of the time series is based on the one-step prediction error, the model with small AIC is a simple one with the high accuracy of the one-step prediction.

4.3 Baseline Methods

4.3.1 Models Used for Comparison with the Trend Shift Model

We compare our trend shift model with the following existing models.

- **Local model** proposed in [8]. It is a model with \( \Sigma_{t} \bar{\xi}_{t} \bar{\eta}_{t} \) removed from Eq. (2).
- **Linear model** is a variation of the local linear trend model [12], which extends the local model by introducing a slope term. In short, the linear model modifies Eq. (3) of the trend shift model as follows.

\[
\mu_{t+1} = \mu_{t} + \nu_{t} + \xi_{t}, \quad \xi_{t} \sim NID(0, \sigma_{\xi}^{2}) \tag{10}
\]

\[
\nu_{t+1} = \nu_{t} \tag{11}
\]

- **Trend model** is also a variation of the local linear trend model [12]. In the trend model, Eq. (3) is modified as follows.

\[
\mu_{t+1} = \mu_{t} + \nu_{t} \tag{12}
\]

\[
\nu_{t+1} = \nu_{t} + \xi_{t}, \quad \xi_{t} \sim NID(0, \sigma_{\xi}^{2}) \tag{13}
\]

4.3.2 Comparative Method for TP-TBSM

To evaluate TP-TBSM, we construct trend shift models with different trend detection methods: our TP-TBSM and the dynamic programming (DP) method [8]. The method of detecting trends using DP was used by Onishi [8]. For each trend, the DP method prepares a straight line connecting the boundary points of the trend, and calculates the root mean square error by comparing with the NAV. The DP method dynamically changes the trend points to minimize the error. It is necessary to determine the number of trends.

4.4 Results and Discussion

4.4.1 Trend Shift Model

The local model, linear model, trend model, and trend shift model (TP-TBSM) are compared. As presented in Table 2, the average error of the prediction of the trend shift model is the smallest for eight out of 11 products. This indicates that the trend shift model could accurately estimate the influence coefficient of the factors.

In addition, the prediction errors of the local and linear models are larger for most products. These models do not fully consider the influence of trends. The error variation of the trend model is large. This is because the value of the slope term expressing the trend is largely influenced by the immediately preceding value in the trend model.

As presented in Table 3, the local model exhibits the lowest AIC value for all the products and the linear model exhibits the second lowest value. It is thought that AIC has become smaller because simple random walk is used for these two. Overfittings are caused by random walks. Further details are provided in the case study.

Upon comparing the trend model with the trend shift model, it can be observed that the trend shift model shows a smaller AIC value for eight out of 11 products, and it can be concluded that the trend shift model is a better model than the trend model.

### Table 2: Average error of prediction

<table>
<thead>
<tr>
<th>No.</th>
<th>Local</th>
<th>Linear</th>
<th>Trend</th>
<th>Trend Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0121257</td>
<td>0.0132519</td>
<td>0.0362272</td>
<td>0.00760425</td>
</tr>
<tr>
<td>2</td>
<td>0.01692213</td>
<td>0.01192982</td>
<td>0.02064795</td>
<td>0.00807945</td>
</tr>
<tr>
<td>3</td>
<td>0.0263603</td>
<td>0.0187761</td>
<td>0.01338155</td>
<td>0.0096469</td>
</tr>
<tr>
<td>4</td>
<td>0.0091594</td>
<td>0.01260263</td>
<td>0.0132246</td>
<td>0.0095815</td>
</tr>
<tr>
<td>5</td>
<td>0.02357305</td>
<td>0.0224242</td>
<td>0.01857475</td>
<td>0.01863415</td>
</tr>
<tr>
<td>6</td>
<td>0.01534265</td>
<td>0.0147249</td>
<td>0.0112846</td>
<td>0.0217712</td>
</tr>
<tr>
<td>7</td>
<td>0.019291</td>
<td>0.0176646</td>
<td>0.0425186</td>
<td>0.01621625</td>
</tr>
<tr>
<td>8</td>
<td>0.01504885</td>
<td>0.02040415</td>
<td>0.027809</td>
<td>0.00924125</td>
</tr>
<tr>
<td>9</td>
<td>0.0211324</td>
<td>0.01933145</td>
<td>0.03037835</td>
<td>0.01348215</td>
</tr>
<tr>
<td>10</td>
<td>0.01922795</td>
<td>0.01959765</td>
<td>0.0345798</td>
<td>0.0112273</td>
</tr>
<tr>
<td>11</td>
<td>0.01532515</td>
<td>0.01736885</td>
<td>0.0213364</td>
<td>0.00893415</td>
</tr>
</tbody>
</table>

### Table 3: AIC

<table>
<thead>
<tr>
<th>No.</th>
<th>Local</th>
<th>Linear</th>
<th>Trend</th>
<th>Trend Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.052401</td>
<td>-3.969067</td>
<td>-3.764244</td>
<td>-3.814294</td>
</tr>
<tr>
<td>2</td>
<td>-4.212589</td>
<td>-4.127697</td>
<td>-3.918999</td>
<td>-3.968967</td>
</tr>
<tr>
<td>3</td>
<td>-3.723839</td>
<td>-3.644144</td>
<td>-3.416982</td>
<td>-3.509882</td>
</tr>
<tr>
<td>4</td>
<td>-3.730098</td>
<td>-3.640922</td>
<td>-3.316481</td>
<td>-3.584658</td>
</tr>
<tr>
<td>5</td>
<td>-3.365685</td>
<td>-3.284338</td>
<td>-3.0366</td>
<td>-3.21709</td>
</tr>
<tr>
<td>6</td>
<td>-4.281416</td>
<td>-4.194314</td>
<td>-4.109473</td>
<td>-3.82778</td>
</tr>
<tr>
<td>7</td>
<td>-4.076699</td>
<td>-3.991273</td>
<td>-3.76433</td>
<td>-3.830447</td>
</tr>
<tr>
<td>8</td>
<td>-3.960386</td>
<td>-3.876909</td>
<td>-3.623052</td>
<td>-3.836974</td>
</tr>
<tr>
<td>9</td>
<td>-4.133614</td>
<td>-4.047846</td>
<td>-3.821848</td>
<td>-3.887454</td>
</tr>
<tr>
<td>10</td>
<td>-4.133536</td>
<td>-4.047639</td>
<td>-3.820015</td>
<td>-3.744925</td>
</tr>
<tr>
<td>11</td>
<td>-4.193313</td>
<td>-4.1072</td>
<td>-3.868612</td>
<td>-3.86926</td>
</tr>
</tbody>
</table>

(Please note that the values shown are not actual values but placeholders for formatting purposes.)
Table 4  Average error of prediction. “error” denotes the failed prediction.

<table>
<thead>
<tr>
<th>No.</th>
<th>DP</th>
<th>TP-TBSM (δ₁ = 5)</th>
<th>TP-TBSM (δ₁ = 10)</th>
<th>TP-TBSM (δ₁ = 15)</th>
<th>TP-TBSM (δ₁ = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01541</td>
<td>0.008867</td>
<td>0.006762</td>
<td>0.008439</td>
<td>0.007604</td>
</tr>
<tr>
<td>2</td>
<td>0.01808</td>
<td>0.009293</td>
<td>0.007683</td>
<td>0.009647</td>
<td>0.009647</td>
</tr>
<tr>
<td>3</td>
<td>0.01371</td>
<td>0.07039</td>
<td>0.01172</td>
<td>0.009581</td>
<td>0.009581</td>
</tr>
<tr>
<td>4</td>
<td>0.01048</td>
<td>error</td>
<td>0.02572</td>
<td>0.02795</td>
<td>0.02177</td>
</tr>
<tr>
<td>5</td>
<td>0.01371</td>
<td>0.01199</td>
<td>0.01445</td>
<td>0.01199</td>
<td>0.01445</td>
</tr>
<tr>
<td>6</td>
<td>error</td>
<td>error</td>
<td>0.01445</td>
<td>0.01199</td>
<td>0.01445</td>
</tr>
<tr>
<td>7</td>
<td>0.01997</td>
<td>0.009581</td>
<td>0.009581</td>
<td>0.009581</td>
<td>0.009581</td>
</tr>
<tr>
<td>8</td>
<td>error</td>
<td>error</td>
<td>0.01445</td>
<td>0.01199</td>
<td>0.01445</td>
</tr>
<tr>
<td>9</td>
<td>error</td>
<td>0.006100</td>
<td>0.006323</td>
<td>0.01202</td>
<td>0.008934</td>
</tr>
<tr>
<td>10</td>
<td>0.01738</td>
<td>0.01238</td>
<td>0.01863</td>
<td>0.01863</td>
<td>0.01863</td>
</tr>
<tr>
<td>11</td>
<td>0.009193</td>
<td>0.006015</td>
<td>0.008935</td>
<td>0.01202</td>
<td>0.008934</td>
</tr>
<tr>
<td>12</td>
<td>0.01738</td>
<td>0.01238</td>
<td>0.01863</td>
<td>0.01863</td>
<td>0.01863</td>
</tr>
</tbody>
</table>

Table 5  AIC. “error” denotes the failed prediction.

<table>
<thead>
<tr>
<th>No.</th>
<th>DP</th>
<th>TP-TBSM (δ₁ = 5)</th>
<th>TP-TBSM (δ₁ = 10)</th>
<th>TP-TBSM (δ₁ = 15)</th>
<th>TP-TBSM (δ₁ = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-3.245881</td>
<td>error</td>
<td>-3.584658</td>
<td>-3.584658</td>
<td>-3.584658</td>
</tr>
<tr>
<td>4</td>
<td>-2.816559</td>
<td>-2.669863</td>
<td>-3.217069</td>
<td>-3.217069</td>
<td>-3.217069</td>
</tr>
<tr>
<td>5</td>
<td>error</td>
<td>error</td>
<td>-3.584658</td>
<td>-3.584658</td>
<td>-3.584658</td>
</tr>
<tr>
<td>8</td>
<td>error</td>
<td>error</td>
<td>-3.674589</td>
<td>-3.803849</td>
<td>-3.874545</td>
</tr>
</tbody>
</table>

4.4.2 TP-TBSM

The results (average error of prediction and AIC) of the trend shift models constructed based on DP and TP-TBSM are compared. The parameter δ₁ of TP-TBSM was set to 5, 10, 15, and 20.

As presented in Table 5, the model based on TP-TBSM achieved better results in terms of AIC than the model based on DP. The number of trends in DP is fixed at 9, whereas TP-TBSM detects different numbers of trends.

As presented in Table 4, the smaller the parameter δ₁, the better the result of the prediction. In addition, the pre-
diction error of TP-TBSM is smaller than that of DP for almost all the products. In short, the TP-TBSM method could flexibly determine the number of trends and achieve better results in terms of AIC and prediction error.

4.4.3 Case Study

We discuss the result of the trend shift model on the product No.11, Hihumi trusts.

The prediction results are shown in Fig. 4. The average error of the trend shift model using TP-TBSM is the smallest one among all the models. From this figure, it can be observed that the trend shift model can successfully predict the NAV by considering trend. The NAVs predicted by local models and linear models do not change much from the start point. It is natural that local models cannot predict trends. In addition, we can see that the linear model and the trend model cannot handle trend changes well.

The level term \( \mu_i \) of each model is shown in Fig. 5. As \( \mu_i \) varies owing to random walk, larger variation of \( \mu_i \) indicates that the change of NAV is random and we could not estimate the influence degrees of factors. In the local and linear models, \( \mu_i \) significantly varies every day. In the trend model, this level term fluctuates smoothly, and it is different from the change of NAV of local and linear models. However, the change is still big and then their models cannot handle the change of factors’ fluctuations. In the trend shift model, the variation of \( \mu_i \) is suppressed, and we may conclude that the trend shift model could reduce the effects of chance to yield better results of factor analysis.

As shown in Table 4. TP-TBSM(\( \sigma_i = 5 \)) achieved the lowest average error, and the second and the third places are also achieved by TP-TBSM. It is thought that TP-TBSM can detect a trend change point flexibly and accurately.

5. Conclusion and Future Work

In this paper, we proposed a trend shift model by incorporating the trend change points into a state space model in order to quantitatively analyze factors affecting the NAV and predict future NAVs. To realize the trend shift model, we also proposed a trend detection model, i.e., TP-TBSM. In the TP-TBSM, by repeating the evaluation and segmentation phases, it is possible to reduce the dependence on the parameter, as compared with the conventional method, and to detect the trend more flexibly. The trend shift model enables global analysis across trends. From the experimental results, we observed that the trend shift model incorporating the change point detected using TP-TBSM has higher prediction accuracy than the baseline. We will carry out further extensive experiments to validate and improve our model. We are analyzing it assuming that the trend will continue for a while now, so it will not exert much effect in the period when the trend changes severely. We will plan to extend the TP-TBSM method to analyze in the period when the trend changes severely.

Acknowledgments

This work was partly supported by JSPS KAKENHI (19H04116) and The Kyoto University Foundation.

References

Makoto Kirihata received B.E. degree from Kyoto University in 2018. He is on the Graduate School of Informatics, Kyoto University as a master student. His general research interests are in the area of investment, time series analysis, data mining and machine learning.

Qiang Ma received his Ph.D. degree from Department of Social Informatics, Graduate School of Informatics, Kyoto University in 2004. He was research fellow (DC2) of JSPS from 2003 to 2004. He joined National Institute of Information and Communications Technology as a research fellow in 2004. From 2006 to 2007, he served as an assistant manager at NEC. From October 2007, he joined Kyoto University and has been an associate professor since August 2010. His general research interests are in the area of databases and information retrieval. His current interests include Web information system, Web mining, knowledge discovery, and multimedia mining and search.