Using change point identification in financial data to detect turning points in companies

Satoshi Kawamata*, Yasushi Nagata

Waseda University, 3-4-1 Okubo, Shinjuku-ku Tokyo, 169-8555, Japan

*contact author’s e-mail address : kawamata@moegi.waseda.jp

Abstract:
Change point search methods are often used in quality engineering. These methods detect the points in time when important changes can occur, and are applied to various types of time series data. In this study, we search for change points in financial data from Japanese companies across various industries. By using these methods, we can easily and quantitatively identify companies turning points, helping financial institutions analyze financial indicators over a certain period of time.

To obtain sufficient data for the analysis, we use quarterly financial data over the period 2006 to 2015 to apply methods focusing on average values. For the analysis, we select financial indicators such as return on assets, capital adequacy ratio, and rate of sales growth. For each financial indicator, we use the $t$ test to identify change points in data. Then, we run the analysis by grouping data for multiple financial indicators and employing the Hotelling $T$-square test. In this way, we are able to detect turning points that could not be found from a single financial indicator. Finally, we survey the major economic events that occurred around each turning point.

Keywords
cchange point search, $t$ test, Hotelling $T$-square test, time series analysis, financial analysis

1. Introduction

Financial and economic research institutions carry out credit screening and market research to analyze fluctuations in financial indicator data over a certain period. The aim is to check whether a financial condition has changed significantly and to investigate what kind of event has a big influence on business management. In the previous research in financial analysis, discriminant analysis on corporate rating (Tanaka and Nakagawa (2014)), regression analysis on stock prices (Wen et al. (2014)) have been studied, but we could not find research using statistical methods to calculate change in financial data.

In this study, we apply change point search methods to corporate financial data in order to conveniently extract the time point when characteristic changes have occurred from long-term data. By using these methods, it becomes possible to quantitatively analyze the turning point of the company which was conventionally sensuous. We also examine what kind of event occurred at the change points. In corporate management, it is possible to prioritize measures against events such as infectious disease epidemics, economic depression, and natural disasters. Among the change point search methods, the maximum $t$ method proposed by Hirotsu (1992) defined as a change point only the one in which the test statistic is the largest. Hence, when dealing with data in which there are multiple serious changes, there is the risk of overlooking some points of change. To address this problem, Kawai (2011) proposed the sequential $t$ method, which detects a change point when the test statistic exceeds a rejection limit value.

The structure of this paper is as follows. The outline of the analysis methods is described in Section 2. Data and results are described in Section 3, while Section 4 summarizes the conclusions.

2. Methods
If there is a big difference between financial data of a company before and after an event, it can be inferred that such event had a strong influence on business management. By considering financial data of companies in various industries, we identify the perceived turning points through methods focused on the average value. In addition, we are able to easily detect the change in financial conditions that cannot be read by simply looking at each financial data through methods based on multiple financial indexes. In this section, we introduce the outline of the analysis methods we employ, and explain their accuracy with a simple experiment.

2.1 Change point search using single time series data
2.1.1 Maximum $t$ method

In this study, we select a method that enables accurate change point detection in any sample in order to make it possible to search for change points in each company considered. The idea of using the maximum $t$ method and the sequential $t$ method is based on Kawai (2011), but the test statistic is calculated according to the maximum $t$ method by Hirotsu (1992). This method divides the time series data into a control group and a treatment group, and repeatedly performs the test on the $a$ population averages. Specifically, the hypothesis test is set as follows.

Null hypothesis

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a \quad (2.1)$$

Alternative hypothesis

$$H_1: \mu_1 = \cdots = \mu_{a-1} \neq \mu_{a+1} = \cdots = \mu_a \quad (2.2)$$

In equation (2.2), $\tau$ represents the point at which the change occurs. Define

- $n_{ij}$: the number of samples moving from the $i$-th group to the $j$-th group,
- $T_{ij}$: total data moving from the $i$-th group to the $j$-th group, and
- $V_E$: the error dispersion in one-way ANOVA.

Assume $p = a$ and calculate the test statistic in equation (2.4).

$$\max_i t = \max_{i=1, \ldots, p-1} \left| k_i \right| \quad (2.3)$$

$$t_i = \left( \frac{1}{n_{ij}} + \frac{1}{n_{(i+1)j}} \right) \left( \frac{T_{(i+1)j}}{n_{(i+1)j}} \frac{T_{ij}}{n_{ij}} \right)^{\frac{1}{2}} \sqrt{V_E} \quad (2.4)$$

Normally, the point at which $t_i$ takes the maximum value when $i$ is changed from 1 to $p-1$ in equation (2.4) is set as the change point. However, since the analysis target is expressed as financial data, it is not possible to predict any trend for the average value. Since the test statistic may include both positive and negative values, we define the point at which the absolute value is maximum as the change point. In order to perform the analysis for each sample, the error dispersion is calculated by dividing data from the first group to the $i$-th group and data from the $i+1$-th group to the $a$-th group.

The detection accuracy of the method for the change of the average value is shown through the following experiment. We generate 100 random numbers from a standard normal distribution $N(0,1^2)$ and add the specific value $\mu'$ from the 51st data. The maximum $t$ method is executed 1,000 times and the times in which the test statistic takes the maximum value are counted. The number of counts at each point in time when the value of $\mu'$ is changed from 1 to 3 is shown in Figure 1. The figure confirms that the change point detection number is maximum corresponds to the time point 50. Although the difference in the average value can be detected accurately, in Figure 1 (c), the time point 50 becomes a change point 852 times.
2.1.2 Sequential $t$ method

The presence of multiple change points can be detected by using the sequential $t$ method. We set the null hypothesis and the alternative hypothesis as follows.

Null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$

(2.5)

Alternative hypothesis

$$H_1 : \mu_1, \cdots, \mu_a \text{ At least one of them is different.}$$

The calculation method of the test statistic in the sequential $t$ method is described by equation (2.6). To calculate the error variance, the sequential width is set to 2.

$$t(j, j+1) = \left( \frac{1}{n_{(j-1)j}} + \frac{1}{n_{(j+1)j}} \right)^{-1} \left( \frac{T_{(j+1)j} - T_{(j-1)j}}{n_{(j+1)j} - n_{(j-1)j}} \right) V_n$$

(2.6)

$$\lVert t(j, j+1) \rVert > t(\phi_n, (\alpha/2)/b)$$

(2.7)

The test statistic is calculated by modifying $j$ in equation (2.6) from 2 to $p-2$. If equation (2.7) is satisfied, $j+1$ is regarded as a change point. The calculation of the rejection limit value conforms to the Bonferroni method. The error degree of freedom is 2 and $b$ is 37 as the number of verifications. If the test statistic does not exceed the rejection limit value, the number of change points detected is 0. For the sequential $t$ method, it is common to compare data referring to subsequent periods ($t$ and $t+1$). Hence, there is a possibility that the detection accuracy of the change point may be lowered. Okada (2013) mentions that by making the width larger, it becomes easier to specify the area of the change point. However, we cannot analyze both ends of the observation point.

The detection accuracy is analyzed also for the sequential $t$ method. The simulation setting is the same as the maximum $t$ method, and results are shown in Figure 2. In this study, we set the rejection limit each time and focus on the period in which the test statistic by the sequential $t$ method is the maximum. As shown in Figure 2, the change point detection number is maximum in period 49 for any $\mu'$ value. This means that the difference is maximum when the average value between time points 49 and 50 is compared with the one between time points 51 and 52. The reason for defining the period is that, unlike the maximum $t$ method, two points in time are analyzed at once, and thus it is not possible to specify the change point at a single time point. It is confirmed that the differences in the average value can be accurately detected similarly to the maximum $t$ method, but Figure 2 also shows that the detected period varies greatly when the value of $\mu'$ is small. Figure 2 (c) highlights that, when
\( \mu' \) takes the maximum value, the test statistic is the highest 166 times in period 49. It can be seen that the detection accuracy is lower than the one in the maximum \( t \) method. This is because only few samples are used for the analysis, and the width is sequentially set to 2.

\[ \sum_{i=1}^{m} X_{i}, \cdots , X_{n} \sim \mathcal{N}(\mu, \Sigma) \quad \text{and} \quad Y_{i}, \cdots , Y_{k} \sim \mathcal{N}(\mu, \Sigma). \]

Based on these conditions, we set the hypothesis test as follows.

Null hypothesis

\[ H_{0} : \mu_{1} = \mu_{2} \]  \hspace{1cm} (2.8)

Alternative hypothesis

\[ H_{1} : \mu_{1} \neq \mu_{2} \]  \hspace{1cm} (2.9)

The test statistic is calculated in equation (2.11), where \( D = m + k \).

\[ \max T^{2} = \max_{i=1, \cdots , m} \sum_{i=1}^{k} (X_{i} - \bar{X})'W^{-1}(X_{i} - \bar{X}) \]  \hspace{1cm} (2.10)

\[ T_{\max}^{2} = \frac{mk(D-2)}{D} (\bar{X} - \bar{Y})'W^{-1}(\bar{X} - \bar{Y}) \]  \hspace{1cm} (2.11)

In addition, \( W \) is called a pooled sum-of-products matrix and is calculated in equation (2.12).

\[ W = \sum_{i=1}^{m} (X_{i} - \bar{X})(X_{i} - \bar{X})' + \sum_{i=1}^{k} (Y_{i} - \bar{Y})(Y_{i} - \bar{Y})' \]  \hspace{1cm} (2.12)

In this method, the test statistic can only take positive values. In the maximum method, the analysis is performed by changing \( m \) from 1 to 39 and \( k \) from 39 to 1.

The detection accuracy of the maximum method for the change in the average value is shown below. We generate 100 random numbers from a standard normal distribution \( \mathcal{N}(0,1^{2}) \) for each column and add the specific value \( \mu' \) from the data on line 51. The maximum method is executed 1,000 times. As shown in Figure 3, for each \( \mu' \) value, the time point at which the change point detection number is the maximum is 50. This confirms that the differences in the average value can be accurately detected. In Figure 3 (c), the time point 50 becomes a change point 994 times. Compared with the maximum \( t \) method, the number of variables increases, thus the
detection accuracy is higher when each variable shows the same behavior.

(a) $\mu' = 1$  
(b) $\mu' = 2$  
(c) $\mu' = 3$

Figure 3: Change point detection accuracy of the maximum method using the Hotelling test

2.2.2 Sequential method using the Hotelling $T$-square test

We introduce the sequential method also in the change point search with plural financial indexes. 

\[ \frac{mk(D - 2)}{D} (\overline{X} - \overline{Y}) W^{-1} (\overline{X} - \overline{Y}) > \frac{c(D - 2)}{D - c - 1} F_{c, D-c-1} (\alpha / b) \]  

(2.13)

Similarly to the sequential $t$ method, a rejection limit based on the Bonferroni method is set. Moreover, $m$, $k$, and the degree of freedom $c$ are set at 3, while $b$ is set at 35 as the number of tests. The null hypothesis is rejected when the test statistic exceeds the rejection limit value on the right-hand side of equation (2.13). The sequential width is 3 because the product sum matrix would become 0 if it was set at 1, and the inverse matrix of the sum of products matrix cannot be calculated if it is equal to 2.

The simulation setting is the same as the maximum method using the Hotelling $T$-square test, and the result is shown in Figure 4. As in the sequential $t$ method, we focus on the period in which the test statistic is maximum. The reason for defining the period is that the change point cannot be specified at a single point because, unlike the maximum method, three points at a time are considered. As shown in Figure 4, at any $\mu'$ value, the number of detected change points is maximum in period 48. This means that when the average vector in the period between the time points 48 and 50 and the average vector in the period between the time points 51 and 53 are compared with each other, the difference is the largest.

It is confirmed that, similarly to the maximum method, the differences in the average value can be accurately detected. However, the detected change point varies greatly when the value of $\mu'$ is small. Figure 4 (c) shows that the number of times period 48 becomes the change point is 135. The detection accuracy is thus inferior with respect to the maximum method. It is important to note that the number of samples is small, and the width is set to 3. Since the number of variables is increasing as compared with the sequential $t$ method, the detection accuracy could potentially be higher when each variable shows the same behavior. Nonetheless, the accuracy can lower as the width changes from 2 to 3.
3. Data Analysis

3.1 Data

While applying the change point search methods, if the financial data of a small company are selected, there is a possibility that the test statistic is very high. To avoid this, in each industry we select a company whose sales are ranked higher. After careful consideration of industry bias and company size, 139 Japanese companies are selected. Companies whose financial data are extremely missing are excluded from the analysis.

The following sections search for change points in quarterly financial index data over a 10-years periods (from April 1, 2006 to March 31, 2016). We implemented complement for the missing values in order for the change amount to be uniform. For example, when the data in time point $t$ is 100 and the data in time point $t + 3$ is 175, 125 and 150 are substituted in correspondence of the missing time points $t + 1$ and $t + 2$. Financial data were retrieved Nikkei Value Search. We used the Nikkei Telecom 21 index to investigate corporate activities at the time of change point detection and examined them based on the Nikkei Newspaper articles. We consider sales for the single index analysis in Section 3.2. The unit is 1 million yen.

3.2 Change point search from single financial indicator data

3.2.1 Analysis results of maximum $t$ method

The change points detected by the maximum $t$ method are summarized in Figure 5. Since there are both positive values (increase in sales) and negative ones (decrease in sales), the detected change points are represented with the corresponding sign. As shown in Figure 5, many changes occur at time point 10 (July - September 2008) and time point 29 (April - June 2013). At these points, it seems that not only the specific industry but also the entire Japanese economy had the same influence on companies, as detected change points are all in the same direction.

On September 15, 2008, the Lehman shock triggered the global financial crisis and the yen appreciation suddenly advanced. In April 2013, the Bank of Japan introduced a different dimension relaxation, which resulted in a significant depreciation of the yen.
3.2.2 Analysis results of sequential $t$ method

Figure 6 shows the value of the test statistic for the sequential $t$ method. The significance level $\alpha$ is set at 5%, and the rejection limit value is set at 38.45. Since the time point at which the test statistic has been influenced cannot be specified in sequential methods, the horizontal axis measures the period. As shown in Figure 6, companies exceeding the rejection limit value are A (financial institution) in period 15 (October 2009 - September 2010), B (food company) in period 18 (July 2010 - June 2011), and C (financial institution) in period 19 (October 2010 - September 2011). During the change point detection period, company A was established through a merger while company C through business integration. Company B was promoting restructuring of its affiliated company.

3.3 Change point search with multiple financial indicators

3.3.1 Financial indicators data

Financial analysis is done from various viewpoints. Akimoto (2006) explains the importance of a comprehensive evaluation of the company. He also describes the analytical techniques for the analysis of profitability, safety, fund, growth, and added value/productivity. In the maximum and sequential $t$ methods, the change point search is performed based on the average value in a plurality of homogeneous time series data. Therefore, a weak point is
that they are equated even when dealing with data with different properties. On the other hand, since the analysis methods are change point search methods based on the average vector, it is actually possible to analyze different financial indicators. However, in the sequential method, the sequential width is forced to increase as the (variable) number of financial indicators to be analyzed increases. To accurately grasp the point when the change occurred, the number of target financial indicators should not be high. In this study, we select 3 indicators.

Although financial analysis can take diverse viewpoints, profitability, efficiency, safety, and growth are considered as the objectives of our analysis. The efficiency analysis follows Otomasa (2009). As financial indicators, we choose quarterly data on return on assets, capital adequacy ratio and sales growth rate, all in percentage term. Missing value supplementation was done in the same way as sales.

3.3.2 Analysis result of the maximum $t$ method

The maximum $t$ method was applied to each financial indicator. The detected change points are shown with the corresponding sign. Figure 7 (a) shows that the change point detected as a negative value since time point 7 (October - December 2007) is increasing, overlapping with the occurrence of the subprime mortgage crisis. After that, financial markets have been greatly influenced by the mortgage loan bad debts, leading to the Lehman shock, which has further developed into a global financial crisis. The industries in which many change points were extracted from time 7 (October - December 2007) to time 12 (January - March 2009) were real estate, steel metal, wholesale, transportation, automobiles, and precision equipment. This may be caused by the deteriorating business performance in export due to trade reduction. The real estate company was the proliferation of interest bearing liabilities due to aggressive investment in the period of falling sales price.

The result from capital adequacy data is shown in Figure 7 (b). The change point did not concentrate at a specific point in time as compared with the previous two indicators. This result suggests that the change in the capital adequacy ratio does not directly reflect the influence of foreign exchange and the economy. Many change points of positive values were detected at time 10 (July - September 2008) and at time 29 (April - June 2013). According to the Nikkei newspaper, the corporate statistics show that the capital adequacy ratio reached its historical high at the time point 28 (January - March 2013).

Finally, Figure 7 (c) shows the result from sales growth rate data. The points of change are concentrated at time 1 (April - June 2006) and time 39 (October - December 2015). In 2006, the longest economic expansion after the war occurred, while in 2015 the economy was stagnant due to investment declines and consumption reduction.

![Figure 7: Number of change points detected by the maximum $t$ method](image)
3.3.3 Analysis result of the sequential $t$ method

Subsequently, we perform the sequential $t$ method for each financial indicator. Figure 8 (a) shows the result on return on assets data, with a significance level $\alpha$ of 1% and a rejection limit value of 86.01. Since the sequential method cannot specify the exact point of influence on the test statistic, the horizontal axis measures the period. As shown in Figure 8 (a), companies exceeding the rejection limit are E (chemical company) in period 7 (October 2007 - September 2008), F (financial institution) in period 11 (October 2008 - September 2009), and G (automobile manufacturer) in period 15 (October 2009 - September 2010).

Company E had increased sales due to management diversification, but it has deteriorated profitability because of yen appreciation and rising raw material prices. Company F had a deteriorating profit due to the fall in the Nikkei Stock Average, but the stock price recovered to the 10,000-yen level, which led to the company’s profitability improvement. As for company G, overseas sales increased and the final profit and loss have turned into a surplus at time 17.

Next, the result from capital adequacy ratio data is shown in Figure 8 (b). Specifically, companies exceeding the rejection limit are H (transportation company) in period 4 (January 2007 - December 2007), I (transport company) and J (pharmaceutical company) in period 6 (July 2007 - June 2008), and K (precision equipment company) in period 9 (April 2008 - March 2009). For company I, the capital adequacy ratio declined as a result of the introduction of new vessels due to the strong demand for transportation as well as the lowering of the Baltic Shipping Index. Company J acquired the US pharmaceutical company, but due to other expenses, it missed the eighth consecutive highest profit. For company K, sluggish sales of digital cameras and office machines, the occurrence of amortization of goodwill and intangible fixed assets due to the acquisition of medical device companies in the US.

Figure 8 (c) shows the results from sales growth rate data. Only the real estate company L in period 27 (October 2012 - September 2013) exceeded the rejection limit value. Since the sales growth rate data of many companies repeatedly increases and decreases, it is difficult to detect the change point with the method using the average value. For company L, the fluctuation was small between points 27 and 28 and between points 29 and 30, whereas the value decreased significantly between points 28 and 29.

![Graphs showing change point detection by the sequential $t$ method](image)

(a)Return on assets  (b)Capital adequacy ratio  (c)Sales growth rate

Figure 8: Change point detection by the sequential $t$ method

3.3.4 Maximum method with the Hotelling $T$-square test

Figure 9 summarizes the number of change points detected by the maximum method through Hotelling $T$-square test. There are only positive values on the vertical axis. It can be seen from the figure that many change points
were detected at time points 10 (July - September 2008) and 27 (October - December 2012). It seems that the test statistic is negative at time point 10 and positive at time point 27 for many companies. The two time points above are the points before the exchange and stock prices undergo major changes.

![Figure 9: Number of change points detected by the maximum method with the Hotelling test](image)

3.3.5 Sequential method with the Hotelling $T$-square test

Figure 10 shows the result of performing the sequential method through the Hotelling $T$-square test. The significance level $\alpha$ is set to 1%, while the rejection limit value at 20,995. Companies that greatly exceed the rejection limit value are M (chemical company) in period 1 (April 2006 - September 2007), N (ceramic company) in period 10 (July 2008 - December 2009) and O (metal company) in period 24 (January 2012 - June 2013). Since the sequential width is set at 3 for calculation of the test statistic, the target range of the period is widened.

![Figure 10: Change point detection by the sequential method with the Hotelling test](image)

The test statistic for company N was particularly large, thus it was not possible to confirm a large change with respect to the single financial indicator in the change point detection period. As a result of the factor analysis, the following events occurred during the change point detection period.

- Penalty by the European Commission (the highest amount for a Japanese company at the time);
Thus, during this period company N seemed to be in a state of crisis management. However, from the three financial indicators data and the test statistics of the sequential $t$ method, it is not possible to confirm any sudden change for the company. A similar result is obtained for the other two companies. Specifically, company M was squeezing revenue on amortization of goodwill when acquiring a chemical company, while company O was forced to cut down employees after an aggressive M & A.

4. Conclusions

Financial analysis is commonly conducted not only based on a single financial indicator, but also by observing changes in multiple financial indicators. By using change point search methods using multiple financial indicators in this study, we could quantitatively show changes in corporate management that were not detected by just looking at financial data.

In this study, we analyzed financial data by assuming that they are normally distributed. However, each financial indicator exhibited low fitness test values. Thus, the reliability of the analysis should be further enhanced by applying it to actual data with higher fitness. Moreover, in the analysis method, the abnormality detection of the inspection system itself based on various measurement values could be considered.

References:

Akimoto, T. (2006), "Business Analysis and Corporate Evaluation", Sosei Co.
Hirotsu, C. (1992), "Analysis of experimental data - Beyond analysis of variance", Kyoritsu Publishing.
Kawai, M. (2011), "Proposal of change point search method for large scale data", Master’s thesis at Waseda University graduate school.
Nikkei Telecom 21, https://t21-nikkei-co-jp.waseda.ac.jp/g3/CMNDF11.do, Last accessed : January 19, 2016.
Nikkei Value Search, https://valuesearch-nikkei-co-jp.waseda.ac.jp/home?serverid=002, Last accessed : January 19, 2016.
Okada, A. (2013), "Proposal of change point searching method considering variance for large scale data", Master’s thesis at Waseda University graduate school.
Otomasa, S. (2009), "Analysis of Financial Statement", Dobunkan Publishing.
Takemura, A. (1991), "Basis of Multivariate Inference Statistics", Kyoritsu Publishing.
Tanaka, K. and Nakagawa, H. (2014), "A method of corporate credit rating classification based on support vector machine and its validation in comparison of sequential logit model", The Operations Research Society of Japan, vol.57, 2014, pp.92-111.
Wen, F. and Xiao, J. and He, Z. and Gong, X. (2014), "Stock Price Prediction based on SSA and SVM", Procedia Computer Science, vol.31, 2014, pp.625-631.

Acknowledgement:

We would like to thank the anonymous referees for their valuable comments. This work was partly supported by JSPS Grants-in-Aid for Scientific Research Grant Number 15K00056.
Authors’ biographical notes

Satoshi Kawamata is a graduate student in the Department of Industrial and Management System Engineering of the Graduate School of Creative Science and Engineering at Waseda University. Yasushi Nagata is a professor in the Department of Industrial and Management System Engineering of the School of Creative Science and Engineering at Waseda University.

[DOI: 10.17929/tqs.4.41]

Received, March 21, 2017
Revised: June 9, 2018
Accepted: June 17, 2018