# Analysis of Investors' Behavior through Non-Time Series Analysis of Stock Prices

Kenichi YOSHIDA, University of Tsukuba

April 1, 2019

After the war began with the special demand economy in the early 1950's, Japan was followed by a reactionary recession and repeated business cycles such as the "izanami economy" and the "Global Financial Crisis." In this study, we attempt to analyze the investors' behavior behind these business cycles by analyzing TOPIX price movements from 1954 to 2016 using simple data mining methods. This working note reports the results of the analysis:

1. We use the gradient boosting decision tree (GBDT) as the data mining method to analyze TOPIX price movements; this method can analyze mixed distribution. Furthermore, as a method to detect the change in investors' behavior, we optimize the analyzing period of the data mining process by considering only the fluctuation $d_t$ in the monthly TOPIX price and the standard deviation $s_t$ of the daily price within the month.

2. The academic contributions of this paper include:

   (a) The training periods used to create the prediction model suggest the existence of a long-term investors' behavior over the business cycles (Fig.3).
   Although various studies have been made to clarify the mechanism behind business cycles, our study shows the stable investors' behavior over the business cycles solely on stock price information (Fig.6).

   (b) We show the existence of anomaly where stock prices can be predicted in multiple markets (Fig.2,13-17). As shown in Fig. 7 & 8, prior to 1990, the Sharpe ratio of the proposed method is worse than that of index investment. After 1990 the results are reversed and the method used in this study outperforms index investment.
   Since the results shown in Fig. 8 seem to suggest a change in the investors' behavior after 1990, we applied data from 2001 onwards to the CAC, DAX, SPX, and UKX indexes, with the results shown in Fig. 13-17. It is seen that the proposed method earns profit on all these indexes during this time period with Sharpe ratios that are universally higher than the corresponding index investment results.

   (c) Note that an out-of-sample testing framework is used to produce the results shown in Fig. 2,13-17. In other words, we always use past data to generate a model for predicting future prices. However, intermediate models generated in this process have always shown similar results if interpreted using a retrofitting method (Fig. 4 & 5). Although Fig. 4 & 5 only show the retrofitted interpretation of the results, they can be interpreted as follows:

   - When prices crash, the price in the next month will go down.
   - In the adjustment phase, an increase in price will go down in the next month. A decrease in price will go up in the next month.
   - Otherwise, the price will go up.

   where the terms "price crash" and "adjustment phase" are defined numerically as follows:

   | | |
   |---|---|
   | price crash: | $d_t < -0.05$ and $s_t > 0.029$ |
   | adjustment phase: | $\lvert d_t \rvert < 0.05$ and $s_t < 0.029$ |

   In the application to past TOPIX data, the accuracy of this model is about 60%. This relatively high accuracy seems to be the cause of the high return shown in Fig. 2.

   (d) Fig. 4 & 5 also indicate a mixed distribution underlying the data.
   The found regularity requires the analysis of discrete time points, and we contend that linear differential relations are not well suited to this task. For example, the simple rule "**If** $d_t < -0.05$ **and** $s_t > 0.029$ **then** price will go down" cannot be simply handled by a linear differential equation. The non-time series analysis used in this paper seems to be more appropriate for handling this type of regularity.

1) The relationship between the identified training period and the business cycles studied by previous researches, and 2) whether the prediction ability found this time will continue, are left as future research issues.