DP

RIETI Discussion Paper Series 19-E-103

Using Machine Learning to Detect and Forecast Accounting Fraud

KONDO, Satoshi
KPMG AZSA LLC

MIYAKAWA, Daisuke
Hitotsubashi University

SHIRAKI, Kengo
KPMG AZSA LLC

SUGA, Miki
KPMG LLP

USUKI, Teppei
KPMG AZSA LLC

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 19-E-103
December 2019

Using Machine Learning to Detect and Forecast Accounting Fraud*

Satoshi Kondo, Daisuke Miyakawa, Kengo Shiraki,

Miki Suga, and Teppei Usuki**

Abstract

This study investigates the usefulness of machine learning methods for detecting and forecasting accounting fraud. First, we aim to "detect" accounting fraud and confirm an improvement in detection performance. We achieve this by using machine learning, which allows high-dimensional feature space, compared with a classical parametric model, which is based on limited explanatory variables. Second, we aim to "forecast" accounting fraud, by using the same approach. This area has not been studied significantly in the past, yet we confirm a solid forecast performance. Third, we interpret the model by examining how estimated score changes with respect to change in each predictor. The validation is done on public listed companies in Japan, and we confirm that the machine learning method increases the model performance, and that higher interaction of predictors, which machine learning made possible, contributes to large improvement in prediction.

Keywords: Accounting fraud, machine learning, fraud detection, fraud forecast, interpretability

JEL classification: M42, C53, C14

*This paper is the result of joint research between Hitotsubashi University and KPMG AZSA LLC, as well as the results of the Research Institute of Economy, Trade and Industry's (RIETI) "Corporate Finance and Behavior Dynamics Research Group" project. For the original draft of this manuscript, Kazuhiko Ohashi, Toshiki Honda, Hidetoshi Nakagawa, Kenji Suzuki, Iichiro Uesugi, Arito Ono, Kaoru Hosono, Tsutomu Ogawa, Takao Shimizu, Eiji Fukami, Jun Miyashita, Atsushi Nakajima, Makoto Yano, Masayuki Morikawa and the RIETI Discussion and Paper Committee all provided many useful comments. We are sincerely grateful for their contributions.
** Usuki, Kondo, Shiraki: KPMG AZSA LLC Digital Innovation, KPMG Ignition Tokyo, OTEMACHI building 7F, 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-0004; Suga: KPMG LLP Financial Due Diligence 1350 Avenue of the Americas, New York, NY, UNITED STATES, 10019; Miyakawa: Corresponding author, Associate Professor, Hitotsubashi University Graduate School of Business Administration, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8439 E-mail: dmiyakawa@hub.hit-u.ac.jp

1. Introduction

Not to mention some cooperate accounting scandals happened in home country or abroad, the misreporting of financial information (so-called "falsification of financial statements") is a serious economic event that should be avoided from practical point of view. The misreporting of listed firms' financial information, in particular, distorts the decision-making of various economic entities involved in financial transactions such as stock, bond trading as well as bank lending, resulting in inefficient resource allocation. In real business relationships, such misreporting may also result in excessive risk-taking that neither the customer nor the supplier recognizes. As a result, when this risk becomes apparent, unintended stagnation of economic activities may occur. Even more serious, when such misreporting is intentional (so-called "accounting fraud") and occurs frequently, financial activities and real economic activities could not be properly initiated from the outset (i.e., market breakdown).

These problems caused by accounting fraud are not necessarily limited to business activities but, rather, extends to policy management. Corporate financial information is referenced in various policy interventions such as the provision of subsidy to small and medium size enterprises. If a company's information observed from outside does not represent the reality, its intended policy may not be implemented correctly.

Furthermore, when evaluating policy effects after the policy intervention, the existence of such accounting fraud can lead to serious errors in the policy evaluation.

Based on the awareness of these issues, theoretical research has been conducted on fraudulent accounting mechanisms, mainly in the accounting field (e.g., Dechow et al. [1996]). The first type of such theoretical research relates to the "reasons" companies engage in accounting fraud. For example, a theoretical study points out that a company faced with sluggish business performance is likely to attempt to realize better procurement terms by intentionally falsifying financial information. By describing the optimal behavior of a company given certain objective functions and constraints, these theoretical studies have specified determinants of accounting fraud.

The second type of theoretical research seeks to identify financial information that is "correlated" with the occurrence of accounting fraud by referring to practical knowledge in the accounting field. For example, those studies have been looking at “discretionary accounting accruals”. With the second type of theoretical research, should such a discretionary accounting accrual—thought to be linked to management's profit adjustment behavior—be observed, there is a high probability that some accounting fraud has occurred.

Based on these theoretical considerations, in recent years further empirical efforts have been made to identify companies likely to engage in accounting fraud (detection of accounting fraud in the financial statements of each period). Specifically, statistical models have been developed to detect the occurrence of accounting fraud; Dechow et al. [2011] and Song et al. [2016] are the primary research in this area. In these two empirical studies, based on the theoretical arguments therein, the independent variables are manually selected and put into parametric models to predict concurrent event. These studies confirm that the variables are showing consistent results with theoretical discussions, and the models show a good in-sample fit.

We can point out at least two issues. First, with current efforts aimed at detecting accounting fraud, there are a huge number of variables yet to be considered. For example, Song et al. [2016] introduced a total of six variables into the model based on theoretical assumptions; however, other vast amount of information about the companies of our interest may bring additional prediction power. Although limiting the number of variables is suitable for testing theoretical hypotheses, there is room for improvement in model performance. It is understandable that existing research uses limited number of variables simply because parametric models can only handle so many. We expect better prediction with machine learning models, which can overcome this limitation.

The first purpose of this study is to build an accounting fraud detection model using machine-learning methods. This allows us to incorporate a large number of explanatory variables. Something to add is that, as discussed in the recent well-known article by Perols [2011] and Perols et al. [2017], detecting accounting fraud is a needle in a haystack. (so-called "imbalance problems"). Appropriate handling is done to tackle this issue. Based on the results of theoretical examinations that existing research has referenced, we use the broadest possible swath of corporate information as explanatory variables. These additional variables include not only financial indicators constructed from corporate financial information, but corporate governance-related variables with a focus on the shareholder, executive and employee information, and bank transaction variables based on banking information.

Next, most existing research focuses explicitly on the detection of accounting fraud in current statements (so called nowcasting), while the occurrence of future accounting fraud (so called forecasting) is not explicitly addressed. Unlike, for example, a bankruptcy event, where the event is transparent to public, observing an accounting fraud in real time is not necessarily possible. In order to detect hidden fraud events, nowcasting can bring a great practical value in the field of auditing. From the view point of the audit business, it would also be prudent to screen in advance those companies for which financial accounting misreporting is highly likely to occur.

The second objective of this research is, thus, to apply the machine learning-based analysis framework to forecast accounting fraud (corporate level forecasting of fraudulent accounting events at a future time), which existing research does not sufficiently address, and to verify the performance of this forecasting model.[1]

The way we verify the results of the "detection" performance for accounting fraud, which was obtained from the analysis using the data of listed companies in Japan, is summarized as follows. Pursuant to the verification of detection performance and given the aim of verifying the degree of improvement in performance, we compared the detection performance of a parametric model that relies only on limited variables that emulate the results of existing research (Model 1); a non-parametric model that uses machine learning and relies only on limited variables (Model 2); and a non-parametric model constructed using machine learning after variables were expanded (Refer to Models 3-16 and, especially, Model 12 as complete models where all variables have been inputted).

First, we confirmed that both the use of machine learning techniques (comparing Models 1 and 2) and high-dimensional feature space (comparing Models 2 and 12) improved detection performance. Since the expansion of variables in Model 12 is

[1] See West and Bhattacharya [2016] for a survey of recent prediction and forecast model development trends targeting fraud (credit cards, securities, insurance) in the financial sector, including accounting fraud.

possible only with the use of machine learning, it is not easy to measure how much "application of machine learning methods" and "variable expansion" actually contributed to the fraud detection performance independently. The above results suggest there is room for ingenuity when it comes to the methods of constructing a detection model even when using a variable group similar to that of the existing model.

Second, we build a machine learning-based model to "forecast" accounting fraud, aiming to predict fraud events happening one year after the time of scoring. We validate the performance in hold-out sample. Then, we confirm that a sufficient level of performance can be achieved. This result means that it is worthwhile to use machine learning model with high-dimensional data to forecast future accounting fraud, which existing research has yet dealt with explicitly.

Third, in constructing these detecting and forecasting models, based on theoretical assumptions, it is found that variables (e.g., corporate governance-related variables) other than those referenced by existing research contributes to a certain degree. Specifically, the average length of employee service and the percentage of outstanding shares held by company executives rank high in terms of the degree of importance to forecasting the occurrence of future accounting fraud. We also confirm how the estimated forecast score fluctuates when these variables fluctuate. These results imply that, besides those variables suggested by theoretical assumptions presented in existing studies, there may be many other variables that contribute significantly to detecting and forecasting fraud. In this sense, the results obtained in this study indicate that there is more room for the practical use of the models when feature space is expanded, while also providing information useful for the future theoretical consideration of accounting fraud mechanisms.

The composition of this paper is as follows. Section 2 outlines the existing research that is the premise of this study. Section 3 explains the institutional background of accounting fraud, which is the premise of this study's analysis. Sections 4 and 5 explain the verification strategies and data used for analysis. Section 6 summarizes the results of our empirical analysis. Sections 7 and 8 discuss and conclude the study.

2. Literature Review

Theoretical research on fraud in the accounting field consists of a series of studies that have modeled accounting fraud and examine its determining factors, as well as discussions around fraud patterns mainly from accounting perspective. An example of the former is Dechow et al. [1996], which describes one of company's incentives to make fake profit is sustaining finance. An example of the latter is a series of studies that discuss the correlation between accruals quality and the occurrence of accounting fraud (e.g., Dechow et al. [2010]).

Dechow et al. [2011] is a renowned study on the construction of accounting fraud detection models. They chose variables such as accruals quality, financial performance, nonfinancial measures, off-balance-sheet activities, and market-based measures that are theoretically seemed to be related to the occurrence of accounting fraud. Then, they came to understand the correlation between the predicters and the concurrent accounting fraud, and, thus, estimated a score representing the likelihood of accounting fraud occurring at present. Similarly, in Song et al. [2016], in addition to the variables referenced in Dechow et al. [2011], variables related to real-activities manipulation, conservatism, and Japanese-specific factors are used in parametric models. They confirm that variables established by theoretical arguments are indeed correlated with the accounting fraud, and the sign of the estimates are as expected.

Regarding financial fraud in general, West and Bhattacharya [2016] provide a survey of recent empirical studies, targeting to detect/forecast fraud. Perols [2011] and Perols et al. [2017], in particular, are the prominent studies in recent years that involve constructing accounting fraud detection models using machine learning methods. They made use of a support vector machine to detect accounting fraud, while also discussing how best to deal with data imbalance issues.[2]

Contributions of this study based on prior research can be summed up by the following two points. First, very few studies have used machine learning-based models to comprehensively detect and forecast misstatement on financial statements while responding to imbalance problems. The study aims at contributing to the accumulation of empirical findings in the field. Second, as discussed in Song [2018], in addition to information that existing research refers to, it is highly possible that there are more variables that contribute to the prediction of accounting fraud. The present study accounts explicitly for this issue by using variables such as the ones frequently referred to in audit practice, governance-related variables, and lender bank-related variables.

3. The Practical Background of Accounting Fraud

3.1 Types of Accounting Fraud

In this study, accounting fraud is defined as "the act of disclosing financial accounting information, the contents of which do not reflect the actual situation." There are several types of fraudulent accounting cases that fall under this definition.

First, there is the overstatement of revenue (e.g., sales) through the recording of bogus sales. If cost is recorded accurately, an overstatement of sales, which is the top

[2] There are many examples of machine learning-based models being constructed to predict corporate bankruptcy and the like (e.g. Miyakawa 2019a).

line of the profit and loss statement, will lead to an overstatement of pretax income equivalent to the overstated amount. Most cases of past fraudulent accounting have been of this type. Next, we have an understatement of expenses centered on the cost of sales. The past examples also include understating sales expenses and general administrative expenses such as advertising expense. Such an understatement of costs can also be seen for valuation cost items involving top managements' estimates, such as impairment losses, allowance for loan losses, and provision for loss on guarantees. Last, we have overstatements of assets such as goodwill. Types of typical assets that fall in the category of "overstatement of assets" are current assets, such as inventories, property, plant and equipment (as represented by production facilities), and intangible assets (as represented by software). Given liabilities, an overstatement of assets leads to an overstatement of net worth.[3]

In this study, all of these types are treated collectively as the targets of our interest; construction of the models classified by the type are tasks worthy of future attention. In the context of model construction for predicting accounting fraud, there have also been discussions on the need to build models for each different type (e.g., Perols et al. [2017]). In fact, existing research that utilizes machine learning-based

[3] In addition to overstatement of sales, understatement of expenses, and overstatement of assets, we can also mention an understatement of debt. However, looking at past examples, these seem to be minor.

models (e.g. Miyakawa [2019a]) states the following. Take, for example, the case of bankruptcy and business closure. They are both an event of closing the door, however, there is a big difference in the list of variables that contribute to the estimations, and the scores corresponding to the occurrence of each event do not necessarily correlate highly.

3.2 Accounting Fraud Systems

This section provides an overview of the systems in measuring fraudulent accounting events, which will be referred throughout in this study. First, in this study, a company is considered as committed an accounting fraud if it received an order to pay a fine issued by the Japanese Financial Services Agency (hereinafter referred to as FSA) due to misstatements on an annual securities report or similar incident. In an attempt to ensure fairness and transparency in the stock market, accounting fraud systems in Japan were first introduced in April 2005 as an administrative measure against illegal acts that damage trust in the stock market, effectively imposing financial penalties on those who have committed illegal acts. Illegal activities covered by the civil penalties system include fraudulent transactions such as insider training, as well as misstatement on securities reports (such as violations of continuous disclosure obligations) and similar documents, and the Commissioner of the Financial Services Agency issues a fine payment order based on the civil penalties system. Second, a company is also considered as committed an accounting fraud if the Japanese Securities and Exchange Surveillance Commission recognizes the company's engagement in false securities report and files a change to the Public Prosecutors Office.

The source for identifying false statements is securities reports and similar documents, and a public inspection period of five years exists for these. Therefore, when making corrections to misstatements identified by the above system, the possible period for this is limited to the public inspection period (i.e., five years). As later described, this study first targets the accounting fraud events detected only by the above two systems. However, as an alternative measurement of accounting fraud, we next target events that also include corrections in financial statement response to those accounting frauds.

Note that only events discovered by the above system are our target in this study. For this reason, if the above system does not discover or has not yet discovered accounting frauds, there will be measurement errors related to our outcome. In other words, we could be constructing our models by using the data potentially with measurement errors in the outcome, and testing the performance of those models by using the data with the errors. In this sense, the evaluation of our models is conservative.

4. Methodology

In this section, we will take a look at the two-step process of machine learning model build, training and testing, respectively.

4.1 Methodology: Training

The structure of the model to be built does not depend on whether the purpose of the analysis is to detect an event occurring at present or to forecast an event in the future. The objective is to calculate Score $\boldsymbol{S}$, which corresponds to the probability of an event now or in the future, given information vector $\boldsymbol{X}$ available at the time of scoring.

$$S = \mathrm{F}(\boldsymbol{X})$$

Modeling data needs to mimic the real snapshot data. For detection, $\boldsymbol{X}$ and our target are prepared as of the same point-in-time. For forecast, $\boldsymbol{X}$ is prepared as of prior point-in-time than our target. How far in advance $\boldsymbol{X}$ needs to go back depends on how far in advance we want to forecast the event. For example, to forecast the occurrence of accounting fraud within the next one year, $\boldsymbol{X}$ needs to be observed at the beginning of the one-year prediction window.

The most existing empirical studies in the accounting domain treat the target as a binary (i.e. 1/0) variable – event observed or event non-observed. With appropriate assumptions made with regards to the error term ε, we can employ the models such as *logit* and *probit*. Then, we can estimate the parameters $(\alpha, \boldsymbol{\beta})$, and predict $\boldsymbol{S}$ for given information vector $\boldsymbol{X}$.[4]

$$S = \mathrm{F}(\boldsymbol{X}; \alpha, \boldsymbol{\beta}) = \mathrm{Prob}(L^* > 0) \quad \text{where } L^* = \alpha + \boldsymbol{X\beta} + \varepsilon$$

It should be noted that such an approach has several problems in the context of prediction. First, there is a possibility that what exactly this information vector $\boldsymbol{X}$ should be is not always clear. As stated above, historical empirical efforts to estimate account fraud has involved the selection of this information vectors based on underlying theoretical arguments in the accounting field. However, as discussed in Song [2018], there are many suitable references from a practical or economic point of view even though existing theoretical studies in the accounting field do not adequately take these into account (e.g., governance variables). If the purpose is to detect fraudulent accounting events with higher performance, high-dimension information should be introduced to the model. Unfortunately, parametric models such as *logit* and *probit* are not ideal for such purposes.

Second, the model may exhibit complex non-linearity, which violates the assumption of *logit* and *probit*. To introduce complex non-linearity into a model, one can create variables that are made up of multiple attributes in $\boldsymbol{X}$ (e.g., interaction terms) or

---

[4] The practical procedures for detecting and forecasting consist of (1) using a sample to estimate these parameters, (2) using hold-out data not used in such estimates to evaluate detection and forecast performance and confirm that sufficient performance has been ensured, and (3) calculating S using information X of the company targeted for detecting/forecasting.

transform specific attributes (e.g., higher order terms, discretization of variables). However, this may result in the issue of the variable dimensions mentioned in the first problem once again.

Based on these discussions, this study constructs a machine learning-based model. Specifically, we use Random Forest, which ensembles many decision tree-based models, each of which is built on a subsample of the modeling data. Pursuant to constructing a decision tree, first, starting with a dataset containing both fraud and non-fraud records, all possible split-points for each attribute in ***X*** will be considered. Next, using a certain indicator, we measure the information gain (the decrease in the degree of fraud and non-fraud mix) which we get by dividing the data by those split-points. After identifying which attribute at which split-point maximize this information gain, the branching rule will be established. Thereafter, we construct a decision tree by repeating this procedure until we finally reach a state in which only fraud or non-fraud is included. Upon doing so, rather than constructing each tree using all of the data, constructing each tree using a subset of the data and/or the variables is called "Random Forest" (Breiman [2001]). It should be noted that, as is evident by this analytical logic, the issues of variable dimension in classical parametric models is not a major problem here.

However, another issue requires a consideration due to the characteristics of the accounting fraud of our interest – i.e. rare event. Such problems occur in various situations; one example is bankruptcy. This study employs Weighted Random Forest, an extended version of the Random Forest proposed by Chen et al. [2004]. When constructing individual decision trees and ensembles, this approach involves imposing a relatively large weight on rare events.

The step where we construct a classifier composed of individual decision trees using the above procedure is hereinafter referred to as the training step. By completing this step, a nonparametric function (classifier) can be obtained, where analysts can input $\boldsymbol{X}$ and $\boldsymbol{S}$ will be outputted.

4.2 Methodology: Tests

Before using the classifier constructed in the training step in actual practice, we must first confirm how much performance the classifier can achieve. For this, we calculate $\boldsymbol{S}$ given $\boldsymbol{X}$ on the hold-out data, which never used in the training process, and validate the score against the actual fraud and non-fraud outcome in the hold-out-data. A typical way is to set a threshold for $\boldsymbol{S}$ with some criteria; this is done with reference to performance indicators such as accuracy, precision, recall, and F-measures. While these methods have the advantage of making it possible to intuitively understand the performance results, they are disadvantageous in that the result is dependent on a specific threshold.[5] Therefore, this study refers to a performance metrics based on the ROC curve as a method of more robust performance evaluation. Here, the ROC curve is a locus drawn by a collection of coordinate points (consisting of False-positive and True-positive rate), each of which is obtained by changing thresholds; by referring to the area under this curve, performance can be expressed as a single number called AUC (Area Under the Curve).[6]

4.3 The "Interpretability" of the Trained Model

One criticism of machine learning-based non-parametric models is that it is difficult to interpret changes in model core S in line with changes in specific information in $\boldsymbol{X}$. It is by no means easy to associate forecast results obtained from an ensemble of multiple decision trees with specific information in $\boldsymbol{X}$, as demonstrated by the explanation of the training steps in the previous section. However, in practice, there are many situations where it is necessary to answer the question, "Why does the company score so high for this particular event?" It sure is not desirable in a practical sense when the forecasting classifier is completely a black box.

[5] For example, the recall metrics accounts for how much of events currently occurring (or will occur in the future) are correctly detected (or forecasted) under a specific threshold. This indicator places as much importance as possible on being alert to occurrences of the actual event. It is, for example, useful in detecting serious diseases. However, if the threshold is set as low as possible, the recall metrics will always be 100%. Thus, it is problematic to us only one specific indicator.
[6] See Appendix 1 for more information on ROC Curves and AUC.

Therefore, this study responds to these criticisms in part by referring to variable importance (Janitza et al. [2018]), which accounts for what degree a particular variable contributes to improvements in the classification performance. Moreover, we identify the qualitative implications of predictions by fluctuating specific information in $\boldsymbol{X}$. Beginning with a hypothetical company having the average of all the attributes, we plot visually how much $\boldsymbol{S}$ changes as we move each information in $\boldsymbol{X}$ away from the average by a small amount at a time, keeping all others fixed.

Note that these attempts may not necessarily explain "causal effects" that a particular factor brings on the accounting fraud, when all the other factors are fixed. For examples, let us assume that it has been confirmed that a specific variable in $\boldsymbol{X}$ indicated high variable importance, and increase in the variable is linked to a large increase in $\boldsymbol{S}$. In the context of prediction, this result is meaningful in that, with increase in the variable, there is a higher probability that accounting fraud occurs currently (nowcast) or in the future (forecast). However, this result does not preclude the possibility that any other variable(s) that is (are) highly correlated with this variable will also be highly correlated with fraudulent accounting events. We need to carefully interpret the results after recognizing the possibility of such spurious correlations.

In recent years, various methods have been proposed to estimate the causal impact that certain variables in $\boldsymbol{X}$ have on S (Chernozhukov et al. [2018]), and there is also a gradual increase in the number of applicable cases (Miyakawa [2019b]). If these causal relationships can be accurately measured in addition to just the predicted scores, more proactive approach becomes possible. For example, if it is confirmed that active efforts such as adopting a specific governance structure or introducing regulations to prevent the accounting fraud from happening, this information will be important from a practical and strategic standpoint. We will consider this point for further study.

5. Data

This section provides an overview of the data sets for analysis. In this analysis, we will analyze companies that are listed in the Japanese securities market and employing Japanese accounting standards. We exclude financial companies.

5.1 Accounting Fraud Flags

To identify the firms committing accounting fraud, we look at whether the FSA has ordered the company to pay a fine for misstatements on financial statements (annual securities reports) and other documents, and/or whether the Securities and Exchange Surveillance Commission have filed reports of false securities reports to the Public Prosecutors Office. By analyzing historical revisions of those reports, we can further measure what corrections have been made to the past annual securities reports, making it possible for us to comprehensively capture misstatement of financial information by Japanese companies. Note that those findings lead us to identify what company and which fiscal year for which such a serious fraud, which led to public sanctions, had occurred. Pursuant to this analysis, we introduce two types of Fraud Flags; Fraud Flag 1 (only main event), which indicates public sanction, and Fraud Flag 2 (main + ancillary event), which additionally indicates revised records not subject to these sanctions. All flags are labeled at company x fiscal year level.

Predicting Fraud Flag 2 can be a fairly difficult task in the sense that they are indicating relatively less serious events from accounting fraud perspective. One of the goals of analysis that utilizes the above two flags is that this study examines how much performance can be expected with tasks that are difficult to predict.

Accounting Fraud Flag 1*:*

If, for a given company and a given fiscal year, at least one of the following two conditions is met, this dummy flag equals 1 (positive data); otherwise, it is 0 (negative data).

1. *FSA issued an order to pay fines for misstatements on annual securities reports or similar documents to the company for the fiscal year.*
2. *Securities and Exchange Surveillance Commission accused the company of*

misstatements on annual securities reports or similar documents for the fiscal year where the annual securities reports were corrected.

Accounting Fraud Flag 2*:*

If, for a given company and a given fiscal year, either of the two conditions for Accounting Fraud Flag 1 (1, 2), or condition 3 below (so all together 1, 2, or 3) is met, this dummy flag equals 1 (positive records); otherwise, it is 0 (negative records).

3. *The companies corrected their annual securities reports for the fiscal year other than the years for which companies are fined or accused by the FSA or the Securities and Exchange Surveillance Commission*

As is apparent from the above definition, Accounting Fraud Flag 1 measures only serious accounting fraud events, while Accounting Fraud Flag 2 also includes misstatements incidental to the fraudulent accounting event and which requires correction. In practice, predicting an event corresponding to Accounting Fraud Flag 1 is considered to be the most important matter. However, should it be possible to predict misstatement associated with such important fraudulent accounting events even in years when there are no orders to pay charges or indictments related to misstatement, this would no doubt be thought of as relevant information in terms of business management and audit practices.

Figure 1 shows the number of observations of accounting fraud by these flags (vertical axis) for each year (horizontal axis). Looking at this, we can see that the number of flags decreases in the second half of the analysis target period. This is considered to be due to the fact that the detection of accounting fraud based on the above-mentioned system generally takes a certain length of time from the actual time the fraud occurs, and that the time until such discovery cannot be ensured. This means that, compared to the training data, which uses the first half of the analysis period, the number of fraudulent events (the number of positive records) is relatively small in the testing data, which uses the second half of the period. Thus, to make sure the possible qualitative differences in flags between training and testing data do not have a critical effect on the results, we also validate the robustness by splitting the data at company level within the same period.

5.2 Variables

The candidate variables for the prediction in this analysis are created from the following three data sources. The first source is financial information in the financial statements of each publicly-traded company for each fiscal year, which are extracted from the data described in Section 5.1. Pursuant to constructing the variables, we first prepare the variables used in Song et al. [2016], which is representative of prior studies on accounting fraud detection for Japanese companies. Specifically, these include the *%*

Soft assets calculated by subtracting PP&E (Property, Plant and Equipment), cash and short-term investments from total assets and dividing this by total assets; the absolute value of the CFO discretionary accruals (*CFO discretionary accruals AB*); dummy variables for the issuance of stocks and bonds (*Actual issuance*); scores for abnormal returns (*C score AR*); substantial discretionary behavior (*AB cash flow*); and the ratio of stock held by non-financial corporations (*CORP*). These variables are brought to the table based on degree of manipulation of actual economic activity, accounting conservatism and factors unique to Japan, in addition to those thought to be theoretically related to the occurrence of fraudulent accounting, such as accruals quality, financial performance, nonfinancial measures, off-balance-sheet activities, and market-based measures as discussed by Dechow et al. [2011] and others. When these variables are missing for some reason, the missing value is replaced by 0, after which a dummy variable has been created to indicate the lack of the value.[7]

While these are plausible variables based on theoretical examinations of the accounting domain, from the viewpoint of the prediction that this paper addresses, there is a possibility that better predicting performance can be achieved by including higher-dimensional corporate information as covariates. This study, using the financial

[7] Other methods such as linear interpolation or interpolation with the mean or median value of the observed sample may be considered as methods of complementing such missing values. The same analysis was performed using the data set actually supplemented with sample mean values. See Appendix 2 for analysis results.

statement data described above, includes about 40 types of financial variables and their corresponding missing dummy variables. These variables include sales and accounts receivable, earnings, cash flow, inventories, fixed assets, deferred tax assets and liabilities, total and net assets, and other variables. In specific terms, these include various changes in sales and accounts receivables; accounts receivable turnover period; overseas sales ratios levels, changes and profit ratios for each profit and loss stage; cash flow level; relationship with operating profit; rotation period and net asset or total asset ratio for inventories; tangible fixed asset level and total asset ratio for fixed assets; goodwill net asset ratio; relationship with retained earnings for deferred tax assets and liabilities; levels and changes in total assets and net assets; and variables indicating whether it is a new listing company or not.

Furthermore, considering the potential importance of the governance variables pointed out by Song [2018] and others, the following two types of variable groups were constructed from Nikkei NEEDs Financial Quest data. The first is governance variables consisting of the ratio of shareholder stock held by foreign corporations, executives and majority shareholders, and average employee service time. For the second source, in addition to the following – number of commercial banks, borrowing ratio of megabanks, and the Herfindahl Index for borrowed shares, which are constructed from the financial institution relationships of each company, we also utilized any changes to these. Table 1 summarizes these variables. In constructing accounting fraud predictive models, in addition to these variables, as the third source, we added industry dummies indicating 32 categories of industries. Table 2 shows these 32 industry categories.

As a result of all the data preparation, the final dataset for, for example, constructing a predictive model for Accounting Fraud Flag 1, contains 34,923 observations (company x fiscal year) for 4,094 companies from the January 2006 financial period to the fiscal year ending March 2016. In this dataset, there are 126 positive observations for Accounting Fraud Flag 1. We use 25,401 observations from the January 2006 to March 2013 period for the model training, while for the model testing, we use 9,522 observations from the period from April 2013 to March 2016. The training and testing data for Accounting Fraud Flag 1 contains of 107 and 19 positive observations, respectively. Table 3 summarizes the overall observations, including the above Flag 1 case.

6. Model Build and Interpretation

This section describes the model build, performance evaluations, and the interpretation of results.

6.1 Training

First, we train the model that predicts the account fraud. We run the framework of Weighted Random Forest, using company information $\boldsymbol{X}$ measured in the same year as the fraud occurrence. Pursuant to this, Models 1 through 16 are built in accordance with the different type of variable group to be inputted.

Table 4 provides a detailed summary of the input variable groups; of particular note are Models 1, 2, 6, and 12. Model 1 is the baseline performance to be compared against in this paper; we constructed the model by using a theoretically-based parametric model with limited variables based on the setup of Song et al. [2016], a typical prior research effort aimed at detecting accounting fraud. Model 2 uses exactly the same set of variables as Model 1, but Model 2 applies Weighted Random Forest instead of probit used in Model 1. Model 6 is an application of high-dimensional feature space that can be achieved by machine learning-based model and utilizes both existing research variables and additional financial variables. By comparing the performance of Model 6 with that of Model 2, one can capture the effects of using additional variables. Finally, we have Model 12, which, in addition to the added financial variables described above, includes governance and banking variables.

Next, using the same setup, we carry out the forecast of accounting fraud one fiscal year ahead. This is different from the training for detection models in that it uses the company information $\boldsymbol{X}$ that was already available as of one year prior to the year in which the accounting fraud occurred. For each of these model training frameworks, Accounting Fraud Flag 1 and Accounting Fraud Flag 2 are used separately.

6.2 Testing

Table 5 shows AUC, which is a performance evaluation indicator, of all the models trained using Accounting Fraud Flags 1 and 2, and scored on the test data, along with its standard error. The upper and lower panels in Figure 2 represent the AUC of each model, together with its 95% confidence interval for detection and forecast for serious + ancillary events.

First, a slight increase in performance is observed from Model 1 to Model 2 for both nowcast and forecast. This suggests that the prediction performance, in hold-out data, will be improved by using a non-parametric model that allows for complex non-linearity, even when the amount of input information is limited. It should be noted that while such an increase can be seen regarding the measured value of AUC, we cannot reject the null hypothesis that there is no difference between the AUCs of the two models if the standard error is taken into consideration.

Second, a significant increase in AUC has been confirmed from Model 2 to Model 6. This result means that the use of high-dimensional information (additional variables) plays an important role in improving nowcast and forecast performance, which made possible by changing the estimation method of the model.

Third, despite these results, it can be confirmed that, for Models 6 (existing research variables and additional financial variables) through 12 (existing research variables, additional financial variables, governance variables, and bank variables), increases in nowcast and forecast performance is limited. It is important to note that these results need to be interpreted together with the improvement gain we see in the performance of Model 7, which is based on the variables of existing research and governance variables, and of Model 8, which is based on the variables of existing research and banking variables, compared to that of Model 2. One interpretation is that, most information in governance variables and bank variables that are related to fraud can be explained by the additional financial variables (they are highly correlated), however, they can be thought of as containing useful information to predict fraud.

Fourth, as we can see by the results in Table 5, performance is significantly improved as compared with the conventional model for Accounting Fraud Flag 2, which is considered to be relatively difficult in terms of predicting. This result suggests that it is a great deal to use higher dimensional data for the purpose of predicting difficult target.

Figure 3 illustrates the corresponding relationship between the score measured based on Model 12 and the actual rate of the occurrence of accounting fraud. When creating this Figure, we calibrate the Weighted Random Forest score by regressing the actual Accounting Fraud Flag on Weighted Random Forest score. The result is shown together with the actual rate of fraud occurring. First, we found that the score-based fraud probability and the actual fraud rate generally behave similarly, and it is possible to interpret the economic implications of the score output from the machine learning-based model. Second, we found that the probability of fraud occurring increases significantly in groups with particularly high scores.

Finally, to confirm the robustness of the results shown in this section, we established training and test data using a different method than the current splitting method, which divides training data and test data in the time direction. The model construction and the results of model performance are shown. The reason we conducted this analysis is that since there are companies (particularly those with Accounting Fraud Flag 2) whose Accounting Fraud Flags have been positive over multiple fiscal year periods, time-invariant factors exist for companies. If the company is included in both training data and test data, we may be falsely measuring model performance. Figure 4 shows the empirical distribution of AUC by dividing the data multiple times (100 times), and each time, each company appears only in the training data or the test data. We can confirm that the same implications as that of the previous results were obtained.

6.3 Interpreting the Results

Given the results in the previous section, we confirm that the model achieves sufficiently high performance on the hold-out data, when trained using financial and governance variables and used machine learning-based algorithm. In this section, we will interpret these scores of the trained models by examining variable importance and the changes in score due to the fluctuation of predictors.

To begin with, Table 6 shows variable importance for Models 2, 6, and 12. First, it can be confirmed that the contribution to the prediction power coming from the additional financial variables is relatively high. Second, it can be seen that some of the governance variables that were not sufficiently considered in the existing research are ranked high in terms of variable importance.

Next, using a hypothetical company having the average of all the attribute as a starting point, Figure 5 illustrates changes in the score when each attribute is changed either upward or downward. First, we can confirm that the variables emphasized by existing research, such as soft asset ratio, have implications consistent with the results verified by the existing research. Second, we can see that governance-related variables represented by average years of service and additional financial variables represented by the ratio of inventory to equity will affect the score in an intuitively consistent manner. Third, as represented by sales and operating income, it can also be confirmed that the influence of those variables on the scores is not simply linear, suggesting that machine learning is highly suited to handle complex non-linearity.

6.4 Dynamics of the score

How does the model we construct in the present paper work in the actual cases of accounting fraud? The four panels in Figure 6 illustrate the dynamics of the score in the case of accounting fraud detection for the four companies in our test data. We choose the firms experiencing the incident of accounting fraud denoted by Accounting Fraud Flag 2) during the periods of test data so that we can exactly see how the detection score evolves over the course of the incident. First, in the top two panels, we can see the score increased from the beginning of the fraud incident. In both the panels, firms score started to increase at the beginning of the fraud periods then decline after the period (note: the firm corresponding to the upper-left panel was delisted 2016). Second, the firm in the lower-left panel showed a hike in the score at the beginning of the fraud periods but the score decreased during the fraud periods, then increased again after the period. Third, in the lower-right panel, the score does not working as we expected.

Although these examples illustrate the limitation of our detection/forecasting model, we think it is still useful to employ the scores provided by our model. As another exercise, we apply our model to a firm recently fined by the Japanese FSA due to accounting fraud and see the dynamics of the score, which is shown in Figure 7. As most of the figure sin Figure 6, the detection score increased exactly in the period when the firm started accounting fraud. From the practical viewpoint, it is still useful to rely on the model like ours to obtain the information associated with the incident.

7. Discussion

This section holds a discussion based on the results of this study, which we confirmed in the previous section. First, it should be noted that the construction of the machine learning-based model used in this study requires the establishment of various hyperparameters, especially in the training step. Some examples include the number of trees to be constructed, the minimum number of branches, and the type of statistics referred to in setting the branching rules. In Appendix 3, we verify the robustness of these and confirm that, while the results obtained in this study are for the most part not dependent on the establishment of these parameters, it is necessary to avoid the misunderstanding that machine learning methods can be used to automate all the tasks involved in model construction.

Second, it should be recognized that, in addition to the machine learning method adopted in this study, many methods can be used to detect and predict accounting frauds. In fact, a good performance was achieved in analysis conducted by some of the authors of this study, even using different machine learning methods, which brings up an idea that more robust model can be obtained in the future by assembling scores of different methods. Conversely, by utilizing the knowledge obtained in this study, we can limit the number of variables to those considered relatively important from prediction viewpoint. No question that it is beneficial to build a compact and easy-to-use model. Interesting to note that, as shown in Appendix 4, even if the target flags are the same, there are differences between detecting and forecasting in the variable groups that ranked high in the variable importance. Also, even when performing the same forecast, the different target flag settings can cause differences in the variable groups that indicate high variable importance. These findings mean that the variable selection should be conducted in accordance with the purpose when building such compact and easy-to-use model.

Third, the predictive model developed in this study assume constant social-economic environments and accounting standards. In operating a predictive model, it is always necessary to be cognizant of the effects of these changes on the model. For example, if there is a completely different kind of fraud in a new business, the current model may not be able to detect or forecast this. Moreover, even if the realities of a company remain the same, financial variables may take different values due to changes in accounting standards, which may affect fraud detection and fraud forecast results. Thus, to respond to such issues, it is important to establish a mechanism to regularly monitor the effectiveness of the model and conduct this monitoring regularly.

Fourth, when a satisfying predictive model has been built, it is critical to fully be aware that how to use the score in reality is the next important step to be raised. Given that there are companies in which accounting fraud is currently taking place, or is likely to occur, professional staff with specialized knowledge needs to handle with good judgment promptly, considering at which point the probability of accounting fraud is substantial, and what type of response is therefore needed. In connection with this point, account-level anomaly detection is considered to be an important research theme, complementing the company-level fraudulent model conducted in this study. In addition, it is also important that there is a concern regarding what degree the on-site professional staff performs business operations based on such scores when the model scores are provided (e.g., Kleinberg et al. [2018]).

Fifth, from a policy perspective, there is an issue of how to adopt the development of such predicting technology. For example, if the details of a model for detecting fraudulent accounting become widely known, companies that are attempting some type of accounting fraud may do so in a way that it's difficult for the model to detect. In reality, it is unlikely that the model details will become widely known, and it is not necessarily easy to engage in fraud without being detected by the model; thus, it is unclear how practical this never-ending game is, but at the very least we should be cognizant that, from a policy perspective, these issues do exist. For example, one possible solution would be to work on continuing to improve models that target companies engaging in malicious efforts to, in a single step, wipe out predictive models.

Sixth, we need deeper discussions regarding causal reasoning, which has not been fully addressed in this paper. The question of whether it is possible to suppress the occurrence of fraudulent events as causal relationships by adopting a specific audit approach or governance device is important in the sense of considering advance measures to prevent accounting fraud.

8. Conclusion

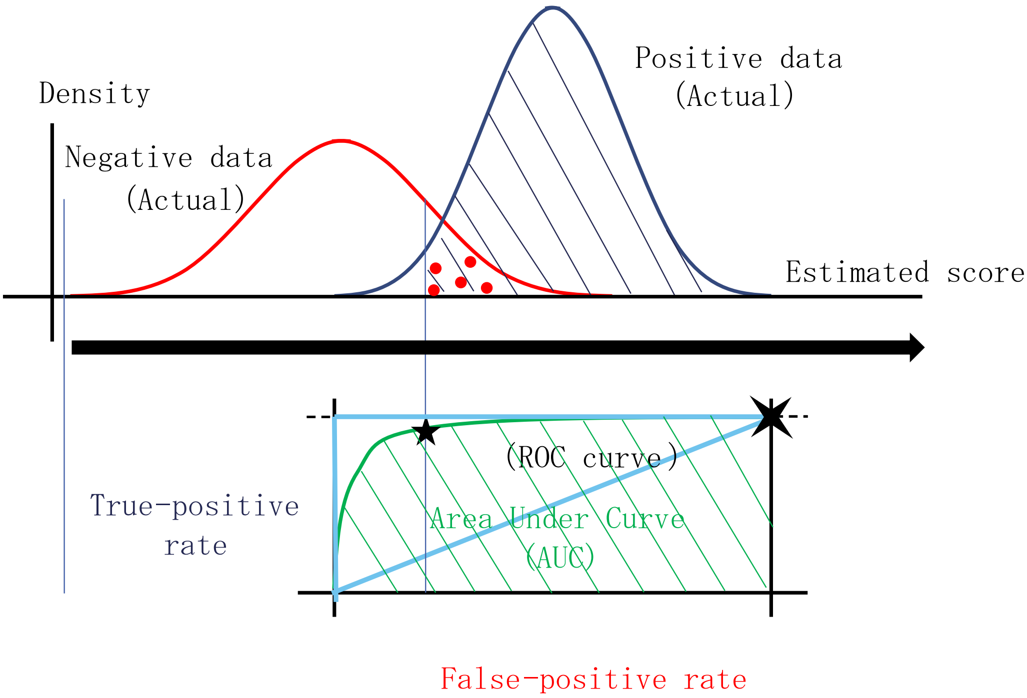
This study empirically examined the usefulness of machine learning in detecting and forecasting accounting fraud. First, rather than the parametric approach that existing research took, which was based on limited variables, we took machine learning approach with high-dimensional feature space to detect accounting fraud. We verified the improvements in performance. Second, we forecast accounting fraud using the same framework, which was not sufficiently considered in existing research. Third, to interpret the constructed model, we consider what kind of information was mainly responsible for the model scores. From empirical analysis targeting fraudulent accounting events of publicly traded companies in Japan, we confirmed that the use of machine learning methods contributed to some improvement in detection performance and that the use of the high-dimensional space contributed to a significant improvement in detection performance. We also confirmed that the machine learning-based model could achieve a sufficient level of forecast performance from a practical point of view. Moreover, this research also confirmed that a certain degree of useful information for detecting and forecasting of accounting fraud was included in the variables other than those that existing research has referred to, based on theoretical assumptions (for example, governance-related variables). These results also provided useful information for future theoretical studies on the mechanism of accounting fraud.

Appendix A: The Concept of Performance Metrics

Figure A1 below shows the procedure for evaluating model performance using a ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) in the case where a sample dataset is scored based on a given model and each record in the dataset is labeled whether the actual event occurred or not. First, we draw the distribution of estimated scores for positive group (event occurred) and negative group (event not occurred). Second, we plot True-positive rates (the percentage of the positive group that are correctly estimated for the event to occur) and False-positive rates (the percentage of the negative group that are wrongly estimated for the event to occur) at different thresholds on the vertical and horizontal axes respectively. Third, after obtaining the locus (ROC) by connecting those points described above, fourth, we calculate AUC by finding the area corresponding to the underneath of this curve.

Since this AUC approaches 1 when the score-based classifications are highly accurate and takes a score of 0.5 for completely random classifications, prediction performance can be evaluated by calculating this numerical value. Although, in reality, the level varies depending on the type of business, there are many cases where the goal is to exceed the value of 0.8.

Figure A1.



Density
Negative data
(Actual)
Positive data
(Actual)
Estimated score
True-positive
rate
(ROC curve)
Area Under Curve
(AUC)
False-positive rate

Appendix B: Confirming the Robustness of the Missing Value Treatment

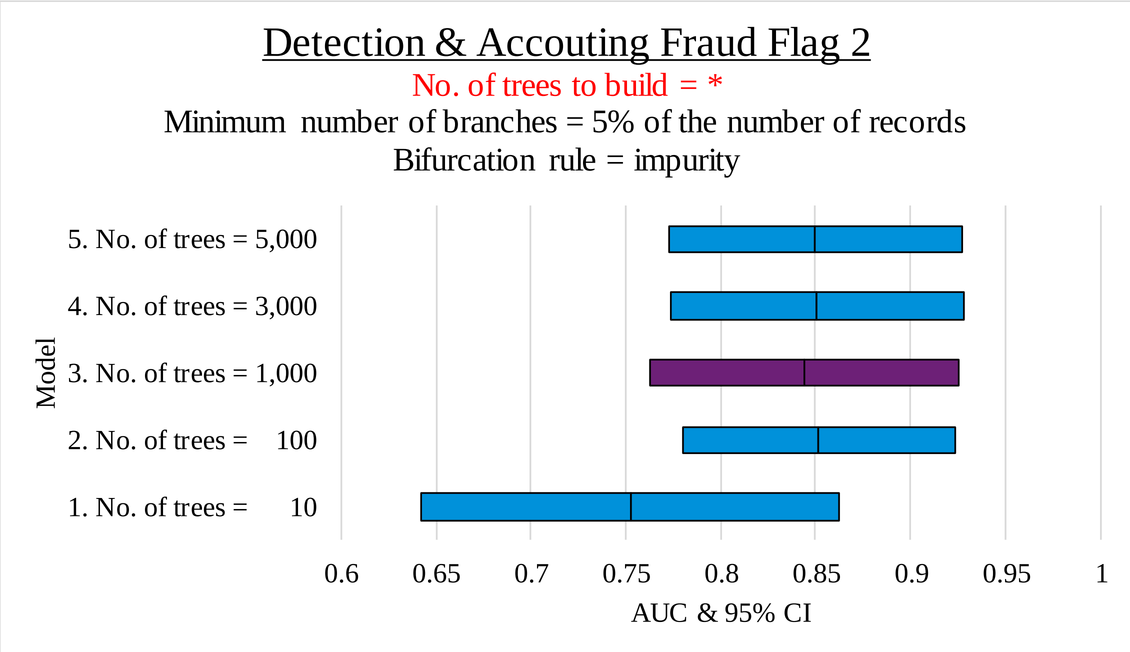
The table below shows the results of the analysis done in the same way as in this study except that the missing values are treated with the sample mean. The upper panel shows the results of the performance of the fraud detection model, while the lower panel shows the results of the performance of the fraud forecast model. The “before change” column corresponds to the case where a missing value is replaced by 0, and a dummy variable is established to indicate the missing. Separately, the “after change” column corresponds to the case where the same method is taken whereby the average value replaces the missing value. In calculating the average value, we calculate the average on a per-company basis, and should the average on a per-company basis is not possible, we use the entire average.

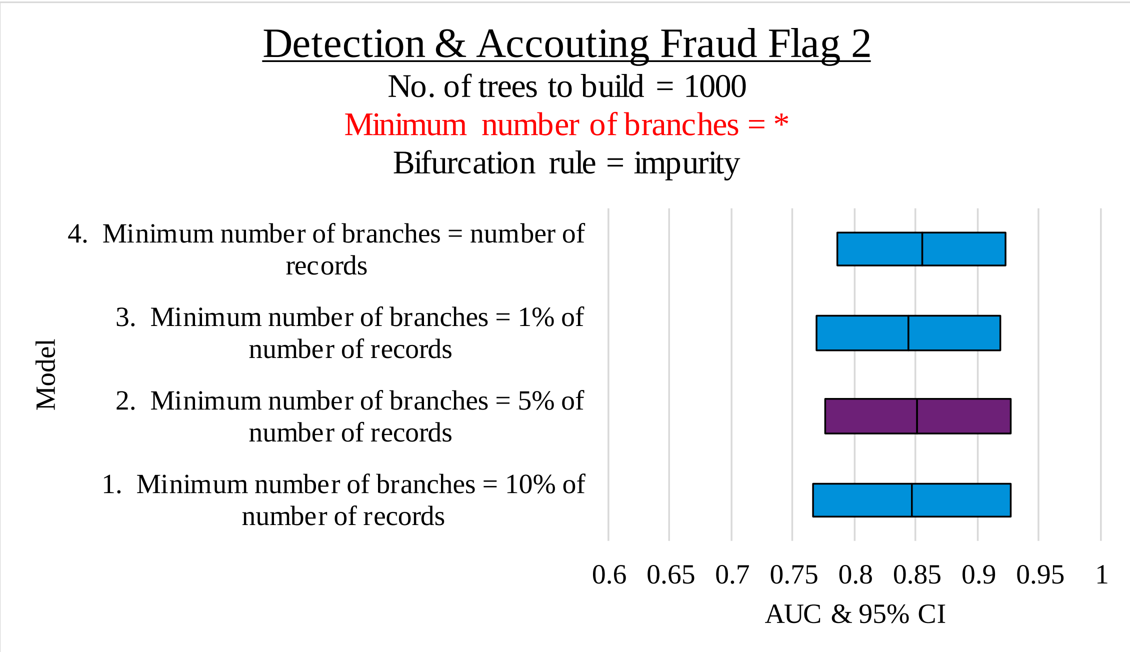
Table B1.

			Detection & Accounting Fraud Flag 1						Detection & Accounting Fraud Flag 2					
			Before change		After change		Difference		Before change		After change		Difference	
			AUC	se	AUC	se	AUC	se	AUC	se	AUC	se	AUC	se
1	Basic	Prbt	0.71	0.06	0.71	0.06	0.00	0.00	0.73	0.04	0.75	0.04	0.02	0.00
2	Basic	WRF	0.79	0.05	0.80	0.05	0.01	0.00	0.77	0.05	0.79	0.05	0.02	0.00
3	Additional	WRF	0.84	0.04	0.83	0.05	-0.01	0.01	0.86	0.03	0.87	0.04	0.01	0.00
4	Governance	WRF	0.64	0.07	0.65	0.07	0.00	0.00	0.67	0.06	0.67	0.06	0.00	0.00
5	Bank Relation	WRF	0.68	0.06	0.62	0.07	-0.06	0.01	0.75	0.05	0.68	0.05	-0.07	0.01
6	Basic + Add	WRF	0.86	0.04	0.86	0.05	0.00	0.01	0.88	0.03	0.89	0.04	0.01	0.00
7	Basic + Gov	WRF	0.81	0.05	0.80	0.05	0.00	0.00	0.78	0.05	0.79	0.05	0.00	0.00
8	Basic + BK	WRF	0.86	0.03	0.81	0.04	-0.05	0.01	0.82	0.04	0.80	0.04	-0.02	0.00
9	Basic + Add + Gov	WRF	0.86	0.04	0.86	0.05	0.00	0.01	0.88	0.04	0.89	0.03	0.02	0.00
10	Basic + Add + BK	WRF	0.88	0.04	0.86	0.05	-0.02	0.01	0.89	0.04	0.88	0.04	-0.01	0.00
11	Basic + Gov + BK	WRF	0.86	0.03	0.83	0.05	-0.03	0.01	0.84	0.04	0.82	0.05	-0.02	0.01
12	Basic + Add + Gov + BK	WRF	0.87	0.04	0.86	0.05	-0.01	0.01	0.90	0.04	0.88	0.04	-0.01	0.01
13	Add + Gov	WRF	0.82	0.05	0.82	0.05	0.00	0.00	0.86	0.03	0.88	0.04	0.02	0.00
14	Add + BK	WRF	0.86	0.04	0.85	0.05	-0.01	0.01	0.88	0.04	0.87	0.05	-0.01	0.01
15	Add + Gov + BK	WRF	0.85	0.04	0.83	0.05	-0.01	0.01	0.88	0.04	0.87	0.05	-0.01	0.01
16	Gov + BK	WRF	0.74	0.06	0.73	0.06	0.00	0.00	0.78	0.05	0.73	0.06	-0.05	0.01

			Prediction & Accounting Fraud Flag 1						Prediction & Accounting Fraud Flag 2					
			Before change		After change		Difference		Before change		After change		Difference	
			AUC	se	AUC	se	AUC	se	AUC	se	AUC	se	AUC	se
1	Basic	Prbt	0.68	0.07	0.69	0.07	0.01	0.00	0.65	0.06	0.70	0.06	0.05	0.00
2	Basic	WRF	0.68	0.06	0.75	0.05	0.07	-0.01	0.68	0.05	0.75	0.04	0.07	-0.01
3	Additional	WRF	0.80	0.05	0.79	0.06	0.00	0.01	0.83	0.04	0.83	0.04	0.01	0.00
4	Governance	WRF	0.64	0.07	0.64	0.07	0.00	0.00	0.70	0.05	0.69	0.05	-0.01	0.00
5	Bank Relation	WRF	0.60	0.07	0.55	0.06	-0.05	0.00	0.67	0.06	0.62	0.06	-0.04	0.00
6	Basic + Add	WRF	0.80	0.05	0.80	0.06	0.01	0.01	0.82	0.04	0.84	0.04	0.02	0.00
7	Basic + Gov	WRF	0.72	0.06	0.75	0.06	0.03	0.00	0.73	0.04	0.77	0.04	0.04	0.00
8	Basic + BK	WRF	0.73	0.05	0.73	0.04	-0.01	-0.01	0.75	0.05	0.75	0.05	-0.01	0.00
9	Basic + Add + Gov	WRF	0.80	0.06	0.80	0.06	0.01	0.01	0.82	0.04	0.84	0.04	0.02	0.00
10	Basic + Add + BK	WRF	0.83	0.04	0.83	0.06	-0.01	0.02	0.85	0.04	0.84	0.04	-0.01	0.01
11	Basic + Gov + BK	WRF	0.75	0.05	0.76	0.05	0.01	0.00	0.76	0.05	0.75	0.05	-0.01	0.00
12	Basic + Add + Gov + BK	WRF	0.82	0.05	0.82	0.06	0.00	0.01	0.85	0.04	0.85	0.04	0.00	0.01
13	Add + Gov	WRF	0.79	0.06	0.79	0.07	-0.01	0.01	0.82	0.04	0.84	0.04	0.02	0.00
14	Add + BK	WRF	0.84	0.05	0.81	0.06	-0.03	0.01	0.85	0.04	0.84	0.05	-0.02	0.01
15	Add + Gov + BK	WRF	0.82	0.05	0.81	0.06	-0.01	0.01	0.85	0.04	0.84	0.05	-0.01	0.01
16	Gov + BK	WRF	0.67	0.07	0.65	0.06	-0.03	0.00	0.72	0.05	0.69	0.05	-0.03	-0.01

Appendix C: Robustness Check of Tuning Parameter-Related Results





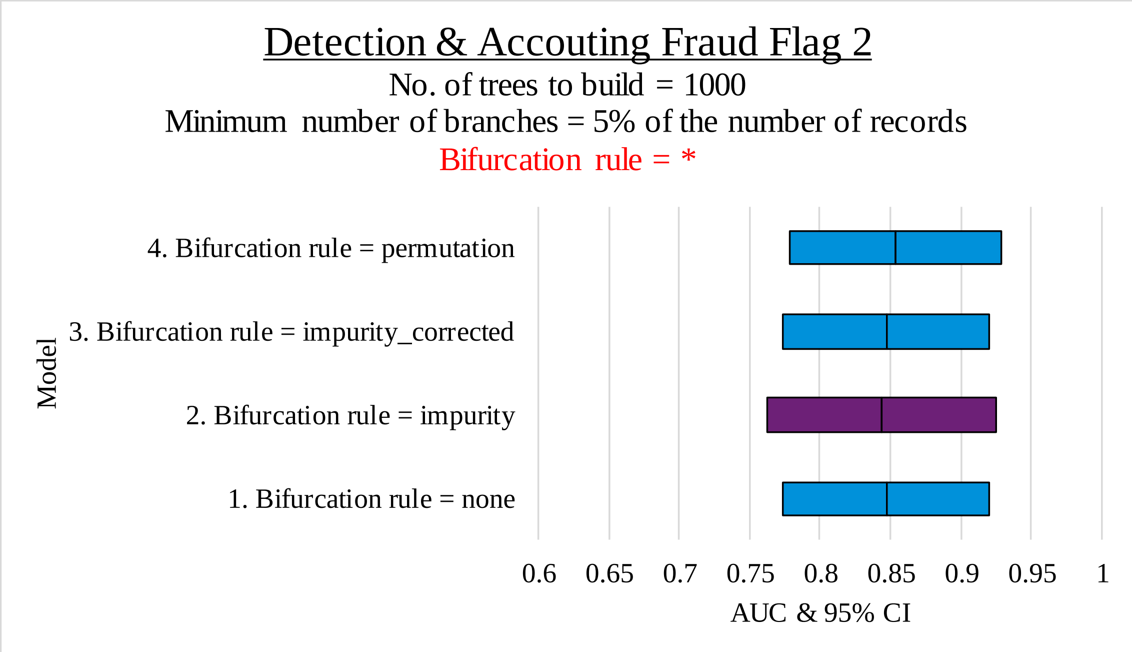
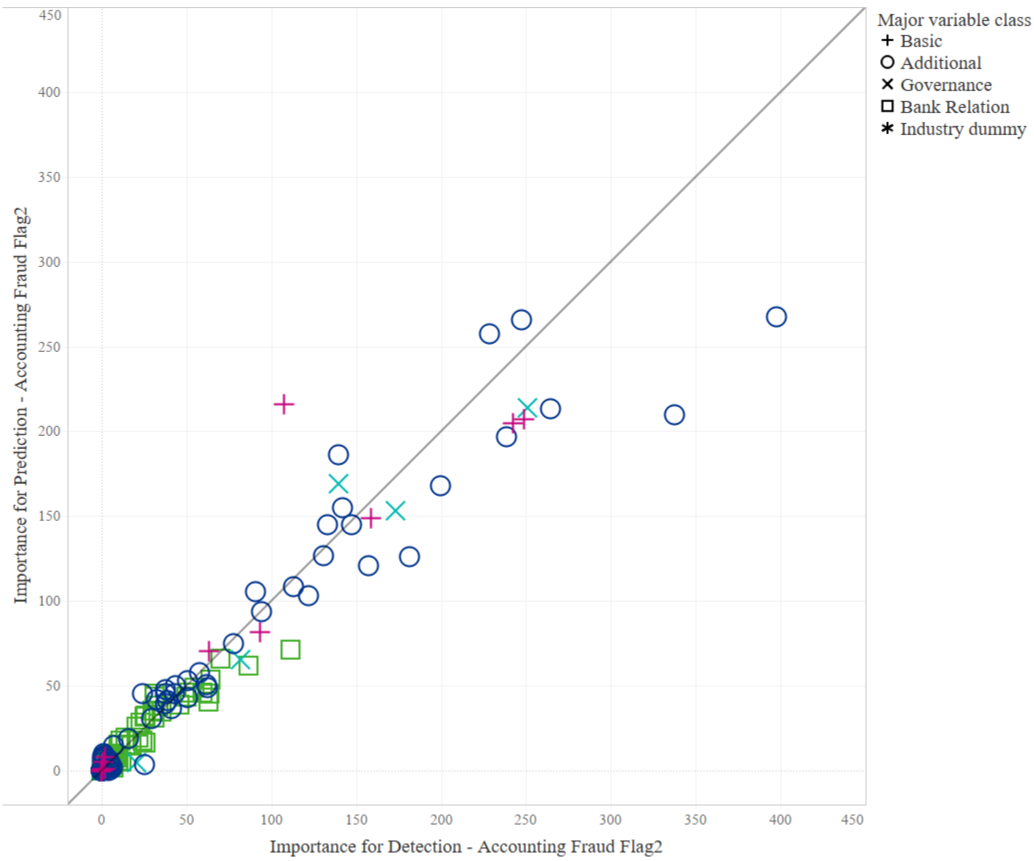

Detection & Accouting Fraud Flag 2
No. of trees to build = 1000
Minimum number of branches = 5% of the number of records
Bifurcation rule = *
4. Bifurcation rule = permutation
3. Bifurcation rule = impurity_corrected
2. Bifurcation rule = impurity
1. Bifurcation rule = none
Model
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
AUC & 95% CI

Figure C1.

Appendix D: Differences in the High Variable Importance

The upper panel plots the variable importance of each variable contributing to forecast (vertical axis) and detection (horizontal axis) for Accounting Fraud Flag 2. The lower panel plots the variable importance of each variable contributing Accounting Fraud Flag 2 (vertical axis) and Accounting Fraud Flag 1 (horizontal axis).



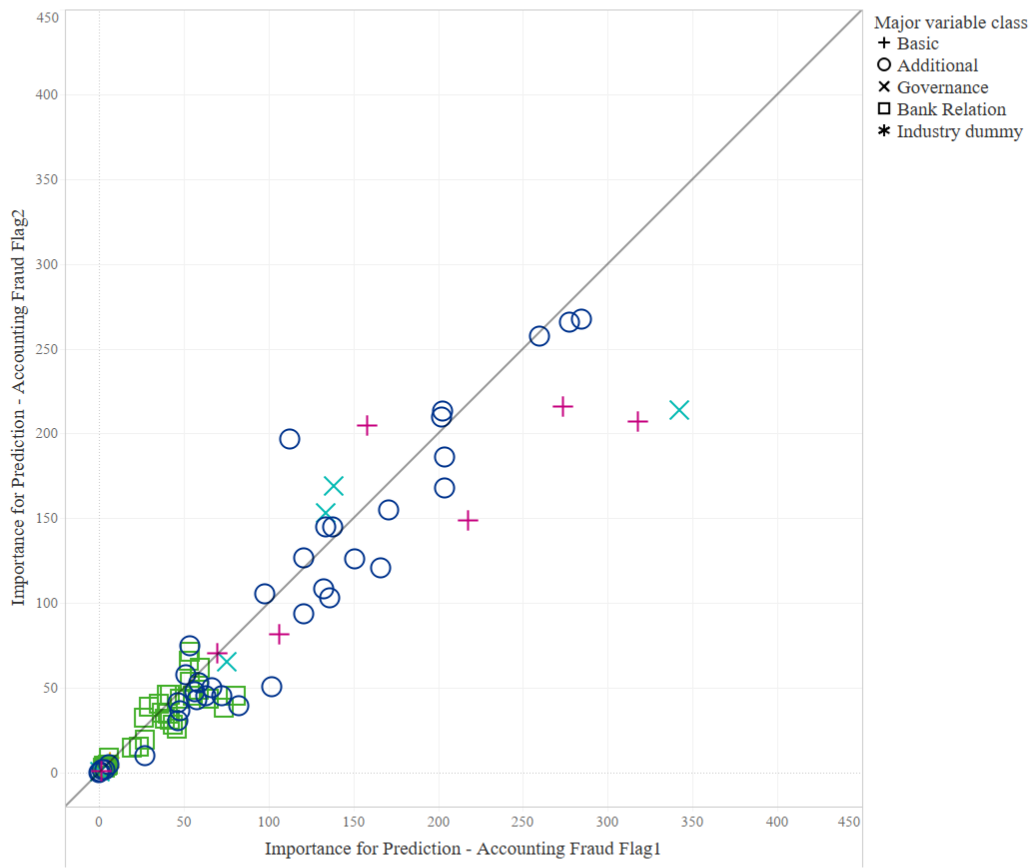

Major variable class
Basic
Additional
Governance
Bank Relation
Industry dummy
Importance for Prediction - Accounting Fraud Flag2
Importance for Prediction - Accounting Fraud Flag1
0
50
100
150
200
250
300
350
400
450

Figure D1.

References

Breiman, Leo. Random Forests. *Machine Learning* 45 (1): 5–32. 2001.

Chen, C., A. Liaw, and L. Breiman. Using Random Forest to Learn Imbalanced Data. Technical Report 666 Statistics Department of University of California at Berkley. 2004.

Chernozhukov, V., D. Chetverikov, M. Demirer, E Duflo, C. Hansen, W. Newey, and J. Robins. Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal* 21(1): 1-68. 2018.

Dechow, P. M., Ge, W., Larson, C. R. and R. G. Sloan. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28 (1): 17-82. 2011.

Dechow, P. M., Ge, W., and C. Schrand. Understanding Earnings Quality: A Review of the Proxies, their Determinants and their Consequences. *Journal of Accounting and Economics* 50 (2)(3): 344-401. 2010.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC. *Contemporary Accounting Research* 13 (1): 1-36. 1996.

Janitza, S., E. Celik, Anne-Laure Boulesteix. "A Computationally Fast Variable Importance Test for Random Forests for High-Dimensional Data," *Advances in Data Analysis and Classification* 12(4): 885-915. 2018.

Perols, J. "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms," *Auditing: A Journal of Practice & Theory* 30(2): 19-50. 2011.

Perols, J., Bowen, B., and C. Zimmerman. "Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Forecast," *Accounting Review* 92(2): 221-245. 2017.

Song Survey and Research on Early Detection of Improper Accounting in Response to Recent Changes in the Economic Environment. FSA Institute Discussion Paper Series, DP 2017-6. 2018.

Song, M., N. Oshiro, and A. Shuto. Predicting Accounting Fraud: Evidence from Japan. *The Japanese Accounting Review* 6: pp. 17-63. 2016.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human Decisions and Machine Forecasts. *Quarterly Journal of Economics* 133(1): 237-293. 2018.

Miyakawa, D. "Forecasting Firm Dynamics with Machine Learning," mimeo. 2019a.

Miyakawa, D. "Shocks to Supply Chain Network and Firm Dynamics: An Application of Double Machine Learning," forthcoming in RIETI Discussion Paper Series. 2019b.

West, J. and M. Bhattacharya. “Intelligent Financial Fraud Detection: A Comprehensive Review,” *Computer & Security* 57: 47-66. 2016.

Tables and Figures

Table 1. Variable Definitions and Classifications

Major variable classification	Minor variable classification	Variable
Basic		% Soft assets
		CFO discretionary accruals AB
		Actual issuance
		C score AR
		AB cash flow
		CORP
Additional	Sales and accounts receivable	Sales
		Sales compared to previous period
		Comparison of sales over three periods
		Sales growth rate
		Overseas sales ratio
		Difference with previous period's accounts receivable balance
		Dummy variable that takes the value '1' when accounts receivable balance increases from previous period and '0' otherwise
		Accounts receivable turnover period
		Previous accounts receivable turnover period
		Accounts receivable turnover compared to previous period
		Rate of accounts receivable change
		Ratio of change in accounts receivable divided by sales growth
	Profits	Operating profit
		Operating profit margin
		Dummy variable that takes the value '1' when there are operating losses for two consecutive periods and '0' otherwise
		Ordinary profit
		Ordinary profit margin
		Non-operating income and expenses
		Non-operating income and expenses ratio
		Comparison of Profit/loss attributable to owners of parent over three periods
		Dummy variable that takes the value '1' when there is ordinary loss with net profit or ordinary profit with a net loss and '0' otherwise
	Cash flows	Cash flows from operating activities
		Dummy variable that is '1' when the cash flows from operating activities for 2 consecutive periods is negative and '0' otherwise
		Ratio of operating profit to operating cash flow
	Inventories	Inventory turnover rate
		Inventory net asset ratio
		Inventory total asset ratio
	Fixed assets	Tangible fixed assets
		Ratio of tangible fixed assets to total assets
		Goodwill net asset ratio
		Dummy variable that is '1' when goodwill balance > operating profit and '0' otherwise
	Deferred tax assets and liabilities	Dummy variable that is '1' deferred tax assets are recorded when retained earnings are negative, and '0' otherwise
		Retained earnings ratio for deferred tax assets/liabilities (net)
	Total assets/net assets	Comparison of net assets over three periods
		Total asset balance
	Other	Dummy variable that is '1' for new listings, and '0' otherwise

Table 1. Variable Definitions and Classifications (Continued)

Major variable classification	Minor variable classification	Variable
Governance		Ratio of shareholder stock held by foreign corporations
		Ratio of shareholder stock held by executives
		Average length of employment
		Ratio of shareholder stock held by majority shareholders
Bank Relation		Number of banks from which has received long-term borrowings
		Number of banks from which has received short-term borrowings
		Number of banks from which has received all borrowings
		Percentage of long-term borrowings held by megabanks
		Percentage of short-term borrowings held by megabanks
		Percentage of all borrowings held by megabanks
		Herfindahl index of long-term borrowings
		Herfindahl index of short-term borrowings
		Herfindahl index of total borrowings
		Herfindahl index of long-term borrowings adjusted for company size
		Herfindahl index of short-term borrowings adjusted for company size
		Herfindahl index of total borrowings adjusted for company size
		Dummy variable that is '1' when long-term borrowings are procured from one bank and '0' otherwise
		Dummy variable that is '1' when short-term borrowings are procured from one bank and '0' otherwise
		Dummy variable that is '1' when all borrowings are procured from one bank and '0' otherwise
		Number of banks from which long-term borrowings procured compared to the previous period
		Number of banks from which short-term borrowings procured compared to the previous period
		Number of banks from which all borrowings procured compared to the previous period
		Percentage of long-term borrowings held by megabanks compared to the previous period
		Percentage of short-term borrowings held by megabanks compared to the previous period
		Percentage of all borrowings held by megabanks compared to the previous period
		Herfindahl index of long-term borrowings compared to the previous period
		Herfindahl index of short-term borrowings compared to the previous period
		Herfindahl index of all borrowings compared to the previous period
		Herfindahl index of long-term borrowings adjusted for company size compared to the previous period
		Herfindahl index of short-term borrowings adjusted for company size compared to the previous period
		Herfindahl index of all borrowings adjusted for company size compared to the previous period
		Dummy variable that takes the value '1' when top bank providing long-term borrowings is replaced and '0' otherwise
		Dummy variable that takes the value '1' when top bank providing short-term borrowings is replaced and '0' otherwise
		Dummy variable that takes the value '1' when top bank providing all borrowings is replaced and '0' otherwise

Table 2. The 32 Industry Classifications

No.	Industry	No.	Industry	No.	Industry	No.	Industry
1	Food products	9	Iron & steel	17	Other manufacturing	25	Land transport
2	Fibers	10	Nonferrous metal products	18	Marine products	26	Ocean transport
3	Pulp/paper	11	Machinery	19	Mining	27	Air transport
4	Chemicals	12	Electrical instruments	20	Construction	28	Warehousing
5	Pharmaceuticals	13	Shipbuilding	21	Commercial business	29	Communications
6	Petroleum	14	Automobiles	22	Retail business	30	Electricity
7	Rubber	15	Transportation equipment	23	Real estate	31	Natural gas
8	Ceramics	16	Precision equipment	24	Railway/bus	32	Services

Table 3. List of Observed Values

		Detect			Predict		
		Train	Test	All	Train	Test	All
Accounting Fraud Flag 1 (Only main event)	Number of financial statements observed	25,401	9,522	34,923	25,401	9,522	34,923
	Number of companies	3,858	3,424	4,094	3,858	3,424	4,094
	Number of positive observations	107	19	126	107	19	126
Accounting Fraud Flag 2 (Main + Ancillary event)	Number of financial statements observed	25,401	9,522	34,923	25,401	9,522	34,923
	Number of companies	3,858	3,424	4,094	3,858	3,424	4,094
	Number of positive observations	150	23	173	150	23	173

Table 4. Model List

		Model															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Variable group	Definition	Probit	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF
Basic	6 variables (from existing research)	○	○				○	○	○	○	○	○	○				
Additional	144 variables			○			○			○	○		○	○	○	○	
Governance	16 variables				○			○		○		○	○	○		○	○
Bank relation	30 variables					○			○		○	○	○		○	○	○
industry dummy	32 variables	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Table 5. Performance Evaluation Results

	Model															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Probit	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF	WRF
Outcome pattern																
Detect + Accounting Fraud Flag 1 (Only main event)																
AUC	0.71	0.79	0.84	0.64	0.68	0.86	0.81	0.86	0.86	0.88	0.86	0.87	0.82	0.86	0.85	0.74
s.e.	0.06	0.05	0.04	0.07	0.06	0.04	0.05	0.03	0.04	0.04	0.03	0.04	0.05	0.04	0.04	0.06
Detect + Accounting Fraud Flag 2 (Main+Ancillary event)																
AUC	0.73	0.77	0.86	0.67	0.75	0.88	0.78	0.82	0.88	0.89	0.84	**0.90**	0.86	0.88	0.88	0.78
s.e.	0.04	0.05	0.03	0.06	0.05	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.05
Predict (1year-ahead) + Accounting Fraud Flag 1 (Only main event)																
AUC	0.68	0.68	0.80	0.64	0.60	0.80	0.72	0.73	0.80	0.83	0.75	0.82	0.79	0.84	0.82	0.67
s.e.	0.07	0.06	0.05	0.07	0.07	0.05	0.06	0.05	0.06	0.04	0.05	0.05	0.06	0.05	0.05	0.07
Predict (1year-ahead) +Accounting Fraud Flag 2 (Main+Ancillary event)																
AUC	0.65	0.68	0.83	0.70	0.67	0.82	0.73	0.75	0.82	0.85	0.76	0.85	0.82	0.85	0.85	0.72
s.e.	0.06	0.05	0.04	0.05	0.06	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.05

Remarks: Items in bold type are AUC 0.9 or greater.

Table 6. Variable importance

Model 2 Basic + WRF

Rank	Variable class	Variable	Importance
1	Basic	CORP	972.3
2	Basic	% Soft assets	946.1
3	Basic	Actual issuance	790.7
4	Basic	AB cash flow	636.2
5	Basic	C score AR	608.5
6	Basic	CFO discretionary accruals AB	534.5

Model 6 Basic & Add + WRF

Rank	Rank (Model 2)	Variable class	Variable	Importance
1	-	Additional	Non-operating income and expenses ratio	363.7
2	-	Additional	Ratio of tangible fixed assets to total assets	334.4
3	-	Additional	Tangible fixed assets	333.0
4	3	Basic	Actual issuance	298.4
5	-	Additional	Non-operating income and expenses	290.1
6	-	Additional	Sales	288.6
7	-	Additional	Inventory net asset ratio	286.4
8	2	Basic	% Soft assets	279.3
9	1	Basic	CORP	264.5
10	-	Additional	Ratio of operating profit to cash flows from operating activities	260.5
11	-	Additional	Inventory turnover rate	212.4
12	-	Additional	Cash flows from operating activities	208.4
13	-	Additional	Total asset balance	207.2
14	-	Additional	Inventory total asset ratio	201.6
15	4	Basic	AB cash flow	190.5
16	-	Additional	Ordinary profit margin	175.2
17	-	Additional	Retained earnings ratio for deferred tax assets/liabilities (net)	171.5
18	-	Additional	Ordinary profit	161.3
19	-	Additional	Accounts receivable turnover period	147.2
20	-	Additional	Operating profit margin	142.6
21	-	Additional	Goodwill net asset ratio	136.9
22	-	Additional	Operating profit	125.6
23	5	Basic	C score AR	120.6
24	-	Additional	Previous accounts receivable turnover period	115.7
25	6	Basic	CFO discretionary accruals AB	108.4

Model 12 Basic, Add, Gov & BK + WRF

Rank	Rank (Model 2)	Rank (Model 6)	Variable class	Variable	Importance
1	-	1	Additional	Non-operating income and expenses ratio	267.3
2	-	3	Additional	Tangible fixed assets	265.7
3	-	2	Additional	Ratio of tangible fixed assets to total assets	257.4
4	3	4	Basic	Actual issuance	216.1
5	-	-	Governance	Average length of employment	213.8
6	-	6	Additional	Sales	213.4
7	-	5	Additional	Non-operating income and expenses	209.5
8	1	9	Basic	CORP	207.2
9	2	8	Basic	% Soft assets	204.6
10	-	7	Additional	Inventory net asset ratio	196.9
11	-	10	Additional	Ratio of operating profit to cash flows from operating activities	186.3
12	-	-	Governance	Percentage of shares held by directors	169.0
13	-	12	Additional	Cash flows from operating activities	167.6
14	-	13	Additional	Total asset balance	154.7
15	-	-	Governance	Ratio of shareholder stock held by majority shareholders	153.4
16	4	15	Basic	AB cash flow	149.0
17	-	11	Additional	Inventory turnover rate	145.0
18	-	14	Additional	Inventory total asset ratio	144.9
19	-	17	Additional	Retained earnings ratio for deferred tax assets/liabilities (net)	126.4
20	-	16	Additional	Ordinary profit margin	126.3
21	-	18	Additional	Ordinary profit	120.6
22	-	22	Additional	Operating profit	108.1
23	-	19	Additional	Accounts receivable turnover period	105.3
24	-	20	Additional	Operating profit margin	103.0
25	-	21	Additional	Goodwill net asset ratio	93.4

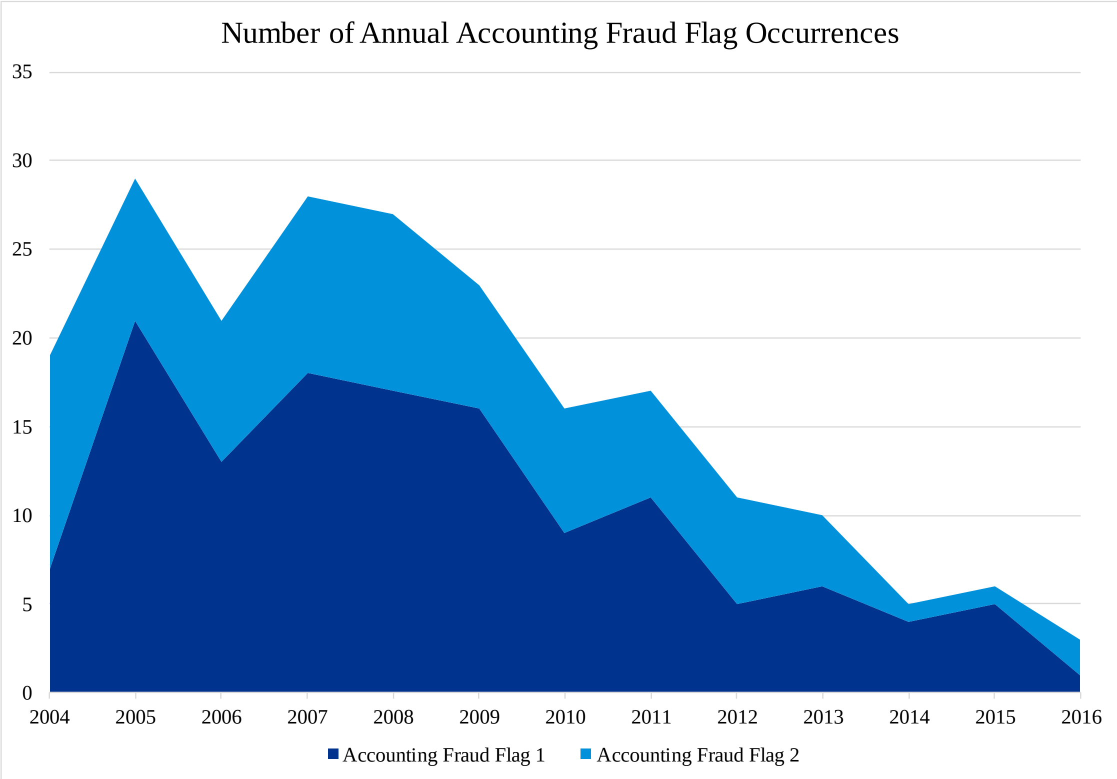

Number of Annual Accounting Fraud Flag Occurrences
35
30
25
20
15
10
5
0
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Accounting Fraud Flag 1
Accounting Fraud Flag 2

Figure 1

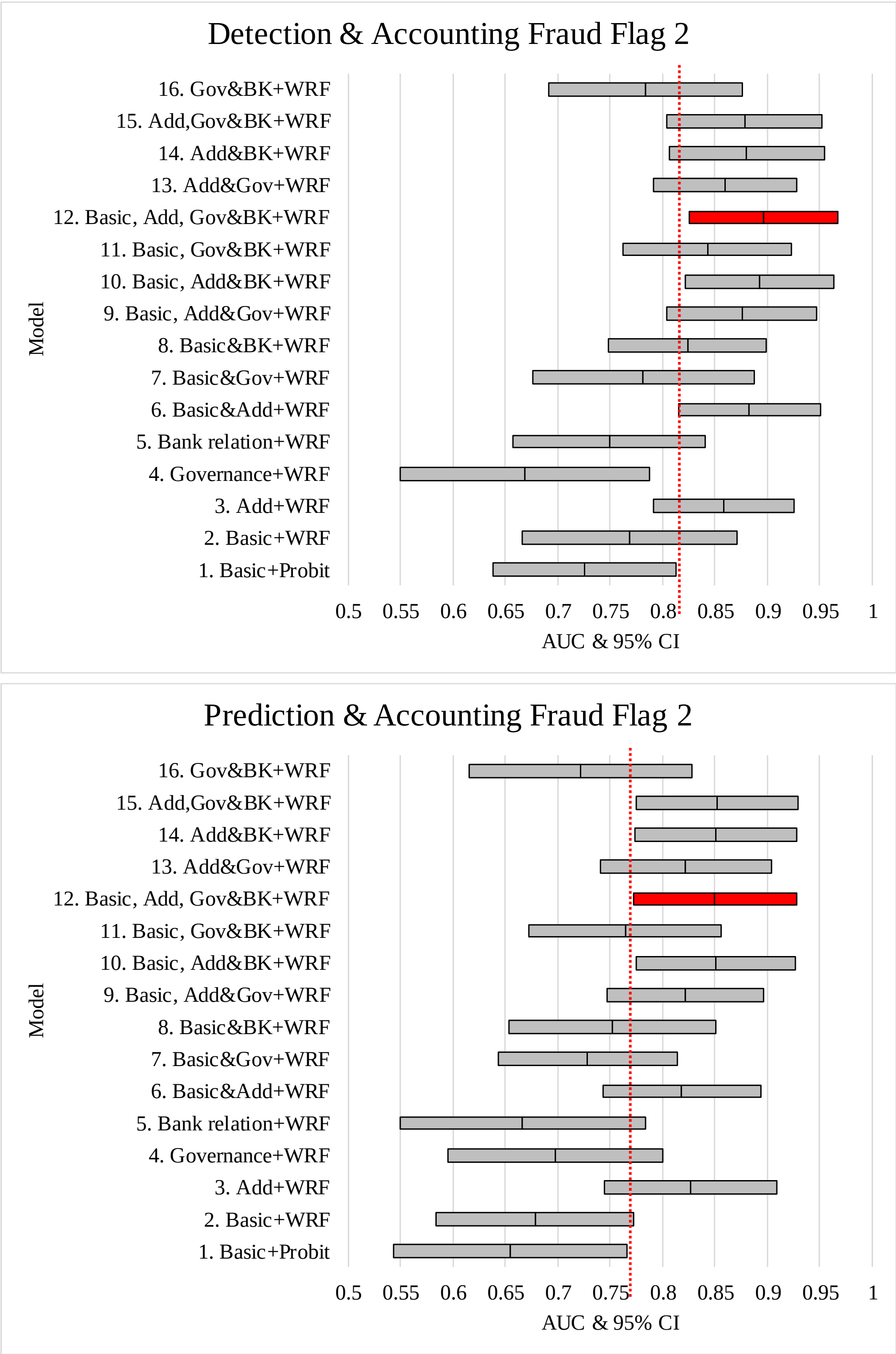


Figure 2. Performance Evaluation Results

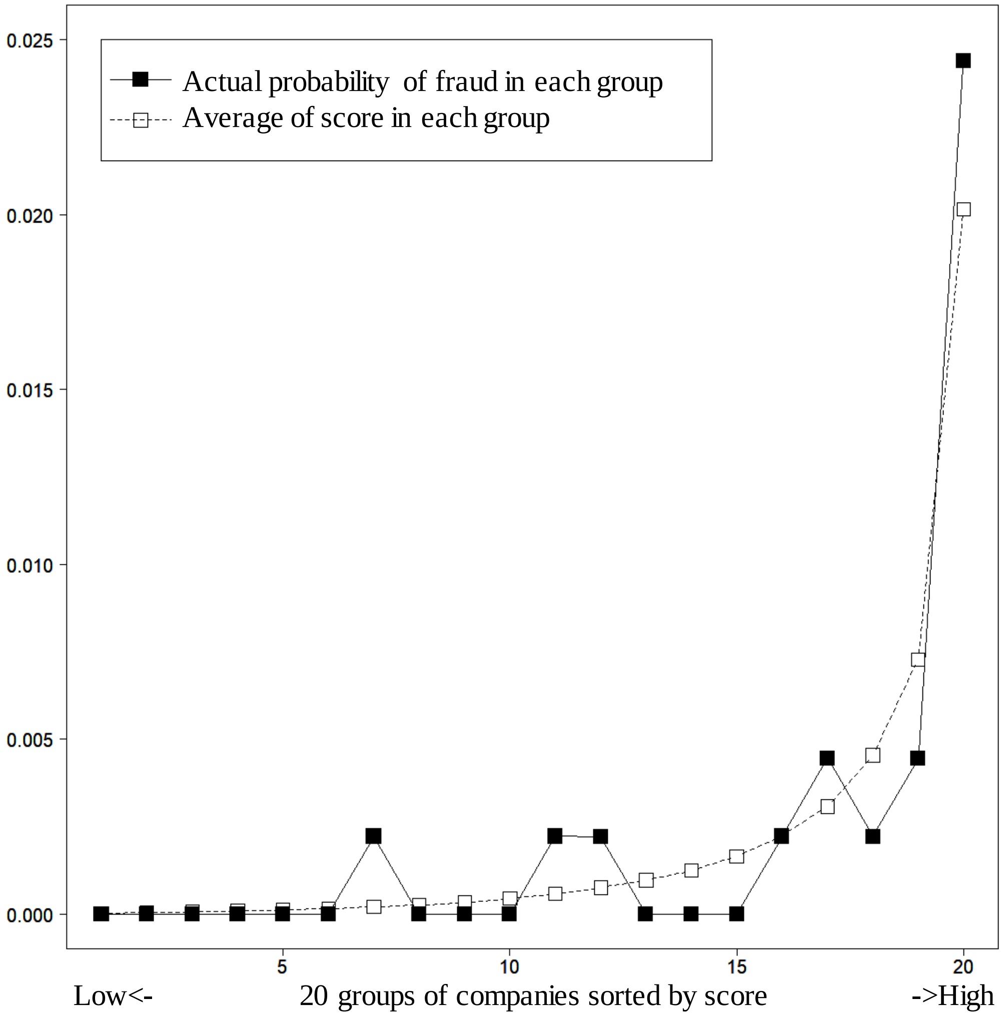


Figure 3. The Corresponding Relationship Between Score and Probability of Accounting Fraud

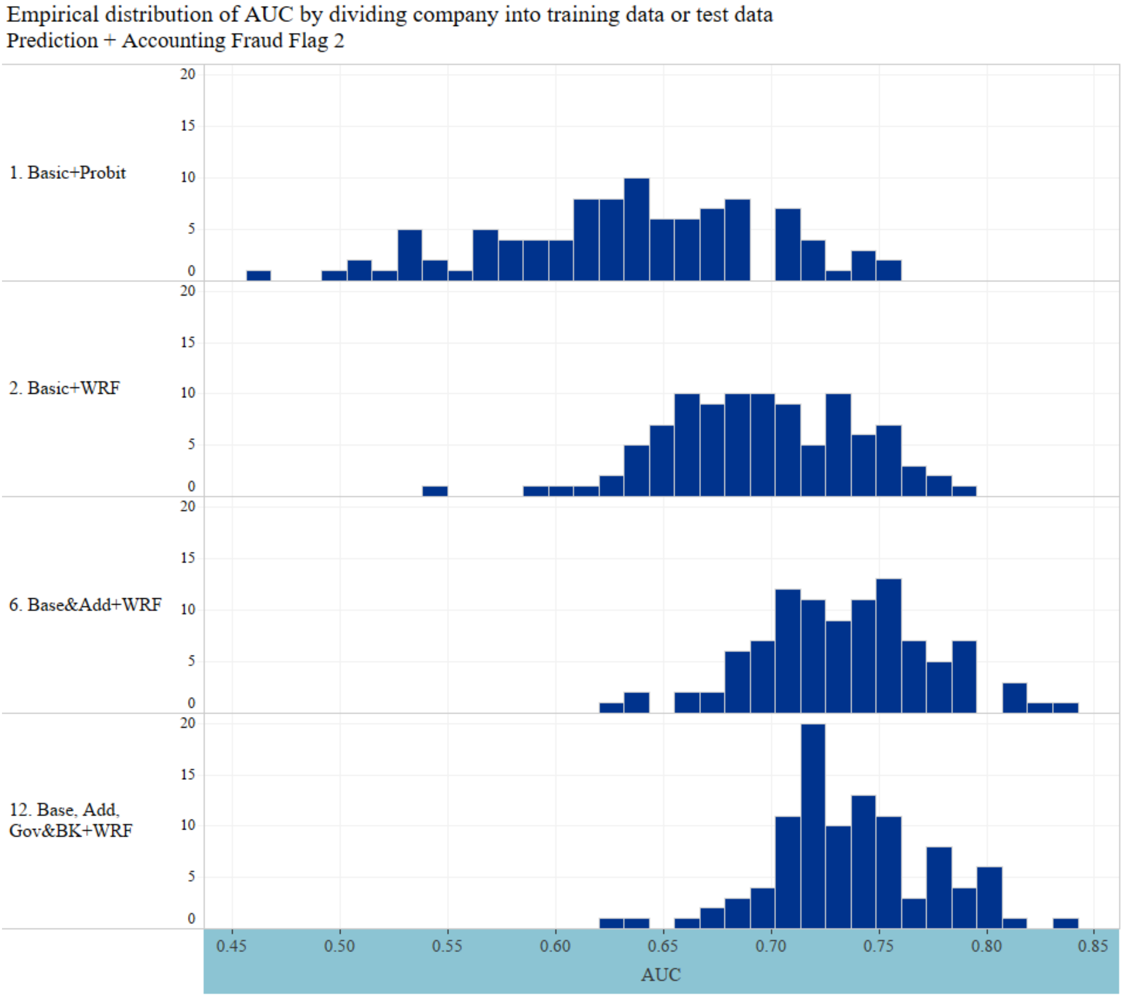

Empirical distribution of AUC by dividing company into training data or test data
Prediction + Accounting Fraud Flag 2
1. Basic+Probit
2. Basic+WRF
6. Base&Add+WRF
12. Base, Add, Gov&BK+WRF
20
15
10
5
0
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
AUC

Figure 4.

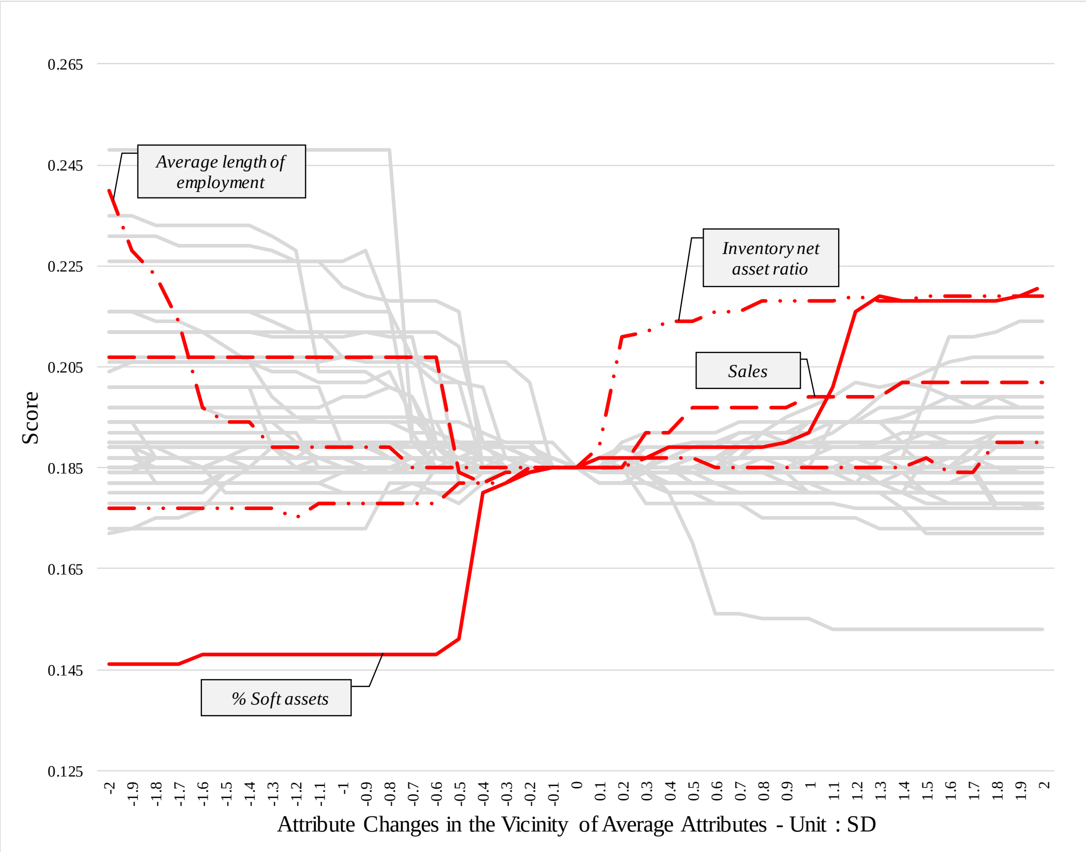


Figure 5. The Corresponding Relationship Between Attribute Changes in the Vicinity of Average Attributes and Changes in Score

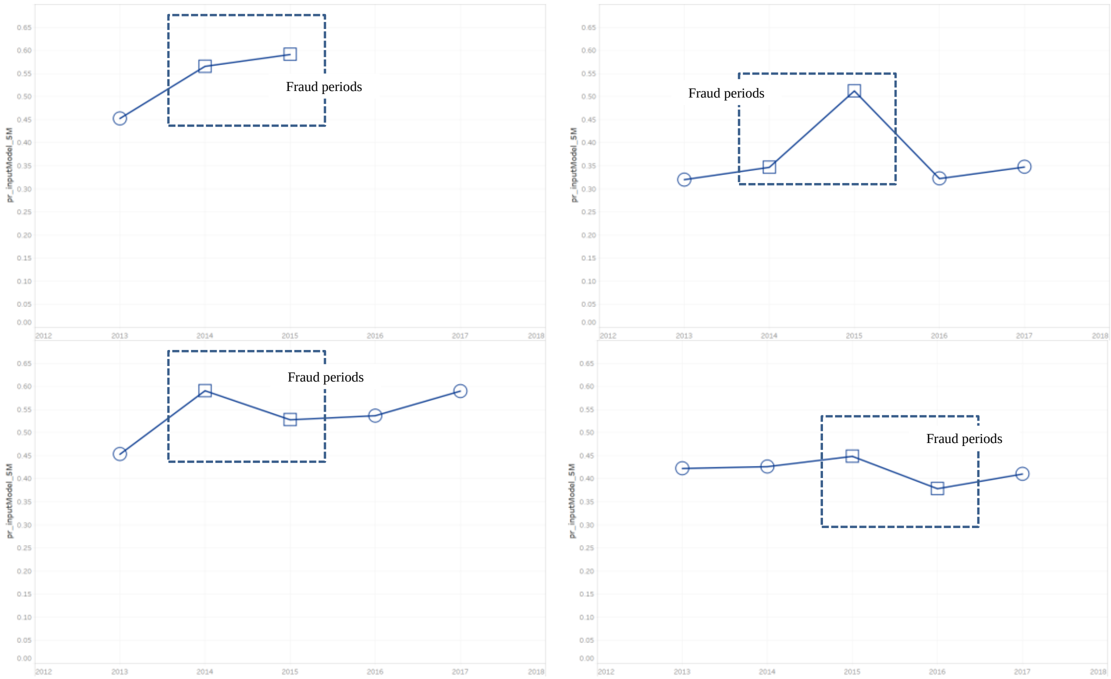


Figure 6. Dynamics of the detection score

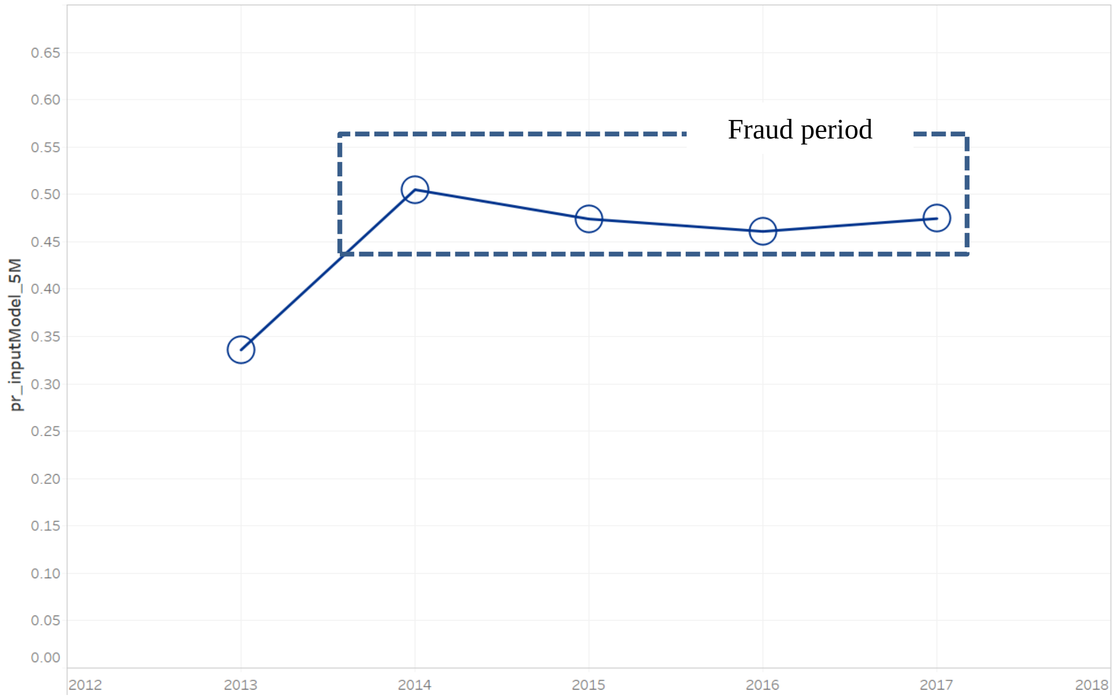


Figure 7. Dynamics of the detection score for the pure hold-out data