

Title:**A Dynamic Fraud Alarm Model for Corporate Fraud Detection: Evidence from China****Author:****Chang Chuan Goh, Anthony Bellotti, Xiuping Hua****Abstract**

We propose a fraud alarm model for corporate fraud detection for Chinese firms. We build our model using corporate governance information and financial indices of Chinese firms, which can be obtained from annual reports. In our study we consider (1) the dynamic aspect of the fraud detection problem, i.e., what is a suitable use of past data to build a good model; and (2) class imbalance given that fraud is a rare event. As part of model build, we carry out data drift analysis by investigating dataset shift to select the best time window for the training set. To model the diminishing effects of older data, adaptive forgetting factors are included when training the model. We then adopt oversampling techniques such as random oversampling and SMOTE to deal with the class imbalance problem. Five machine learning algorithms are used to build the model. They are logistic regression, support vector machine, artificial neural network, random forest, and gradient boosting. The fraud probability is computed for each firm-year observation, at both firm and sector level. Instead of predicting a binary outcome of fraud, the model triggers an alarm for potential fraud events if there is a rise in fraud probability based on model predictions. The alarm rate is flexible in the sense that we can opt for a higher true positive rate, at the expense of a higher false positive rate, depending on the fraud type. When applying the fraud alarm model to a real data set in China, our experiments show that despite training on data from a period with lower fraud occurrences, it is able to capture the rise in corporate fraud in a later period with higher fraud occurrences, at both firm and sector level. We find that random forest outperforms the remaining algorithms in fraud probability estimation.