Using Machine Learning to Predict Extreme Earnings Changes and Abnormal Returns: Evidence from Random Forest Models

Wai Ho (Wilson) Chan
wilson-wai-ho.chan@connect.polyu.hk
School of Accounting and Finance
Faculty of Business
The Hong Kong Polytechnic University

Mingming Ji
ming-ming.ji@connect.polyu.hk
School of Accounting and Finance
Faculty of Business
The Hong Kong Polytechnic University

Jian Kang
jian.kang@dufe.edu.cn
School of Finance
Dongbei University of Finance & Economics

Jingran Zhao
jingran.zhao@polyu.edu.hk
School of Accounting and Finance
Faculty of Business
The Hong Kong Polytechnic University

March 2025

We would like to thank comments from Ilia Dichev, Dave Ding, the participants at the 2023 JAAF conference, the 14th TJAR conference, and 2024 JCAE Annual Symposium. We also thank for the financial support from the Research Grant Council of the Hong Kong Special Administration Region, China (Project No. GRF 15510121).

Using Machine Learning to Predict Extreme Earnings Changes and Abnormal Returns: Evidence from Random Forest Models

Abstract: This paper uses machine learning techniques to predict one-year-ahead earnings jumps and crashes. We focus on extreme changes in earnings, as they can lead to substantial revaluations of firm value and attract considerable investor attention. Additionally, extreme earnings events are inherently difficult to predict. Overall, we find that a random forest (RF) model outperforms linear models in predicting earnings jumps and crashes. Moreover, we show that the hedge portfolio based on random forest model generates 14.13% size-adjusted annual return. The RF model outperforms in forecasting extreme earnings events by effectively capturing complex nonlinear relationships, reducing information noise, and managing skewed distributions and outlier biases. Our analysis further demonstrates that the RF model outperforms analysts' predictions for smaller firms, during periods of high economic uncertainty, and when reversal signals are present.

Keywords: Earnings Forecast; Extreme Earnings Change; Machine Learning; Random Forest

JEL Classification: G11; G17; M41.