A MACHINE LEARNING APPROACH OF MEASURING AUDIT QUALITY: EVIDENCE FROM CHINA

Authors: Hanxin Hu (Rutgers), Ting Sun (TCNJ), Miklos A. Vasarhelyi (Rutgers), Min Zhang(Renmin University, China)

This version: Nov 7, 2020

Contributions

- This study proposes a machine learning approach to predict audit quality with a wide range of variables (72) describing characteristics of accounting firm, individual audit partners, and public companies in China.
- It measures audit quality by constructing “surprise” scores based on traditional proxies of audit quality, including audit adjustments, non-clean audit opinion, and restatement(misstatement).
- The proposed approach is able to capture the “unobservable” factors that affect audit quality.
- This is the first archival research applying multiple audit adjustments (6) to measure audit quality
- We evidence the effectiveness and demonstrate the application of the proposed audit quality measures by
  - *examining the relationship between audit failures (as proxied by penalties) and our audit quality measure*

Prior literature and motivations(1)

- Existing studies used a variety of proxies to measure audit quality, the vast majority of them are derived from externally **observable** data (Bell, Causholli, and Knechel, 2015).
  - *abnormal accruals (Lee, Nagy, Zimmerman, 2019),*
  - *meet or beat analyst forecasts (Carey and Simnett, 2006),*
  - *accounting restatements (Myers et al., 2004)*
  - *going concern audit opinions (Reichelt and Wang, 2010)*
  - *audit fee*
  - *audit size*
- Those traditional proxies are based merely on “observable” attributes of companies and accounting firms
  - *indirectly measure audit quality*
  - *are “pertaining to the types of financial reporting and other audit-related outcomes” (Bell, Causholli, and Knechel, 2015)*
  - *using them individually or collectively cannot provide a relatively complete picture of audit quality.*

Prior literature and motivations(2)

DeFond and Zhang (2014):

"No single category paints a complete picture of audit quality. We therefore recommend that when possible, researchers use multiple proxies from different categories to take advantage of their strengths and attenuate their weaknesses"

Prior literature and motivations (3)

- Can we find an alternative way to measure audit quality from the angle of those **unobservable** factors?
- Those characteristics may come from the auditor's side (i.e., auditor's competence and independence) or company's side (i.e., competence and incentive). They are unobservable to us because they are internal information
  - *management incentive to falsify financial records,*
  - *deficiencies in the internal control of the company,*
  - *social or economic bonding between the auditor and the client management,*
  - *sufficiency and relevance of audit evidence supporting the audit conclusion (Hansen, 2014; Bell, Causholli, and Knechel, 2015; Aobdia, 2019),*
  - *accounting firm culture, skills and personal qualities of partners and staff in the accounting firm (FRC, 2008),*
  - *audit industry and markets, institutional background, and economic consequences of audit outcomes (Francis, 2011)*

Prior literature and motivations (4)

- Gul et al. (2013) : “aggressiveness (ARAgg)” to measure the difference between the predicted probability of modified (nonclean) opinion and the actual value of modified (unclean) opinion, and use the “aggressiveness” as an audit quality proxy.
  - *They define an indicator variable, MAO, which equals 1 if a client receives a modified audit opinion, and 0 otherwise.*
  - *They then estimate the predicted probability of issuing MAOs by running a logistic model, with MAO as the dependent variable and a set of client characteristics as explanatory variables.*
  - (-1,1) (0,1) (0 or 1)
    Their audit reporting aggressiveness measure (ARAgg) =predicted probability - the actual value of MAO.
  - *A higher ARAgg value means that an auditor’s propensity to issue MAOs is lower than what would be predicted from the data.*
  - *This measure reflects the differences between the prediction (expectation) and the actual case, which is considered a proxy of audit quality (in terms of the issuance of modified audit opinion) driven by certain “unobservable” factors*

Research Questions

1. What factors are predictable for audit quality as proxied by audit adjustments, nonclean opinion, restatement?
2. Among those factors, what is(are) the most important category(categories) of predictors?
3. How to construct an alternative audit quality measure that could capture the unobservable factors driving audit quality?
4. The usefulness of these measures?
5. How to apply the measure?

Research design

- **Step 1:**

Develop **machine learning** models to predict audit quality (proxied by the existence of six types of audit adjustments, non-clean opinions, and restatements) with independent variables consisting of attributes that measure the companies, audit firms, and individual audit partners characteristics

  - *The prediction of the probability that a certain event will occur is based on the past behaviors of the training data*
  - *We test the accuracy of the predictions with the test set and find those predictions are accurate, as high as 83% of AUC*
  - *The prediction is what are expected to happen, given the values of the independent variables.*

- **Step 2: Compute the "surprise" score, which is our proposed measure of audit quality**

Surprise= predicted probability of each of the above events- actual value of the event (dummy variable)

- *The surprise score is the difference between the prediction (what we expect) and the actual case*

- *There are some "unobservable" factors that drive the actual value of a certain event deviates from the expectation, the surprise score is the outcome due to those unobservable factors*

For example, for the existence of net income audit adjustments,

Surprise=predicted probability of the existence of net income adj- actual value of the existence of net income audit adjustments.

Assuming the predicted probability of the existence of net income adj is 0.8; the actual value=0 (no adjustments), so surprise=0.8-0=0.8

A higher "surprise" means that an auditor's propensity to adjust the net income is lower than what would be predicted from the entire sample (suggesting a lower audit quality)

- **Step 3: An application of this measure : Penalties=f(surprise score, controls)**

Specifically, we compare the result of three groups of audit quality measures:

  - surprise score,
  - traditional audit quality proxy,
  - aggressiveness(Gul et al 2013)

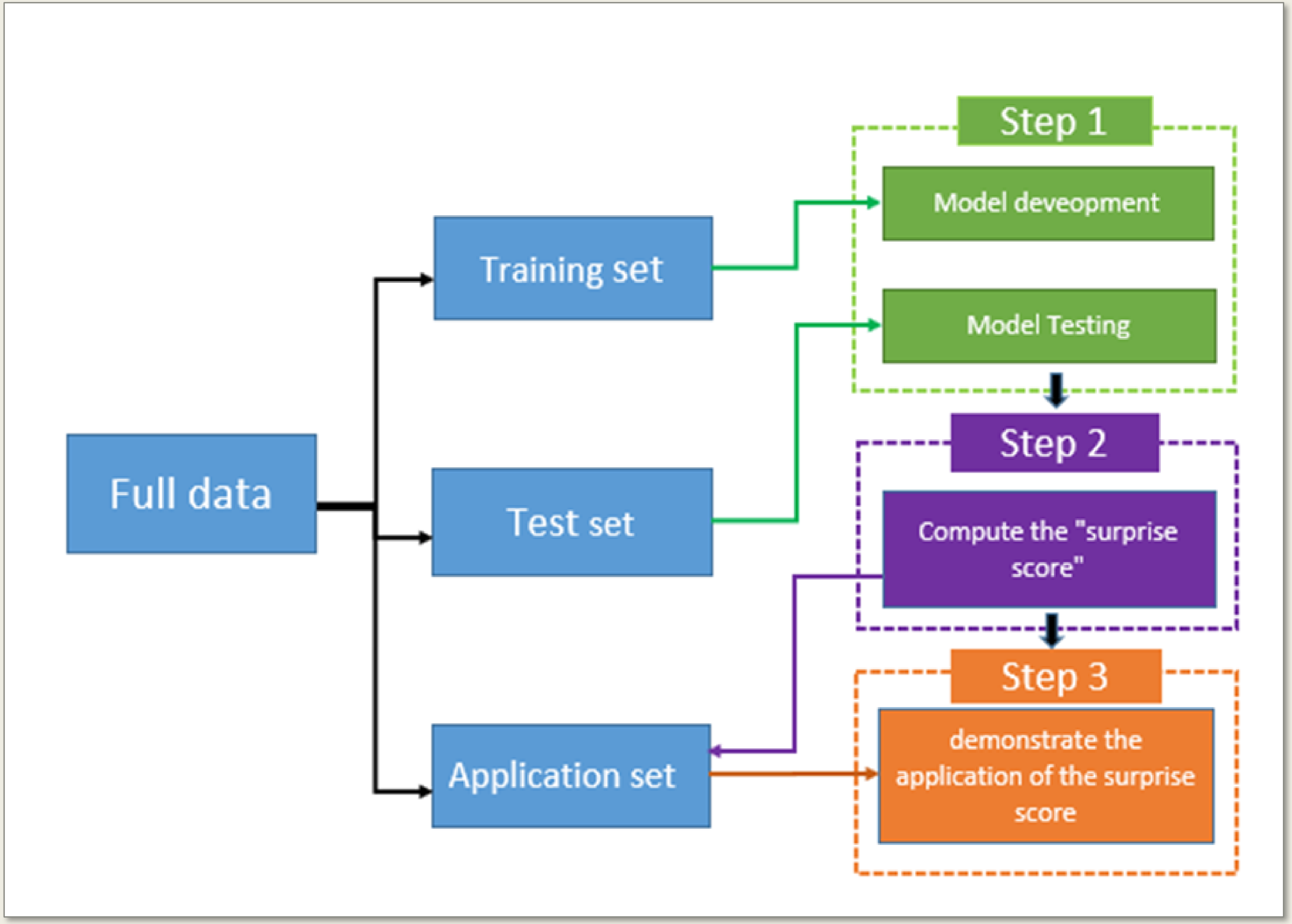
Data

- **Dependent variables (we developed a prediction model for each dependent variable)**:
  - *Net income adj*
  - *Total assets adj*
  - *Total liability adj*
  - *Stockholders' equity adj*
  - *Income before income tax adj*
  - *Income tax adj*
  - *nonclean opinion*
  - *Restatement(misstatement)*
- **Independent variables (example: using net income adjustment as the dependent variable):**

  Companies' characteristics (27 variables)

  Audit firms' characteristics (28 variables), e.g., revenue, subsidiaries, net assets

  Audit partners' characteristics (15 variables), e.g., education, age, gender, birthplace, title
- **Data sources:** Chinese Ministry of Finance, CICPA (Chinese Institute of CPAs), CSMAR(China Stock Market & Accounting Research Database)
- **Research period:** 2010-2017
- **Data size:** 11574
- **Data splitting:** training (6626 observations) /test (1325 observations)/application (3546 observations)


Full data
Training set
Test set
Application set
Step 1
Model deveopment
Model Testing
Step 2
Compute the "surprise score"
Step 3
demonstrate the application of the surprise score

Machine learning algorithms

- Random Forest
- SVM
- Gradient boosting
- XGBoosting
- Deep neural networks
- Logistic regression (the bench mark algorithm)
- RUSboosting
- Balanced Random Forest

Prediction Results (1) —— nonclean opinion

- Examples: prediction results when separately treating nonclean opinion, net income adjustment and liability adjustment.
- Treating nonclean opinion as dependent variable:

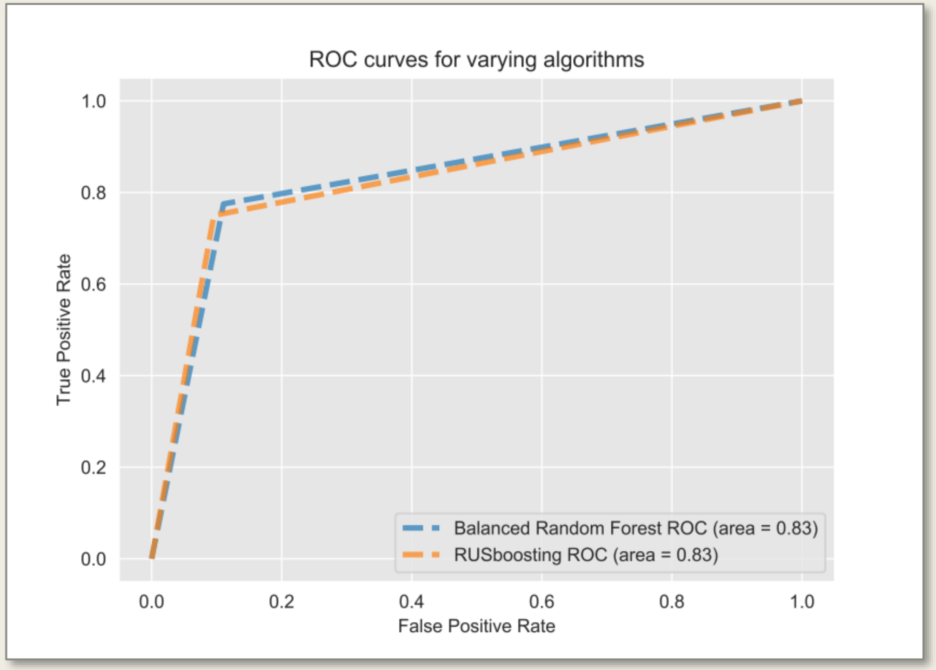


Figure 1: ROC AUC for varying algorithms when "unclear opinion" is the target variable

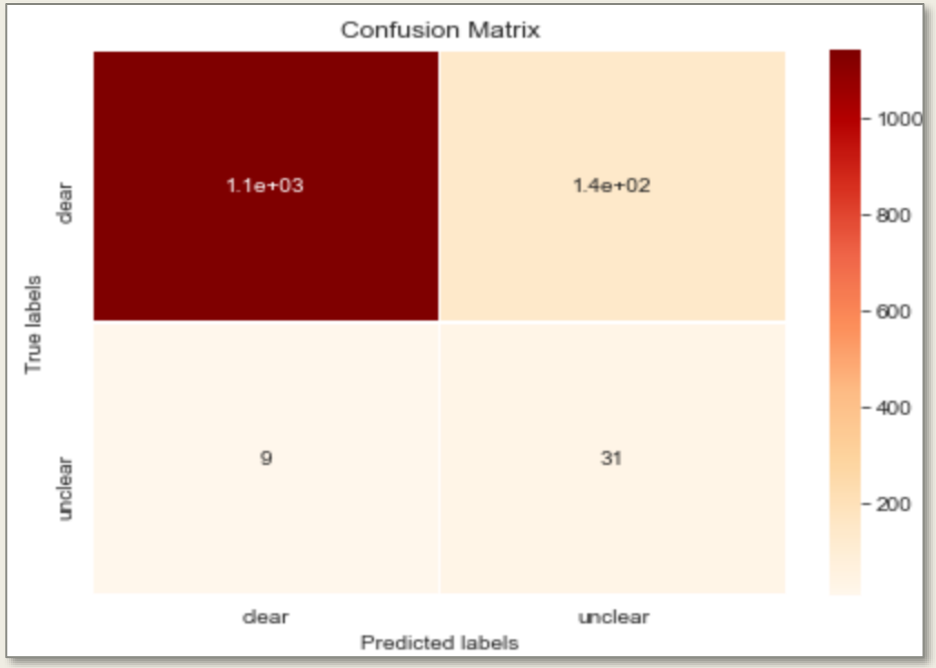


Figure 2: Confusion matrix when using balanced random Forest algorithm

Prediction Results (2) —— Net income adj

- Examples: prediction results when separately treating nonclean opinion, net income adjustment and liability adjustment.
- Treating the dummy variable whether net income adjustment exists as dependent variable:

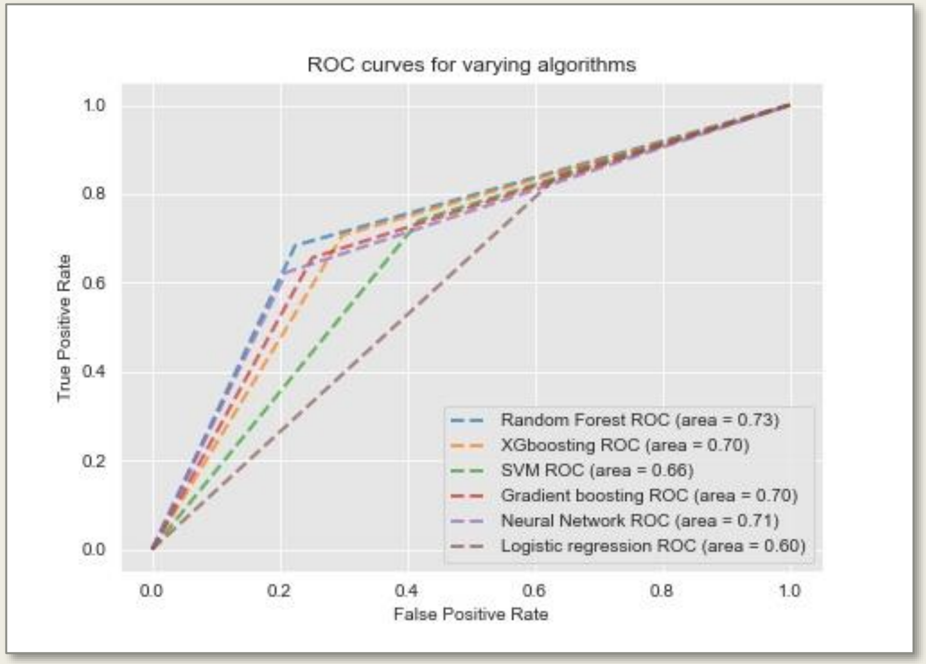


Figure 3: ROC AUC for varying algorithms when "net income adj" is the target variable

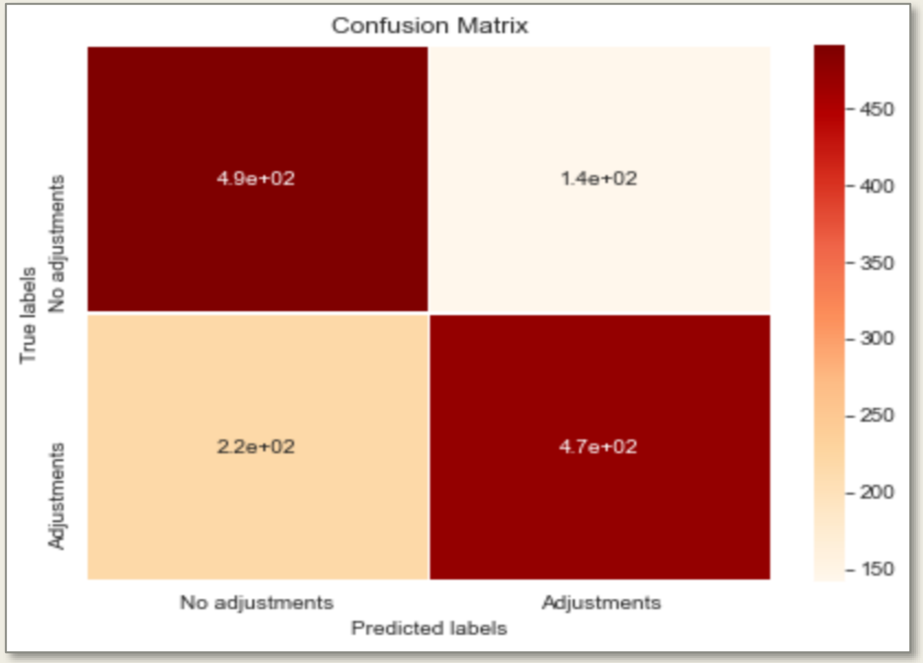
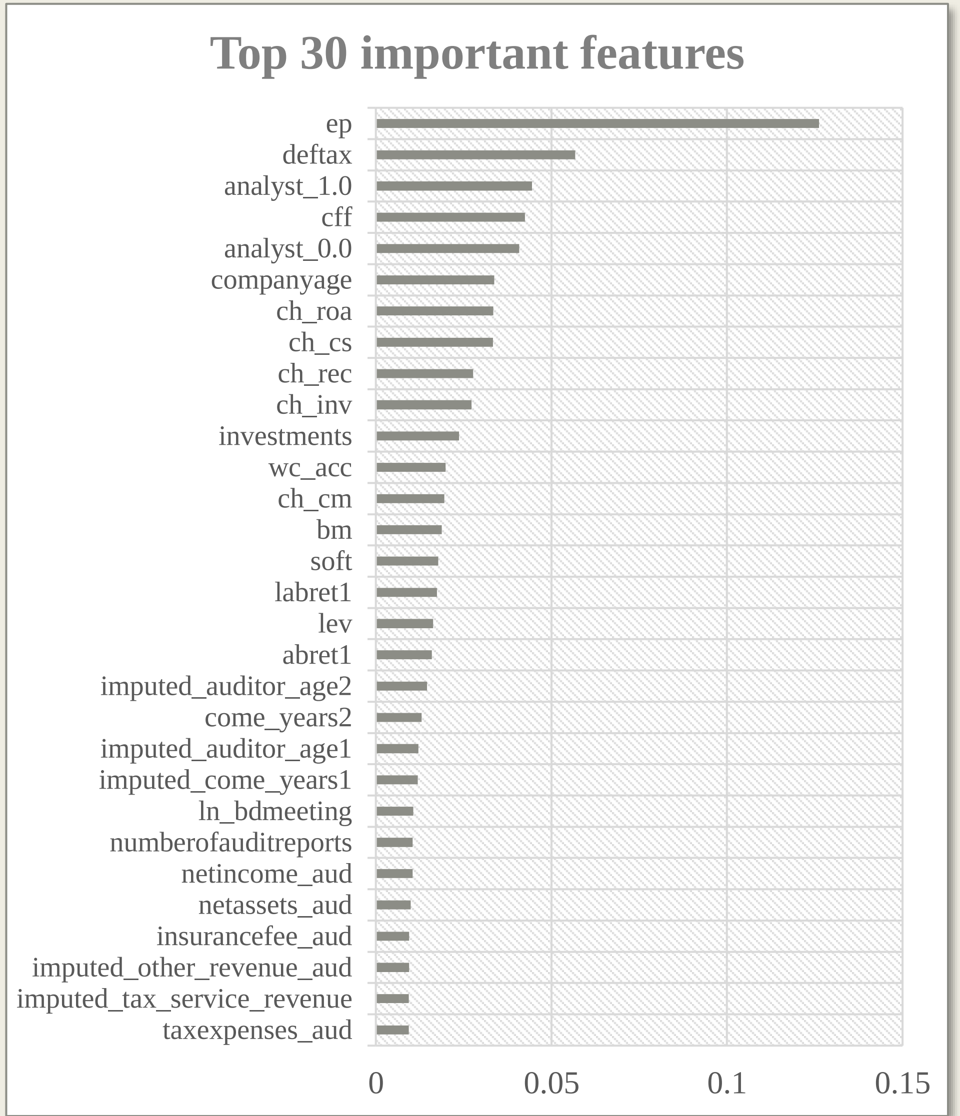


Figure 4: Confusion matrix when using random forest algorithm

Important predictors (1) —— non-clean opinion

- When treating nonclean opinion as the dependent variable and using Balanced random forest algorithm (106 variables in total):



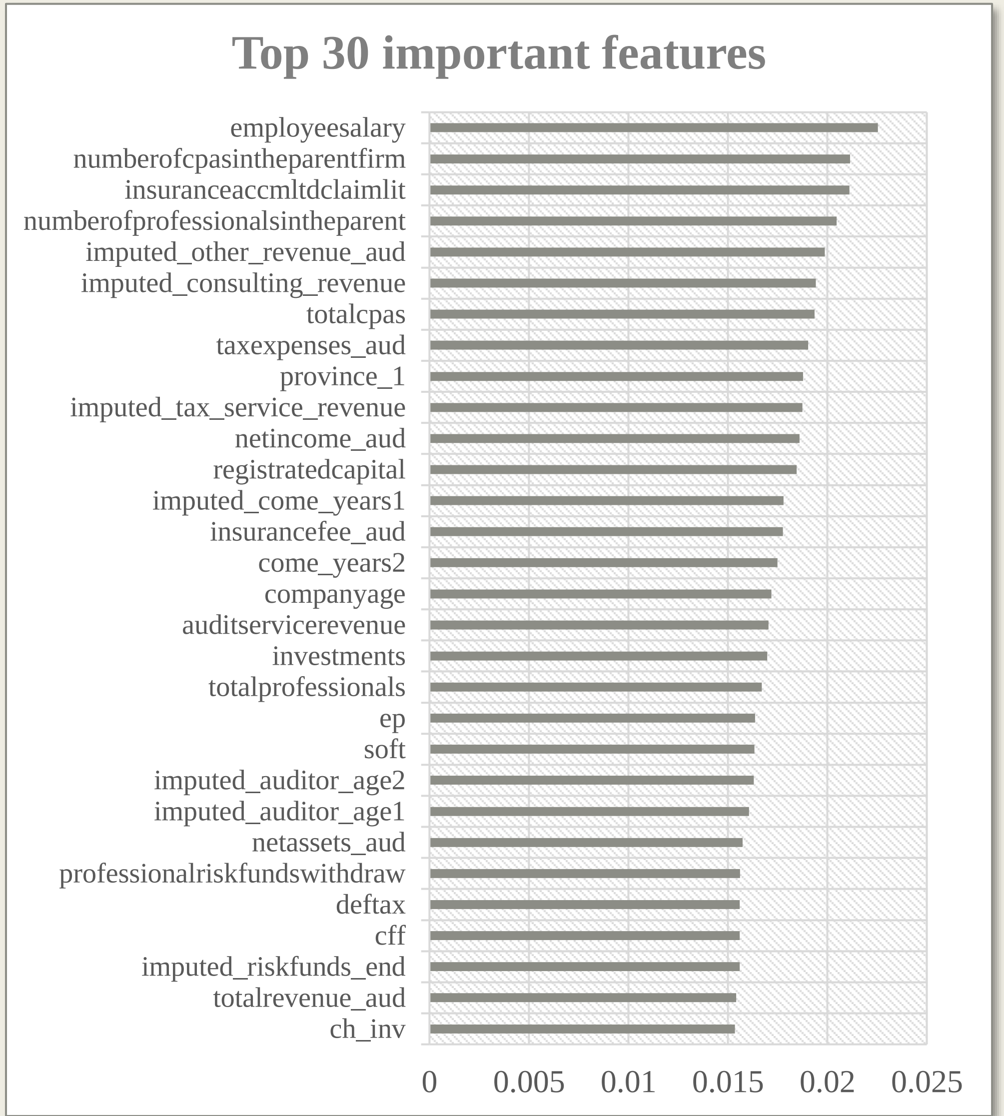
※ Which group of factors is the most important:

Categories of factors	Proportions among top 30 features
Audit firms' characteristics	7/30 = 23.33%
Audit partners' characteristics	4/30 = 13.33%
Companies' characteristics	19/30 = 63.33%

Obviously, companies' innate characteristics are the most important features when predicting the issuance of nonclean opinion. Additionally, partners' experience and audit firms' financial situation are also associated with the dependent variable.

Important predictors (2) —— Net income adj

- When treating net income adjustment as the dependent variable and using random forest algorithm (106 variables in total):



※ Which group of factors is the most important:

Categories of factors	Proportions among top 30 features
Audit firms' characteristics	19/30 = 63.33%
Audit partners' characteristics	4/30 = 13.33%
Companies' characteristics	7/30 = 23.33%

Audit firm characteristics are the most important features when predicting whether there is net income adjustment. Additionally, partners' experience and companies' financial situation are also associated with the dependent variable.

Application

- Purpose: explore whether our proposed surprise scores are more indicative to audit failures as compared to traditional audit quality proxies.
- Empirical design:

	Model 1	**Model 2**	**Model 3**	**Notes**
Dependent variable	Penalty (used as an indicator of audit failure)	Penalty	Penalty	Penalty =1 if the company received a penalty for accounting violations, and 0 otherwise (we excluded the observations that the auditor provided an non-clean opinion for the problematic financial statements)
Independent variable	Our audit quality measure(i.e., the surprise score)	The traditional observable audit quality measure	The aggressiveness score based on Gul , et al. (2013)	For each audit quality measure (e.g., net income adjustment, nonclean opinion etc.) we created a set of two models for comparison purpose.
Controls	Same controls	Same controls	Same controls	All controls were derived from previous research

Application: Results

1. Nonclean audit opinion; Nonclean opinion surprise score; The aggressiveness score

Independent variable	Estimated coefficient	P value
Nonclean audit opinion	-27.78	0.960933
Nonclean opinion surprise score (ML)	*3.467*	*7.84e-13* ***
The aggressiveness score (Logistic)	*0.3958*	0.008126 **

2. Net income adjustment; Net income adjustment surprise score; The aggressiveness score

Independent variable	Estimated coefficient	P value
Net income adjustment	-0.01700	0.853533
Net income adjustment surprise score (ML)	*0.01105*	*0.911363*
The aggressiveness score (Logistic)	*0.02298*	*0.812480*

Note: the regression results for other audit adjustment variables are similar to those for net income adjustment.

Application: Result explanation

- **Why the result of audit adjustment / audit adjustment surprise score / the aggressiveness score is not significant:**
- A potential reason for this is the fact that the existence and the amount of audit adjustment are based on a negotiation process between the auditor and the client company;
- When both parties could not reach an agreement of adjustment, the auditor will issue a non-clean opinion. This is the final outcome as well as the worst case of the negotiation.
- In other words, a higher score of net income audit adjustment signals poor audit quality, *but it does not necessarily result in an audit failure that is so severe that the company will receive a penalty*, while a higher surprise score of non-clean audit opinion (which means the auditor should issue an non-clean opinion but he/she did not) signals a relatively severe audit failure which will result in a penalty.

Conclusion and discussion

- This research applies machine learning technology to construct a surprise score based on traditional audit quality proxies and we use the proposed surprise score to measure audit quality
- In our application section, we find a significant association between our proposed quality measure and audit failure proxied by penalties, while there is no significant association between traditional audit quality measures and the audit failure.
- This research highlights the importance of the audit firms’ characteristics for audit quality predictions
- It calls for the disclosure of audit adjustments and audit partners information etc.
- Regulators may find our approach useful to evaluate the audit quality and better govern the market
- Investors may find the measures useful to support decision-making
- Researcher may find them useful to investigate the relationship between audit quality and other factors, such as industry specialization, mandatory rotation, audit fee etc.

Thank you!

Dr. Ting Sun
sunt@tcnj.edu