Synergizing Auditors and Machines: Evidence from Going Concern Assessment

Yu Gu
Ph.D. Candidate
Rutgers, The State University of New Jersey
Southwestern University of Finance and Economics
yg431@scarletmail.rutgers.edu

Chanyuan (Abigail) Zhang Parker
Assistant Professor
University of Texas at San Antonio
abigail.zhang@utsa.edu

Miklos Vasarhelyi*
KPMG Distinguished Professor
Rutgers, The State University of New Jersey
miklosv@business.rutgers.edu

*Corresponding author

Acknowledgement: we thank feedback from Michael Alles, Chelsea Anderson (discussant), Jeff Boone, Allen Blay, Graham Gal, Jake Krupa (discussant), Miguel Minutti-Mezza, Valeri Nikolaev, Dan Palmon, Aaron Saiewitz, Lanxin Jiang, an anonymous associate editor, an anonymous reviewer, participants from the 2022 Accounting Information System midyear meeting, 2022 American Accounting Association annual year meeting, 2023 Auditing midyear meeting, 2023 Accounting Information Systems midyear meeting, and 2023 PCAOB conference.

Synergizing Auditors and Machines: Evidence from Going Concern Assessment

Abstract

Amid recent attention over auditors' ability to properly assess client firms' going concerns, this study examines the potential for auditor and Machine Learning (ML) collaboration in assessing going concerns. We first examine the comparative performance of auditors and ML. Consistent with bounded rationality, we find that errors in auditor-issued Going Concern Opinions (GCOs) are associated with auditors' capability, cognitive bias, and constraints. In contrast, we find that ML tends to make mistakes when unobservable information exists that cannot be directly incorporated into ML models. Using a novel approach, we derive a "team" assessment that integrates auditors and ML. Our results show that the team assessments can synergize the relative advantages of auditor and ML, and overcome their relative disadvantages, serving as a potential way of generating higher-quality going concern assessments. Our research has implications for accounting practitioners and regulators regarding auditor-machine collaboration in the "AI era."

Keywords:
Going Concern Opinion, Audit Quality, Professional Judgment, Cognitive Constraints, Man-Machine Collaboration, Machine Learning, Artificial Intelligence

JEL Classification: M41, M42

1. Introduction

This study examines the potential for collaboration between auditors and machine learning (ML) to improve the quality of going concern assessments.[1] Under U.S. securities laws and U.S. public auditing standard AS 2415, auditors are required to issue a Going Concern Opinion (GCO) when there is substantial doubt about the entity's ability to continue as a going concern for not exceeding one year beyond the date of the financial statements being audited. The auditor's evaluation of a company's ability to continue as a going concern is a topic of interest to investors and other users of financial statements as it serves as an independent early signal to investors of company financial distress (PCAOB, 2023).

Despite existing research on the quality of GCOs (e.g., Carson et al., 2013), revisiting this topic is warranted for several reasons. Recent high-profile corporate failures have raised questions about auditors' effectiveness in assessing companies' going concerns, highlighting the need for improved going concern assessment methods.[2] Additionally, regulatory bodies like the Public Company Accounting Oversight Board (PCAOB) and the International Auditing and Assurance Standards Board (IAASB) are considering revising auditing standards to enhance the reliability, effectiveness, and transparency of going concern evaluations (IAASB 2024; PCAOB 2023a). Moreover, the advent of the "AI era," characterized by significant advancements and lower barriers of entry in ML, presents an opportunity to transform traditional checklist-based audit methodologies in assessing going concerns (Gutierrez et al., 2020).

[1] In this research, going concern assessments refers to the evaluations of client firms' going concern status. Going concern opinions refer to going concern evaluations from auditors when there is substantial doubt about the entity's ability to continue as a going concern for not exceeding one year beyond the date of the financial statements being audited (AS 2415).

[2] For instance, Silicon Valley Bank collapsed in early 2023, merely two weeks after the auditor gave the bank a clean bill of health despite having a 13 percent decline in client deposits during the last month of the audit (Weil and Eaglesham, 2023).

This study aims to explore whether ML can complement the auditor's skill set in evaluating a company's going concern status. The process of evaluating client firms' going concerns requires auditors to synthesize multiple sources of information gathered throughout the audit process (AS 2415) – a task that is limited by human cognitive capacities as per bounded rationality (Simon, 1955). ML, on the other hand, excels at detecting patterns within vast data sets (Alpaydin, 2020), offering a promising tool for enhancing the auditor's ability to integrate and analyze various pieces of data. Additionally, the issuance of GCOs involves significant auditor judgment, which can be subject to influences from auditor capabilities, cognitive biases, and constraints (Carson et al., 2013; Kahneman et al., 2021). ML models, by learning directly from data, can potentially mitigate these judgment flaws (Kahneman et al., 2021).

While ML presents opportunities to enhance the auditing process, it is not without limitations, and the judgment of auditors might still remain essential. Auditors can integrate soft information into their analyses, such as nuanced insights gained from interactions with client's managers, while ML systems, which primarily rely on quantifiable, codified data, cannot. This capability is particularly crucial in going concern assessments, where auditors must evaluate management's plans to address financial distress (AS 2415), relying on qualitative judgments and intentions that are not easily encoded into data. Furthermore, auditors often have access to a company's confidential information, offering a depth of insight that might not be available in the public datasets typically utilized by ML models. This proprietary information could provide a more comprehensive view of a company's financial health and prospects, thus contributing to the auditor's evaluation of firms' going concerns.

Therefore, while ML excels at processing and analyzing large volumes of data to identify patterns, the incorporation of soft information and proprietary insights by auditors could be equally crucial. This synergy between ML's data-processing capabilities and the auditor's judgment suggests that a collaborative approach could yield more effective assessments of a company's ability to continue as a going concern.

We follow Gutierrez et al. (2020) and compile a sample of distressed public firms in the United States from 2003 to 2021 with 2010 to 2021 as the hold-out testing period. We adopt a commonly-used advanced ML algorithm called Extreme Gradient Boosting and train this model to utilize companies' financial indicators to evaluate whether there is substantial doubt about the entity's ability to continue as a going concern for not exceeding one year beyond the fiscal year-end date. We follow literature convention to use Chapter 7 or 11 bankruptcy as a proxy for substantial doubt about the entity's ability to continue as a going concern (e.g., Berglund, 2020; Carson et al., 2013; Chava and Jarrow, 2004; Desai et al., 2020; Hardies et al., 2018; Mossman et al., 1998; Xu and Kalelkar, 2020). Then, we compare ML's assessments with GCOs in how accurately they provide early warnings of corporate failure.

Our analysis shows that ML is effective in providing early warnings to companies at risk of bankruptcy. Specifically, in the hold-out testing sample, ML achieves an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.9176, indicating an "outstanding" level of predictive accuracy (Hosmer et al., 2013; Mandrekar, 2010). Furthermore, its Precision-Recall AUC is 0.2101, significantly outperforming random predictions in a context where the actual bankruptcy rate is approximately 1 percent. When comparing the effectiveness of ML predictions with auditor-issued GCOs, we find that ML doubles the F-score, which measures the balance between precision (the proportion of correct early warnings of bankruptcies in all early warnings issued) and recall (the proportion of correct early warnings of bankruptcies in all bankruptcies), suggesting a substantial improvement over GCO.

Before exploring the potential for auditor-ML collaboration, we first examine what factors are associated with errors made by auditors or ML. Prior research often regards any discrepancy between bankruptcy occurrences and the issuance of GCO as errors in GCO (e.g., Berglund, 2020; Carson et al., 2013; Desai et al., 2020; Xu and Kalelkar, 2020). However, this practice may oversimplify the complexities of auditor judgment, as auditors might issue a GCO based on valid concerns at the time of audit, or fail to predict a company's collapse due to unforeseeable future events. To avoid unjustly attributing errors to auditors, we refine our analysis by excluding cases where both ML and auditor assessments agree but do not match the actual outcome. This approach recognizes that under such circumstances, an effective ML model would make the same decision. Then, for the *remaining* data, we categorize auditor errors as "false negative" when they do not issue a GCO for companies that go bankrupt the following year, and as "false positive" when they issue a GCO for companies that survive. Similarly, for the *remaining* data, ML errors are identified as "false negative" when it fails to provide an early warning to an impending failure, and "false positive" when it incorrectly forecasts a failure that does not materialize.

With this refined approach to identifying GCO errors, we find that GCO errors are associated with auditors' capability, cognitive biases, and constraints. Specifically, Big Four auditors issue GCOs with fewer false positive and false negative errors, consistent with the conclusion that Big Four auditors provide higher-quality audits (e.g., Che et al., 2020; DeFond et al., 2017; Jiang et al., 2019). However, we find that when auditor offices are smaller or have lower industry expertise, auditors tend to be more conservative by issuing more GCOs, increasing false positive errors, and decreasing false negative errors. Furthermore, consistent with the notion that when the independence of the auditor is hampered, reflected in higher client influence, auditors are less likely to issue GCO, thus decreasing false positive errors and increasing false negative errors. Lastly, we find that auditors' tendency to "anchor" current-year opinion on the previous year's (i.e., anchoring bias) also positively relates to GCO errors.

Regarding ML errors, we find that when unobservable information (e.g., proprietary or soft information) signals the need to issue GCO, there is a higher chance of ML false negative error. This happens because ML does not have access to such unobservable information that indicates the necessity to issue non-clean going concern opinion; on the contrary, when unobservable information suggests no need to issue GCO, ML tends to have more false positive errors. Furthermore, we find that the increased interaction between auditors and clients – evidenced by changes in the audit report lag and length – correlates with more false negatives from ML. Taken together, these findings underscore ML's limitations in integrating subtle and uncodified information into its assessments.

Given the distinct strengths and weaknesses of both auditors and ML in evaluating going concerns, we propose a collaborative approach that leverages the unique advantages of each. This approach involves integrating auditors' qualitative insights into ML, potentially bridging the gap in ML's ability to process contextual information. To test this concept, we enhance the ML model by including GCO as an input variable, given that it encompasses auditors' processing of contextual information (Gutierrez et al., 2020). Then we retrain the new model and obtain its assessments, which we term the "team assessments."

Next, we examine whether this team assessment can mitigate the inherent limitations encountered when auditors and ML operate independently. We first investigate scenarios that we find to be challenging for auditors due to their own limitations, including audit offices with lower industry expertise, smaller sizes, heightened client influence, and the presence of anchoring bias. Our findings indicate that in these specific circumstances, team assessments provide more informative evaluations of firms' going concerns than GCOs. We then shift our focus to situations where ML typically struggles, particularly when dealing with subtle and uncodified information that it cannot effectively process. This includes conditions characterized by high levels of unobservable information and significant changes in audit report lag and length. In these cases, we find that team assessments outperform sole ML evaluations. These results collectively underline the benefits of integrating auditor insights with ML capabilities. By doing so, the combined approach not only addresses the specific weaknesses of each method but also enhances the overall quality of going concern assessments. This integrated strategy shows potential in overcoming the individual constraints of both auditors and ML.

Effective ML depends on stable and consistent training data. When this stability is disrupted – such as during unpredictable macroeconomic conditions – the performance of ML typically suffers. This vulnerability presents an opportunity for auditors, who are better equipped to adapt their decision-making processes to include real-time information flexibly. To demonstrate this, we utilize the period affected by COVID-19 pandemic (i.e., fiscal years 2020-2021) as one marked by significant economic disturbance. We find that, during this period, ML assessment is no longer effective in providing early warnings of next-year bankruptcy, whereas GCO and the team assessment remain effective. The continued effectiveness of the team assessments during such a turbulent period further supports the value of integrating auditor insights with ML predictions.

This study makes the following contributions to the accounting literature. First, we expand upon existing research that explores the comparative advantages of algorithms and human judgment (Coleman et al., 2022; Liu, 2022; Cao et al. 2024) by focusing specifically on auditors, whose decision-making quality is crucial for the healthy functioning of capital markets. Amid concerns about increasing audit deficiency rates (PCAOB, 2023b), and in light of efforts by the PCAOB and the IAASB to enhance audit quality, our study provides critical insights into how ML can improve auditor decision-making processes.

Second, we advance the discussion on the role of auditor judgment in the quality of GCO, a topic extensively studied by previous literature (e.g., Carson et al. 2013; Gramling et al., 2011; Joe, 2003; Kida, 1980; Lehmann and Norman, 2006; Tucker et al., 2003). By using ML as a "judgment-free" benchmark for comparison, we delve deeper into understanding the limitations of auditors' judgment in going concern assessments. This approach allows us not only to utilize ML for its predictive capabilities but also to gain deeper conceptual insights into the factors influencing auditor decision-making. To identify auditor errors more accurately, we refine the measurement of errors in GCOs by comparing them to actual outcomes of bankruptcy, excluding cases where auditors could not reasonably have predicted the future. Our findings indicate that auditors' decisions are often affected by their capabilities, cognitive biases, and constraints.

Third, we introduce a novel yet straightforward method to empirically demonstrate the benefits of collaboration between auditors and ML. While ML generally exhibits higher accuracy than GCOs in assessing going concerns, it can falter in situations involving unobservable information that is accessible to auditors but not directly available to ML. This emphasizes the need for a collaborative approach that leverages the strengths of both ML and auditors. Our study demonstrates that a "team assessment", which integrates the advantages of both auditors and ML, can effectively mitigate the limitations inherent when either party works in isolation. As the auditing industry and standard setters explore ways to improve going concern assessments, our findings suggest that auditor-ML collaboration could be a viable strategy to enhance the quality of going concern assessments.

2. Background Literature

2.1. Going concern opinions

In the United States, auditors are required to evaluate whether there is substantial doubt about the client entity's ability to continue as a going concern for not exceeding one year after the financial statement being audited (AS 2415). The going concern assessment is inherently a judgment decision-making process, and the evaluation heavily relies on the auditor's knowledge gained via audit processes, as well as knowledge about conditions and events that existed at or before the completion of fieldwork relating to going concern (Carson et al., 2013).

GCO serves as a warning for investors when the firm is in financial distress (Chen and Church, 1996; Geiger et al., 2014; Jones, 1996; Knechel and Vanstraelen, 2007; Lennox, 1999). A large body of literature has examined the accuracy of GCO, and the descriptive analyses show that GCO has limited effectiveness in assessing client firms' going concerns (Carson et al., 2013). Previous studies primarily use Type I (false positive) and Type II (false negative) errors of GCO to measure its quality (Carcello et al., 2009; Carson et al. 2013; Geiger and Rama, 2006; Hossain et al., 2020; Myers et al., 2014). Type I error occurs when the firms receive GCO but do not subsequently go bankrupt. Type II error occurs when the firms enter bankruptcy or default but do not receive GCOs. Approximately half of the firms going bankrupt or defaulting in the U.S. do not receive GCOs the prior year, and more than 90 percent of the firms that receive GCOs survive the following year (Barnes, 2004; Carson et al., 2013; Gutierrez et al., 2020). Type I errors of GCOs happen potentially due to the prudent nature (i.e., conservatism) of auditors to avoid litigation costs for failing to warn investors of a soon-to-fail firm (Carson et al., 2013). Type II errors of GCOs happen potentially because there are changes in the management's mitigation plans or because auditors lack the competence to identify failing firms on time (Carson et al., 2013).

The determinants of GCO issuance have been widely researched (Carson et al., 2013; Brunelli, 2018; Geiger et al., 2021; PCAOB, 2012). The factors could be categorized into client factors and auditor factors. Client factors include accounting-based financial ratios to measure the financial distress situation, such as calculated Z-Score (Altman, 1968), mitigation plans (Carson et al. 2013; PCAOB, 2012), and industrial economic impact (Custodio et al., 2022). Auditor factors are based on the auditor-client relationship, auditors' characteristics, and judgment. The relationship considers the factors of litigation cost, reputation loss, and the possibility of losing the client (Kida, 1980). GCO decisions can also be influenced by an auditor's experience and knowledge (Gramling et al., 2011; Lehmann and Norman, 2006), such as the audit plan preparation (PCAOB, 2012), auditor type (Big4 or non-Big4) (DeFond and Lennox, 2011), audit partner gender (Hardies et al., 2016; Hossain et al., 2018), mainstream religion of the auditor office location (Omer et al., 2018), and auditor tenure (Read and Yezegel, 2016). Furthermore, mood affects the audit decision-making processes (Curtis, 2006). In summary, the issuance of GCO not only relies on the client's financial distress status and potential mitigation plans, but also depends on the auditor's judgment which is influenced by their personal characteristics and external environment factors.

2.2. Auditors' cognitive constraints and behavioral biases

During the assessment of a company's going concern status, auditors' judgments may be influenced by their cognitive limitations and behavioral biases. Simon (1955) introduces bounded rationality, which suggests that while individuals aim to make rational decisions, their ability to do so is limited by various constraints. According to bounded rationality, auditors are expected to use information in a "satisfying" manner in forming judgments rather than in an "optimizing" manner (Carpenter and DirSmith, 1992). Similarly, when assessing clients' going concerns, auditors may not be able to process or mentally represent the rich set of information. Instead, they may resort to a simpler model of going concern risk assessment and cannot issue GCO in an optimal way.

Prior literature has identified factors that can influence auditors' judgement and decision making, and we broadly classify them into categories of capability, cognitive bias, and constraints. Capability includes factors that represent auditor's ability to perform the audit, such as industry specialization, office size, and Big Four brand name. The majority of literature finds that a higher level of industry specialization relates to higher audit quality, reflected in lower chances of restatements and lower levels of abnormal accruals (Balsam et al., 2003; Romanus et al., 2008; Reichelt and Wang, 2010). Audit quality also tends to be higher when audit is performed by auditors with larger offices (Choi et al., 2010; Francis and Yu, 2009; Francis et al., 2013; Newton et al., 2013) or by Big Four auditors (Becker et al., 1998; Che et al., 2020; DeFond et al., 2017; Eshleman and Guo, 2014; Jiang et al., 2019).

Kahneman et al. (2021) posit that bias in judgment and decision-making is the systematic deviation of the assessments from the actual outcome. Accounting literature has identified a few factors that may introduce bias in auditors' decision-making, including non-audit service, client influence, tenure, and new clients. Regarding non-audit fee services, the literature presents mixed evidence on its impact on audit quality. On one hand, close and repeated interaction may prevent impartial audit and impair audit quality (Bazerman et al., 1997; Francis, 2006). On the other hand, such service may lead to knowledge spillover, potentially enhancing audit quality (Wu, 2006; Lim and Tan, 2008). Research also suggests that a higher client's financial significance (i.e., influence) to the audit office, the more likely the auditor reports favorably to retain the influential client (Francis and Yu, 2009). In addition, auditor tenure could impact auditor outcome, but with mixed evidence. Longer tenure might impair independence due to social boning, potentially reducing audit quality (Bell, Causholli, and Knechel, 2015). On the contrary, longer relationships can enhance audit efficiency through better client knowledge (Gul et al., 2009). Moreover, audit quality for the first year of auditor-client engagement may be negatively impacted due to the limited knowledge of a new client's business operations and financial practices (Myers et al., 2003; Liu et al., 2017).

Another potential contributor to auditors' cognitive bias, which prior literature has not explored, is auditors' tendency to base current-year decision on prior years. Tversky and Kahneman (1974) proposed the Anchoring Heuristic, which states that when making a numerical estimate, we are frequently influenced by the number we begin with. Anchoring bias is a cognitive bias that arises in later development when we place an excessive amount of reliance on the first piece of information we are given about a topic (Kahneman et al., 2011). As the anchoring bias suggests, auditors may rely on the prior year's decision of whether to issue GCO in this year. This tendency is partially driven by the common practice of comparing current year financial reports with prior years in audit planning and substantive analytical procedures (e.g., Zhang et al., 2022; Rozario et al., 2023).

Other factors that might affect auditors' judgment and decision-making are constraints such as workload compression. The PCAOB has raised concerns regarding how time and budget constraints could negatively impact audit outcomes (PCAOB, 2015). Prior literature indicates that the pressure to complete a high number of audit engagements within a short timeframe can lead to dysfunctional behavior and lower audit quality (Lopez and Peters, 2012).

As will be shown in the Section "What Factors Are Associated with Auditor GCO Error", we will examine whether and how the above factors are associated with auditors' errors in assessing going concerns.

2.3. Statistical models that predict bankruptcy

The literature on using statistical models to predict bankruptcy started with the Multivariate Discriminant Analysis (MDA) (Altman, 1968; Mckee, 1976) and then continued with the Logit (e.g., Menon and Schwartz, 1987; Ohlson, 1980; Sohn and Kim, 2007) and Probit Analyses (e.g., Koh, 1987). Shumway (2001) proposes a hazard model that considers the dynamic development of bankruptcy. Later, machine learning (ML) was proposed to be an alternative solution to bankruptcy prediction, given that advanced estimation models have shown to have better predictive capability than conventional statistical methods (Martens et al., 2008; Yeh et al., 2014). For example, prior research employed sophisticated models like Neural Networks (Koh and Tan, 1999) and Decision Trees (Koh and Low, 2004). Meanwhile, the accuracy of the prediction model is improved as more predictor variables are included. In recent years, researchers have been trying state-of-the-art algorithms to build bankruptcy prediction models, such as classification and regression tree recurrent neural network (CART-RNN) (Jan, 2021), ensemble methods (e.g., Random Forest) (Hsu and Lee, 2020), and Gradient Boosting using a high dimensional analysis (Jones, 2017).

Despite a vast body of research investigating models for going concern prediction and bankruptcy prediction (e.g., Koh, 1991; Hopwood et al., 1994; Gutierrez et al., 2020), audit firms do not systematically employ the models as a diagnostic tool in going concern assessments, according to conversations with practitioners and standard setters (Gutierrez et al., 2020). The most typical method for going concern assessment is to use a checklist that guides auditors to check the indicators suggested by standards step by step (Gutierrez et al., 2020). While there is a debate on the superiority of statistical models over auditors (Hopwood et al., 1994; Yeh et al., 2014; Barboza et al., 2017), in general, statistical models outperform the GCOs in providing early warnings of bankruptcy (e.g., Bellovary et al., 2007).

However, prior literature has not studied the comparative performance of statistical models and auditors in going concern assessment, i.e., when do auditors tend to make mistakes and when are models prone to errors. Furthermore, prior literature has not explored the potential for auditor-ML collaboration. This study sheds light on these matters. Thus, the purpose of this study is not to find the best possible model for assessing firms' going concerns but to examine the potential for collaboration between auditors and ML to improve the quality of going concern assessments.

3. Data and Sample

Following Gutierrez et al. (2020), we constructed our sample in the following steps. First, we obtained an intersection of COMPUSTAT, CRSP, and Audit Analytics from fiscal years 2003 to 2021. We chose the starting year to be 2003 to focus on the post-SOX period. The dataset ends in 2021, ensuring it encompasses the complete set of bankruptcy data up to 2022, the most recent year as of this research.[3] We utilize Chapter 7 or 11 bankruptcy as a proxy for substantial doubt about firms' existence as a going concern, following literature common practice (e.g., Berglund, 2020; Carson et al., 2013; Desai et al., 2020; Hardies et al., 2018; Xu and Kalelkar, 2020).[4] Since AS 2415 requires auditors to assess clients' going concerns within one year after the fiscal year

[3] The bankruptcy data was collected at the end of 2023 when the bankruptcy filing date was up to May 2023 in Audit Analytics, thus incomplete for 2023.

[4] Some literature also includes other forms of firm failures, such as debt defaults, debt restructurings, merger and acquisition, and delisting (Nogler, 1995; Carson et al., 2013; Gutierrez et al., 2020). For example, Gutierrez et al. (2020) utilize "default", which contains bankruptcy and bankruptcy alternatives. Using the default observations from 2003-2015 in Gutierrez et al. (2020) (we thank Gutierrez et al. (2020) for providing the data), we do not observe substantial discrepancy between bankruptcy and default observations (specifically, there are 262 bankruptcy and 268 default observations from 2003-2015). Thus, in this research, we directly utilize bankruptcy.

end, we did a precise matching based on bankruptcy begin date and fiscal year end date to ensure that we capture client firms entering into bankruptcy within one year following the date of the financial statements. Also, we made sure that the predictor variables (introduced in the section "Machine Learning Model Setup") provided to the ML models are available as of fiscal year end to avoid look-ahead bias. Next, we removed observations with missing values to compute the required variables. Lastly, we excluded firms not defined as distressed per Gutierrez et al. (2020) because auditors consider the issuance of GCO to distressed firms (AS 2415).[5] [6] In total, our dataset has 29,054 firm-year observations spanning from 2003 to 2021. The detailed sample derivation procedure is presented in Panel A, Table 1. Our sample contains 3,187 GCOs, and 381 bankruptcies within one year.

Panel B, Table 1 presents the sample distribution by years. We observe that the highest percentages of bankruptcy at t+1 occurred in the fiscal years 2007, 2008, and 2019. These peak years correspond to the onset of bankruptcies in 2008, 2009, and 2020, coinciding with the global financial crisis and the COVID-19 pandemic, consistent with documentations in the Cornerstone (2022) report.[7] Panel C, Table 1 presents the sample distribution by industry. Life science industry receives the highest percentage of GCO while the energy and transportation industry reports the highest bankruptcy percentage. Panel D, Table 1 presents a comparison of GCO against bankruptcy. Consistent with descriptive statistics from prior literature (e.g., Carson et al. 2013; Gutireez et al. 2020), there are 93.3 percent (2,972/3,187) of client firms received GCOs in the

[5] Gutierrez et al. (2020) define distressed firms as companies with either negative net income or negative operating cash flows.

[6] Unlike prior studies that drop observations from the financial industry, we made a conscious decision to keep these observations due to the increasing concern of firm failures in the financial industry. Dropping financial industry does not impact our results. Financial firms are defined as those firms having SICs between 6000 and 6999.

[7] Notably, the number of firms declaring bankruptcy surged after the World Health Organization (WHO) designated COVID-19 as a pandemic in March 2020, reaching its peak in July 2020 and continuing through the first quarter of 2021. From the second quarter of 2021 onward, the rate of bankruptcy filings fell below historical averages, largely because many firms had already filed for bankruptcy in 2020.

current year but did not enter into bankruptcy within one year, and 43.6 percent (166/381) of client firms went bankrupt within a year without receiving GCOs.

4. Establish Machine Learning Models for Going Concern Assessment

4.1. Machine learning model setup

In establishing ML models that assess whether substantial doubt exists for the client firm to continue as a going concern, we utilize a representative ML algorithm called Extreme Gradient Boosting (XGBoost) (e.g., Ding et al., 2020; Perols, 2011; Perols et al., 2017; Hunt et al., 2021).[8] XGBoost works by iteratively aggregating predictions from multiple simpler models (Sagi and Rokach, 2018). XGBoost is extensively applied across various sectors in both academic research and practical industry applications (Bentéjac et al., 2021; Chen and Guestrin, 2016), highlighting its effectiveness and reliability in handling complex predictive tasks.

The outcome variable of interest is $BANKRUPTCY$, which is a dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end, and zero otherwise. In terms of predictors, we follow Gutierrez et al. (2020) to use the following variables: ROA is return on assets. $LEVERAGE$ is the percentage of total liabilities over total assets. $WCAP$ is the percentage of working capital over total assets. $CURRENT$ is the percentage of current assets over current liabilities. $CASH$ is the percentage of cash and cash equivalents over total assets. CFO is the cash flow from operating activities over total assets. $SIZE$ is the natural logarithm of total assets. $NEGEQUITY$ is an indicator variable that equals to one if total liabilities exceed total assets, and zero otherwise. REL_MKTCP captures the relative market capitalization of the firm. EX_RET

[8] Details of each algorithm can be found in James et al. (2013). Although commonly used as an econometrics methodology, logistic regression is also a type of ML algorithm. It is often considered a basic ML algorithm and used as a benchmark.

measures firm return. $SIGMA$ represents the firm's volatility.[9] Table 2 presents descriptive statistics of these predictor variables. In general, they are comparable to prior literature (Gutierrez et al., 2020). Table 2 also presents descriptive statistics of other variables which will be introduced in later sections. Variable definitions are included in Appendix A.

Next, we follow prior studies to adopt a rolling-window design to train and test ML models (Bao et al., 2020; Brown et al., 2020). We organize our dataset into six sequential batches. The initial batch utilizes data from fiscal years 2003 to 2009 for training, with the year 2010 designated as the out-of-sample testing period. Starting with the base year of 2003 for training, each subsequent batch advances one year forward for training and testing – meaning, after the first batch, we train on data from 2003 to 2010 and test on data from 2011, and so on. [10]

4.2. Economic cost imbalance

In assessing whether substantial doubt exists for a client firm to continue as a going concern within one year beyond the financial statement being audited, it is necessary to consider the economic cost imbalance between false positive and false negative errors. In going concern assessments, false negative errors are, in general, more costly than false positive errors because failing to provide timely warnings of a client firm that go bankrupt within one year may result in high litigation costs and investor loss (Carson et al., 2013). We define a *misclassification cost* as the cost ratio between false negative errors and false positive errors (Perols, 2011). Given the limited exploration of misclassification costs in existing literature on going concern assessments,

[9] Note that the purpose of this study is not to find the best possible model for assessing firms' going concerns. While adding more predictors could likely improve ML performance, our study uses a simpler ML model to establish a baseline. This means that our current ML setup represents a conservative estimate of ML's capabilities in this context.
[10] When constructing the models, we opt for the default hyperparameters, which are the settings or configurations that govern the behavior of the learning algorithm but are not learned from the data. These parameters need to be set before the learning process begins. For XGBoost, an example of a hyperparameter is the "max_depth," which determines the maximum depth of the trees used in the model. This choice simplifies the process and enhances the replicability of our findings. With default parameters, the overall predictive power of the XGBoost model is already reaching at about 92% of ROC-AUC, leaving little room for incremental improvement even if hyperparameters are tuned.

we test a series of cost ratios (10, 20, 30, 40, and 50), similar to Perols (2011), to determine the most appropriate one. To do so, we use the last year of training data as a hold-out validation set while adjusting the rest of the training data to reflect different levels of misclassification costs. For each level, we measure performance on the hold-out validation set, identifying the cost ratio that maximizes predictive accuracy. Once the optimal ratio is determined, we apply it across the entire training dataset and assess the final model performance on the separate, untouched testing period.[11]

4.3. Evaluation metrics

We adopt multiple complementary evaluation metrics to assess model performance, including recall, precision, F-score, the Area under the Receiver Operating Characteristic curve (ROC-AUC), the Area Under the Precision-Recall Curve (PR-AUC), and the expected cost of misclassification (ECM). Details of these measures are provided in Appendix B. The higher the recall, precision, and F-score, the better the model performs.[12] The value of ROC-AUC ranges from 0 to 1, with value below 0.5 considered worse than random guessing, 0.5 to 0.7 as poor performance, 0.7 to 0.8 as acceptable, 0.8 to 0.9 as excellent, and 0.9 to 1 as outstanding (Hosmer et al. 2013; Mandrekar, 2010). The value of PR-AUC ranges from 0 to 1, the higher the better. The baseline of PR-AUC is the ratio of positives to the total number of observations (Saito and Rehmsmeier, 2015). Compared to ROC-AUC, PR-AUC can better reflect model performance

[11] To integrate misclassification costs into our model training, we employ a cost-sensitive learning approach that adjusts the balance between positive (bankruptcy) and negative (non-bankruptcy) cases in the training data, as detailed in prior studies (Perols et al., 2017). Specifically, we utilize the Observation Under-sampling (OU) method from Perols et al. (2017), which generates multiple subsets of the training data. Each subset contains all instances where bankruptcy occurred and an equal number of non-bankruptcy instances selected at random. If the training set includes P positive and N negative instances, the OU method creates subsets each with P positive and P randomly chosen negative instances, leading to N/P under-sampled subsets. This under-sampling is applied solely to the training data, leaving the validation and testing datasets unchanged. We train a model using each subset and average their out-of-sample performance.

[12] In classifying the probabilistic ML prediction into a binary assessment, we use the conventional 50 percent as the classification threshold. Results hold when we choose alternative classification thresholds (i.e., 40 percent or 60 percent).

when class imbalance exists in the dataset (Krupa and Minutti-Meza, 2022). ECM incorporates the economic costs associated with false positive and false negative errors. The lower the ECM, the more cost-effective. Replacing ML assessment with GCO, we can calculate the performance metrics for GCO. However, note that since GCO is binary, we can only generate recall, precision, F-score, and ECM for GCO as ROC-AUC and PR-AUC requires probabilistic prediction.

5. Results

5.1. ML assessment vs GCO

Panel A, Table 3 presents the comparison of performance between ML and GCO in assessing going concerns in the hold-out testing period of 2010 to 2021.[13] We observe that ML assessments have a higher recall of 0.5420 compared to 0.5294 for GCO, indicating that ML assessments can provide early warnings to more firms that actually go bankrupt within a year than GCOs. Additionally, ML assessments demonstrate a significantly higher precision (0.1514) than GCO (0.0569), indicating that ML is more precise in its assessments. The F-score, which balances recall and precision, is also higher for ML (0.2367) than for GCO (0.1028), suggesting that ML provides a more balanced performance between recall and precision. Lastly, ML has a lower ECM of 0.1036 compared to 0.1788 for GCO, indicating that ML assessments are more cost-effective than GCO. Additionally, ML reaches an ROC-AUC of 0.9176, considered "outstanding" per academic standards (Hosmer et al. 2013; Mandrekar, 2010). Turning to PR-AUC, ML assessments reaches 0.2101, equivalent 20 times better than a random guess given that the rate of bankruptcy in population is about 1 percent.

[13] Based on our analysis of the hold-out validation set, we have determined that a misclassification cost ratio of 10 yields the most balanced and optimal performance overall. Further details about this selection process can be found in Appendix C. Moving forward, we train the ML model using this chosen misclassification cost ratio of 10. This trained model is then be applied to generate predictions for the out-of-sample testing period covering the years 2010 to 2021, which has not been previously seen by our model.

5.1.1. Identifying false positive and false negative errors of ML and GCO

Previous studies often interpret any discrepancy between bankruptcy occurrences and the issuance of GCO as errors in GCO issuance (e.g., Berglund, 2020; Carson et al., 2013; Desai et al., 2020; Xu and Kalelkar, 2020). This interpretation, however, may not account for instances where auditors have made careful judgments with the information available at the fiscal year-end. For instance, there could be situations where a GCO was properly issued at the date of audit report, yet the firm managed to survive due to unexpected changes in circumstances. In contrast, there might be cases where no GCO was issued, but unforeseen events later caused the firm to declare bankruptcy. By using ML as a benchmark, we can exclude scenarios where auditors might not necessarily be held accountable. This is because an effective ML model would likely have made the same decision under those circumstances.

To illustrate, we provide Panel B, Table 3 that tabulates scenarios of bankruptcy, ML assessment, and GCO. In scenario 4, both ML and auditor express substantial doubt about client firms' going concerns whereas the client firm did not enter into bankruptcy within one year from the date of financial statement. In scenario 5, both ML and auditor give a clean opinion but the firm filed for bankruptcy within one year after fiscal year end. We posit that although in both of these situations, auditor assessments do not align with the actual outcome of bankruptcy, they might not be directly considered as auditor error because the ML model would have made the same decision. Thus, we refine measures of auditor error by focusing on scenario 2 and 7.[14]

We regard scenario 2 as GCO false positive error since auditors issued GCO but ML does not, and the client firm did not enter into bankruptcy within the next fiscal year. Similarly, we consider scenario 7 as GCO false negative error since auditors didn't issue GCO but ML did, and

[14] Our results presented in Section "What Factors Are Associated with Auditor GCO Error?" and "What Factors Are Associated with ML Error?" are similar but weaker after including these two scenarios.

the client entered into bankruptcy within the next fiscal year. In the same vein, we regard scenario 3 as ML false positive error and scenario 6 as ML false negative error. In the next sections, we explore what factors are associated with these refined measures of GCO and ML errors.

5.1.2. What factors are associated with auditor GCO error?

Using scenarios identified above of GCO false positive and false negative errors, we examine what factors are associated with these errors. Based on the theories of cognitive constraints and bounded rationality (e.g., Liu, 2022), we use Model (1) to examine how factors that represent auditors' capability, cognitive bias, and constraints are related to their errors in assessing client firms' going concerns.

$$\begin{aligned} &GCO\_FP\_ERROR \text{ or } GCO\_FN\_ERROR \\ &\quad = b_0 + b_1 AUDITOFFICEEXPERTISE + b_2 AUDITOFFICESIZE + b_3 BIG4 + b_4 \\ &\quad NONAUDITFEERATIO + b_5 INFLUENCE + b_6 TENURE + b_7 NEWCLIENT + b_8 \\ &\quad ANCHORINGGCO + b_9 ANCHORINGNOGCO + b_{10} BUSY + b_{11} \\ &\quad WORKLOADCOMPRESSION + Industry\ Fixed\ Effects + Year\ Fixed\ Effects \\ &\quad + \varepsilon \qquad (1) \end{aligned}$$

GCO_FP_ERROR is a dummy variable that equals one for scenario 2 presented in Panel B, Table 3, zero otherwise. GCO_FN_ERROR is a dummy variable that equals one for scenario 7 presented in Panel B, Table 3, zero otherwise.[15] The variables that represent auditors' capability include office-level industry expertise ($AUDITOFFICEEXPERTISE$), office size ($AUDITOFFICESIZE$), and Big4 brand name ($BIG4$). We use the following variables to capture auditors' cognitive bias: non-audit fee ratio ($NONAUDITFEERATIO$), client influence ($INFLUENCE$), tenure ($TENURE$), new client ($NEWCLIENT$), and anchoring bias ($ANCHORINGGCO$ and $ANCHORINGNOGCO$). Anchoring bias refers to auditors' tendency to rely on previous years' opinions when issuing current year's GCO. Anchoring to a prior year's

[15] Scenarios 4 and 5 presented in Panel B, Table 3 are excluded in this analysis.

clean opinion is labeled as $ANCHORINGNOGCO$ and anchoring to a prior year's non-clean opinion is labeled as $ANCHORINGGCO$. Lastly, variables that represent auditors' cognitive constraints include whether the audit is done in busy season ($BUSY$) and workload compression ($WORKLOADCOMPRESSION$). The variables are defined in Appendix A and descriptive statistics are provided in Table 2. We discuss our expectations on how these variables are associated with auditor GCO errors in Appendix D.

We estimate Model (1) using Ordinary Least Squares (OLS) regression.[16] Industry and year fixed effects are included and standard errors are clustered by client firm. Panel A, Table 4 presents the results. We observe that the higher industry expertise and the larger the audit office, the lower the GCO false positive error and the higher the false negative error, suggesting that auditors with more city-level industry expertise and larger office size tend to be less conservative, thus issuing less GCOs, decreasing false positive error and increasing false negative error.[17] In other words, audit offices with lower industry expertise and smaller sizes tend to be more conservative by issuing more GCO.

Big Four auditors are associated with lower both types of errors, consistent with the Big Four effect where they provide higher-quality audits (e.g., DeFond et al., 2017). Additionally, a negative association exists between non-audit fee ratio and GCO false negative error, consistent with spillover effects from non-audit services (Lim and Tan, 2008; Wu, 2006). Furthermore, client influence has a negative association with GCO false positive error and a positive association with GCO false negative error, in line with the notion that when independence of the auditor is hampered, auditors are less likely to issue GCO. Moreover, auditors' tendencies to align with

[16] Using logistic regression generates similar results. However, sample size will be smaller due to dropped singletons.
[17] In un-tabulated analysis, we confirm that, all else equal, higher audit office industry expertise and larger audit office size reduces the chance of GCO issuance.

previous years' opinions significantly relate to GCO errors – anchoring to non-clean opinions increases false positives, while anchoring to clean opinions decreases false positives and raises false negatives. Lastly, we observe that audits conducted during busy season have a lower chance of GCO false negative error, suggesting that auditors tend to be more prudent during busy season.

In sum, this section shows that GCO error is associated with multiple factors related to auditors' capability, cognitive bias, and constraints.[18] Next, we examine what factors are associated with ML errors.

5.1.3. What factors are associated with ML error?

Using scenarios identified above of ML false positive and false negative errors, we examine what factors are associated with these errors. We posit that ML is prone to false negative errors when it fails to capture unobservable factors – such as proprietary and soft information – that indicate a firm's risk. On the other hand, ML tends to generate false positive errors in scenarios where these unobservable factors would suggest that issuing a non-clean opinion is more justified. We use Model (2) to examine how factors that represent unobservable information might be associated with ML error.

$$\begin{aligned} ML\_FP\_ERROR\ or\ ML\_FN\_ERROR \\ = b_0 + b_1 UNOBSERVABLEISSUE + b_2 UNOBSERVABLENOTISSUE + b_3 \\ AUDITREPORTLAGCHANGE + b_4 AUDITREPORTLENGTHCHANGE \\ + Industry\ Fixed\ Effects + Year\ Fixed\ Effects + \varepsilon \qquad (2) \end{aligned}$$

[18] We provide further evidence on the difference between auditors and ML in their decision-making process. Humans have limited ability to comprehend high dimensional information and thus tend to process information in a more linear fashion (Kahneman et al., 2021). In contrast, ML purportedly excels at identifying complex nonlinear relationship and interactions among variables. To illustrate the difference in the ways auditors and ML process information, we follow Liu (2022) and regress GCO and ML assessments on bankruptcy determinants. We expect to see that the R squared is higher when the dependent variable is GCO than when the dependent variable is ML prediction, indicating that auditors process information following a linear manner. Consistent with our expectation, we observe that the same set of variables explain more variation in GCO (adjusted R squared 0.3264) than in ML prediction (adjusted R squared 0.2513), suggesting that auditors process information in a more linear way than ML.

ML_FP_ERROR is a dummy variable that equals one for scenario 3 presented in Panel B, Table 3, zero otherwise. ML_FN_ERROR is a dummy variable that equals one for scenario 6 presented in Panel B, Table 3, zero otherwise.[19] We include a variety of proxies of unobservable information. First, we follow Liu (2022) to establish a separate ML model to estimate the chance of auditors issuing GCO. Unlike our main ML model that predicts bankruptcy, this model predicts auditors' decision to issue GCOs. Details of this model are provided in Appendix E. The probability prediction from this model represents an auditor's likelihood of issuing GCO, and it represents a counterfactual scenario of the auditor's decision had she only considered observable hard information without impact of proprietary and soft information, such as private communications with the managers. Like Liu (2022), we take the difference between the actual GCO and the likelihood of GCO to represent the impact of unobservable information on auditors' decision to issue GCO.

Given that the real-world issuance of GCO is a binary outcome, whereas the model's probability prediction is a continuous value ranging between 0 and 1, the resulting difference can be either negative or positive. Negative discrepancies occur when no GCO is issued (GCO is 0), and positive discrepancies occur when a GCO is issued (GCO is 1). We interpret the absolute value of positive discrepancies (labelled as $UNOBSERVABLEISSUE$) as the degree to which unobservable information influences auditors to issue a GCO. We expect a positive correlation between $UNOBSERVABLEISSUE$ and ML false negative errors because ML does not have access to such unobservable information that indicates the necessity to issue non-clean going concern opinion. On the contrary, we label the absolute value of negative discrepancies as $UNOBSERVABLENOTISSUE$, indicating the extent to which such information leads auditors

[19] Scenarios 4 and 5 presented in Panel B, Table 3 are excluded in this analysis.

against issuing a GCO. Because the ML model lacks access to this unobservable information, we anticipate a positive correlation between $UNOBSERVABLENOTISSUE$ and the ML model's false positive errors.

We further assess the impact of unobservable information by examining changes in the audit report's timeliness (i.e., change of audit report lag, $AUDITREPORTLAGCHANGE$) and changes in the length (number of sentences) of the audit report ($AUDITREPORTLENGTHCHANGE$). These changes reflect the auditors' increased interactions with their clients during audits. A significant increase in both the time taken to complete the audit report (audit report lag) and the detail provided within the report (audit report length) suggests that auditors are investing more effort in discussions with the client and are incorporating more information about the client's situation into the audit report. Therefore, we expect a positive correlation between the changes in audit report lag and length with both types of ML errors, indicating that these metrics capture the extent of critical, yet unobservable, information that influences the auditors' risk assessment. The variables are defined in Appendix A and descriptive statistics are provided in Table 2.

We estimate Model (2) using OLS regression. Industry and year-fixed effects are included and standard errors are clustered by client firm. Panel B, Table 4 reports the results. Consistent with our expectations, we observe that unobservable information that contains signals of "should issue GCO" ($UNOBSERVABLEISSUE$) is significantly and positively associated with ML false negative errors; in contrast, unobservable information that indicates "no need to issue GCO" ($UNOBSERVABLENOTISSUE$) is significantly and positively associated with ML false positive errors. Furthermore, the more additional interactions between the auditor and the client, reflected in greater change of audit report lag ($AUDITREPORTLAGCHANGE$) and audit report length

($AUDITREPORTLENGTHCHANGE$), the more the ML false negative error, consistent with our expectation that ML falls short in utilizing the unobservable information in situations like this to make better decisions.

Overall, we conclude that ML tends to make mistakes in situations where there is useful unobservable information that ML does not have direct access to but is available to auditors. In the next section, we examine whether it is possible to synergize ML and auditors to produce higher-quality going concern assessments.

5.2. Synergizing ML and auditor

In the previous section, we find that auditors' GCO errors are associated with factors related to auditors' capability, cognitive bias, constraints, whereas ML makes more mistakes when useful unobservable information exists that it cannot directly access. In this section, we explore the potential of an auditor-ML collaboration by integrating ML's competitive advantage for pattern recognition and auditors' competitive advantage of accessing proprietary information and processing soft information.

To mimic the scenario of auditor-ML collaboration in assessing going concerns, we directly input GCO as an additional predictor variable in the ML model that predicts next year bankruptcy. This approach is justified because the GCO represents auditor's final judgment after incorporating all necessary information, including proprietary and soft information that is unobservable to ML (Gutierrez et al., 2020). Moreover, including GCO as a predictor makes it empirically feasible to test the potential of auditor-ML collaboration. We term the prediction from the ML model with GCO as an additional predictor variable the "team assessment". We find that adding GCO as an additional predictor further improves performance metrics as shown in the footnote[20], consistent with the conclusion that GCO contains useful proprietary and soft information in assessing going concerns (Gutierrez et al., 2020). Next, we explore whether the team prediction can synergize the relative advantages of auditor and ML, and overcome their relative disadvantages.[21]

5.2.1. Can team assessments overcome auditor disadvantage?

In Section "What Factors Are Associated with Auditor GCO Error", we conclude that audit offices that have lower industry expertise and that are smaller tend to be more conservative by issuing more GCO. Additionally, when the client has a higher influence on the auditor, the auditor tends not to issue GCO. Furthermore, when auditors anchor their current year opinion on the previous year's, they are more likely to make mistakes. In this section, we explore whether team assessment is more informative than GCO in these situations. We utilize the model specified in Model (3) to examine the relative performance of team assessment and GCO in assessing going concerns.

$$BANKRUPTCY = b_0 + b_1 TEAM + b_2 GCO + \boldsymbol{bCONTROLS} + Industry\ Fixed\ Effects + Year\ Fixed\ Effects + \varepsilon \quad (3)$$

$BANKRUPTCY$ is a dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end, and zero otherwise. GCO is a dummy variable that equals one if client firm i receives a going concern opinion for fiscal year t, zero otherwise. $TEAM$ is a dummy variable that equals one if the probability prediction from the ML model with GCO as an additional predictor is greater than 0.5, zero otherwise.[22] CONTROLS contains all predictor

[20] Compared to the ML model without GCO as an additional predictor, the one with GCO improves recall from 0.5420 to 0.5462; precision from 0.1514 to 0.1563; F-score from 0.2367 to 0.2430; ROC-AUC from 0.9176 to 0.9217; and PR-AUC from 0.2101 to 0.2295. And it reduces ECM from 0.1036 to 0.1019.

[21] Note that in the following analysis, since we are no longer interested in identifying false positive and negative errors of auditor or ML, we do not drop scenarios 4 and 5 presented in Panel B, Table 3.

[22] Inference remains similar if directly using the probability prediction from the ML model with GCO as an additional predictor (untabulated).

variables we used in the ML model to predict next year bankruptcy. We estimate Model (3) using OLS regression. Industry and year fixed effects are included and standard errors are clustered by client firm.

Table 5 reports results of applying Model (3) to four different subsamples: Column (1) presents results of Model (3) for the subsample with lower audit office industry expertise, defined as audit office industry expertise ($AUDITOFFICEEXPERTISE$) lower than the median; Column (2) presents results for subsample with smaller audit office size, defined as audit office size ($AUDITOFFICESIZE$) smaller than median; Column (3) presents results when anchoring bias exists, defined as either $ANCHORINGGCO$=1 or ANCHORINGNOGCO=1[23]; Column (4) presents results for subsample when the client firm has higher influence on the auditor, defined as influence ($INFLUENCE$) greater than median.

In all columns, we observe that $TEAM$ is significantly and positively associated with $BANKRUPTCY$ and that it has significantly higher coefficient than GCO, consistent with our expectation that team assessments are more informative than GCO in situations where auditors may be more conservative, impacted by previous years' decisions, or influenced by clients.

5.2.2. Can team assessments overcome ML disadvantage?

In this section, we examine whether team assessments are better than ML assessments in situations where ML alone has limited capability – that is, when useful unobservable information exists that ML does not have direct access to. We use Model (4) to examine the relative performance of team assessments and ML assessments.

$$BANKRUPTCY = b_0 + b_1 TEAM + b_2 ML + \boldsymbol{bCONTROLS} + Industry\ Fixed\ Effects + Year\ Fixed\ Effects + \varepsilon \qquad (4)$$

[23] We cannot separate the two as ANCHORINGGCO=1 perfectly correlates with GCO=1 and ANCHORINGNOGCO=1 perfectly correlates with GCO=0.

Compared to Model (3), in Model (4), we replace GCO with ML, which is a dummy variable that equals one if the probability prediction from the ML model is greater than 0.5, zero otherwise. [24] Model (4) is estimated using OLS regression. Industry and year fixed effects are included and standard errors are clustered by client firm.

Table 6 reports results of running Model (4) on three different subsamples: Column (1) presents results of Model (4) for the subsample with high level of unobservable information, defined as the difference between actual GCO and likelihood prediction of GCO ($UNOBSERVABLE$) at the 1st decile or the 9th decile[25]; Column (2) and (3) present results when the audit report lag change is large ($AUDITREPORTLAGCHANGE$ at the top decile) and when the change of audit report length is large ($AUDITREPORTLENGTHCHANGE$ at the top decile).

In all columns, we observe that when there exists potentially useful unobservable information not available to ML, ML assessment lost its informativeness to provide early warnings to next-year bankruptcy, whereas team prediction, which supposedly synergize ML and auditor, remains informative. To conclude, we show that team prediction continues to be effective where ML alone lost its ability to properly assess client firms' going concerns.

Taken together, this section illustrates that auditor-ML collaboration has the potential of generating higher quality going concern assessments that can compensate either party in situations where their ability is limited.[26]

---

[24] Inference remains similar if directly using the probability prediction from the ML model (untabulated).

[25] We exclude the 10th decile because the Appendix E show that this rank of unobservable information more likely captures auditors' conversative tendency of issuing GCO than the proprietary or soft information that has more useful signal that could have been provided to ML by auditor.

[26] In un-tabulated analysis, we explore the additional benefit of having a continuous probability prediction that ranges from 0 to 1 rather than a binary assessment. We posit that the probability prediction not only maintains the original information of assessment of firms' going concerns, but also is not dependent on any threshold (e.g., 50 percent) based on which a binary value is derived. To test this proposition, we utilize team assessment and perform a head-to-head comparison of the binary version of the team assessment and the probabilistic version of the assessment. We observe that the coefficient of the probabilistic team prediction is significantly larger than that of the binary prediction. Including both versions of the prediction, we observe that the information in the binary version is subsumed by its

5.2.3. The need of auditor during macroeconomic instability

Since our hold-out testing period spans from fiscal years 2010 to 2021, we have a unique opportunity to explore the comparative effectiveness of auditor judgment and ML assessments during the economic volatility triggered by the COVID-19. The pandemic was officially declared by the World Health Organization on March 13, 2020, prompting us to utilize the data from fiscal years 2020 to 2021 as a distinct subsample influenced by this global event. The effectiveness of ML relies on the premise that past data, on which it is trained, can reliably forecast future outcomes (Alpaydin, 2020). However, the unprecedented circumstances of the COVID-19 pandemic disrupt this premise, as they introduce new factors that significantly alter the business environment. ML models, due to their reliance on historical data, may struggle to adapt to such rapid changes, unlike auditors who can immediately incorporate new information into their assessments. Therefore, we anticipate a reduction in the effectiveness of ML assessments during the period affected by the pandemic, while expecting auditor-issued GCO to retain their relevancy. Additionally, we expect that the team assessment will maintain its efficacy, as it benefits from the strengths of both auditors and ML.

Table 7 presents the results of the effectiveness of GCO, ML assessment, and team assessment in evaluating client firms' going concerns during COVID-19 period. Consistent with our expectations, the results in Column (1) indicate that the ML assessments alone do not show a significant correlation with the future bankruptcy of the client firms. This suggests that ML's predictive capability was compromised during the pandemic. However, Columns (2) and (3) show that auditors' GCO and the collaborative team's assessment continue to be informative of potential client bankruptcy, especially GCO.

---

probability counterpart. Overall, this finding suggests that utilizing probabilistic assessment can offer additional power to evaluate firms' going concerns.

These findings highlight the critical role of auditor judgment under circumstances where ML falls short, specifically in its inability to adjust to sudden changes in the macroeconomic environment brought on by the pandemic. The sustained effectiveness of the combined team assessment, even in such challenging times, further shows the potential of integrating auditor insights with ML predictions.

6. Concluding Remarks

We examine the potential for auditor and ML collaboration in assessing going concerns. We show that auditors' capability, cognitive bias, and constraints are associated with errors in GCOs, and ML tends to make mistakes when unobservable information exists that cannot be directly incorporated into ML models. To demonstrate auditor-ML collaboration, we construct a "team" assessment that integrates auditors and ML. Our results show that the team assessments can synergize the relative advantages of auditor and ML, and overcome their relative disadvantages, serving a potential way of generating higher quality going concern assessments.

This study is subject to a few limitations. First, the list of factors that are associated with GCO error or ML error may not be complete. Future research can continue exploring conditions under which either auditors or models are more prone to errors. Second, we utilize an archival approach to derive team assessment. Future research can consider using field experiments to gauge the quality of assessments from actual auditor-ML collaborations.

We add to the accounting literature by exploring the interplay between ML and auditor judgment in enhancing auditors' decision-making quality. Amid rising audit deficiencies and efforts by the regulators to improve audit quality, our research provides critical insights into how ML can augment auditor decision-making processes. We also advance the discussion on the role of auditor judgment in issuing GCO by using ML as a benchmark, allowing us to delve into the limitations but also advantages of auditor judgment. Furthermore, we introduce a novel method demonstrating the benefits of combining auditors with ML, suggesting that a collaborative approach can effectively address the limitations of each method working in isolation. This collaborative strategy has the potential to enhance the accuracy and reliability of going concern assessments, offering a promising direction for future audit practice enhancements.

References

Alpaydin, E., 2020. Introduction to machine learning. MIT Press.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finance. 23(4), 589-609. https://doi.org/10.2307/2978933.

AS 2415: Consideration of an Entity's Ability to Continue as a Going Concern. https://pcaobus.org/oversight/standards/auditing-standards/details/AS2415 (accessed 11 February 2025).

Balsam, S., Krishnan, J., Yang, J.S., 2003. Auditor industry specialization and earnings quality. Audit. J. Pract. Theory. 22(2), 71-97. https://doi.org/10.2308/aud.2003.22.2.71.

Bao, Y., Ke, B., Li, B., Yu, Y.J., Zhang, J., 2020. Detecting accounting fraud in publicly traded US firms using a machine learning approach. J. Account. Res. 58(1), 199-235. https://doi.org/10.1111/1475-679X.12292.

Barboza, F., Kimura, H., Altman, E., 2017. Machine learning models and bankruptcy prediction. Expert Syst. Appl. 83, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006.

Barnes, P., 2004. The auditor's going concern decision and Types I and II errors: The Coase Theorem, transaction costs, bargaining power and attempts to mislead. J. Account. Public Policy. 23(6), 415-440. https://doi.org/10.1016/j.jaccpubpol.2004.10.003.

Bazerman, M.H., Morgan, K.P., Loewenstein, G.F., 1997. The impossibility of auditor independence. MIT Sloan Manag. Rev.

Becker, C.L., DeFond, M.L., Jiambalvo, J., Subramanyam, K.R., 1998. The effect of audit quality on earnings management. Contemp. Account. Res. 15(1), 1-24. https://doi.org/10.1111/j.1911-3846.1998.tb00547.x.

Bell, T.B., Causholli, M., Knechel, W.R., 2015. Audit firm tenure, non-audit services, and internal assessments of audit quality. J. Account. Res. 53(3), 461-509. https://doi.org/10.1111/1475-679X.12078.

Bellovary, J.L., Giacomino, D.E., Akers, M.D., 2007. A review of going concern prediction studies: 1976 to present. J. Bus. Econ. Res. 5(5). https://doi.org/10.19030/jber.v5i5.2541.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54, 1937-1967.

Berglund, N.R., 2020. Do client bankruptcies preceded by clean audit opinions damage auditor reputation? Contemp. Account. Res. 37(3), 1914-1951.

Brown, N.C., Crowley, R.M., Elliott, W.B., 2020. What are you saying? Using topic to detect financial misreporting. J. Account. Res. 58(1), 237-291. https://doi.org/10.1111/1475-679X.12294.

Brunelli, S., 2018. Audit reporting for going concern uncertainty: Global trends and the case study of Italy. Springer.Carcello, J. V., A. Vanstraelen, and M. Willenborg. 2009. Rules rather than discretion in audit standards: Going-concern opinions in Belgium. The Accounting Review 84(5): 1395-1428. https://doi.org/10.2308/accr.2009.84.5.1395

Carpenter, B.W., Dirsmith, M.W., 1992. Early debt extinguishment transactions and auditor materiality judgments: A bounded rationality perspective. Account. Organ. Soc. 17(8), 709-739.

Carson, E., Fargher, N.L., Geiger, M.A., Lennox, C.S., Raghunandan, K., Willekens, M., 2013. Audit reporting for going-concern uncertainty: A research synthesis. Audit. J. Pract. Theory. 32, 353-384. https://doi.org/10.2308/ajpt-50324.

Cao, S., Jiang, W., Wang, J., Yang, B., 2024. From man vs. machine to man + machine: The art and AI of stock analyses. J. Financ. Econ. 160, 103910. https://doi.org/10.1016/j.jfineco.2024.103910.

Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. Rev. Finance. 8(4), 537-569. https://doi.org/10.1093/rof/8.4.537.

Che, L., Hope, O.K., Langli, J.C., 2020. How Big-4 firms improve audit quality. Manag. Sci. 66(10), 4552-4572.

Chen, K.C., Church, B.K., 1996. Going concern opinions and the market's reaction to bankruptcy filings. Account. Rev., 117-128.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 785-794.

Choi, J.H., Kim, C., Kim, J.B., Zang, Y., 2010. Audit office size, audit quality, and audit pricing. Audit. J. Pract. Theory. 29(1), 73-97. https://doi.org/10.2308/aud.2010.29.1.73.

Coleman, B., Merkley, K., Pacelli, J., 2022. Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations. Account. Rev. 97(5), 221-244. https://doi.org/10.2308/TAR-2020-0096.

Cornerstone, 2022. Trends in large corporate bankruptcy and financial distress. https://www.cornerstone.com/wp-content/uploads/2022/09/Trends-in-Large-Corporate-Bankruptcy-and-Financial-Distress-Midyear-2022-Update.pdf (accessed 11 February 2025).

Curtis, M.B., 2006. Are audit-related ethical decisions dependent upon mood? J. Bus. Ethics. 68, 191-209. https://doi.org/10.1007/s10551-006-9066-9.

Custodio, B.C.H., Gu, Y., González, P.J., 2022. Decision tree tool for auditors’ going concern assessment in Spain. Int. J. Digit. Account. Res. 22, 193-226. https://doi.org/10.4192/1577-8517-v22_7.

Desai, V., Desai, R., Kim, J.W., Raghunandan, K., 2020. Are going-concern issues disclosed in audit reports associated with subsequent bankruptcy? Evidence from the United States. Int. J. Audit. 24(1), 131-144. https://doi.org/10.1111/ijau.12183.

DeFond, M., Erkens, D.H., Zhang, J., 2017. Do client characteristics really drive the Big N audit quality effect? New evidence from propensity score matching. Manag. Sci. 63(11), 3628-3649. https://doi.org/10.1287/mnsc.2016.2528.

DeFond, M.L., Lennox, C.S., 2011. The effect of SOX on small auditor exits and audit quality. J. Account. Econ. 52(1), 21-40. https://doi.org/10.1016/j.jacceco.2011.03.002.

Ding, K., Lev, B., Peng, X., Sun, T., Vasarhelyi, M.A., 2020. Machine learning improves accounting estimates: Evidence from insurance payments. Rev. Account. Stud. 25, 1098-1134. https://doi.org/10.1007/s11142-020-09546-9.

Eshleman, J.D., Guo, P., 2014. Do Big 4 auditors provide higher audit quality after controlling for the endogenous choice of auditor? Audit. J. Pract. Theory. 33(4), 197-219. https://doi.org/10.2308/ajpt-50792.

Francis, J.R., 2006. Are auditors compromised by nonaudit services? Assessing the evidence. Contemp. Account. Res. 23(3), 747-760. https://doi.org/10.1506/4VD9-AE3K-XV7L-XT07.

Francis, J.R., Michas, P.N., Yu, M.D., 2013. Office size of Big 4 auditors and client restatements. Contemp. Account. Res. 30(4), 1626-1661. https://doi.org/10.1111/1911-3846.12011.

Francis, J.R., Yu, M.D., 2009. Big 4 office size and audit quality. Account. Rev. 84(5), 1521-1552. https://doi.org/10.2308/accr.2009.84.5.1521.

Geiger, M.A., Gold, A., Wallage, P., 2021. Auditor going concern reporting: A review of global research and future research opportunities. Routledge.

Geiger, M.A., Raghunandan, K., Riccardi, W., 2014. The global financial crisis: US bankruptcies and going-concern audit opinions. Account. Horiz. 28(1), 59-75. https://doi.org/10.2308/acch-50659.

Geiger, M.A., Rama, D.V., 2006. Audit firm size and going-concern reporting accuracy. Account. Horiz. 20(1), 1-17. https://doi.org/10.2308/acch.2006.20.1.1.

Gramling, A.A., Krishnan, J., Zhang, Y., 2011. Are PCAOB-identified audit deficiencies associated with a change in reporting decisions of triennially inspected audit firms? Audit. J. Pract. Theory. 30(3), 59-79. https://doi.org/10.2308/ajpt-10048.

Gul, F.A., Fung, S.Y.K., Jaggi, B., 2009. Earnings quality: Some evidence on the role of auditor tenure and auditors' industry expertise. J. Account. Econ. 47(3), 265-287. https://doi.org/10.1016/j.jacceco.2009.03.001.

Gutierrez, E., Krupa, J., Minutti-Meza, M., Vulcheva, M., 2020. Do going concern opinions provide incremental information to predict corporate defaults? Rev. Account. Stud. 25, 1344-1381. https://doi.org/10.1007/s11142-020-09544-x.

Hardies, K., Breesch, D., Branson, J., 2016. Do (fe)male auditors impair audit quality? Evidence from going-concern opinions. Eur. Account. Rev. 25(1), 7-34. https://doi.org/10.1080/09638180.2014.921445.

Hardies, K., Vandenhaute, M.L., Breesch, D., 2018. An analysis of auditors' going-concern reporting accuracy in private firms. Account. Horiz. 32(4), 117-132. https://doi.org/10.2308/acch-52297.

Hopwood, W., McKeown, J.C., Mutchler, J.F., 1994. A reexamination of auditor versus model accuracy within the context of the going-concern opinion decision. Contemp. Account. Res. 10(2), 409-431. https://doi.org/10.1111/j.1911-3846.1994.tb00400.x.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression, Vol. 398. John Wiley & Sons.

Hossain, S., Chapple, L., Monroe, G.S., 2018. Does auditor gender affect issuing going-concern decisions for financially distressed clients? Account. Finance. 58(4), 1027-1061. https://doi.org/10.1111/acfi.12242.

Hossain, M., Raghunandan, K., Rama, D.V., 2020. Abnormal disclosure tone and going concern modified audit reports. J. Account. Public Policy. 39(4), 106764. https://doi.org/10.1016/j.jaccpubpol.2020.106764.

Hsu, Y.F., Lee, W.P., 2020. Evaluation of the going-concern status for companies: An ensemble framework-based model. J. Forecast. 39(4), 687-706. https://doi.org/10.1002/for.2653.

Hunt, J.O., Rosser, D.M., Rowe, S.P., 2021. Using machine learning to predict auditor switches: How the likelihood of switching affects audit quality among non-switching clients. J. Account. Public Policy. 40(5), 106785. https://doi.org/10.1016/j.jaccpubpol.2020.106785.

International Auditing and Assurance Standards Board (IAASB), 2024. Audit evidence. https://www.iaasb.org/iaasb-consultations/latest-our-projects#:~:text=Audit%20Evidence&text=The%20key%20matters%20discussed%20up,in%20audits%20of%20financial%20statements. (accessed 11 February 2025).

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning (Vol. 112, p. 18). Springer, New York.

Jan, C.L., 2021. Using deep learning algorithms for CPAs' going concern prediction. Information 12(2), 73. https://doi.org/10.3390/info12020073.

Jiang, J., Wang, I.Y., Wang, K.P., 2019. Big N auditors and audit quality: New evidence from quasi-experiments. Account. Rev. 94(1), 205-227. https://doi.org/10.2308/accr-52106.

Joe, J.R., 2003. Why press coverage of a client influences the audit opinion. J. Account. Res. 41(1), 109-133. https://doi.org/10.1111/1475-679X.00098.

Jones, F.L., 1996. The information content of the auditor's going concern evaluation. J. Account. Public Policy. 15(1), 1-27. https://doi.org/10.1016/0278-4254(95)00062-3.

Jones, S., 2017. Corporate bankruptcy prediction: A high dimensional analysis. Rev. Account. Stud. 22, 1366-1422. https://doi.org/10.1007/s11142-017-9407-1.

Kahneman, D., Lovallo, D., Sibony, O., 2011. Before you make that big decision.

Kahneman, D., Sibony, O., Sunstein, C.R., 2021. Noise: A flaw in human judgment. Hachette UK.

Kida, T., 1980. An investigation into auditors' continuity and related qualification judgments. J. Account. Res., 506-523. https://doi.org/10.2307/2490590.

Knechel, W.R., Vanstraelen, A., 2007. The relationship between auditor tenure and audit quality implied by going concern opinions. Audit. J. Pract. Theory. 26(1), 113-131. https://doi.org/10.2308/aud.2007.26.1.113.

Koh, H.C., 1987. Prediction of going-concern status: A probit model for the auditors (Doctoral dissertation, Virginia Polytechnic Institute and State University).

Koh, H.C., 1991. Model predictions and auditor assessments of going concern status. Account. Bus. Res. 21(84), 331-338. https://doi.org/10.1080/00014788.1991.9729848.

Koh, H.C., Low, C.K., 2004. Going concern prediction using data mining techniques. Manag. Audit. J. 19(3), 462-476. https://doi.org/10.1108/02686900410524436.

Koh, H.C., Tan, S.S., 1999. A neural network approach to the prediction of going concern status. Account. Bus. Res. 29(3), 211-216. https://doi.org/10.1080/00014788.1999.9729581.

Krupa, J., Minutti-Meza, M., 2022. Regression and machine learning methods to predict discrete outcomes in accounting research. J. Financ. Report. 7(2), 131-178.

Lehmann, C.M., Norman, C.S., 2006. The effects of experience on complex problem representation and judgment in auditing: An experimental investigation. Behav. Res. Account. 18(1), 65-83. https://doi.org/10.2308/bria.2006.18.1.65.

Lennox, C., 1999. Identifying failing companies: A re-evaluation of the logit, probit and DA approaches. J. Econ. Bus. 51(4), 347-364. https://doi.org/10.1016/S0148-6195(99)00009-0.

Lim, C.Y., Tan, H.T., 2008. Non-audit service fees and audit quality: The impact of auditor specialization. J. Account. Res. 46(1), 199-246. https://doi.org/10.1111/j.1475-679X.2007.00266.x.

Liu, M., 2022. Assessing human information processing in lending decisions: A machine learning approach. J. Account. Res. 60(2), 607-651. https://doi.org/10.1111/1475-679X.12427.

Liu, L.L., Xie, X., Chang, Y.S., Forgione, D.A., 2017. New clients, audit quality, and audit partner industry expertise: Evidence from Taiwan. Int. J. Audit. 21(3), 288-303. https://doi.org/10.1111/ijau.12095.

Lopez, D.M., Peters, G.F., 2012. The effect of workload compression on audit quality. Audit. J. Pract. Theory. 31(4), 139-165. https://doi.org/10.2308/ajpt-10305.

Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. 5(9), 1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.

Martens, D., Bruynseels, L., Baesens, B., Willekens, M., Vanthienen, J., 2008. Predicting going concern opinion with data mining. Decis. Support Syst. 45(4), 765-777. https://doi.org/10.1016/j.dss.2008.01.003.

McKee, T., 1976. Discriminant prediction of going concern status: A model for auditors. In Sel. Pap. AAA Annu. Meet.
Menon, K., Schwartz, K.B., 1987. An empirical investigation of audit qualification decisions in the presence of going concern uncertainties. Contemp. Account. Res. 3(2), 302-315. https://doi.org/10.1111/j.1911-3846.1987.tb00640.x.
Mossman, C.E., Bell, G.G., Swartz, L.M., Turtle, H., 1998. An empirical comparison of bankruptcy models. Financ. Rev. 33(2), 35-54. https://doi.org/10.1111/j.1540-6288.1998.tb01367.x.
Myers, J.N., Myers, L.A., Omer, T.C., 2003. Exploring the term of the auditor-client relationship and the quality of earnings: A case for mandatory auditor rotation? Account. Rev. 78(3), 779-799.
Myers, L.A., Schmidt, J., Wilkins, M., 2014. An investigation of recent changes in going concern reporting decisions among Big N and non-Big N auditors. Rev. Quant. Finance Account. 43, 155-172. https://doi.org/10.1007/s11156-013-0368-6.
Newton, N.J., Wang, D., Wilkins, M.S., 2013. Does a lack of choice lead to lower quality? Evidence from auditor competition and client restatements. Audit. J. Pract. Theory. 32(3), 31-67. https://doi.org/10.2308/ajpt-50461.
Nogler, G.E., 1995. The resolution of auditor going concern opinions. Audit. 14(2), 54.
Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res., 109-131. https://doi.org/10.2307/2490395.
Omer, T.C., Sharp, N.Y., Wang, D., 2018. The impact of religion on the going concern reporting decisions of local audit offices. J. Bus. Ethics. 149, 811-831. https://doi.org/10.1007/s10551-016-3045-6.
Perols, J., 2011. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Audit. J. Pract. Theory. 30(2), 19-50. https://doi.org/10.2308/ajpt-50009.
Perols, J.L., Bowen, R.M., Zimmermann, C., Samba, B., 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction. Account. Rev. 92(2), 221-245. https://doi.org/10.2308/accr-51562.
Public Company Accounting Oversight Board (PCAOB), 2012. Going Concern Consideration and Recommendations. https://pcaobus.org/News/Events/Documents/03282012_IAGMeeting/Going_Concern_Considerations_and_Recommendations.pdf (accessed 11 February 2025).
Public Company Accounting Oversight Board (PCAOB), 2015. Concept release on audit quality indicators. https://assets.pcaobus.org/pcaob-dev/docs/default-source/rulemaking/docket_041/release_2015_005.pdf (accessed 11 February 2025).
Public Company Accounting Oversight Board (PCAOB), 2023a. Standards and emerging issues advisory group meeting discussion - going concern. https://assets.pcaobus.org/pcaob-dev/docs/default-source/about/advisory/documents/seiag-march-2023/going-concern---march-2023-seiag-briefing-paper-(final).pdf (accessed 11 February 2025).
Public Company Accounting Oversight Board (PCAOB), 2023b. PCAOB Report: Audits with deficiencies rose for second year in a row to 40% in 2022. https://pcaobus.org/news-events/news-releases/news-release-detail/pcaob-report-audits-with-deficiencies-rose-for-second-year-in-a-row-to-40-in-2022 (accessed 11 February 2025).

Read, W.J., Yezegel, A., 2016. Auditor tenure and going concern opinions for bankrupt clients: Additional evidence. Audit. J. Pract. Theory. 35(1), 163-179. https://doi.org/10.2308/ajpt-51217.

Reichelt, K.J., Wang, D., 2010. National and office-specific measures of auditor industry expertise and effects on audit quality. J. Account. Res. 48(3), 647-686. https://doi.org/10.1111/j.1475-679X.2009.00363.x.

Romanus, R.N., Maher, J.J., Fleming, D.M., 2008. Auditor industry specialization, auditor changes, and accounting restatements. Account. Horiz. 22(4), 389-413.

Rozario, A.M., Vasarhelyi, M.A., Wang, T., 2023. On the use of consumer tweets to assess the risk of misstated revenue in consumer-facing industries: Evidence from analytical procedures. Audit. J. Pract. Theory. 42(2), 207-229.

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 10(3), e0118432.

Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8(4), e1249. https://doi.org/10.1002/widm.1249.

Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. J. Bus. 74(1), 101-124. https://doi.org/10.1086/209665.

Simon, H.A., 1955. A behavioral model of rational choice. Q. J. Econ., 99-118.

Sohn, S.Y., Kim, H.S., 2007. Random effects logistic regression model for default prediction of technology credit guarantee fund. Eur. J. Oper. Res. 183(1), 472-478. https://doi.org/10.1016/j.ejor.2006.10.006.

Tucker, R.R., Matsumura, E.M., Subramanyam, K.R., 2003. Going-concern judgments: An experimental test of the self-fulfilling prophecy and forecast accuracy. J. Account. Public Policy. 22(5), 401-432. https://doi.org/10.1016/j.jaccpubpol.2003.08.002.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science. 185(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124.

Van Rijsbergen, C.J., 2004. The geometry of information retrieval. Cambridge University Press.

Weil, J., Eaglesham, J., 2023. KPMG gave SVB, Signature Bank clean bill of health weeks before collapse. https://www.wsj.com/articles/kpmg-faces-scrutiny-for-audits-of-svb-and-signature-bank-42dc49dd (accessed 11 February 2025).

Wu, M.G., 2006. An economic analysis of audit and nonaudit services: The trade-off between competition crossovers and knowledge spillovers. Contemp. Account. Res. 23(2), 527-554. https://doi.org/10.1506/4DA7-D9YU-P9HQ-PX82.

Xu, Q., Kalelkar, R., 2020. Consequences of going-concern opinion inaccuracy at the audit office level. Audit. J. Pract. Theory. 39(3), 185-208.

Yeh, C.C., Chi, D.J., Lin, Y.R., 2014. Going-concern prediction using hybrid random forests and rough set approach. Inf. Sci. 254, 98-110. https://doi.org/10.1016/j.ins.2013.07.011.

Zhang, C., Thomas, C., Vasarhelyi, M.A., 2022. Attended process automation in audit: A framework and a demonstration. J. Inf. Syst. 36(2), 101-124.

Tables

Table 1. Sample Overview

Panel A. Sample Derivation

	Total Obs.	# of GCO at t	# of Bankruptcy at t+1
The intersection of COMPUSTAT, CRSP, and Audit Analytics from fiscal year 2003 to 2021 (Bankruptcy is from Audit Analytics up to 2022)	105872	3865	477
Less: Firms with missing data to compute control variables	(35673)	(532)	(60)
Less: Firms not defined as distressed	(41000)	(124)	(28)
Less: Duplicated firm-year observations	(145)	(22)	(8)
Main Sample	29054	3187	381

This table shows the sample selection for the observations in the main analysis using financial and market variables. Financial distressed firms are defined as companies with either negative net income or negative operating cash flows, following Gutierrez et al. (2021).

Panel B. Sample Distribution by Year

Fiscal Year	Obs.	# of GCO at t	% of GCO at t	# of Bankruptcy at t+1	% of Bankruptcy at t+1
2003	1633	125	7.7%	23	1.4%
2004	1445	115	8.0%	17	1.2%
2005	1439	131	9.1%	14	1.0%
2006	1417	116	8.2%	11	0.8%
2007	1439	151	10.5%	31	2.2%
2008	1697	218	12.8%	44	2.6%
2009	1517	131	8.6%	10	0.7%
2010	1138	103	9.1%	15	1.3%
2011	1140	89	7.8%	17	1.5%
2012	1209	104	8.6%	10	0.8%
2013	1273	95	7.5%	19	1.5%
2014	1399	113	8.1%	19	1.4%
2015	1575	184	11.7%	32	2.0%
2016	1529	205	13.4%	18	1.2%
2017	1536	207	13.5%	17	1.1%
2018	1612	255	15.8%	26	1.6%
2019	1769	319	18.0%	42	2.4%
2020	2126	214	10.1%	4	0.2%
2021	2161	312	14.4%	12	0.6%
Total	29054	3187	11.0%	381	1.3%

This table presents the sample distribution by fiscal year, detailing the number of observations, the number and percentage of auditor-issued going concern opinions, and the number and percentage of bankruptcies for each fiscal year.

Panel C. Sample Distribution by Industry

Industry	**Obs.**	**# of GCO at t**	**% of GCO at t**	**# of Bankruptcy at t+1**	**% of Bankruptcy at t+1**
Industrial Applications & Services	3367	483	14.3%	30	0.9%
Energy & Transportation	3877	435	11.2%	113	2.9%
Real Estate & Construction	1082	143	13.2%	18	1.7%
Manufacturing	5167	465	9.0%	64	1.2%
Life Sciences	6327	1069	16.9%	42	0.7%
Technology	5587	363	6.5%	46	0.8%
Trade & Services	3330	212	6.4%	66	2.0%
Finance	275	14	5.1%	2	0.7%
Other non-specified	42	3	7.1%	0	0.0%
Total	29054	3187	11.0%	381	1.3%

This table presents the distribution of the sample by industry, detailing the number of observations, the number and percentage of auditor-issued going concern opinions, and the number and percentage of bankruptcies for each industry. Industry classification is from SEC at https://www.sec.gov/corpfin/division-of-corporation-finance-standard-industrial-classification-sic-code-list.

Panel D. GCO and Bankruptcy

	GCO at t = 0	**GCO at t = 1**	**Total**
Bankruptcy at t+1 = 0	25701	2972	28673
Bankruptcy at t+1 = 1	166	215	381
Total	25867	3187	29054

This table presents a confusion matrix of GCO against bankruptcy.

Table 2. Descriptive Statistics

	count	mean	std	1%	25%	50%	75%	99%
Outcome Variables								
BANKRUPTCY	29054	0.013	0.114	0.000	0.000	0.000	0.000	1.000
GOING_CONCERN	29054	0.110	0.313	0.000	0.000	0.000	0.000	1.000
Predictor Variables								
NEGEQUITY	29054	0.075	0.263	0.000	0.000	0.000	0.000	1.000
ROA	29054	-0.275	0.429	-2.543	-0.353	-0.118	-0.029	0.175
LEVERAGE	29054	0.512	0.349	0.027	0.245	0.467	0.697	1.919
WCAP	29054	0.309	0.326	-0.691	0.073	0.280	0.553	0.941
CURRENT	29054	4.082	5.261	0.193	1.320	2.243	4.490	32.760
CASH	29054	0.335	0.306	0.000	0.068	0.227	0.573	0.978
CFO	29054	-0.144	0.329	-1.807	-0.212	-0.030	0.043	0.232
SIZE	29054	5.308	1.938	1.452	3.892	5.133	6.555	10.424
EX_RET	29054	0.091	1.294	-1.004	-0.493	-0.216	0.163	8.621
SIGMA	29054	0.233	0.419	0.004	0.099	0.142	0.206	3.253
REL_MKTCP	29054	-11.441	1.782	-15.233	-12.724	-11.454	-10.234	-6.984
Other Variables								
GCO_FN_ERROR	18109	0.003	0.050	0.000	0.000	0.000	0.000	0.000
ML_FN_ERROR	18109	0.002	0.049	0.000	0.000	0.000	0.000	0.000
GCO_FP_ERROR	18109	0.098	0.298	0.000	0.000	0.000	0.000	1.000
ML_FP_ERROR	18109	0.023	0.149	0.000	0.000	0.000	0.000	1.000
ML	18467	0.045	0.208	0.000	0.000	0.000	0.000	1.000
TEAM	18467	0.044	0.206	0.000	0.000	0.000	0.000	1.000
TEAMPROB	18467	0.065	0.170	0.000	0.000	0.001	0.020	0.889
AUDITOFFICEEXPERTISE	15892	0.490	0.383	0.002	0.120	0.408	1.000	1.000
AUDITOFFICESIZE	15912	16.206	1.974	11.039	14.836	16.492	17.864	19.417
BIG4	18467	0.634	0.482	0.000	0.000	1.000	1.000	1.000
BUSY	15917	0.834	0.372	0.000	1.000	1.000	1.000	1.000
WORKLOADCOMPRESSION	15851	0.760	0.298	0.012	0.663	0.884	1.000	1.000
NONAUDITFEERATIO	15917	0.127	0.163	0.000	0.002	0.069	0.191	0.753
INFLUENCE	15912	0.175	0.273	0.002	0.018	0.053	0.182	1.000
TENURE	15917	5.743	4.368	1.000	2.000	4.000	8.000	19.000
NEWCLIENT	18467	0.111	0.314	0.000	0.000	0.000	0.000	1.000
ANCHORINGGCO	18467	0.052	0.223	0.000	0.000	0.000	0.000	1.000
ANCHORINGNOGCO	18467	0.515	0.500	0.000	0.000	1.000	1.000	1.000
UNOBSERVABLE	18467	0.015	0.264	-0.762	-0.026	-0.004	-0.001	0.971
UNOBSERVABLEISSUE	18467	0.067	0.211	0.000	0.000	0.000	0.000	0.971
UNOBSERVABLENOTISSUE	18467	0.051	0.136	0.000	0.001	0.004	0.026	0.762
AUDITREPORTLENGTHCHANGE	10652	2.055	7.630	-17.000	0.000	0.000	3.000	33.000
AUDITREPORTLAGCHANGE	15344	-0.014	0.203	-0.840	-0.074	-0.012	0.058	0.674

Continuous variables are winsorized at the 1% and 99% levels. The “Other Variables” are only available for the hold-out sample period 2010-2021 where ML assessments can be obtained.

Table 3. ML Assessments vs GCOs

Panel A. Performance Metrics Comparison

	ML	GCO
Recall	0.5420	0.5294
Precision	0.1514	0.0569
F-Score	0.2367	0.1028
ECM	0.1036	0.1788
ROC-AUC	0.9176	Not Available
PR-AUC	0.2101	Not Available

This table presents a comparison of ML assessments and GCOs in their ability to provide early warnings to bankruptcy within one year after the financial statement being audited. Recall (or true positive rate) is the percentage of firms correctly assessed by the model to go bankrupt within a year from all bankrupt firms. Precision is the percentage of bankruptcy firms correctly detected by the model from all positive predictions that it made. F-score combines both precision and recall. ECM is the expected cost of misclassification. ROC-AUC is the area under the Receiver Operating Characteristic curve. PR-AUC is the area under the precision-recall curve, which plots precision on the y-axis against recall on the x-axis for different classification thresholds. ROC-AUC and PR-AUC are not available for GCO since their calculation requires continuous probability prediction. For Recall, Precision, F-Score, and ECM, we set the classification threshold for ML probability prediction to be 0.50.

Panel B. Identifying ML error and GCO error

Scenario	Bankruptcy at t+1	ML Assessment at t	GCO at t	Obs.	Note
1	0	0	0	15747	Both ML and GCO are correct
2	**0**	**0**	**1**	**1779**	**GCO false positive error**
3	**0**	**1**	**0**	**413**	**ML false positive error**
4	0	1	1	297	Either ML or GCO can necessarily be considered wrong
5	1	0	0	61	Either ML or GCO can necessarily be considered wrong
6	**1**	**0**	**1**	**44**	**ML false negative error**
7	**1**	**1**	**0**	**46**	**GCO false negative error**
8	1	1	1	80	Both ML and GCO are correct

This table presents different scenarios of ML assessments, GCO, and bankruptcy within one year. The total number observations for this table is 18467 for the hold-out testing period of 2010-2021.

Table 4. What Factors are Associated with GCO Error and ML Error

Panel A. What Factors are Associated with GCO Error

	(1) GCO_FP_ERROR	(2) GCO_FN_ERROR
AUDITOFFICEEXPERTISE	-0.0170***	0.0027**
	(-2.97)	(2.13)
AUDITOFFICESIZE	-0.0143***	0.0010***
	(-7.59)	(2.90)
BIG4	-0.0215***	-0.0027*
	(-3.47)	(-1.84)
NONAUDITFEERATIO	-0.0202	-0.0026*
	(-1.57)	(-1.68)
INFLUENCE	-0.0522***	0.0033*
	(-4.79)	(1.65)
TENURE	-0.0006	-0.0001
	(-1.40)	(-1.41)
NEWCLIENT	0.0004	-0.0008
	(0.06)	(-0.77)
ANCHORINGGCO	0.7879***	-0.0006
	(93.14)	(-0.90)
ANCHORINGNOGCO	-0.1439***	0.0015*
	(-30.30)	(1.81)
BUSY	0.0058	-0.0032*
	(0.70)	(-1.82)
WORKLOADCOMPRESSION	-0.0055	0.0001
	(-0.54)	(0.03)
Industry and Year Fixed Effects	Yes	Yes
Adj. R2	0.4794	0.0136
Observations	15540	15540

This table shows the OLS regression results of the association between GCO error and factors related to auditors' capability, cognitive bias and constraints. In this test, scenarios 4 and 5 listed in Table 3 Panel B are dropped. GCO_FP_ERROR represents scenario 2 and GCO_FN_ERROR represents scenario 7. Variable descriptions are provided in Appendix A. T-stats are shown in the parentheses. ***, **, and * represent statistical significance at 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered by client firm.

Panel B. What Factors are Associated with ML Error

	(1) ML_FP_ERROR	(2) ML_FN_ERROR
UNOBSERVABLEISSUE	-0.004	0.0451***
	(-1.47)	(5.06)
UNOBSERVABLENOTISSUE	0.2878***	-0.0021
	(9.14)	(-1.49)
AUDITREPORTLAGCHANGE	-0.0021	0.0115***
	(-0.19)	(3.00)
AUDITREPORTLENGTHCHANGE	0.0003	0.0002***
	(1.13)	(2.81)
Industry and Year Fixed Effects	Yes	Yes
Adj. R2	0.0864	0.0379
Observations	10268	10268

This table shows the OLS regression results of the association between ML assessment error and factors related to auditors' capability, cognitive bias and constraints. In this test, scenarios 4 and 5 listed in Table 3 Panel B are dropped. ML_FP_ERROR represents scenario 3 and ML_FN_ERROR represents scenario 6. Variable descriptions are provided in Appendix A. T-stats are shown in the parentheses. ***, **, and * represent statistical significance at 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered by client firm.

Table 5. Can Team Assessments Overcome Auditor Disadvantage?

	DV = BANKRUPTCY			
	(1) When audit office has lower industry expertise	(2) When audit office is smaller	(3) When anchoring bias exists	(4) When client has higher influence on auditor
TEAM	0.0787***	0.0986***	0.0609***	0.1121***
	(4.82)	(5.87)	(3.73)	(6.65)
GCO	0.0281***	0.0184***	0.0049	0.0237***
	(4.31)	(3.22)	(0.83)	(3.37)
CONTROLS	Yes	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes	Yes
Adj. R2	0.0892	0.0975	0.0526	0.1227
Observations	7937	7944	9162	7959
T-stat of coefficient test between TEAM and GCO	8.61***	19.50***	9.44***	23.19***

This table shows the OLS regression results of comparing the abilities of team assessments and GCOs in assessing client firms' going concerns. BANKRUPTCY is a dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end, and zero otherwise. GCO is a dummy variable that equals one if client firm i receives a going concern opinion for fiscal year t, zero otherwise. TEAM is the binary team assessment which is the output of a ML model with GCO as an additional predictor. It equals 1 if the "team" casts substantial doubt about the client firm's ability to continue as a going concern within the next year, 0 otherwise. CONTROLS contains all predictor variables we used in the ML model to predict next year bankruptcy. Variable descriptions are provided in Appendix A. Column (1) presents results for the subsample with lower audit office industry expertise, defined as audit office industry expertise (AUDITOFFICEEXPERTISE) lower than the median; Column (2) presents results for subsample with smaller audit office size, defined as audit office size (AUDITOFFICESIZE) smaller than median; Column (3) presents results when anchoring bias exists, defined as either ANCHORINGGCO=1 or ANCHORINGNOGCO=1 ; Column (4) presents results for subsample when the client firm has higher influence on the auditor, defined as influence (INFLUENCE) greater than median. T-stats are shown in the parentheses. ***, **, and * represent statistical significance at 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered by client firm.

Table 6. Can Team Assessments Overcome ML Disadvantage?

	DV = BANKRUPTCY		
	(1) High level of useful unobservable information	(2) Large increase of audit report length	(3) Long audit report lag
TEAM	0.1080***	0.1051*	0.1236***
	(3.95)	(1.76)	(3.40)
ML	0.0032	-0.0164	0.0460
	(0.15)	(-0.31)	(1.24)
CONTROLS	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes
Adj. R2	0.1546	0.1456	0.2190
Observations	3167	976	3065
T-stat of Coefficient Test between TEAM and ML	14.27***	1.77	1.28

This table shows the OLS regression results of comparing the abilities of team assessments and ML assessments in assessing client firms' going concerns. BANKRUPTCY is a dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end, and zero otherwise. ML equals 1 when ML has substantial doubt about the client firm's ability to continue as a going concern within the next year, 0 otherwise. TEAM is the binary team assessment which is the output of a ML model with GCO as an additional predictor. It equals 1 if the "team" casts substantial doubt about the client firm's ability to continue as a going concern within the next year, 0 otherwise. CONTROLS contains all predictor variables we used in the ML model to predict next year bankruptcy. Variable descriptions are provided in Appendix A. Column (1) presents results of Equation (4) for the subsample with high level of unobservable information, defined as the difference between actual GCO and likelihood prediction of GCO (UNOBSERVABLE) at the bottom (1st) or the 9th decile (Appendix E explains why 10th decile is not used); Column (2) and (3) present results when the audit report lag change is large (AUDITREPORTLAGCHANGE at the top decile) and when the change of audit report length is large (AUDITREPORTLENGTHCHANGE at the top decile). T-stats are shown in the parentheses. ***, **, and * represent statistical significance at 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered by client firm.

Table 7. Effectiveness of Assessments During COVID-19 Pandemic

	DV = BANKRUPTCY		
	(1)	(2)	(3)
ML	0.0286		
	(1.50)		
GCO		0.0264***	
		(3.58)	
TEAM			0.0388*
			(1.93)
CONTROLS	Yes	Yes	Yes
Industry and Year Fixed Effects	Yes	Yes	Yes
Adj. R2	0.0590	0.0668	0.0632
Observations	3884	3884	3884

This table shows the OLS regression results of comparing the abilities of GCO, ML, and team assessments in assessing client firms' going concerns. BANKRUPTCY is a dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end, and zero otherwise. GCO is a dummy variable that equals one if client firm i receives a going concern opinion for fiscal year t, zero otherwise. ML equals 1 when ML has substantial doubt about the client firm's ability to continue as a going concern within the next year, 0 otherwise. TEAM is the binary team assessment which is the output of a ML model with GCO as an additional predictor. It equals 1 if the "team" casts substantial doubt about the client firm's ability to continue as a going concern within the next year, 0 otherwise. CONTROLS contains all predictor variables we used in the ML model to predict next year bankruptcy. Variable descriptions are provided in Appendix A. This test is for the sample period fiscal years 2020-2021. T-stats are shown in the parentheses. ***, **, and * represent statistical significance at 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered by client firm.

Appendix A. Variable Definitions

Variable	Definition	Source
Outcome Variables		
BANKRUPTCY	A dummy variable that equals one if client firm i enters into bankruptcy within one year after the fiscal year end date of t, and zero otherwise.	Audit Analytics
GOING_CONCERN	A dummy variable that equals one if client firm i receives a going concern opinion for fiscal year t, zero otherwise.	Audit Analytics
Predictor Variables		
NEGEQUITY	An indicator variable that equals to one if total liabilities exceed total assets, and zero otherwise	COMPUSTAT
ROA	Net income divided by total assets.	COMPUSTAT
LEVERAGE	The percentage of total liabilities over total assets.	COMPUSTAT
WCAP	The percentage of working capital over total assets.	COMPUSTAT
CURRENT	The percentage of current assets over current liabilities.	COMPUSTAT
CASH	The percentage of cash and cash equivalents over total assets.	COMPUSTAT
CFO	The cash flow from operating activities over total assets.	COMPUSTAT
SIZE	Natural logarithm of total assets.	COMPUSTAT
EX_RET	Cumulative Firm Returns – Cumulative Market Returns over the 12 months.	CRSP
SIGMA	Standard deviation of the residuals from a monthly return model of firm returns on market returns for the 12 months.	CRSP
REL_MKTCP	Log (Firm Market Capitalization / Index Market Capitalization).	CRSP
Other Variables		
GCO_FN_ERROR	A dummy variable that equals one for scenario 7 presented in Panel B, Table 3, zero otherwise.	Calculated
ML_FN_ERROR	A dummy variable that equals one for scenario 6 presented in Panel B, Table 3, zero otherwise.	Calculated
GCO_FP_ERROR	A dummy variable that equals one for scenario 2 presented in Panel B, Table 3, zero otherwise.	Calculated
ML_FP_ERROR	A dummy variable that equals one for scenario 3 presented in Panel B, Table 3, zero otherwise.	Calculated
ML	A dummy variable that equals one if the probability prediction from the ML model is greater than 0.5, zero otherwise.	Calculated
TEAM	A dummy variable that equals one if the probability prediction from the ML model with GCO as an additional predictor is greater than 0.5, zero otherwise.	Calculated
AUDITOFFICEEXPERTISE	An audit office's annual market share based on audit fees within a two-digit SIC category	Audit Analytics
AUDITOFFICESIZE	Natural logarithm of one plus total annual audit fees of an audit office.	Audit Analytics

BIG4	Indicator variable equal to one if the audit firm is a Big 4, and zero otherwise	COMPUSTAT
BUSY	Indicator variable equal to one if a company has a fiscal year-end date of December 31, and zero otherwise	COMPUSTAT
WORKLOADCOMPRESSION	The relative level of workload compression of an auditor office during the fiscal year end month of the auditee	Audit Analytics
NONAUDITFEERATIO	Non-audit fees deflated by total fees paid (audit plus non-audit fees). Non-audit fee equals to the sum of benefit fee, IT fee, Tax fee, audit related fee, and other fees.	Audit Analytics
INFLUENCE	The ratio of a company's total fees (i.e., audit fees plus non-audit fees) relative to the aggregate annual total fees generated by the local office that audits the company	Audit Analytics
TENURE	Number of years that the company is audited by the same audit firm	Audit Analytics
NEWCLIENT	Indicator variable equal to one if the auditor-client relationship is in its first year, and zero otherwise	Audit Analytics
ANCHORINGGCO	A dummy variable equal to one if the client firm receives an GCO in both the current year and the prior year, zero otherwise.	Audit Analytics
ANCHORINGNOGCO	A dummy variable equal to one if the client firm does not receive an GCO in both the current year and the prior year, zero otherwise.	Audit Analytics
UNOBSERVABLE	The discrepancy between GCO and the probability prediction of a ML model that predicts GCO issuance.	Calculated
UNOBSERVABLEISSUE	The absolute value of the positive discrepancy between GCO and the probability prediction of a ML model that predicts GCO issuance. It takes value of zero of the discrepancy if negative.	Calculated
UNOBSERVABLENOTISSUE	The absolute value of the negative discrepancy between GCO and the probability prediction of a ML model that predicts GCO issuance. It takes value of zero of the discrepancy if positive.	Calculated
AUDITREPORTLENGTHCHANGE	The change of number of sentences in the audit report from current year to prior year.	SeekInF.com
AUDITREPORTLAGCHANGE	The change of audit report lag from current year to prior year. Audit report lag is the natural logarithm of days between fiscal year end and auditor signature date.	Audit Analytics

Appendix B. Evaluation Metrics

Recall (or true positive rate) is the percentage of firms correctly assessed by the model to go bankrupt within a year from all bankrupt firms. Precision is the percentage of bankruptcy firms correctly detected by the model from all positive predictions that it made. F-score, defined as $\frac{2 * (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$ (Van Rijsbergen 2004), is a comprehensive measure of the model's performance in terms of both precision and recall. The higher the recall, precision, and F-score, the better the model performs. In classifying the probabilistic ML prediction into a binary assessment, we use the conventional 50 percent as the classification threshold.[27]

The Area under the Receiver Operating Characteristic curve (ROC-AUC) is the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the recall on the y-axis against the false positive rate (the number of non-bankruptcy firms assessed as bankruptcy divided by the number of non-bankruptcy firms) on the x-axis for different possible classification thresholds. It indicates the overall accuracy of a model at various thresholds used to discriminate between positives and negatives. The value of ROC-AUC ranges from 0 to 1, with value below 0.5 considered worse than random guessing, 0.5 to 0.7 as poor performance, 0.7 to 0.8 as acceptable, 0.8 to 0.9 as excellent, and 0.9 to 1 as outstanding (Mandrekar 2010; Hosmer, Lemeshow, and Sturdivant 2013).

The Area Under the Precision-Recall Curve (PR-AUC) is the area under the precision-recall curve, which plots precision on the y-axis against recall on the x-axis for different classification thresholds. The value of PR-AUC ranges from 0 to 1, the higher the better. The baseline of PR-AUC is the ratio of positives to the total number of observations (Saito and Rehmsmeier 2015). Compared to ROC-AUC, PR-AUC can better reflect model performance when class imbalance exists in the dataset (Krupa and Minutti-Meza 2022)

Lastly, we consider the expected cost of misclassification (ECM). As discussed in the previous section, there is a cost imbalance between false positive and false negative errors. To incorporate the considerations of the economic costs, we follow prior literature and use the ECM to measure economic costs generated from model prediction (Hopwood et al., 1994; Perols, 2011). Specifically, the ECM is calculated as:

$$ECM = C^{FN} \times P(Bankruptcy) \times \frac{n^{FN}}{n^{P}} + C^{FP} \times (1 - P(Bankruptcy)) \times \frac{n^{FP}}{n^{N}} \qquad (1)$$

C^{FN} and C^{FP} are the estimates of the cost of false negative and false positive errors, respectively, deflated by the lower of C^{FN} and C^{FP} (Perols et al. 2017). In our setting, C^{FN} is greater than C^{FP}. $P(Bankruptcy)$ is the prior probability of bankruptcy. We utilize the percentage of bankruptcy in the training period as $P(Bankruptcy)$. Lastly, n^{P} and n^{N} are the number of bankruptcy and non-bankruptcy observations, respectively; and n^{FN} and n^{FP} are the numbers of false negative and false positive observations. The lower the ECM, the better the performance.

[27] Results hold when we choose alternative classification thresholds (i.e., 40 percent or 60 percent).

Appendix C. Performance Metrics on Validation Set

Cost	ROC-AUC	PR-AUC	Precision	Recall	F1	ERC
10	0.928	0.260	0.165	0.577	0.247	0.099
20	0.927	0.250	0.106	0.732	0.181	0.157
30	0.927	0.251	0.086	0.795	0.152	0.200
40	0.927	0.247	0.074	0.848	0.134	0.227
50	0.926	0.246	0.066	0.856	0.120	0.266

This table presents ML performance on hold-out validation set averaged across 2009-2020. Recall (or true positive rate) is the percentage of firms correctly assessed by the model to go bankrupt within a year from all bankrupt firms. Precision is the percentage of bankruptcy firms correctly detected by the model from all positive predictions that it made. F-score combines both precision and recall. ECM is the expected cost of misclassification. ROC-AUC is the area under the Receiver Operating Characteristic curve. PR-AUC is the area under the precision-recall curve, which plots precision on the y-axis against recall on the x-axis for different classification thresholds. For Recall, Precision, F-Score, and ECM, we set the classification threshold for ML probability prediction to be 0.50.

Appendix D. What Factors Are Associated with Auditor GCO Error?

The variables that represent auditors' capability include office-level industry expertise ($AUDITOFFICEEXPERTISE$), office size ($AUDITOFFICESIZE$), and Big4 brand name ($BIG4$). It is unclear how these measures are associated with GCO error. First, it is likely that auditors with more industry expertise, that are larger, and that have Big Four brand can provide higher-quality audits (e.g., Reichelt and Wang 2010; DeFond et al. 2017), thus reducing both types of errors. It is also likely that they may feel less pressure when issuing unfavorable opinions, such as GCOs. This could lead to increased false positive errors and decreased false negative errors; on contrary, these auditors might adopt a less conservative approach due to their strong market standing and perceived credibility. This might result in issuing fewer GCOs, which could decrease false positive errors and increase false negative errors.

We use the following variables to capture auditors' cognitive bias: non-audit fee ratio ($NONAUDITFEERATIO$), client influence ($INFLUENCE$), tenure ($TENURE$), new client ($NEWCLIENT$), and anchoring bias ($ANCHORINGGCO$ and $ANCHORINGNOGCO$). Higher the non-audit fee ratio raises concerns about auditors' independence (Bazerman et al.1997; Francis 2006), potentially making them less likely to issue GCOs. This could lead to decreases in false positive errors and increases in false negative errors. However, prior literature also suggests spillover effects from non-audit services that may enhance overall audit quality Lim and Tan 2008; Wu 2006), potentially reducing both false negatives and false positives. Longer tenure could compromise auditor independence, reducing the likelihood of issuing GCOs (Bell et al. 2015), thereby increasing false negative errors and reducing false positive errors. However, longer tenure also equips auditors with deeper insights into their clients' operations (Gul, Fung, and Jaggi 2009), possibly enhancing their ability to accurately assess going concern risks and thus reducing both types of errors. As for new client, on one hand, auditors may exhibit higher independence with new clients, potentially reducing false negatives and increasing false positives. On the other hand, auditors may hesitate to issue unfavorable opinions in their first year with a new client, which could increase false negatives and decrease false positives. Anchoring bias refers to auditors' tendency to rely on previous years' opinions when issuing current year's GCO. Anchoring to a prior year's clean opinion ($ANCHORINGNOGCO$) likely increases false negative errors and reduces false positive errors. In contrast, anchoring to a prior year's non-clean opinion ($ANCHORINGGCO$) tends to decrease false negative errors and increase false positive errors.

Lastly, variables that represent auditors' cognitive constraints include whether the audit is done in busy season ($BUSY$) and workload compression ($WORKLOADCOMPRESSION$). On the one hand, operating under cognitive constraints, such as during the busy season or under compressed workloads, auditors might adopt a more conservative stance. This could lead to a higher issuance rate of GCOs, thereby increasing false positive errors and reducing false negative errors. On the other hand, these same cognitive constraints might impair auditors' ability to effectively analyze and recognize patterns of distress within the firms they audit. This limitation can result in a decreased likelihood of issuing GCOs when they are indeed necessary, thereby decreasing false positive errors and increasing false negative errors.

Appendix E. ML Model to Predict GCO

Following Liu (2022), we establish a separate ML model to predict the issuance of GCO. The probability prediction from this ML model will represent an auditor's likelihood of issuing GCO and it serves as a counterfactual scenario of the auditor's decision had she only consider the observable hard information without the impact of unobservable factors. We adopt the same ML setup as specified in Section "Machine Learning Model Setup". The only difference is that instead of predicting $BANKRUPTCY$, we predict GCO. For simplicity, we do not adopt the Observation Under-sampling (OU) method from Perols et al. (2017) to adjust training data. The model performance is outstanding by academic standards (Mandrekar 2010; Hosmer et al. 2013) with ROC-AUC of 90.9 percent and PR-AUC of 58.9 percent.

The discrepancy between GCO and the probability prediction of GCO issuance from the ML model ($UNOBSERVABLE$) represents the extent to which unobservable factors influence auditors' decisions. Table AC presents the proportion of GCO_FP_ERROR, GCO_FN_ERROR, ML_FP_ERROR, and ML_FN_ERROR at each quintile of $UNOBSERVABLE$. We observe that at the top (10th) quintile, there is a disproportionally large amount of GCO false positive error, indicating that this quintile may capture more of auditors' conservatism tendency than useful unobservable information. Thus, in Table 6, Column (1), we do not consider the top quintile.

GCO and ML Errors at Different Levels of Unobservable Information

Quintile of UNOBSERVABLE	GCO_FP_ERROR	GCO_FN_ERROR	ML_FP_ERROR	ML_FN_ERROR
1	0.000	0.018	0.124	0.008
2	0.000	0.012	0.055	0.003
3	0.000	0.014	0.024	0.010
4	0.000	0.005	0.014	0.004
5	0.000	0.005	0.004	0.005
6	0.000	0.002	0.002	0.002
7	0.000	0.001	0.001	0.001
8	0.000	0.001	0.000	0.001
9	0.169	0.000	0.069	0.003
10	0.955	0.000	0.092	0.021