2022

Do Different Data Analytics Impact Auditors' Decisions?

J F. Brazel

Efosa Ehimwenma
Trinity University, eehimwen@trinity.edu

Jared Koreff
Trinity University, jkoreff@trinity.edu

Do Different Data Analytics Impact Auditors' Decisions?

Joseph F. Brazel
North Carolina State University
jfbrazel@ncsu.edu

Efosa Ehimwenma
Trinity University
eehimwen@trinity.edu

Jared Koreff*
Trinity University
jkoreff@trinity.edu

Running Head: Do Different Data Analytics Impact Auditor's Decisions

We thank Jhace Alston-Kirby, Prem Gandhi, and Sara Beth Jewell for research assistance.

*Corresponding Author

Do Different Data Analytics Impact Auditors' Decisions?

SUMMARY: Global stakeholders have expressed interest in increasing the use of data analytics throughout the audit process. While data analytics offer great promise in identifying audit-relevant information, auditors may not use this information to its full potential, resulting in a missed opportunity for possible improvements to audit quality. This article summarizes a study by Koreff (2022) that examines whether conclusions from different types of data analytical models (anomaly vs. predictive) and data analyzed (financial vs. non-financial), result in different auditor decisions. Findings suggest that when predictive models are used and identify a risk of misstatement, auditors increase budgeted audit hours more when financial data is analyzed than when non-financial data is analyzed. However, when anomaly models are used and identify a risk of misstatement, auditors' budgeted hours do not differ based on the type of data analyzed. These findings provide evidence that different data analytics do not uniformly impact auditors' decisions.

Key Words: Anomaly Models; Auditor Decisions; Data Analytics; Non-Financial Data; Predictive Analytics

I. INTRODUCTION

In this article, we summarize a study that attempts to explain how auditors' prior experience using different types of analyses impacts how they respond to conclusions drawn from different data analytical models, depending on the type of data analyzed (Koreff 2022).

Despite the advances in technology enabling accounting firms to develop more sophisticated data analytics to identify audit relevant information, and potentially improve audit quality, use of these tools by auditors is often inconsistent for a variety of reasons, including concerns over inspection risk (Eilifsen, Kinserdal, Messier, and McKee 2020), the Public Company Accounting Oversight Board (PCAOB) not explicitly requiring the use of these tools (PCAOB 2021b), and the restrictive nature of the technology (Dowling and Leech 2014). Koreff (2022) shows that even when the same output is presented, auditors' experience (familiarity) with the combination of the type of model and data used to arrive at the same conclusion can result in inconsistent decision making.

The interview data in Koreff (2022) shows that auditors report a comparable amount of experience analyzing financial and non-financial data when using anomaly models, which explains why decisions do not differ when auditors use anomaly models that analyze different data. Thus, when firms develop more advanced anomaly-based analytics, the type of data analyzed is not expected to result in inconsistent auditor decision making. However, the same cannot be said for predictive analytics, as interviewees reported that predictive analytics tend to focus on financial data relative to non-financial data. Accordingly, Koreff (2022) demonstrates that, when predictive analytics are used, auditors are more likely to incorporate the findings into their decisions when financial data is analyzed as compared to non-financial data. These findings are in line with the PCAOB's data and technology research project expressing a concern that

auditor experience and understanding of analytics represent important factors to the effective use of these tools (PCAOB 2021a), and ultimately improvement of audit quality.

Taken together, Koreff (2022) observes that two attributes of analytics, model and data, do not impact auditors' decisions individually. However, the *combination* of these two attributes impact auditors' decisions.

II. MOTIVATION AND EXPECTATIONS

Advances in technology have resulted in the development of data analytical tools that can perform a list of analyses such as population testing, identifying outliers based on a criteria, predictive modeling, and analysis of non-traditional unstructured data. In fact, the American Institute of Certified Public Accountant's (AICPA) Assurance Services Executive Committee (ASEC) has developed an "Audit Data Analytics Guide" that suggests that data analytics are an outgrowth and expansion of analytical procedures (AICPA 2015; Appelbaum, Kogan, and Vasarhelyi 2017; AICPA 2017). Furthermore, Statement on Auditing Standards (SAS) 142 (entitled "Audit Evidence") permits auditors to use automated tools and techniques to enhance the evaluation of audit evidence, including the analysis of non-financial data.[1]

Although data analytics can be seen as an extension of analytical procedures (Appelbaum et al. 2017), auditors do not always use analytical procedures effectively (PCAOB 2007; PCAOB 2014; PCAOB 2008; PCAOB 2013; Barr-Pulliam, Brazel, McCallen, and Walker 2020; Brazel, Leiby, and Schaefer 2022a; Cao, Dug, Tan, and Xu 2022). As an additional barrier to consistent implementation of data analytics, PCAOB standards do not require the use of data analytics (PCAOB 2021b). Shortcomings of analytics include users not considering risks beyond what the analytics identified (Seow 2011), and not properly evaluating false positives (Koreff, Weisner,

[1] For examples on permitting the use of automated tools, see paragraphs A3, A4, A43, A45, A46, A47 and A61. See paragraph A59 for permitting the use of analysis of non-financial data.

and Sutton 2021). Auditors have a preference for simpler analytics, including comparing current year balances to prior year balances, and thus may be reluctant to use more sophisticated analytics (Ameen and Strawser 1994; Trompeter and Wright 2010; Schmidt, Riley, and Church 2020b; Schmidt, Church, and Riley 2020a; Brazel, Jones, and Lian 2022b). Yet, the PCAOB encourages the use of these tools in order to improve the audit process and audit quality (PCAOB 2016; PCAOB 2018). One way to promote auditors use of analytics may be to provide auditors with analytics that use familiar analyses.

When auditors use familiar analyses, it is expected to induce cognitive fit. Cognitive fit refers to the congruence between a process used by a decision maker and the decision aiding tool (Vessey and Galletta 1991; Al-Natour, Benbasat, and Cenfetelli 2008). Auditors will experience greater cognitive fit with data analytics that use combinations of data analytic models and data types that they are more familiar with, since cognitive fit is correlated with experience (Dunn and Grabski 2001; Goodhue and Thompson 1995).[2] Data analytics can be used to analyze a multitude of data types, but auditors will experience different levels of cognitive fit depending on experience using the analyses utilized by the analytics (i.e., the combination of model and data). Thus, when auditors view the results of an analytic that uses familiar analyses, auditors will experience greater cognitive fit with the analytic and therefore be more likely to incorporate the results of the analytic into their decision making process.

Two analytical *models* were examined by Koreff (2022): Anomaly and Predictive models. Anomaly models perform a distributional (bell curve) analysis to identify outliers

[2] While we acknowledge that these studies are not from recent years, these findings are echoed by interviewees in Koreff (2022) that experience using analyses increases the likelihood of using the analysis by stating "…it all comes down to experience using it… so I'd say those are probably the largest one [resistance to using analytics] is the lack of experience…" and "anytime there's new data, I'm a little bit nervous … If the auditor has experience with the process or with the client I think there can probably be higher willingness to use certain analytics." See Appendix A for a complete list of quotes from interviewees in Koreff (2022) discussing cognitive fit and prior experience impacting use of analytics.

(Statistical Analysis System Institute 2014). Predictive models analyze patterns of previously identified issues and compare them with current patterns (Kuenkaikaew and Vasarhelyi 2013). Koreff (2022) illustrates that auditors experience using these two types of models does not differ substantially. As a result, Koreff (2022) predicts that the auditor's cognitive fit will depend not only on the analytical model used, but also the data analyzed by the model.

Two types of *data* were assessed by Koreff (2022): financial data and non-financial data. Predictive models focus primarily on analyzing financial data (Dechow, Ge, Larson, and Sloan 2011; Sinclair 2015; Perols, Bowen, Zimmermann, and Samba 2017), whereas anomaly models are more capable of analyzing both types of data (Glover, Prawitt, and Wilks 2005; Hobson, Mayew, and Venkatachalam 2012; Brazel, Jones and Prawitt 2014). See Figure 1 for a graphical depiction of auditors' experiences using the four combinations of the different types of analyses. The depiction in Figure 1 suggests that auditors have comparable experience using predictive and anomaly models (hence the two bars rising to the same level), yet they overwhelmingly use predictive analytics to analyze financial data, as compared to non-financial data. The lack of experience using predictive analytics to analyze non-financial data is expected to result in auditors resisting the incorporation of results from this combination of model and data into their decisions. Yet, the same cannot be said for anomaly models as auditors experience using financial and non-financial data is approximately the same (hence a more balanced amount of time in the bar on the right side of the graph).

As a result, considering only the type of model or type of data individually, rather than a combination of these two factors, used by analytics could paint an incomplete picture of auditors' willingness to use the findings of analytics in their decisions. This difference in experience is expected to impact auditors cognitive fit and, in turn, decision making. When

predictive models identify a risk of misstatement, auditors will increase budgeted audit hours more (and presumably see a greater improvement in audit quality) when financial data is analyzed, as opposed to non-financial data. Yet, when anomaly models are used and identify a risk, no such difference is expected.

[Insert Figure 1 here]

III. THE EXPERIMENT

Participants

Koreff (2022) employed an experiment to test the aforementioned expectations, where the participants consisted of 98 auditors of all ranks employed by a variety of sized firms.[3] Follow-up interviews were conducted with 26 of the auditors that completed the experiment to obtain insights on their experiences using different types of analytics (described in Figure 1).

Description of Experimental Context

Participants were provided with background information related to their role as an in-charge auditor of a privately held, mid-sized sporting equipment manufacturer. Participants were told that their firm's Central Data Analytics Group had identified a potential misstatement with an estimated range that just exceeded performance materiality of $304,000. The conclusion stated that the use of predictive/anomaly models to analyze journal entries/emails presented a 56% risk that revenue was overstated by an amount between $270,000 and $310,000. As such, the risk identified was held constant, however the process used to arrive at that risk varied.[4]

[3] On average, participants had 9.0 years of audit experience. Sixty of the auditors were employed by national or international firms.
[4] The Central Data Analytics Group was described as consisting of non-CPAs without an accounting background. The likelihood of someone without an accounting background identifying an accounting misstatement is low. To make for a more realistic case, a risk of misstatement (as opposed to an actual misstatement) was said to have been identified by the Central Data Analytics Group.

Variables

The experiment manipulated the type of analytical model used (predictive or anomaly) and the type of data analyzed (financial or non-financial).[5] See Appendix B for specific descriptions of these manipulations. The participants were asked:

> Assume 30 hours were initially budgeted to audit revenue. How would you adjust the budgeted hours for the revenue account in percentages (every 5% change results in a change of 1.5 hours)

Results

Koreff (2022) illustrated that auditors with experience using analytics report comparable experiences using anomaly models and predictive models when answering "How experienced are you in using data analytics that identify statistical outliers such as unusually high/low fluctuations or ratios (anomaly models) as part of your job function?" and "How experienced are you in using data analytics that compare current data against previously identified issues/occurrences to identify similarities (predictive models) as part of your job function?" Both questions were measured on five-point Likert scales with endpoints of 1 = "Not at all experienced", and 5 = "Extremely experienced." No significant difference was identified between these measures with means of 2.590 (for anomaly models) and 2.559 (for predictive models).

Results in Koreff (2022) also showed that the type of model used and the type of data analyzed did not individually impact auditors' determination of budgeted audit hours, however budgeted audit hours were impacted by the combined impact of these two factors. See Figure 2 for a graphical depiction of the results. The results demonstrated that, when employing predictive analytics, auditors increased their budgeted hours more when financial data was used as

[5] In both manipulations the background information provided was limited in an effort to keep the case short. Future research may seek to examine the impact of providing additional detailed information.

compared to non-financial data (19.48% increase versus 11.38% increase, $p = 0.01$). However, when anomaly models were used, Koreff (2022) observed no statistically significant difference in the responses of auditors to the two data types (18.42% versus 14.16% increase, $p > 0.10$).[6] Additionally, when financial data was analyzed, auditors increased budgeted audit hours more when predictive models were used (19.48% increase versus 14.16% increase, $p = 0.09$). On the other hand, when data analytics used non-financial data, auditors were more likely to increase budgeted audit hours when anomaly models were used (18.42% increase versus 11.38% increase, $p = 0.07$).

[Insert Figure 2 here]

For additional insights, we conducted additional analyses replicating the primary results presented in Koreff (2022), while adding control variables for auditor age, years of audit experience, years of professional experience, title, and prior experience using data analytics. In all cases, the primary results of Koreff (2022) hold. We also considered the possibility that industry-expertise impacted auditors' use of the analytics as we controlled for auditors' percentage of time auditing manufacturing clients and a variable measuring if the auditor audits any manufacturing clients. These variables did not significantly impact results, and the results are consistent with the main results of Koreff (2022). Finally, we conducted analysis including only auditors employed by national and international firms in the sample. The primary results remained supported, consistent with the results reported by Koreff (2022).

IV. FOLLOW UP INTERVIEWS

Koreff (2022) conducted interviews of auditors that completed the experiment to provide additional insights into auditors' varying levels of experience using different types of analytics.

[6] Statistical analyses (i.e., ANCOVA results) documented in Koreff (2022) confirm that the evidence supports these conclusions.

When asked about prior experience using predictive analytics (specifically, "How would you describe the amount of experience you had using predictive analytics that analyzed financial vs. non-financial data?"), interviewees generally reported greater experience analyzing financial vs. non-financial data. When asked about prior experience using anomaly analytics (specifically, "How would you describe the amount of experience you had using anomaly analytics that analyzed financial vs. non-financial data?") auditors generally reported comparable experience analyzing financial and non-financial data.

V. IMPLICATIONS FOR PRACTICE

Despite the promise that data analytics have to improving the audit process, simply providing these tools to decision makers is insufficient to induce adoption (Messier 1995; Venkatesh et al. 2003; Schmidt et al. 2020a, b). Although firms are developing more advanced analytics, the results of Koreff (2022) suggest that auditors may not use these tools consistently. Effective implementation of these data analytics should account for auditors' prior experiences related to combinations of analytical models and the data processed by these models.

Koreff (2022) findings suggest that if new analytics are deemed effective by the firm, they still need to be cognizant of auditor's lack of experience using the analysis as a barrier to adoption (and potentially improving audit quality). Although auditors have comparable experience using the two types of analytics examined by Koreff (2022), consideration of the type of data these models tend to analyze revealed a disparity in the amount of time auditors spend analyzing different data by these types of models. While predictive analytics tend to focus on analyzing financial data, auditors reported anomaly models incorporating a more balanced amount of financial and non-financial data. This disparity ultimately impacts auditors' decisions. Therefore, public accounting firms should train their employees on how predictive models can be

effective using both financial and non-financial data to encourage consistent decision making. Firms should consider appropriate matching of analytic models to the data being analyzed or determine ways to ensure that auditors' experiences with different model/data combinations employed in practice do not vary substantially (e.g., through training sessions illustrating the use of analytic tools).

REFERENCES

Al-Natour, S., I. Benbasat, and R. T. Cenfetelli. 2008. The effects of process and outcome similarity on users' evaluations of decision aids. *Decision Sciences* 39 (2): 175–211.

Ameen, E. C., and J. R. Strawser. 1994. Investigating the use of analytical procedures: an update and extension. *Auditing: A Journal of Practice & Theory* 13 (2): 69–76.

American Institute of Certified Public Accountants (AICPA). 2015. *Audit data standards - Base standard*.

American Institute of Certified Public Accountants (AICPA). 2017. Guide to audit data analytics.

Appelbaum, D., A. Kogan, and M. A. Vasarhelyi. 2017. Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory* 36 (4): 1–27.

Barr-Pulliam, D., J. F. Brazel, J. McCallen, and K. Walker. 2020. Data analytics and skeptical actions: The countervailing effects of false positives and consistent rewards for skepticism. *Working paper*.

Brazel, J. F., K. L. Jones, and Q. Lian. 2022. Auditor use of benchmarks to assess fraud risk: The case for industry data. *Working paper*.

Brazel, J. F., K. L. Jones, and D. F. Prawitt. 2014. Auditors' reactions to inconsistencies between financial and nonfinancial measures: The interactive effects of fraud risk assessment and a decision prompt. *Behavioral Research in Accounting* 26 (1): 131–156.

Brazel, J. F., J. Leiby, and T. Schaefer. 2022. Do rewards encourage professional skepticism? It depends. *The Accounting Review* (Forthcoming).

Cao, T., D. Rong-Ruey, H.-T. Tan, and T. Xu. 2022. Enhancing auditors' reliance on data analytics under inspection risk using fixed and growth mindsets. *The Accounting Review* (Forthcoming).

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28 (1): 17–82.

Dowling, C., and S. A. Leech. 2014. A Big 4 firm's use of information technology to control the audit process: How an audit support system is changing auditor behavior. *Contemporary Accounting Research* 31 (1): 230–252.

Dunn, C., and S. V. Grabski. 2001. An investigation of localization as an element of cognitive fit in accounting model representations. *Decision Sciences* 32 (1): 55–94.

Eilifsen, A., F. Kinserdal, W. F. Messier, and T. McKee. 2020. An exploratory study into the use of audit data analytics on audit engagements. *Accounting Horizons* 34 (4): 75–103.

Glover, S. M., D. F. Prawitt, and T. J. Wilks. 2005. Why do auditors over-rely on weak analytical procedures? The role of outcome and precision. *Auditing: A Journal of Practice & Theory* 24 (Supplement): 197–220.

Goodhue, D. L., and R. L. Thompson. 1995. Task-technology fit and individual performance. *MIS Quarterly* 19 (2): 213–236.

Hobson, J. L., W. J. Mayew, and M. Venkatachalam. 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50 (2): 349–392.

Koreff, J. 2022. Are auditor's reliance on conclusions from data analytics impacted by different data analytic inputs? *Journal of Information Systems* (Forthcoming).

Koreff, J., M. Weisner, and S. G. Sutton. 2021. Data analytics (AB) use in healthcare fraud audits. *International Journal of Accounting Information Systems* 42: 100523.

Kuenkaikaew, S., and M. A. Vasarhelyi. 2013. The predictive audit framework. *The*

International Journal of Digital Accounting Research 13: 37–71.

Messier, W. F. 1995. Research in and development of audit decision aids. In *Judgment and Decision-Making Research in Accounting and Auditing*, edited by R. H. Ashton and A. H. Ashton, 207–227. Cambridge: Cambridge University Press.

Perols, J., R. M. Bowen, C. Zimmermann, and B. Samba. 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review* 92 (2): 221–245.

Public Company Accounting Oversight Board. 2007. *Report on the PCAOB's 2004, 2005, and 2006 inspections of domestic triennially inspected firms*. Washington, DC.

Public Company Accounting Oversight Board. 2008. *Report on the PCAOB's 2004, 2005, 2006, and 2007 inspections of domestic annually inspected firms*. Washington, DC.

Public Company Accounting Oversight Board. 2013. *Report on 2007-2010 inspections of domestic firms that audit 100 or fewer public companies*. Washington, DC.

Public Company Accounting Oversight Board. 2014. *In the matter of KPMG LLP's quality control remediation submissions*. Washington, DC.

Public Company Accounting Oversight Board. 2016. *Preview of observations from 2015 inspections of auditors of issuers*. *Staff Inspection Brief*. Washington, DC.

Public Company Accounting Oversight Board. 2018. *Strategic plan 2018 - 2022*. Washington, DC.

Public Company Accounting Oversight Board. 2021a. *Spotlight: Data and technology research project update spotlight*. Washington, DC.

Public Company Accounting Oversight Board. 2021b. *Data and technology research project update*. Washington, DC.

Schmidt, P. J., K. S. Church, and J. Riley. 2020. Clinging to excel as a security blanket: Investigating accountants' resistance to emerging data analytics technology. *Journal of Emerging Technologies in Accounting* 17 (1): 33–39.

Schmidt, P. J., J. Riley, and K. S. Church. 2020. Investigating accountants' resistance to move beyond excel and adopt new data analytics Technology. *Accounting Horizons* 34 (4): 165–180.

Seow, P. S. 2011. The effects of decision aid structural restrictiveness on decision-making outcomes. *International Journal of Accounting Information Systems* 12: 40–56.

Sinclair, N. 2015. How KPMG is using Formula 1 to transform audit. *Instutue of Chartered Accountants of Scotland*.

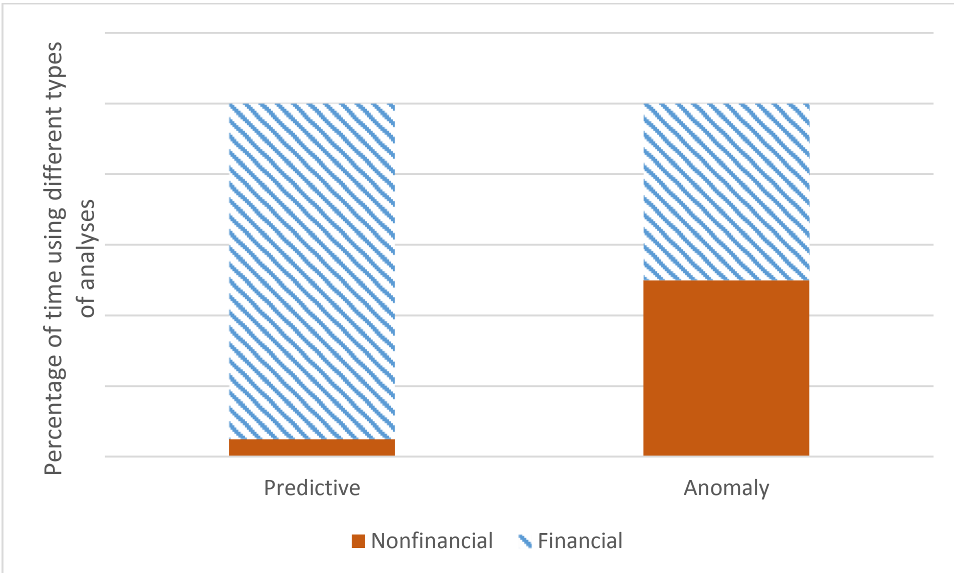
Statistical Analysis System (SAS) Institute. 2014. How a hybrid anti-fraud approach could have saved government benefit programs more than $100 million.

Trompeter, G., and A. Wright. 2010. The world has changed - Have analytical procedure practices? *Contemporary Accounting Research* 27 (2): 669–700.

Venkatesh, V., M. Morris, G. Davis, and F. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* 27 (3): 425–478.

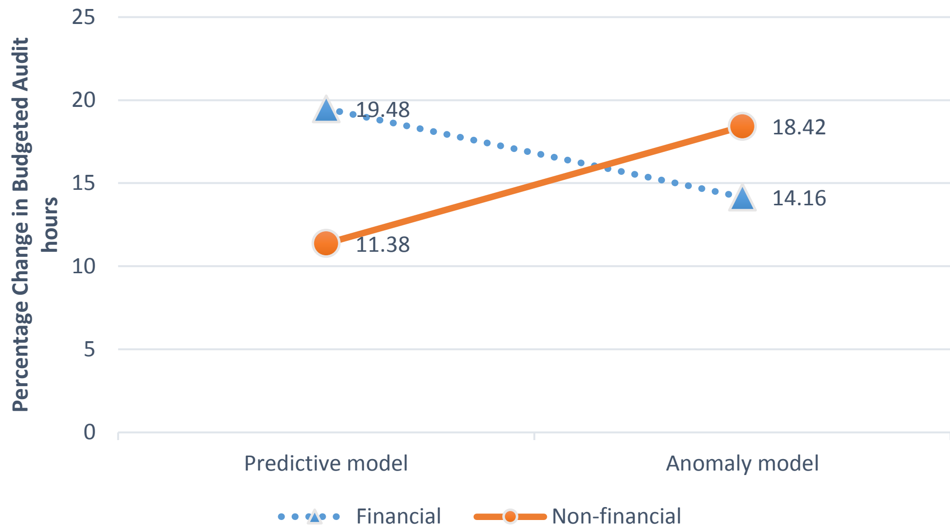
Vessey, I., and D. Galletta. 1991. Cognitive fit: An empirical study of information acquisition. *Information Systems Research* 2 (1): 63–84.

FIGURE 1
Depiction of Auditors Approximate Experience Using Different Models



This Figure provides a graphical depiction of the percentage of time spent using different types of data analytical models that analyze different types of data from the auditor participants in Koreff (2022). The auditors reported comparable levels of use of predictive and anomaly models (as shown by equal heights of the bar graphs). Anomaly models are used to analyze both financial and non-financial data, whereas predictive models focus primarily on analyzing financial information. In follow up interviews with 26 auditors that completed the initial experiment, Koreff (2022) confirmed this graphical depiction of experiences using different analytics. Interviewees also highlighted the tendency for predictive analytics to focus more on the analysis of financial data, whereas anomaly models use mixed data.

FIGURE 2
Graphical Depiction of Results for the Percentage Change in Budgeted Audit Hours



The dependent variable is auditors' percentage change in budgeted audit hours. The auditor participants used a slider scale ranging from -100% to 100% to indicate their desired change in budgeted audit hours. See the Method section for a description of the study's independent variables.

Appendix A – quotes from interviewees discussing prior experience and process familiarity impacting cognitive fit

Interviewee	Prior experience	process familiarity
1		
2	The hardest thing with the data analytics, would just be the clients like all of the firms have their own tool	
3		
4		
5		
6		
7		So I might not necessarily trust the first outcome of that. I would want to make sure that I had the right thought process or have the right data and want to make sure everything is right before I jump to any conclusions
8		Resistance would be if you can't rely on those statistics or if you don't understand what it's saying. Like when I think of Bedford's analysis. Know very much has been used sometimes as like a special procedure. Auditors have a basic understanding of what it is, but don't really understand deep down the statistics and how it works. So I think when you don't have a deep understanding of some of this information, some of this data that you use, it's less important.
9	... there's like a learning curve for sure and just getting used to the tool and kind of like it sounds good in theory but you want to make sure it obviously works right? You want to see it live in action	Where's this data going to get me and is this a waste of time
10		... the most resistance at first right, they didn't really understand how it how it was working until they like kind of did it themselves
11		
12		
13	... another thing is training. So, there were definitely times where it was just drop off this new tool and have fun. Not helpful. Give me a training on it, show me an example of how it could work. And then maybe I can think of how it might apply and how I could incorporate it.	
14		
15		

Appendix A (continued)

16	anytime there's new data, I'm a little bit nervous. … I think a big part of willingness has to do … honestly has it been done before? Is it tried and true? If it's tried and true or if it's been used and it's worked … I think if it's been tried before, an auditors willingness will increase. If the auditor has experienced with the process or with the client I think there can probably be higher willingness to use certain analytics.	Um, so using new data for me is, I feel like I have to go through a few steps before I feel comfortable with the data… An auditors' willingness [to using new analytics] will increase If the auditor has experienced with the process
17		
18		I think part of its [willingness to use analytics is] understanding for sure. I think if someone can understand how the analytics are getting to their certain conclusion then I think that that's really helpful
19		
20		
21		
22		
23		
24	Two is previous experience utilizing them if you know the tests that are being generated don't come up with results. You might not find the value in the use of data analytics.... So I think you know It all comes down to your experience using it … So I'd say those are probably the largest ones is the lack of experience, previous experience and I think that's why people would push away from using data analytics … There's a large learning curve there and with the learning curve becomes more time and with more time tends to become a principal pushing away from these procedures	
25		I think it's [auditors willingness to using data analytics is] partly understanding, like how they're used and what information can be gleaned from an actual data analytic
26		

Appendix A (continued)

This Table shows the interviewees prior experience using a technology enabled tool (e.g., data analytics) and process familiarity of the analysis inducing use of that tool. The first column ("Interviewee") represents the interviewee number. The second column ("prior experience") includes the quote that best depicts the interviewee's discussion of how prior experience induces use of a tool. The third column ("process familiarity") includes the quote that best depicts the interviewee's discussion of how prior experience using a certain analysis induces use of a tool.

Appendix B – Manipulation descriptions

In the predictive models condition, participants were provided a description of the predictive models used that read:

> The Central Data Analytics Group employs predictive analytical models to identify patterns that are similar to previously identified issues. Predictive models rely on prior historical data to identify patterns and predict future events. Predictive models compare information in the data collected from clients associated with previously identified events/occurrences to current information. Predictive models may be used in the audit process to identify a pattern over several years associated with a previously identified material misstatement that may be indicative of a current material misstatement.

In the anomaly models condition, participants were provided a description of anomaly models used that read:

> The Central Data Analytics Group employs anomaly analytical models to identify statistical outliers. Anomaly models rely only on current year (non-historical) data to identify statistical outliers. Anomaly models compare information in the data collected from your firm's client base to identify very high or low amounts or ratios. Anomaly models may be used in the audit process to identify very high or low ratios (i.e. gross margin, debt to equity, current ratio) that may be indicative of a current material misstatement.

The second variable manipulated between participants was the type of data analyzed. In the financial data condition, participants were told:

> The Central Data Analytics Group is capable of identifying journal entries that affect revenue. For the Madison audit, the Central Data Analytics Group used this financial information to identify the number of journal entries that include revenue and were made just below the performance materiality threshold. Although the Central Data Analytics Group has explained what criteria they use for "just below the performance materiality" for the journal entries, this explanation contained substantial statistical jargon and was not well understood by your audit team. Several of your colleagues have reported similar issues with explanations received from the Central Data Analytics Group.

In the non-financial data condition, participants were told:

> The Central Data Analytics Group is capable of identifying sentences in the e-mails that discuss revenue. For the Madison audit, the Central Data Analytics Group used this non-financial information to identify optimistic language used in internal and external e-mails for sentences that discuss revenue. Although the Central Data Analytics Group has explained what criteria they use for "optimistic language" in the e-mails, this explanation contained substantial statistical jargon and was not well understood by your audit team. Several of your colleagues have reported similar issues with explanations received from the Central Data Analytics Group.

Appendix C – quotes from interviewees discussing experience using anomaly and predictive models

Interviewee	Predictive	Anomaly	difference
1	As much as we could on financial to obviously we tried to make it as scientific as possible…	none	(cant compare)
2	that's more of a financial thing. So, say, okay, based on Q4 retakes we would establish like an expectation	I would say financial is more	I would say that they're fairly comparable
3	The financial data is what matters	they usually start out looking at financial and then you can use the non-financial to dig in further and get a better understanding	I would lean towards anomalies using more non-financial data. Predictive I feel like you can do that just using financial data
4	you have to kind of look at both of them [financial and nonfinancial data] together because when you're doing an analytic, you have to understand what some of the qualitative factors are … 80% are a number, are based on financial data and then the other 20% is is uh, is is non financial	none	(cant compare)
5	none	none	(cant compare)
6	Primarily financial data. I mean, you take a look at sub ledgers also. Say your inventory sub ledger is a piece of your ledgers. You will look at your payroll reports if there's a split there, if it's an identified a specific factor that's not necessarily financial in nature. You may probably use that as a point of inquiry	it's kind of hard to detach the two from each other. It's because you kind of use your knowledge of any non financial circumstances to kind of frame what would be your anomalies. With that point of reference in mind. And that's kind of what helps guide your review of the financial data	you have your anomaly analytics that uses non financial data to help identify what would be considered an anomaly, but also within the frame of your financial data … Predictive analytics as you would be primarily using the financial
7	none	It's probably more towards financial data, I would say. … I mean, maybe it's like a 75-25% ratio	
8	I mean most of it tied into financial data, but maybe number of employees, volume of clients … We definitely 90% of the time would use financial data	90% financial, 10% non financial	
9	I feel that the non financial data will will run in line with the actual financial data … I mean, I probably 100% of time rely on non financial data.	none	
10	none	primarily, it was financial data	
11	I tend to do more with financial data, it's easier and more reliable for me to grasp and understand rather than bringing in a specialist…	trend is usually the easiest way sometimes we'll use a ratio if we're looking at like a gross margin assessment year over year… I wouldn't say it's [nonfinancial data] a significant input into that analytic	I use the more predictive with a financial input on a regular basis as that tends to be secondary audit evidence to like a sample or a confirmation and it does help drive a conclusion… primarily the financial that we're using in the anomaly analytic prior year values or kind of a ratio of prior year to current year
12	we have some like keywords or things that will trigger fraudulent entries. So the tool is built to recognize those kinds of keywords or phrases and then we use those keywords/ phrases to select huge amounts of journal entries for additional review…Finding the data for the financial data because a lot of times we use historical data, which, you know, which is obviously you know more than one or two years, the more data you have, the better the more reliable your expectation is. So we spend a lot of time gathering the data, gathering the historical information. If you're worried about non financial data or trying to develop an expectation for payroll expenses this year, for example. You know, that might even be more straightforward to keep	It's more financial data	
13	I would say my experience with the non financial piece is definitely much higher than the financial piece	For the bell curve piece with the journal entries, I think they go hand in hand there … For the anomaly analytics, I don't think that there was a disparity within inputs	For the anomaly analytics, I don't think that there was a disparity within inputs. I don't think there was for either one now I think about it
14	"I think it really depends"	financial data, looked @ non-fin	
15	I would say financial	[discusses the use of both]	I say the predictive were more financial based… I'd say anomaly based ones, they're probably more non financial

Appendix C (continued)

16	at [firm name] definitely more financial for sure	usually with financial data, we are just looking at what's popping out in terms of a percentage. Whereas I guess we would also compare that to the non financial aspect of that would be where do we have the most acquisitions. So I think whenever we were using financial data. We had kind of like a non, more of a non biased view of looking at it. But then we did when we did compare that financial data to non financial data … So we did put a lot of weight I think into the non financial	I think we use a little bit more of the building expectations and comparing them to actual than we did the comparing financial data to non financial data … I think we would probably, thinking back to that inventory example of using what we know versus using Actual numbers for inventory balances. I mean, I would say probably a 50/50 Mix. We do need to look at the numbers, but we also need to use professional judgment and determine where the risk is. And that usually comes from Non financial data so I would. Yeah, I would say probably 50/50
17	Primarily financial data was the focus. In order to interpret the results we certainly needed to have non financial data, but all of the inputs were financial	That was pretty much financial data	I would say they both relied on financial data pretty heavily
18	My experience personally with non financial data is probably limited. So I think in terms of financial data, I think that it's more prevalent	Probably financial	I would say in a general sense that I think that predictive analytics is more accessible to both. Whereas I would say that if you're looking at certain anomalies I could see it being more qualitative [nonfinancial]
19	we definitely lean towards using numbers, rather than explanations and qualitative… I don't really think I did [use nonfin for pred analytics]	none	(cant compare)
20	I try to stick to financial data… more often than not, I would say I would use financial data	I would say again, mostly financial data that we use	I would say it's pretty consistent among the two so I would say 50/50
21	Probably about comparable	none	(cant compare)
22	Probably more on financial data for the most part	none	(cant compare)
23	I definitely don't think I used non financial data in the way that I started to in [year of data collection]	none	(cant compare)
24	it was still a big large financial component that there's certainly a ton of non financial components included but In terms of data analytics It was not a ton of non financial components	doing a lot more non financial information as part of the procedures and now I would just say that it's incorporated into the risk assessment of using more non financial information to understand you know, specific risk that are out there within the financial statement data itself	
25	They [predictive analytics] were more financial in nature … I would say it was mostly financial	we would always tie it back to something financial and then you know if there's a related line. If there's a non financial metrics like number of employees or, you know, hours worked or something like that that we can tie it back to for non financial we would do that but I guess it's difficult for me to compare financial and non financial … It just depends on the line item and it was, you know, just sort of whatever we needed to do to get a relevant metric to look at.	with the anomaly analytics, we would do more like ratios and more ratios, which would be based on financial data, more so than non financial whereas the predictive if you're looking at different revenue line items or cost lines, you would have more non financial data in there to just look at trends
26	… generally we would stick to historical financial information where possible	They would all be driven by financial information that's included on the accounting records	

Appendix C (continued)

This Table shows the interviewees description of their experience using different inputs for different analytics. The first column ("interviewee") represents the interviewee number. The second column ("predictive") includes the quote that best depicts the interviewee's experience using predictive analytics to analyze financial vs. non-financial data. The third column ("anomaly") includes the quote that best depicts the interviewee's experience using anomaly analytics to analyze financial vs. non-financial data. The fourth column ("difference") includes the quote that best depicts the interviewee's comparison of the proportion of time predictive vs. anomaly analytics use financial vs. non-financial data.