Data analytics in auditing: Opportunities and challenges

Christine E. Earley

School of Business, Providence College, 1 Cunningham Square, Providence, RI 02918, U.S.A.

KEYWORDS
Data analytics; Auditing profession; Auditor skills; Future of auditing

Abstract In this article, I provide background regarding a hot topic in the public accounting profession: the rise of big data and the related field of data analytics (DA). The tax and advisory practices of public accounting firms have embraced the use of DA, and firms have made significant investments in growing these practice areas. Although DA holds great promise for the auditing practice as well, the use of widespread DA on audit engagements has lagged behind other practice areas. This is due to the fact that auditing presents unique challenges in the adoption of DA that are not relevant for other practice areas. Despite the impression that DA is not being embraced as readily in auditing, public accounting firms are continuing to make significant investments in developing audit-related DA, and it is only a matter of time before we start to see the transformational impact of these efforts. The purpose of this article is (1) to explain how DA applies to financial statement audits and why it could represent a game changer in how audits are conducted, and (2) to provide a context for researchers in terms of problems to be addressed related to DA.

© 2015 Kelley School of Business, Indiana University. Published by Elsevier Inc. All rights reserved.

1. Data analytics: A game changer for public accounting

The term big data and the related approaches to analyzing data, often referred to as data analytics (hereafter, DA) or predictive analytics, have been discussed at length in the popular press and academic journals—to the point of oversaturation in recent years. University programs have been developed to address DA competencies, seemingly overnight (Briggs, 2013). Indeed, at the American Accounting Association (AAA) annual meeting in August 2014, a panel session co-sponsored by PricewaterhouseCoopers and the University of Illinois was held to discuss how the accounting curricula must adapt to incorporate more data analysis courses (PwC, 2015). The message of the panel was that in order for students to be competitive in both audit and tax, they must learn to become data scientists. Big data is seen as the wave of the future in business, and any organization that falls behind in its development of DA capabilities is expected to lag.
its competitors and could experience dire consequences to the bottom line. In a recent survey of CFOs and CIOs conducted by KPMG (2014), 99% of respondents noted that data and analytics is at least somewhat important to their business strategy, and 96% expressed that they could make better use of big data within their organizations. Service organizations, such as public accounting firms, are in the race to provide better and more comprehensive DA services to their clients, but the question still remains as to how they will actually accomplish this. Another question is whether the core business of public accounting—that is, auditing—will benefit from an investment in DA capabilities or whether DA is ultimately more in the domain of consulting.

The purpose of this article is (1) to explain how DA applies to public accounting firms and why it could represent a game changer in how audits are conducted, and (2) to provide a context for researchers in terms of problems to be addressed related to DA. The potential of DA to improve the practice of auditing is quite significant, but there are many challenges that need to be overcome before widespread use of DA in auditing becomes a reality. The role of DA in auditing has been discussed in several recent academic articles (e.g., Alles, 2014; Alles & Gray, 2014; Brown-Liburd, Issa, & Lombardi, in press), but many of these are thought pieces, the purpose of which is to develop frameworks for approaching the literature related to DA (e.g., Alles & Gray, 2014) and provide specific research questions that can be addressed in future studies (e.g., Brown-Liburd et al., in press; Gray & Debreceny, 2014; Wang & Cuthbertson, in press). The Emerging Assurance Technologies Task Force of the AICPA Assurance Services Executive Committee (ASEC) has also authored a white paper describing in great detail how DA could be used on audits (AICPA, 2014). Despite an interest in DA by academics, empirical academic research related to DA in auditing is still in its infancy, due in part to the lack of information being provided by public accounting firms about their approaches to DA. This article will discuss broad areas of emphasis intended to drive research questions related to DA that are of interest to public accounting firms, users of financial statements, and regulators.

This article is organized as follows. Section 2, which appears next, provides an overview and defines big data—and related DA—in general. Section 3 discusses DA in the context of public accounting firms for all three service lines of audit, tax, and consulting. Section 4 discusses the benefits of DA within auditing in particular. Section 5 discusses the challenges of adopting DA in auditing. Section 6 explores how academic research can inform the adoption of DA in auditing. Finally, Section 7 provides an overall conclusion.

2. What are big data and data analytics?

Big data has been defined as “high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making” (Gartner, 2013). The characteristics of volume, velocity, and variety, which describe the features that make big data unique, are often referred to as the Three Vs of big data (Alles & Gray, 2014). However, as Gartner explains, big data must be analyzed or processed in an innovative way in order to be relevant and useful for decision making. Indeed, big data as a concept is often discussed in conjunction with analysis of the data. For example, as noted by Alles and Gray (2014), big data in the accounting literature is often defined by the types of analysis that can be performed with the data, such as DA or predictive analytics, rather than as a type of data source. In addition, the source of the data can vary. Alles and Gray (2014, p. 5) note specifically:

To auditors, the data in (or contents of) big data refers to collections of multiple types of data, which could include some mix of traditional structured financial and non-financial data, logistics data, sensor data, emails, telephone calls, social media data, blogs, as well as other internal and external data.

The availability of large amounts of computerized data in companies has been steadily increasing over the years, but recent advances in processing speed, cloud storage, and the rise of social networks has changed the ease of access to data and the nature of data that can be captured and stored for later use. At the same time, software used to analyze large volumes of data (i.e., data mining tools) as well as more sophisticated data visualization tools can potentially increase the ability of individuals to understand the story that the data is telling them (AICPA, 2014; Capriotti, 2014; Whitehouse, 2014). DA approaches are very similar to the methods used by academic researchers in conducting empirical research. Crawley and Whelan (2014) acknowledge the overlap between DA techniques and academic research, and discuss how academic researchers could expand their use of DA techniques to enhance and improve the types of research questions they can address. In DA, as in academic research, large amounts of data are collected and tested to ensure
they are sufficient for addressing a particular research question of interest and then analyzed using statistical packages or other software programs that identify patterns or relationships among data items. The steps that follow—analyzing and interpreting the output from these tools—require a high level of expertise on the part of the researcher—or in the case of public accounting firms, the consultant or auditor. The analysis of the data output is only able to be completed by individuals with the ability to engage in expert behaviors such as pattern recognition and critical thinking; therefore, it is impossible with current technological capabilities to fully automate the DA process.

3. How do public accounting firms plan to leverage DA?

DA has been a significant area of investment for public accounting firms, mainly in the advisory (consulting) practice, but also in tax and more recently in auditing. Many companies collect massive amounts of data about customers, the external environment, competitors, etc., and often don’t know how to take the next step of analyzing and applying the data to operating their businesses. An overwhelming 85% of managers surveyed by KPMG (2014) noted that one of their biggest challenges was figuring out how to best analyze the data they have collected. Therefore, public accounting firms’ advisory practices have found a niche in DA that builds upon the expertise they have gained in business analysis over the years. DA has been used in fields outside of accounting (e.g., marketing, human resources) for quite some time, but has only recently been applied to financial transaction data and external data that allows management to make strategic decisions about the business. Although consulting services related to DA represent a positive growth opportunity for firms, issues arise when audit clients seek advisory services related to DA from their auditors, as most of such projects fall under the umbrella of services prohibited by the Sarbanes-Oxley Act of 2002 (SOX).

In the tax area, DA has been used mainly by accounting firms to help their clients engage in tax management approaches (Blanchard, 2014; Deloitte, 2013). For example, as noted by Blanchard (2014), companies that are faced with navigating the increasingly complex U.S. tax code as well as high U.S. corporate tax rates can use DA tools to track tax rates and calculations for multiple tax jurisdictions and automate the monitoring of tax compliance. Deloitte (2013) explains that proper application of DA can enable companies to reduce tax errors, find tax-saving opportunities, and cut administrative costs. PwC (2015) suggests that companies can better use their Enterprise Resource Planning (ERP) systems to analyze financial activity and make decisions that will ease the burden on the tax department when conducting year-end compliance tasks. Since the IRS and state governments are using their own form of DA to find tax fraud and filing errors, accounting firms can help firms to monitor their own compliance to reduce the risk of a tax audit.

DA is being approached differently in the auditing practice of public accounting firms than in the advisory practice. The focus is less on innovation and competing on visually appealing analytics to impress clients, and more on trying to improve the effectiveness and efficiency of audits. The pace of adoption of DA in the audit has been slower than in other fields, such as in advisory or forensic investigation practices (Katz, 2014; Whitehouse, 2014). Given the liability concerns and the highly regulated environment of auditing, it makes sense that firms would be more careful when venturing into DA in the audit practice. Yet DA is being touted as the future of audit (Liddy, 2014; Lombardi, Bloch, & Vasarhelyi, 2014). As Capriotti (2014, p. 38) stated, it “has the potential to be the most significant shift in how audits are performed since the adoption of paperless audit tools and technologies.” As discussed next, use of DA in the audit presents significant opportunities, but also significant challenges to firms.

4. The promise of DA: Why make the investment?

There are four primary benefits of using DA on audits: (1) auditors can test a greater number of transactions than they do now, (2) audit quality can be increased by providing greater insights into clients’ processes, (3) fraud will be easier to detect because auditors can leverage tools and technology that they already use, and (4) auditors can provide services and solve problems for their clients that are beyond current capabilities by utilizing external data to inform audits. Regarding the first benefit, one way in which DA can improve audits is by increasing the sufficiency (i.e., the appropriate amount) of audit evidence. Currently, auditors apply a risk-based model and sample transactions to determine that account balances are fairly stated. DA will allow auditors to automate testing of transactions, and in theory, 100% of the population can be tested. According to Jim Liddy (2014, p. 2), Vice Chair of Audit and Regional Head of Audit for the Americas practice of KPMG, LLP:
In the future, using high powered analytics, auditors will have the capacity to examine 100 percent of a client’s transactions. We will be able to sort, filter, and analyze tens of thousands or millions of transactions to identify anomalies, making it easier to focus in on areas of potential concern and drill down on those areas that may have the highest risks. This will enable us more than ever before to help assess risks and identify trends through the audit process. With smart data each year’s audit will “learn” from prior years, exposing areas of possible risk and building a self-enriching knowledge base to better inform companies and their investors.

In this quote Liddy addresses the increase in number of transactions tested, from the small samples used currently to a possible 100% of the population, but he also addresses the difference in focus from errors in the sample to anomalies in patterns of data about the population. The focus on identification and analysis of anomalies is consistent with other discussions of DA in the literature (e.g., Brown-Liburd et al., in press; Capriotti, 2014; Whitehouse, 2014). Anomalies are instances where the data does not match the auditor’s expectations based on his or her knowledge of the client’s business. One example would be the presence of sales to customers that the client has identified as non-creditworthy or bankrupt; therefore, one would expect no sales to be recorded for those customers. This relates to the second benefit of DA, which is to increase the quality of audits by providing greater insights into clients’ processes. As Liddy (2014) notes, an additional benefit of DA is the ability of auditors to build a database of knowledge about each engagement that can be transferred from year to year; for instance, information about how those anomalous transactions were resolved would inform auditors in the following year as they develop their expectations.

Table 1 presents a comparison of how data is currently used on audits to how data might be used as DA approaches become more prevalent.

A third benefit of DA is improved fraud detection on audits (Capriotti, 2014; Gogtas, Pollner, & Bay, 2007; Gray & Debreceny, 2014; McGinty, 2014). Indeed, at the meeting of the Public Company Accounting Oversight Board (PCAOB) Standing Advisory Group (SAG) in November 2014, where fraud detection by auditors was the subject of three panel sessions, the topic of how DA can be used in fraud detection was discussed by both panelists and SAG members. The general consensus was that DA is promising for fraud detection because software tools that enable auditors to analyze large sets of data in an efficient way can be applied at very low cost to audit firms (AICPA, 2014). These tools, referred to as computer-aided audit techniques, or CAATs, are not new to auditing firms, but their use has been limited, due in part to lack of acceptance by audit personnel (Curtis & Payne, 2014). The increased availability of client data coupled with increased pressure to compete in the DA space with other audit firms may change auditors’ attitudes toward the use of CAATs to detect fraud. For example, as noted by McGinty (2014), auditors are increasingly using CAAT tools such as Benford’s Law, a mathematical principle that can be applied to data sets, that have proven effective in identifying fraudulent transactions.

The fourth benefit of DA in auditing is the possibility of using non-financial data (NFD) and external data to better inform audit planning (particularly in assessing risk), and to more effectively audit

---

1 For more information, see the webcast of the November 2014 SAG meeting at: http://pcaobus.org/News/Webcasts/Pages/11202014_SAG.aspx
in accounting programs have been expected to be proficient in understanding how to apply accounting rules and to understand audit risks associated with particular accounts. For example, new auditors would be expected to explain how to account for a sales transaction (i.e., by debiting accounts receivable and crediting sales revenue) and then understand the risks associated with possible overstatement of total sales revenue and total accounts receivable. They are not typically trained to consider whether the transaction itself makes sense or to develop expectations about sales that would then enable them to recognize when an anomaly has occurred, or more importantly, how to follow up on the anomaly once it is discovered. Indeed, one concern of regulators is that auditors will lack the requisite skills to properly apply DA techniques, and firms will begin expanding advisory services in order to attract and hire data scientists with DA skills. These data scientists have a different mindset than traditional auditors, and this potential shift in focus from auditing to advisory services has regulators concerned about audit quality (Katz, 2014).

There are several different approaches firms could take toward addressing gaps in expertise of auditors, in addition to training them in DA techniques. One possibility entails outsourcing the majority of DA to offshore centers and just providing engagement auditors with the output of DA to inform additional audit testing procedures. Another option involves creating tools that automate as much of the process as possible and categorizing anomalies into manageable ‘buckets’ so that auditors can apply judgment in addressing the smaller number of anomalies identified by these tools; for example, the aforementioned AICPA (2014) white paper discusses the possibility of reducing false positives to identify only “exceptional exceptions.” Regardless of the approach taken, the DA environment will result in auditor judgment playing a much more significant role than in sample-based auditing due to the potential of large numbers of anomalies to evaluate. In this scenario, auditors will have to develop a strong understanding of what constitutes proper accounting in the context of their clients’ business in order to perform this evaluation. Additionally, false positives (i.e., tool-identified anomalies that turn out to be appropriate transactions) remain a significant concern of auditing firms regarding the adoption of DA approaches (Wang & Cuthbertson, in press), and it remains to be seen whether automated tools can eliminate false positives or reduce them to a manageable amount. Too many false positives have the potential to focus auditors’ attention in areas

5. The challenges of DA

Despite the promise of DA for improving audit quality, there are numerous challenges to widespread implementation of DA on audits. The challenges fall into three broad categories: (1) training and expertise of auditors; (2) data availability, relevance, and integrity; and (3) expectations of the regulators and financial statement users. Regarding the first challenge, Brown-Liburd et al. (in press) suggest that the proliferation of large amounts of data—a significant amount of which may be non-financial—could overwhelm the information processing capabilities of auditors. Skills such as pattern recognition and understanding how to evaluate anomalies have traditionally not been the main focus of accounting education and training within public accounting firms, and are often acquired through many years of experience in the field. Conventionally, auditors fresh out of undergraduate and masters
that are ultimately not at risk of being misstated (Whitehouse, 2014).

The second major challenge centers on data availability, data ownership, and data integrity. Many clients may lack the ability to capture data in a way that is useful to the auditor, or the data might contain a lot of noise. Furthermore, data could have been collected by the client, but it is unclear what level of access the auditor will be granted and what data the client will share with the auditor. According to Gray and Debreceny (2014, p. 378), this lack of access to client data is a potential drawback to adapting data mining for fraud detection: “Currently, clients do not give auditors direct access to their databases.” In addition, if the auditor must rotate off the engagement due to losing the client, or due to mandatory rotation in the EU, the auditor may be required by law to give up access to the client’s data, which makes it difficult to build industry databases or leverage knowledge across clients in the same industry. Even if the data can be readily provided and the auditors are granted full access, they will have to consider the integrity of the data (Alles & Gray, 2014; Appelbaum, 2014; Liddy, 2014; Whitehouse, 2014).

Since big data may come from both internal and external sources, the auditor needs to assess whether the data originated from a secure source and whether or not it may have been tampered with before the auditor obtained it. The PCAOB has expressed concerns about this in its inspections of public company audits (Whitehouse, 2014). Completeness of the data set is also an issue. As explained by Alles and Gray (2014, p. 28), in fields outside of auditing, a bit of ambiguity or lack of quality of data sets may be tolerated, but for auditors, “allowing some inaccuracies to ‘slip in’ is difficult to reconcile with the focus in auditing on data integrity.”

The third major challenge involves how DA is viewed by investors and regulators. Over the years, the auditing profession has dealt with an expectations gap regarding what outsiders expect of the auditor versus what standards require auditors to deliver. The gap in expectations occurs when users believe that auditors are providing 100% assurance that financial statements are fairly stated, when in reality, auditors are only providing a reasonable level of assurance—which, due to sampling of transactions on a test basis, is somewhat less than 100%. Given the potential to audit 100% of transactions, it is possible that DA could exacerbate the expectations gap issue because users would be even more likely to require 100% assurance. It is possible that boards of directors and users of financial statements would hold auditors to a higher standard of fraud detection and liability for uncovering financial statement misstatement. For example, as noted by Gray and Debreceny (2014, p. 378), under traditional auditing, the auditor has a defense for not uncovering fraud if the sample they selected happens to be free of the smoking gun that would otherwise indicate fraud: “Data mining can be considered the equivalent to taking a 100% sample. If the smoking gun is in that sample, but the auditors missed it, then the auditors no longer have their traditional industry-practice defense.” In addition, given the focus of DA on non-financial information, regulators fear the possibility that “auditors may be paying less attention to auditing their clients and more attention to providing them with non-audit services” (Katz, 2014, p. 2). Finally, current auditing standards are not written taking into account DA approaches to the audit (AICPA, 2014; Whitehouse, 2014), and standard setters would have to consider adapting standards to these new approaches. For example, standards that are based on auditors drawing conclusions based on a sample of evidence might have to change significantly to accommodate the testing of 100% of transactions, or standards might have to be written to focus on testing data integrity.

6. Academic research on DA requires a shift in focus

As mentioned previously, although numerous academic articles have been written about the future of DA in auditing, relatively little empirical academic research has been conducted addressing the issues presented in this article. Several academic pieces, however, present specific research questions to be addressed in DA. For example, Wang and Cuthbertson (in press) interviewed a practitioner with over 30 years of experience developing tools used for analytics for both internal and external auditors. They identify eight categories of research questions academics can address, including the role of DA in risk analysis, what procedures should be performed, the implications of testing 100% of the population, whether external data should be used, the role of internal auditors’ use of DA, the interpretation of DA results, consequences of using DA, and whether the profession needs a DA framework. Brown-Liburd et al. (in press) focus more on the skill sets and characteristics of auditors performing DA work. They suggest that researchers should focus on auditors’ mental models, expertise development (including generational differences between auditor skill sets), information processing issues such as auditors’ ability to deal with large amounts of information, how both firm training and accounting curricula should be adapted to prepare auditors for a DA environment, and how tools can be developed
and/or borrowed from other fields to aid auditors’ decision making. Alles and Gray (2014) suggest research overlapping many of these areas, but they also suggest that research outside of the external audit domain, including research related to internal auditing or research related to other public accounting firm activities—like forensic or big data consulting—may be helpful in informing how DA might impact audits.

Researchers should additionally investigate how adoption of DA impacts the business risk of the audit firm itself from a liability standpoint (Gray & Debreceny, 2014) or from the standpoint of being subject to regulatory sanction, which could range from fines to being driven out of the audit business altogether. Research into investor expectations, juror decision making, and board of directors’ perceptions of the level of assurance provided in DA audits versus traditional audits remain important avenues of future study. Another area that academics may be able to address is how the client’s own transaction information can be analyzed to detect errors in accounts. For example, Vandervelde et al. (2008) modeled the relations between accounts in order to determine if account errors could be detected, and found that relatedness of accounts is an important factor that auditors need to consider. Relatedness refers to how accounts are recorded in the client’s records to ultimately create financial statements. Accounting is based upon double-entry bookkeeping, where accounts are debited on one side and corresponding accounts are credited on the other side to create an entry that is symmetrical, or in balance. Certain debit-credit relationships make sense, like a debit to an asset account and a corresponding credit to a revenue account. However, some relationships might appear unusual, such as a debit to an expense account and a credit to an owner’s equity account. DA tools could flag these suspicious account relationships so that auditors can investigate them further. As DA tools are being developed by firms, modeling of relevant features similar to those in Vandervelde et al. (2008) would be very informative. In addition, researchers should examine whether economic data and social media data from external sources can be modeled to make predictions about factors that impact a client’s business, which will aid auditors in planning and making business risk assessments.

7. What does DA mean for auditors of the future?

This article presented some of the challenges impacting the widespread adoption of DA in auditing. Again, there is significant promise for improving audit quality through the use of DA, but significant hurdles will need to be overcome. Academic research can inform the myriad questions facing firms as they develop these approaches, but time is of the essence since the pressure to apply DA to auditing is growing. Companies are making investments in big data to improve their own decision making, and they expect auditors to be able to leverage big data to improve the effectiveness and efficiency of audits as well.

Educating students who are entering the public accounting profession and providing existing auditors with expanded skill sets to be able to perform data analysis effectively is another way academics can contribute to addressing the skills gap and expertise challenge associated with DA in auditing (PwC, 2015). As noted previously, the message conveyed by a panel of practitioners and academics at the AAA annual meeting was that accounting programs may need to be revamped to train students to be better data scientists. Indeed, a recent white paper by PwC (2015) calls for a major overhaul of the accounting curricula to include more courses in programming (Python or Java), structured and unstructured databases, multivariate and inferential statistics (including the programming language R), and data visualization tools, among other desired skills. However, it seems that higher-level skills (e.g., pattern recognition, critical thinking) and increased training in analytical procedures should equally be encouraged. Some caution needs to be exercised before a wholesale revamp of the accounting curricula is undertaken, particularly since large accounting firms are developing tools that will enable auditors to work with data without having to conduct the underlying programming themselves. Perhaps rather than invest in more statistics courses or hire faculty with backgrounds in big data, programs should focus more deeply on ensuring that students understand relationships between financial statement accounts, business processes, and how external factors drive business risk. Students should also comprehend how patterns of financial information can tell a story about a company’s performance. It seems that a deep understanding of not only how accounting occurs but also why will enable auditors of the future to better analyze the data provided to them through visualizations and to thrive in a big data environment.

Acknowledgments
I wish to thank Kate Jelinek for the invitation to write this article. I would also like to thank Phil McCollough for his helpful comments on an earlier draft.
References


