# THREE ESSAYS ON AUDIT INNOVATION: USING SOCIAL MEDIA INFORMATION AND DISRUPTIVE TECHNOLOGIES TO ENHANCE AUDIT QUALITY

By ANDREA M. ROZARIO

A dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Dr. Miklos A. Vasarhelyi

and approved by

Dr. Miklos A. Vasarhelyi

Dr. Alexander Kogan

Dr. Helen Brown-Liburd

Dr. David A. Wood

Newark, New Jersey

May 2019

**Copyright page:**

## ABSTRACT OF THE DISSERTATION


### THREE ESSAYS ON AUDIT INNOVATION: USING SOCIAL MEDIA INFORMATION AND DISRUPTIVE TECHNOLOGIES TO ENHANCE AUDIT QUALITY

**By Andrea M. Rozario**

**Dissertation Director:**

**Professor Miklos A. Vasarhelyi**

Advances in technology occur at exponential rates and are transforming business practices. Alles (2015) suggests that audit clients' use of advanced technologies is likely to be the driver of adoption of such technologies by auditors. As a result, it is not surprising that the audit community, including academics, regulators, and audit professionals, are debating to what extent the use of technology will impact auditing (IAASB 2016; PCAOB 2017b). However, the impact of technology on auditing remains unclear. To provide insights into this debate, this dissertation explores the evolution of auditing as a result of nontraditional audit evidence, RPA (robotic process automation), and blockchain and smart contracts.

The first essay examines the usefulness of third-party generated information about firms' brands and products from social media in enhancing substantive analytical procedures for the revenue account. The research questions in this study address whether Twitter measures of consumer interest and consumer satisfaction can improve the prediction performance and error detection performance of substantive analytical models. Extant research has documented that external nonfinancial information from Internet

platforms can be useful for predicting firm performance and stock market prices, accordingly, it is important to examine whether such type of information can be used as an external source of audit evidence and enhance the effectiveness of audit procedures. The results of the study suggest that analytical models with Twitter proxies experience improved prediction and error detection performance than models that do not contain this information. Especially the analytical model that contains prior month sales, gross domestic product, and consumer interest as this model produces superior predictions and detects accounting errors for most of the industries that are examined. Collectively, these findings indicate that auditors can benefit from including social media information in analytical models as it can complement macroeconomic information and substitute contemporaneous firm-specific information such as accounts receivable.

The second essay proposes and implements a framework for RAPA (robotic audit process automation) to foresee the evolution of auditing as a production line. The redesigning of the audit process using RPA (robotic process automation) has the potential to enhance audit quality by automating structured audit procedures and offering auditors the opportunity to perform more meaningful work. The research question this study attempts to answer is: how can auditors redesign the audit process using RPA to accomplish a systematic audit approach? The proposed framework consists of six phases, including 1) developing vision and process objectives, 2) process identification, 3) process understanding, 4) audit data standardization, 5) audit apps prototyping, and 6) feedback and evaluation. The loan testing audit sub-process of a public accounting firm is selected as a candidate for automation to demonstrate the viability of the framework. The automation of this sub-process provides insights into the usefulness of the framework in guiding the application of RAPA to achieve near end-to-end audit process automation.

The third essay proposes an external audit blockchain that benefits from the reliability of the auditee's blockchain records and smart audit procedures that autonomously execute audit procedures on behalf of the auditor. The research question this study aims to answer is: how will blockchain and smart contracts disrupt the audit profession? More specifically, if blockchain is widely adopted across industries, how can auditors leverage blockchain and smart contracts as audit data analytic tools to enhance audit quality? To address this question this study proposes an external audit blockchain supported by smart audit procedures. Blockchain and smart audit procedures have the potential to enhance audit quality and audit reporting and thus help narrow the expectation gap that exists between auditors, financial statement users and regulators by proactively performing audit tests and disseminating their results on the blockchain ledger. A holistic audit framework comprised of on-the-blockchain and off-the-blockchain audit procedures for the revenue account is proposed. The holistic audit framework takes into consideration the revenue risks that blockchain-based audits can potentially address. Additionally, novel functions for the PCAOB to improve their inspection process and issues related to the application of blockchain and smart contracts are discussed.

These three essays aim to inform the debate on the use and impact of technological tools on audit quality. The auditing profession is not immune to technological advances. Accordingly, motivated by the shift in paradigm that the audit profession is experiencing, this dissertation is aimed at providing insights into the evolution of auditing as a result of technology.

## ACKNOWLEDGEMENTS

I would like to thank my dissertation committee members, the mentors and colleagues I have found along my journey as a scholar, and my family. This dissertation would not be possible without their support.

I would like to express my sincere gratitude to my advisor, Dr. Miklos A. Vasarhelyi. He sparked my intellectual curiosity as an undergraduate research assistant and years later gave me the unique opportunity to pursue my doctoral studies in a cutting-edge research institution. His guidance and intellectual spirit further ignited my passion and out-of-the-box thinking for research. Dr. Vasarhelyi, I am thankful for your belief in my ability to become a successful researcher, without your encouragement and support I may not have overcome the challenges I faced during my studies.

I am grateful for the invaluable input from my dissertation committee members. I am very grateful for the support and guidance of Dr. Alexander Kogan. His valuable insights helped me to improve my dissertation. I deeply appreciate the suggestions and guidance from Dr. Helen Brown-Liburd. She encouraged me throughout my doctoral studies and patiently discussed research issues with me. I am also very grateful to Dr. David A. Wood for his helpful comments and guidance.

I would like to thank Dr. Hussein Issa and Dr. Chanta Thomas for reviewing my work and providing many valuable comments. I would especially like to thank Dr. Dan Palmon and Dr. Alexander Sannella for their encouragement and guidance, which I truly appreciate. I would also like to thank Barbara Jensen for her help and kindness.

I would like to express my gratitude to Dr. Bill Kinney and Dr. Theodore Mock for their guidance as I developed research ideas. I would like to thank my colleagues Dr. Brigitte Mueller, Dr. Deniz Appelbaum, Dr. Ting Sun, Dr. Jun Dai, Dr. Kristyn Calabrese, Dr. Mauricio Codesso, Abigail Zhang, Eid Alotaibi, Kelly Duan, Zhaokai Yan, Jiahua Zhou, Abdulrahman Alrefai, Syrena Shirley, Catrina Palmer, Aziza Jones, Jamie Freiman, Xin Xin Wang, Heejae Lee, and Arion Cheong. Your support and research suggestions have helped me become a better scholar. Thank you for the opportunity you offered me to collaborate and engage with you.

I would also like to thank the PhD Project and the AICPA for their support as I pursued my PhD studies. I am also deeply thankful to the many audit professionals that provided thoughtful comments. I am grateful to Likefolio for being so kind to share social media data.

Last but not least. I would like to thank my family. To my parents, thank you for your unconditional love and faith in me. To my siblings, thank you for listening to my research ideas and for your insights, your support is deeply appreciated. To my husband Francis, without your love, support, and patience this dissertation would not have been possible.

IT TAKES A VILLAGE TO RAISE A SCHOLAR.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1. Introduction

This dissertation consists of three essays that explore the evolution of auditing by foreseeing the impact of social media information, RPA, and blockchain and smart contracts to improve the quality of audit services. The first chapter discusses the need for nontraditional sources of information and disruptive technologies in the conduct of external audits, the motivations, and research questions for this dissertation. Chapter 2 investigates the incremental contribution of third-party information generated from Twitter in enhancing analytical models. Chapter 3 presents a framework for RAPA (robotic audit process automation) and applies it to the loan testing audit sub-process of a public accounting firm. Chapter 4 proposes the use of blockchain and smart contracts by external auditors. The concluding chapter, Chapter 5, summarizes the research studies, discusses their limitations, and presents opportunities for future research.

## 1.1 Background: The Need for Nontraditional Sources of Information and Disruptive Technologies in the Conduct of External Audits

The growing use of technology, especially of more sophisticated data analytics, in financial statement audits has led to the development of collaborative research efforts between academia and audit practice. These efforts aim to understand the use and the impact of technology to auditing. The RADAR (Rutgers AICPA Data Analytics Research)[1] initiative, the CPA Canada ADAC (Audit Data Analytics Committee)[2] initiative, and the white papers published by the Big Four accounting firms (Appelbaum et al. 2017) on audit data analytics are examples of efforts in this domain and propose a myriad of ideas and

---

[1] For more information refer to: http://raw.rutgers.edu/radar.html

[2] Refer more information refer to: https://www.cpacanada.ca/en/business-and-accounting-resources/audit-and-assurance/canadian-auditing-standards-cas/publications/cpa-canada-audit-data-analytics-committee

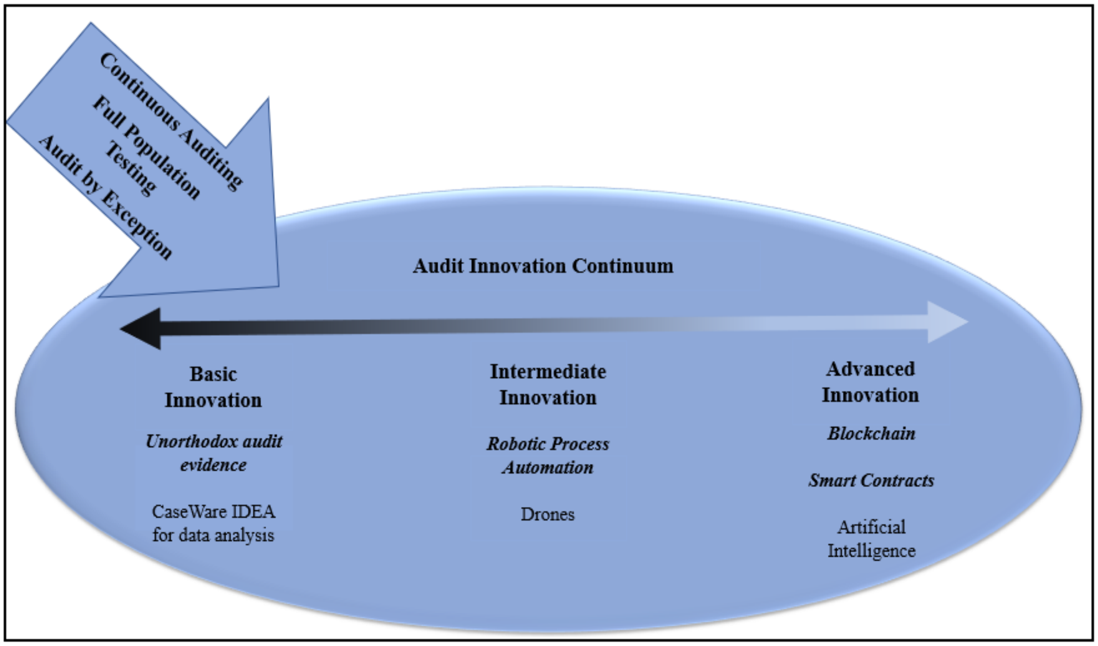
analytic methods that can potentially assist auditors in the performance of more effective and efficient audits. In addition to these collaborative research efforts, academic research in the emerging area of audit analytics can be categorized into three streams:

1) Research that examines sophisticated analytic methods (e.g. Issa 2013; Kogan et al. 2015; Chiu and Jans 2017);
2) Research that analyzes the use of nontraditional sources of information as audit evidence (e.g. Brown-Liburd and Vasarhelyi 2015; Yoon et al. 2015; Yoon 2016);
3) And research that proposes applications of disruptive technologies to auditing (e.g. Dai and Vasarhelyi 2017; Sun 2018).

The potential use of nontraditional sources of information and the applications of disruptive technologies to auditing has recently captured the interest of the audit community (Hamm 2018). Massive and nontraditional amounts of data and disruptive technologies are increasingly being used by audit clients to improve business practices (CPA Canada 2016). Innovations such as social media information, RPA, blockchain and smart contracts can fundamentally transform the financial reporting process and subsequently, the way that financial statement audits are conducted.

These technological innovations can be considered in terms of an audit innovation continuum, shown in Figure 1. The fundamental characteristics of this continuum are continuous auditing, full population testing, and an audit by exception approach (Vasarhelyi and Halper 1991). As financial statement audits evolve to include the use of technology-based data analytics, it is difficult to envision the exclusion of these characteristics in the conduct of audits.

**Figure 1: Audit Innovation Continuum**




On one side of the continuum is basic innovation, which represents the use of existing technologies, or of nontraditional sources of information, by auditors. On this side of the continuum, unorthodox sources of audit evidence and audit data analytic tools, such as CaseWare IDEA, are utilized to perform audit tasks. In the central section of the continuum is intermediate innovation, which can be defined as the use of new technologies to incrementally modify the audit. RPA and drones are examples of this level of audit innovation as these tools can achieve near end-to-end process automation for rules-based tasks [3]. On the opposite side of the continuum is advanced innovation, which consists of the use of new technologies like blockchain, smart contracts, and artificial intelligence to radically redesign the audit. These technologies can substantially transform the audit process by executing rules-based tasks, unstructured tasks, and by storing audit information on a secure and distributed ledger. The described innovations can alter the nature, timing,

---

[3] RPA can collect data and perform matching tests while drones can perform inventory counts.

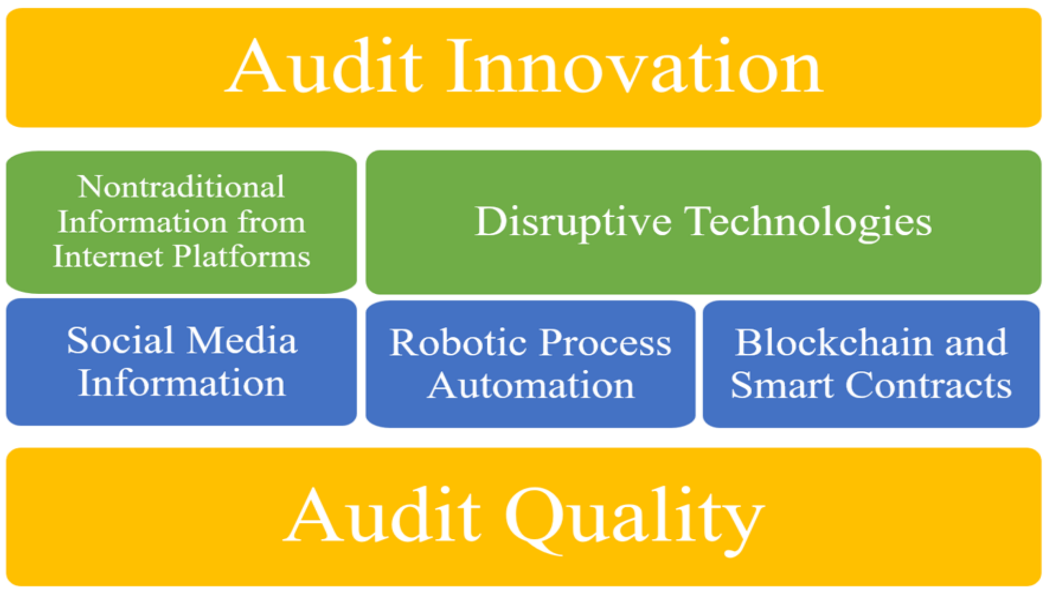
and extent of auditing procedures, potentially leading to improved audit quality, yet, their potential use remains underexplored.

## 1.2 Motivations and Research Questions

Although the utilization of more sophisticated data analytics by auditors is increasingly growing, such techniques are not yet being applied to support the audit opinion. As a result, while there have been substantial technological transformations in the business world, auditing remains largely unchanged. Consequently, the relevancy of the current audit framework, which emphasizes a retroactive, sample-based, and binary, audit opinion (No and Vasarhelyi 2017) is being challenged. In response to the technological evolution in the business environment, audit firms have proposed initiatives that are aimed at understanding the use of technology in auditing (Appelbaum, Kogan, and Vasarhelyi 2017). However, what remains unclear is to what extent auditing can evolve to parallel a largely digital economy.

Technology has the potential to enhance audit quality and transparency in the capital markets (SEC 2015; PCAOB 2016c) and it is vital to consider nontraditional sources of information and disruptive technologies as drivers of higher quality audits. Accordingly, the three essays in this dissertation explore the progressive evolution of auditing in light of basic innovation with social media information, intermediate innovation with RPA, and advanced innovation with blockchain and smart contracts. Social media information that is generated by consumers near real-time can serve as a new source of audit evidence that can assist auditors in the validation of account balances and in the detection of accounting errors. The redesign of the audit process using RPA can help auditors achieve a well-orchestrated audit approach by automating mundane and repetitive audit tasks that do not require audit judgment and by expanding the coverage of audit tests. Finally, blockchain smart contracts can perform autonomous audit tests near real-time on the full population of accounting records and store test results on the secure and distributed blockchain ledger, which can help improve audit quality and the audit reporting process. Figure 2 illustrates the outline for the three essays on audit innovation.

**Figure 2: Three Essays on Audit Innovation**




The first essay of this dissertation, presented in Chapter 2, examines the usefulness of third-party generated information from the Twitter social media platform in enhancing substantive analytical models. As described by SEC commissioner, Kara Stein, financial statement users are "accessing and analyzing massive amounts of information from sources, like social media, unimaginable just a few years ago. This new data may be empowering investors to make smarter investment decisions" (Stein 2015). One of the byproducts of technological advances is the emergence of new sources of information, this new information can potentially serve as a useful indicator for predicting firm performance

(Da, Engelberg, and Gao 2011a; Tang 2017). Accordingly, this essay investigates if consumer postings for individual brands or products that are aggregated at the firm-level have the potential to enhance the prediction, and error detection ability of substantive analytical procedures. Using third-party generated tweets provided by a social media data provider, measures of Twitter consumer interest and Twitter satisfaction are incorporated into analytical models. The results indicate that the models that incorporate Twitter-based measures experience improved prediction and error detection performance compared to the benchmark models that do not contain this information, especially the model that contains prior month sales, prior month GDP information and the Twitter measure of consumer interest.

Automation is not a new concept to auditors, however, exploiting the full power of technology to achieve an audit production line remains underexplored. The application of RPA to the audit process can result in a systematic audit approach where structured audit procedures are automated; thereby offering auditors the opportunity to focus on value-added work that could lead to enhanced audit quality. Presented in Chapter 3, the third essay of this dissertation provides guidance on the redesigning of the audit process using RPA, referred to as RAPA (robotic audit process automation). First, a framework that is based on existing methodologies for RPA and process redesign is proposed. The framework consists of six phases 1) developing vision and process objectives, 2) identification of the process to be automated, 3) understanding of the process, 4) standardization of audit data, 5) prototyping of audit apps, and 6) feedback and evaluation. To validate the feasibility of the framework, a prototype for the loan testing audit sub-process is designed. The results of the framework implementation indicate that it can facilitate near end-to-end automation of this process.

The third essay of this dissertation, presented in Chapter 4, explores the potential use of blockchain and smart contracts as audit data analytic tools that could enhance audit quality and reporting and thus, reduce the expectation gap between auditors and stakeholders. Blockchain and smart contracts are demonstrating to have great potential in improving the quality of business processes (Mainelli and Smith 2015; Vaziri 2016), as a result, it is important to examine the impact of these emerging technologies on auditing. This essay conjectures that financial and nonfinancial blockchain records from an auditee have the potential to be more reliable than records from a traditional accounting system, such as an ERP. Moreover, this essay proposes an external audit blockchain that is supported by what this dissertation defines as “smart audit procedures”. Smart audit procedures are automated audit tests that are executed on the external audit blockchain for the purpose of improving audit quality and audit reporting. In addition, this essay also proposes that the future audit framework, will constitute of on-the-blockchain and off-the-blockchain audit procedures and that the PCAOB can proactively inspect financial statement audits by becoming a node on the auditor’s blockchain. Issues related to the application of blockchain and smart contracts to auditing are also discussed.

Although academics, public accounting firms, and regulators recognize that technology can radically evolve business practices, it is unclear to what extent technology will evolve auditing and thus audit quality. The three essays presented in this dissertation fill a gap in the emerging literature on audit analytics and provide insights into the debate on how and where technology fits in auditing. Although audit standards and extant research posit that nonfinancial information can enhance the effectiveness of analytical procedures, more research is needed to examine the usefulness of new nonfinancial sources of potential audit evidence that emerged as a byproduct of a digital business environment. Moreover, by exploring disruptive technologies such as RPA, and blockchain and smart contracts, this dissertation foresees the evolution of the audit model and the impact of technology on audit quality. Collectively, this dissertation contributes to both, academia and audit practice by foreseeing the evolution of auditing in the presence of technology.

The remainder of this dissertation is organized as follows. Chapter 2 investigates the usefulness of third-party information generated from Twitter in enhancing analytical models. Chapter 3 presents a framework for audit process automation using RPA and applies it to the loan testing audit sub-process. Chapter 4 proposes the use of blockchain and smart contracts by external auditors. The concluding chapter, Chapter 5, summarizes the research studies, discusses their contributions and limitations, and presents future research opportunities.

# Chapter 2. Enhancing Substantive Analytical Procedures with Third-Party Generated Information from Social Media

*"Investors, and others, are accessing and analyzing massive amounts of information from sources, like social media, unimaginable just a few years ago. This new data may be empowering investors to make smarter investment decisions"*

*Kara Stein – SEC Commissioner 2015*[1]

## 2.1. Introduction

Technology has not only decreased the cost to process, store, and analyze business information, it has created new sources of information (Appelbaum 2016). New sources of information include voluminous third-party social media information that is generated about firm activities in real-time and is easily accessible by external parties. Extant research suggests that social media information contains incremental information about firms' stock market prices, and sales performance (e.g. Bollen, Mao, Zheng 2011; Tang 2017). Accordingly, third-party generated social media postings may offer a timely and independent benchmark that can be used to compare sales trends.

More research is needed to advance analytical procedures (Badertscher, Kim, Kinney, and Owens 2017) and examine the usefulness of new nonfinancial sources of potential audit evidence (Yoon 2016). To address this research gap, this paper examines the usefulness of third-party generated social media information for firms' brands and products in enhancing substantive analytical procedures (hereafter referred to as SAPs) for

[1] Refer to: https://www.sec.gov/news/speech/remarks-inst-chartered-acctnts.html

the revenue account[2]. In particular, the prediction performance and error detection performance of traditional and 'continuous'[3] SAPs that incorporate Twitter-based measures of consumer interest and satisfaction are examined. Analytical procedures are defined as "reasonableness tests" where auditors compare their expectation for account balances with those recorded by management (Louwers et al. 2018). Analytical procedures are required in the planning and concluding stages of the audit, and are recommended for substantive testing (PCAOB 2010b, AS No. 2110; PCAOB 2010c, AS No. 2810; PCAOB 2010a, AS No. 2305; AICPA 2012).

Essentially, analytical procedures assist auditors in their ongoing assessment of risk. This objective is achieved by applying analytical procedures to develop an audit plan, collect audit evidence by verifying management's assertions, and to review audit conclusions. Although SAPs may offer a cost-effective alternative compared to test of details, PCAOB inspection findings often note a number of deficiencies in auditors' application of SAPs. These deficiencies include audit firms' failure to develop precise and appropriate expectations and failure to appropriately investigate unexpected differences (PCAOB 2007; PCAOB 2016a). Accordingly, it is of interest to investigate whether social media information can enhance the effectiveness of traditional SAPs and continuous SAPs.

Extant research asserts that analytical procedures that incorporate financial and nonfinancial information tend to be more effective in assessing the reasonableness of

---

[2] Substantive analytical procedures for the revenue account are examined in this study as prior research has documented that Tweets are predictive of upcoming revenue (Tang 2017).

[3] The extant literature suggests that predictive analytics that are performed on a continuous basis could lead to more accurate account expectations (Kogan et al. 2014; Yoon 2016). As a result, a lag length of one month is utilized in this study to illustrate the predictive performance of continuous SAPs, whereas a lag length of 12 months is utilized to illustrate the predictive performance of traditional SAPs.

account balances and the risk of fraud (Brazel, Jones, and Zimbelman 2009; Kogan, Alles, Vasarhelyi, and Wu 2014). A caveat with these assertions is that expectation models that generally include nonfinancial information that is generated by the firm (Hirst and Koonce 1996; Brazel et al. 2009; Trompeter and Wright 2010) could produce less effective expectations as this information could be manipulated by management. In contrast, nonfinancial information that is externally produced, such as customer satisfaction surveys, and economic indicators, has demonstrated to be more useful for predicting firm performance and fraud as it is less susceptible to management manipulation (Lev 1980; Ittner and Larcker 1998). However, this information is only available sporadically and may not be as timely as social media information that is readily available. Consequently, it is important to investigate if contemporaneous nonfinancial information that is independently produced by third-parties on social media platforms can enhance the accuracy and error detection ability of traditional and continuous SAPs.

Empowered by the Internet and electronic commerce, social media has paved the way for a new source of nonfinancial information, third-party generated comments of individual firms' brands and products. Twitter is selected as the setting of this study as it is a simple and popular platform for microblogging (Stieglitz and Dang-Xuan 2013; Paniagua and Sapena 2014). Third-party generated Twitter comments reflect information that is generated outside of the firm and available to the public in real-time. When aggregated at the firm level, Twitter comments about individual firms' brands and products potentially offer a more precise external measure for sales performance than the measure documented in Google search research studies[4].

The PCAOB (2017b) contends that audit firms are designing predictive models with nontraditional indicators to improve risk assessments. However, it is plausible that external nonfinancial information, such as third-party generated information from social media, can contain incremental information that auditors can leverage for substantive testing (Yoon 2016). Social media information created by third-parties is timely, easily accessible, and has the potential to provide a more reliable benchmark than nonfinancial information that is internally produced by management. As a result, social media information could potentially serve as a sufficient, relevant, and reliable source of audit evidence (PCAOB 2010d, AS No. 1105).

This study explores the prediction and error detection performance of traditional and continuous SAPs that include Twitter proxies for consumer interest and satisfaction. Specifically, it utilizes monthly sales data for 24 business-to-consumer industries and aggregated third-party generated Twitter comments of firms' individual brands and products provided by a social media data provider, Likefolio[5]. For prediction performance, the results suggest that continuous SAPs with prior month sales, prior month GDP (gross domestic product) and TCI (Twitter Consumer Interest) or TCS (Twitter Consumer

[4] The measure documented in Google search research is the volume index from search queries that maps one product, or brand, to one firm (Da et al. 2011). Firms may produce thousands of products or provide a variety of services. As a result, aggregating Tweets about individual brands or products at the firm level could provide a more precise measure for consumers' interest to buy or sentiment.

[5] Likefolio is a company that compiles Twitter information about consumers' brands and products for 194 publicly listed firms in order to provide consumer insights to various parties including hedge fund managers and individual investors. Likefolio maps Twitter comments about firms' products or brands to the firms they belong to. For more information please refer to: https://home.likefolio.com/

Satisfaction), or continuous SAPs with prior month sales, prior month GDP, AR (accounts receivable) and TCI or TCS, produce superior sales predictions than the benchmark models that do not incorporate TCI or TCS. However, for auditors to fully exploit the benefits of SAPs, it is also important to examine the error detection performance of the models.

A simulated experiment where errors are seeded into the dependent variable (i.e. overstating sales) is used to evaluate the error detection ability of SAPs with Twitter-based measures. For error detection performance, the results indicate that continuous SAPs that contain prior month sales, prior month GDP and TCI or TCS outperform the benchmark models as they can achieve superior error detection performance under varying cost ratios. Consequently, the more effective model for both prediction and error detection performance is the model with prior month sales, prior month GDP and TCI as it can produce superior prediction and error detection performance for most of the industries that are examined. Taken together, the findings for prediction and error detection performance provide evidence that TCI has incremental value in the absence of contemporaneous firm-specific information and in the presence of macroeconomic information.

This study is closely related to the Da, Engelberg, and Gao (2011a) and Tang (2017) studies in that it examines contemporaneous external nonfinancial information generated from Internet platforms to predict sales, however, it distinguishes itself from several perspectives. First, it aims to advance analytical procedures by examining the prediction performance of expectation models rather than their predictive power, where the emphasis is on the predicted value rather than the statistical significance of the regression coefficients. This is important as it can help inform the debate on the usefulness of nontraditional and external nonfinancial information in audit procedures (IAASB 2016).

Secondly, it examines the expectation models' ability to detect errors thus providing additional information about the effectiveness of analytical procedures that include a new source of nonfinancial data, third-party generated social media comments. Finally, this study expands the scope of the Yoon (2016) study by examining whether a different form of unorthodox audit evidence has the potential to enhance the power of analytical procedures for the revenue account. In summary, this study offers insights that may be useful to audit researchers, practitioners and standard-setters, as they evaluate the relevance of social media information as audit evidence.

The remainder of this essay is organized as follows. The second section presents the literature review and develops research questions grounded on the literature. The third section presents the research design. The fourth and fifth section discuss the results while the last section presents a conclusion and discusses areas for future research.

## 2.2. Literature Review and Research Questions

### 2.2.1. Nonfinancial Information from Internet Platforms

The value-add of external information generated by web platforms is extensively studied in various research disciplines including healthcare, marketing, political science, economics, finance, and accounting. This research generally examines user behavior via Internet search queries or social media postings.

In the healthcare discipline, Ginsberg, Mohabbi, Patel, Brammer, and Smolinski (2009) find that Google queries related to influenza forecast flu outbreaks one to two weeks before the Center for Disease Control and Prevention (CDC) makes a public announcement. Ji, Chun, and Geller (2013) develop a Twitter-based health surveillance tool and find that Twitter users' concerns about illnesses are predictive of health epidemics.

Research in marketing indicates that both Google search volume and Twitter posts are useful tools for understanding consumer interest and behavior. For example, Du and Kamakura (2012) indicate that Google search queries offer a holistic view of consumer preferences and behavior. Burton and Soboleva (2011) find that Twitter is not only useful as a listening tool, but that it also has great potential as a marketing mechanism for external communication with customers because organizations can reach customers directly.

With respect to political predictions, Stephens-Davidowitz (2017) indicates that Google, the “digital truth serum”, predicted that presidential candidate Trump would win the 2016 U.S. election despite the results of traditional polls, which suggested that his contender was the more powerful candidate. Tumasjan, Sprenger, Sandner, and Welpe (2010) study Tweets about German federal elections and find that the volume and sentiment of Tweets about a political party can help predict election outcomes. Gayo-Avello (2013) reviews studies related to the predictive power of Twitter data for political elections and finds that it does not provide strong evidence in predicting election results; this stance is supported by recent research, which documents that Tweets are rather reactive and not predictive of elections (Murthy 2015).

Research in economics finds that search volume is useful for nowcasting economic activities including unemployment rates, sales, and consumer confidence. For example, Askitas and Zimmerman (2009) and Ettredge, Gerdes, and Karuga (2005) demonstrate that search volume is predictive of unemployment rates in Germany and in the U.S., Goel, Hoffman, Lahaie, Pennock, and Watts (2010) study the predictive power of search volume for movies, songs, and video games and indicate that search volume can be used to predict box-office revenue for feature films, ranking of popular songs, and first-month sales of video games. Lastly, Choi and Varian (2012) suggest that volume search is useful in forecasting automobile sales, travel plans, and consumer confidence. The general finding in this literature is that search volume can be used as a timely indicator of economic activity thus offering an advantage over traditional economic indicators that are not readily available.

In finance and accounting, several studies examine the influence of search volume and Twitter sentiment on stock market prices and firm fundamentals. For instance, Da et al. (2011b) suggest that search volume can be used as a direct measure of investor attention and find that higher search volume predicts higher stock market prices. Bollen et al. (2011) apply two mood measuring tools to evaluate collective mood per Twitter posts and find that Twitter mood is predictive of the DJIA (Dow Jones Industrial Average). Lee, Hutton, and Shu (2015) examine the role of social media in the stock market specific to product recalls and demonstrate that firms which are active in social media experienced attenuating benefits to product recall announcements compared to firms with no social media presence. Da et al. (2011a) study the changes in search volume and document that they are a strong predictor of revenue surprises and three-day abnormal returns. Recently, Tang (2017) examined the cross-sectional variation of third-party generated Twitter comments and finds that they are strongly associated with sales from business-to-consumer industries, that tweets by consumers have higher predictive power than the tweets initiated by experts or the media, and that Twitter comments are predictive of upcoming revenue and revenue surprises.

### 2.2.2. Analytical Procedures

The purpose of analytical procedures is to evaluate financial statement information by analyzing plausible relationships between financial and nonfinancial information (AICPA 2012). Analytical procedures are required in the planning stage and review stage of the audit but only recommended for substantive procedures tests (PCAOB 2010b, AS No. 2110; PCAOB 2010c, AS No. 2810; PCAOB 2010a, AS No. 2305; AICPA 2012). Simple year to year account balance comparison, ratio comparison, scanning, and more sophisticated models such as regression, are examples of analytical procedures implemented in audit practice. Auditors perform analytical procedures in three steps: 1) they develop an expectation for an account balance (or ratio) 2) they compare the difference between the expected account balance and actual account balance recorded by management, and 3) they investigate differences that exceed the materiality threshold; if differences do not exceed the materiality threshold, auditors assess whether further audit procedures are needed (Louwers et al. 2018).

Although analytical procedures are executed following the aforementioned method, they serve different purposes. For the planning stage of the audit, analytical procedures assist auditors in enhancing their understanding of the business and its economic events and by highlighting areas that present risks to the audit. Analytical procedures for substantive testing are applied by auditors to collect audit evidence about management's assertions concerning the veracity of account balances or class of transactions. Finally, in the concluding stage of the audit, auditors perform analytical procedures to validate their evaluations of financial statement information (SAS No. 56, AICPA 1988; AS No. 2810, PCAOB 2010c). Collectively, analytical procedures are a tool used by auditors to identify risks, direct their attention to potential irregularities, and obtain confirmatory evidence about the reasonableness of information underlying financial statements.

Research studies relating to improving the effectiveness of analytical procedures spans several decades. Analytical models studied in the prior literature range from simple models, such as ratio analysis, to more sophisticated models including ARIMA, linear regression, SEM, and vector autoregression (Kinney 1978; Kinney 1987; Dzeng 1994; Wild 1987; Hirst and Koonce 1996; Kogan et al. 2014). As an example, Kinney (1978) introduces ARIMA as a possible method for developing expectations for analytical procedures and finds that ARIMA generates better predictions compared to other models, but that it is not as generally applicable as regression models. Wild (1987) introduces a structural model as it can accommodate the interdependencies across related accounts and exogenous variables and indicate that the model does not perform better than multivariate stepwise models.

Wheeler and Pany (1990) evaluate the prediction performance of the Census X-11 time-series model against other models and document that the X-11 model produces superior expectations followed by the predictions of regression models. Dzeng (1994) presents a new forecasting technique, VAR (vector autoregression), for analytical procedures and suggests that VAR predictions are superior to the predictions of other models such as ARIMA and random walks, and that linear regression predictions are second best to those of VAR. Finally, Kogan et al. (2014) introduce a continuity data level auditing system based on four forecasting models, SEM (simultaneous equation model), VAR, BVAR (18ayesian vector autoregression), and LRM (linear regression model) and find that all models perform reasonably well in predicting and detecting errors.

In addition, previous literature finds that disaggregated monthly, or quarterly data, can produce superior account balance predictions. For example, Dzeng (1994) and Chen and Leitch (1998) compare analytical models using monthly and quarterly data and find that monthly data improves the performance of analytical procedures. Using peer sales as the indicator of interest, Hoitash, Kogan, and Vasarhelyi (2006) also find that monthly data can improve the performance of analytical procedures. In contrast, Allen, Beasley, and Branson (1999) do not find that monthly data from multi-locations improve analytical procedures and they attribute this finding to the homogeneity of the services provided by the company, which would reduce the likelihood of finding differences between different levels of aggregation.

### 2.2.3. The Role of Nonfinancial Information in Analytical Models

The emphasis of early research studies on financial indicators to improve the performance of analytical procedures is not surprising since auditors were more likely to rely more on financial information than on nonfinancial information when determining audit scope (Cohen, Krishnamoorthy and Wright 2000; Brazel, Jones and Prawitt 2013). However, as auditors move towards using information from the Internet to enhance their understanding of the business environment and to develop account balance expectations (Trompeter and Wright 2010) it is important to understand the role of nonfinancial information in analytical models.

The relevance of nonfinancial information to predict firm performance has been studied since the 1980s. Lev (1980) documented that economic and industrial indicators can improve the predictive ability of analytical procedures. Amir and Lev (1996) examined the value of nonfinancial information such as market penetration and found that investors prefer nonfinancial information over financial information. Ittner and Larcker (1998) suggest that customer satisfaction has predictive ability for future accounting balances but that it partially impacts current accounting balances. Nonfinancial information is expected to be more objective than financial information produced by a firm as it is less susceptible to management manipulation. Although financial indicators, which provide a historical view of business activities would remain relevant, nonfinancial information is considered to be more valuable as it projects a current and forward-looking view of the business (Lev and Gu 2016).

The academic literature and audit standards suggest that both nonfinancial and financial information can enhance analytical procedures (Dzeng 1994; AICPA 2012). Allen et al. (1999) use financial and nonfinancial information, including the number of pounds serviced and the number of working days of monthly and multi-location data. Brazel et al. (2009) find that the inconsistent pattern between employee growth and sales growth can be used as an indicator to detect financial statement fraud in AAER firms. Parallel to the Brazel et al. (2009) study, Allee, Baik and Roh (2018) use electricity consumption data to proxy for real production activity and find that the inconsistent pattern between electricity consumption growth and sales growth is associated with firms that have higher discretionary accruals. Kogan et al. (2014) design continuity equations based analytical procedures of highly disaggregated procurement data such as purchase orders, receiving documents, and vouchers. While these research studies advance knowledge about the relevance of nonfinancial information in improving the effectiveness of analytical procedures, their main limitation is that their explanatory variables pertain to information that is produced by management or that is static in nature.

A relatively new stream of the auditing literature suggests that big data, which is defined as data that is voluminous, of different types, rapidly changing, and of varying levels of veracity (Buhl 2014), may be used as audit evidence (Vasarhelyi, Kogan and Tuttle 2015; Brown-Liburd and Vasarhelyi 2015; Yoon, Hoogduin and Zhang 2015; Alles and Gray 2016). Brown-Liburd and Vasarhelyi (2015) suggest that exogenous big data can be linked to business activities and therefore be utilized by auditors to arrive at conclusions about account balances. As examples, publicly available big data sources such as search engine query data, weather data, and social media data can support sales balances as they can proxy for consumer demand and satisfaction. These sources of data can improve the precision of audit procedures as they may be less tamperable than information that is generated by the firm's information system.

Using daily and weekly sales of multi-location retail stores, Yoon (2016) is one of the first to explore the usefulness of exogenous big data and analyzes if weather indicators are predictive of sales performance. The findings in her study suggest that weather is correlated with sales and that it is useful in detecting errors. This research study expands the Yoon (2016) study by examining the relevance of social media information generated by third-parties in analytical procedures. Specifically, this paper seeks to examine whether third-party generated information from social media platforms about firm's products and brands, can enhance the accuracy and error detection ability of traditional and continuous SAPs.

Third-party generated Twitter information about firms' products or brands has the potential to advance substantive analytical procedures as it can serve as a valuable source of tertiary audit evidence. The three tenets of audit evidence are 1) sufficiency, 2) relevancy, and 3) reliability (PCAOB 2010d, AS No. 1105). Third-party generated Twitter information, which reflects product and brand interest and satisfaction reviews, has been found to be correlated with firms' sales performance (Tang 2017). This information is voluminous in nature, and it is relevant because it is timely, and can be utilized as a proxy for consumer interest to buy and consumer satisfaction. Furthermore, this information is generated by third parties and can serve as an independent source of information to corroborate actual recorded account balances. This information is also easily accessible to auditors at the time of the audit in contrast to other nonfinancial indicators that are useful for auditing, but only available after a long delay. Consequently, this research study examines the following research questions:

**RQ 1A**: For the revenue account, do traditional substantive analytical models with Twitter information experience improved prediction performance?

**RQ 1B**: For the revenue account, do continuous substantive analytical models with Twitter information experience improved prediction performance?

**RQ 2A**: For the revenue account, do traditional substantive analytical models with Twitter information experience improved error detection performance?

**RQ 2B**: For the revenue account, do continuous substantive analytical models with Twitter information experience improved error detection performance?

## 2.3. Research Design

### 2.3.1. Third-party Generated Information from Twitter

This study investigates the usefulness of social media information, using third-party generated information from Twitter as an example. Twitter comments generated by consumers for interest to buy or satisfaction are extracted by the data provider, Likefolio. The data provider maps thousands of products and brands to the firms that the products belong to (Tang 2017; Likefolio 2018). This mapping makes it feasible to more directly examine the impact of consumer behavior on sales performance compared to the one-to-one Google search product or brand measure, which maps one search query to one firm (e.g. Da et al. 2011a). Another benefit of using the provided data is that it facilitates a time-series analysis spanning six years of Twitter information as daily Twitter information is obtained from 2012 to 2017.

Hence, obtaining Twitter information from the data provider addresses two major limitations identified in prior research. The first limitation is related to the completeness of the search engine queries (Da et al. 2011a) and Tweets (e.g. Moon 2016), while the second limitation is related to the short time periods this data is generally collected for (e.g. Moon 2016; Bollen et al. 2011). Therefore, obtaining a historical dataset that maps thousands of products or brands to individual firms can provide more insights into the correlation between Twitter proxies and sales trends. The data provider uses text mining techniques and machine learning methods to collect recent past and future consumer interest, and positive and negative consumer sentiment. The dataset does not include retweets and is normalized to account for the growth or decline in Twitter usage. Accordingly, TCI (Twitter Consumer Interest) and TCS (Twitter Consumer Satisfaction) measures are incorporated into analytical models to examine their usefulness in predicting firms' sales and detecting accounting errors.

TCI consists of the volume of third-party Twitter comments that indicate a recent past purchase, or a future purchase. Figure 3 presents an example of a Tweet that would be classified as TCI. The second measure, TCS, captures consumer satisfaction (or dissatisfaction) by measuring the ratio of the total number of third-party Twitter comments that express positive sentiment to the total number of third-party Twitter comments that indicate both positive and negative sentiment. Figure 4 presents an example of a Tweet that would be classified as negative sentiment.

**Figure 3: Example of Consumer Interest to Buy**

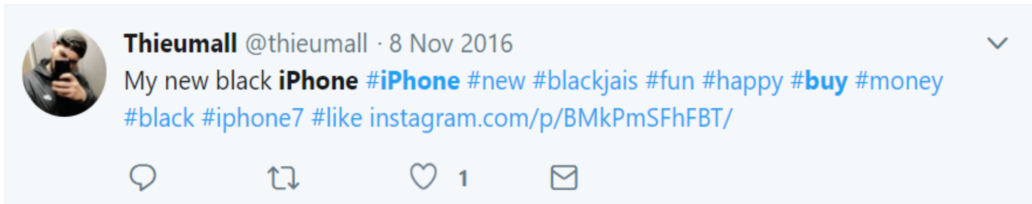



**Figure 4: Example of Negative Consumer Sentiment**

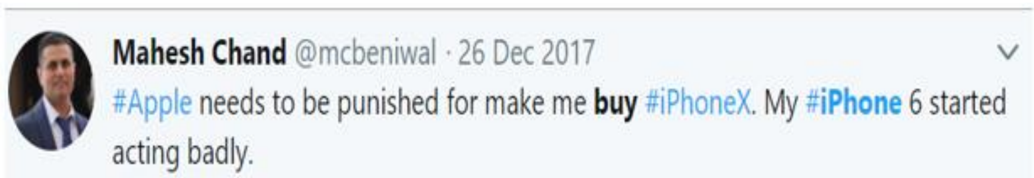



Figure 5 and Figure 6 present the correlation by quarter for correlations between monthly sales and TCI, and that of monthly sales and TCS. The correlation graphs suggest that compared to TCS, TCI is more positively correlated to sales, suggesting that TCI may have more predictive power for sales. Moreover, the graphs also indicate that in general, the correlation between monthly sales and Twitter-based measures marginally increases during Q3.

**Figure 5: Correlation between Sales and Consumer Interest to Buy**

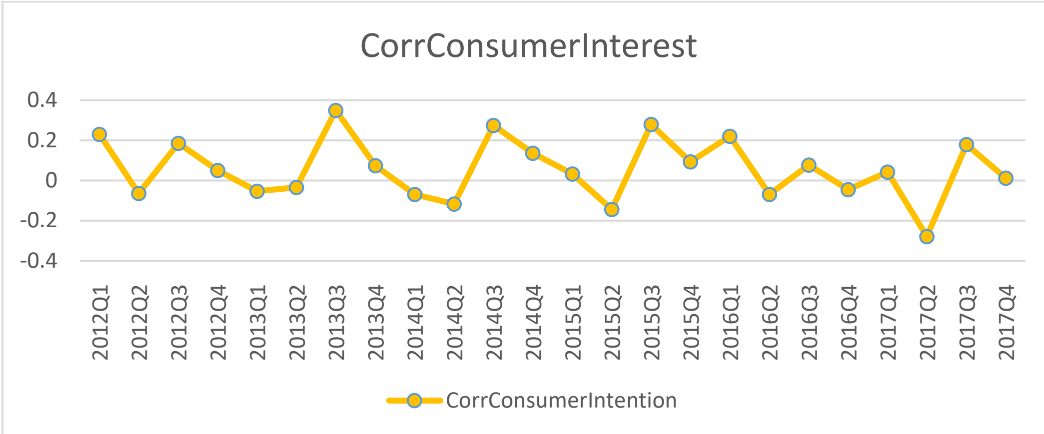
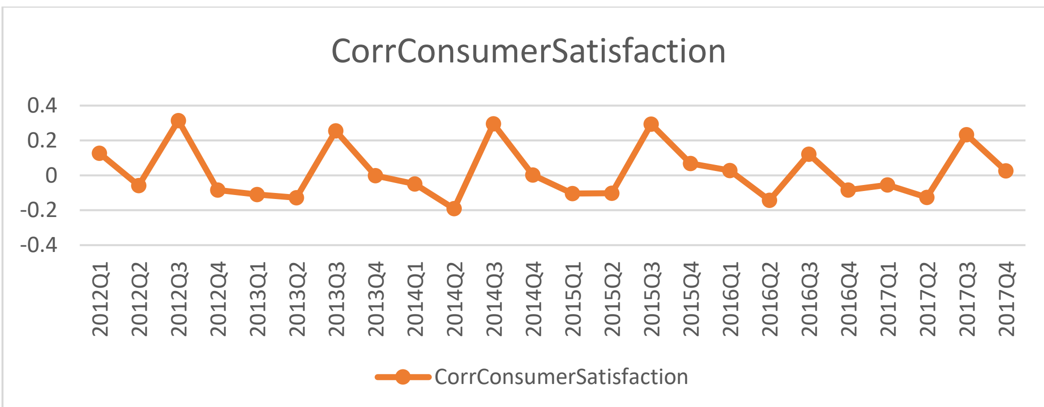



**Figure 6: Correlation between Sales and Consumer Satisfaction**




### 2.3.2. Sample Selection

Financial statement information related to quarterly sales and accounts receivable is collected from the Compustat Fundamentals quarterly database for the 194 publicly listed companies for which the data provider collects Twitter information. Financial data is collected for the period 2012 to 2017. Accounts receivable is selected as it is associated with revenue.

Since financial statement information is available on a quarterly basis, this study uses the cubic splines interpolation method to estimate monthly observations. This method has been applied in prior audit research (Chen and Leitch 1998; Leitch and Chen 1999;

Hoitash et al. 2006; Yin 2018) and estimates monthly observations from quarterly observations. Twitter information is aggregated for respective months by averaging the daily number of Tweets that reflect past or future interest to purchase, and the total number of Tweets that reflect positive and negative sentiment. To estimate predictive models that include both financial and nonfinancial information, the sample had to satisfy certain requirements:

1. Firms should have quarterly financial statement data without missing information or zero values.
2. Twitter information should have daily consumer interest and consumer sentiment information without missing information or zero values.
3. Firms should have quarterly financial information for six years since five years of firm-month observations will be used to train the prediction models and one year of firm-month observations will be used to test the prediction models.
4. Firms in the financial services industry are excluded.

The sample selection process is described in Table 1. There are 88 firms corresponding to 24 consumer-facing industries that satisfy the four sample requirements. The final sample consists of 2,112 firm-quarter observations, which are converted into 6,336 firm-monthly observations.

**Table 1: Sample composition**

| Sample Selection - Firm-Quarter Observations 2012-2017 | | |
|---|---|---|
| | Firms | Firm-Quarter Observations |
| Firms that are publicly listed and have third-party generated Twitter information | 194 | 4,656 |
| Less: Firms with missing financial information or zero values | (9) | (216) |
| Less: Firms with missing information from Twitter for either Consumer Interest or Sentiment | (15) | (360) |
| Less: Firms without four quarters of data | (73) | (1,752) |
| Less: Firms in the Financial Services Industry | (9) | (216) |
| Total | 88 | 2,112 |

The training set consists of 5,280 firm-month observations and the testing set consists of 1,056 firm-month observations. Table 2 presents the 24 business-to-consumer industries that make up the sample by their two-digit SIC codes and their respective average sales and accounts receivable balances.

**Table 2: Descriptive Statistics – Financial Information for Final Sample, from 2012-2017**

**Descriptive Statistics - Financial Information - Firm-Quarter Observations from 2012-2017**

| 2-Digit SIC Code | Industry Name | Number of Firm-Quarter Observations | Revenue | Accounts Receivable |
|---|---|---|---|---|
| 20 | Food and Kindred Products | 288 | 4120.78 | 1951.53 |
| 21 | Tobacco Manufacturing | 24 | 4616.67 | 161.38 |
| 23 | Apparel and Other Textile Products | 72 | 1361.76 | 527.65 |
| 28 | Chemicals and Allied Products | 144 | 4539.45 | 1592.93 |
| 29 | Petroleum and Coal Products | 48 | 29763.38 | 10859.92 |
| 30 | Rubber/Misc. Plastic Products | 24 | 7498.13 | 3441.46 |
| 31 | Leather and Leather Products | 48 | 939.78 | 303.24 |
| 35 | Industrial and Commercial Machinery and Computer Equip | 48 | 1680.86 | 975.20 |
| 36 | Electrical Equipment and Components | 96 | 17770.89 | 7541.23 |
| 37 | Transportation Equipment | 168 | 29908.22 | 36176.99 |
| 39 | Misc. Manufacturing Industries | 72 | 937.04 | 753.35 |
| 42 | Motor Freight Transportation | 24 | 14705.75 | 6303.17 |
| 44 | Water Transportation | 48 | 3034.01 | 423.68 |
| 45 | Transportation By Air | 192 | 6150.74 | 1521.82 |
| 47 | Transportation Services | 24 | 1669.62 | 1090.61 |
| 48 | Communications | 48 | 1052.82 | 398.55 |
| 53 | General Merchandise Stores | 24 | 28794.04 | 1269.33 |
| 55 | Automobile Dealers & Gasoline Service Stations | 48 | 2904.34 | 299.55 |
| 57 | Home Furniture, Furnishings and Equipment Stores | 48 | 5219.94 | 647.29 |
| 58 | Eating & Drinking Places | 360 | 1401.46 | 220.35 |
| 59 | Miscellaneons Retail | 72 | 22690.37 | 5330.04 |
| 70 | Hotels, Rooming Houses, Camps and Other Lodging Places | 48 | 2475.56 | 800.65 |
| 73 | Business Services | 120 | 8571.81 | 5371.53 |
| 75 | Automotive Repair Services & Parking | 24 | 772.17 | 401.16 |

Table 3 displays the average Tweets for consumer interest, positive, and negative sentiment for the 24 business-to-consumer industries that make up the sample. SIC codes 36 and 73, which consist of household appliances, radio, tv equipment (SIC code 36) and computer programming, prepackaged software, and auto rental and leasing (SIC code 73), produce a higher average of Tweets for consumer interest and sentiment.

**Table 3: Descriptive Statistics –Twitter Information for Final Sample, from 2012-2017**

**Descriptive Statistics - Twitter Information - Firm-Quarter Observations from 2012-2017**

| 2-Digit SIC Code | Industry Name | Number of Firm-Quarter Observations | Tweet Consumer Interest | Tweet Positive Sentiment | Tweet Negative Sentiment |
|---|---|---|---|---|---|
| 20 | Food and Kindred Products | 288 | 844.53 | 795.42 | 305.39 |
| 21 | Tobacco Manufacturing | 24 | 7.48 | 22.87 | 13.17 |
| 23 | Apparel and Other Textile Products | 72 | 53.35 | 87.38 | 33.57 |
| 28 | Chemicals and Allied Products | 144 | 85.12 | 214.77 | 89.52 |
| 29 | Petroleum and Coal Products | 48 | 3.63 | 13.27 | 7.10 |
| 30 | Rubber/Misc. Plastic Products | 24 | 773.07 | 1715.88 | 789.19 |
| 31 | Leather and Leather Products | 48 | 35.61 | 68.98 | 20.25 |
| 35 | Industrial and Commercial Machinery and Computer Equip | 48 | 34.53 | 72.03 | 18.30 |
| 36 | Electrical Equipment and Components | 96 | 2677.05 | 3580.18 | 2293.22 |
| 37 | Transportation Equipment | 168 | 113.05 | 368.76 | 170.54 |
| 39 | Misc. Manufacturing Industries | 72 | 168.91 | 435.87 | 162.30 |
| 42 | Motor Freight Transportation | 24 | 52.68 | 119.44 | 117.17 |
| 44 | Water Transportation | 48 | 8.37 | 44.97 | 8.36 |
| 45 | Transportation By Air | 192 | 86.84 | 150.66 | 155.65 |
| 47 | Transportation Services | 24 | 43.35 | 770.62 | 40.76 |
| 48 | Communications | 48 | 63.39 | 92.70 | 28.99 |
| 53 | General Merchandise Stores | 24 | 238.85 | 284.32 | 111.43 |
| 55 | Automobile Dealers & Gasoline Service Stations | 48 | 2.96 | 7.49 | 3.71 |
| 57 | Home Furniture, Furnishings and Equipment Stores | 48 | 288.42 | 303.17 | 43.65 |
| 58 | Eating & Drinking Places | 360 | 851.55 | 1020.33 | 473.72 |
| 59 | Miscellaneons Retail | 72 | 95.58 | 97.85 | 52.81 |
| 70 | Hotels, Rooming Houses, Camps and Other Lodging Places | 48 | 62.60 | 138.94 | 22.49 |
| 73 | Business Services | 120 | 4600.11 | 5333.26 | 2265.66 |
| 75 | Automotive Repair Services & Parking | 24 | 22.08 | 19.48 | 21.92 |

### 2.3.3. Control Variables

Extant literature has indicated that macroeconomic indicators and contemporaneous accounts can be useful in enhancing the effectiveness of analytical procedures (Lev 1980; Hoitash et al. 2006; Minutti-Meza 2011). Accordingly, to measure the effectiveness of Twitter information on analytical procedures, this study controls for GDP (Gross Domestic Product) and accounts receivable (AR) as this information is associated with sales, the predicted variable.

Quarterly GDP information, which is adjusted for seasonality, is obtained from the Bureau of Economic Analysis website[6] and interpolated into monthly observations. The monthly GDP observations are then matched to corresponding firm-monthly observations. Accounts Receivable are included since the literature suggests that the precision and accuracy of analytical models is improved by including concurrent data from relevant accounts.

#### 2.3.4. Analytical Models

While analytical procedures constitute a broad range of audit procedures, prior research has suggested that time-series models, including ARIMA, VAR, and multivariate regression models that are estimated using lagged firm-specific information, lead to more accurate and precise expectations than simple heuristic models (e.g. Minutti-Meza 2011; Kogan et al. 2014). This study uses univariate, and multivariate regression models to predict account balances and detect errors as these models are generally applicable by auditors. Four benchmark models are compared to the models with Twitter-based proxies of TCI and TCS. A univariate expectation model with 1) lagged sales, and multivariate models with 2) lagged sales and lagged GDP, 3) lagged sales and accounts receivable, and 4) lagged sales, accounts receivable and lagged GDP are estimated. These models are compared to the expectation models that contain Twitter-based proxies.

Extant research and audit standards suggest that prior year account balances and contemporaneous firm-specific information is useful in predicting current year account balances (SAS No. 56, AICPA 1988; Hoitash et al. 2006; Minutti-Meza 2011). Furthermore, timelier predictive analysis has the potential to improve the accuracy of

---

[6] https://www.bea.gov/national/

predictions (Kogan et al. 2014; Yoon 2016). Consequently, prior year sales, or prior month sales, and accounts receivable are included as explanatory variables. GDP is expected to impact firms' sales performance (Lev 1980), however, because GDP is not readily available at the time of the audit, lagged GDP for the prior year, or prior month, is incorporated into the benchmark expectation models. Benchmark expectation models for traditional substantive analytical procedures are presented as:

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \epsilon \qquad \mathbf{1}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 GDP_{t-12} + \epsilon \qquad \mathbf{2}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \epsilon \qquad \mathbf{3}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \beta_3 GDP_{t-12} + \epsilon \qquad \mathbf{4}$$

Where $\mathrm{Sales}_{\mathrm{it}}$ represents total sales for firm I and in month t. $\mathrm{Sales}_{\mathrm{it-12}}$ represents total sales for the same month in the last year. $\mathrm{AR}_{\mathrm{it}}$ is total accounts receivable for firm i and in month t. $\mathrm{GDP}_{\mathrm{t-12}}$ is the gross domestic product for the same month in the last year. The expectation models for continuous substantive analytical procedures are depicted in similar form with the exception that lagged sales from prior year, and lagged GDP from prior year, are replaced by lagged sales from prior month, and lagged GDP from prior month.

As this study aims to advance analytical procedures by examining the value of external nonfinancial information from social media platforms using Twitter as an example, models 1 to 4 are compared to the expectation models that contain Twitter-based measures of TCI and TCS. These measures have the potential to be useful sources of information for auditors as they can capture broad consumer views about products or brands (Tang 2017). The specification of the expectation models that contain Twitter-based proxies is as follows:

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 TCI_{it} + \epsilon \qquad \mathbf{5}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 TCS_{it} + \epsilon \qquad \mathbf{6}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 TCI_{it} + \beta_3 GDP_{t-12} + \epsilon \qquad \mathbf{7}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 TCS_{it} + \beta_3 GDP_{t-12} + \epsilon \qquad \mathbf{8}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \beta_3 TCI_{it} + \epsilon \qquad \mathbf{9}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \beta_3 TCS_{it} + \epsilon \qquad \mathbf{10}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \beta_3 TCI_{it} + \beta_3 GDP_{t-12} + \epsilon \qquad \mathbf{11}$$

$$Sales_{it} = \beta_0 + \beta_1 Sales_{it-12} + \beta_2 AR_{it} + \beta_3 TCS_{it} + \beta_3 GDP_{t-12} + \epsilon \qquad \mathbf{12}$$

In models 5 to 12, $TCI_{it}$ represents third-party mentions on Twitter of past, or future, purchases of individual brands or products aggregated at the firm level for firm *i* and month *t*. $TCS_{it}$ represents third-party reviews (positive or negative) of individual brands or products aggregated at the firm level for firm *i* and month *t*. The expectation models for continuous substantive analytical procedures are depicted in similar form with the exception that lagged sales from prior year, and lagged GDP from prior year, are replaced by lagged sales from prior month, and lagged GDP from prior month.

### 2.3.5 Model Comparison

#### *Prediction Performance*

Following the prior literature on analytical procedures, this study first evaluates the prediction performance of the models by generating monthly account balance predictions for the training period and then by generating monthly account balance predictions from out-of-sample observations, the testing period. Each model is trained and validated for each firm in the sample. In this study, the training set comprises observations from the 2012 to

2016 period, and the out-of-sample set comprises observations from the 2017 period. Out-of-sample prediction performance is evaluated as follows: model 1 is compared to models 5, and 6; model 2, is compared to models 7, and 8. model 3, is compared to models 9, and 10; and model 4, is compared to models 11, and 12.

Out-of-sample prediction performance is evaluated using MAPE (Mean Absolute Percentage Error). MAPE is calculated as the absolute difference between the actual value and the predicted value for each firm using each firms' monthly observations:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Actual\ Value_{t}\_Predicted\ Value_{t}|}{Actual\ Value_{t}}$$

A smaller MAPE, indicating a smaller forecast error, is preferable. The average MAPE by industry is then calculated to evaluate the prediction performance of each model at the industry level. The aggregated MAPE can provide a general view of the predictive performance of each model. Moreover, to examine if the results produced by the expectation models with Twitter-based proxies are statistically superior to those produced by the benchmark models, a Wilcoxon Sign-Rank test is separately applied to each industry.

***False Positive and False Negative Errors***

Subsequent to evaluating prediction performance, this study aims to examine the error detection ability of the models. False positive and false negative error percentages are used to measure the error detection ability of each model. Smaller error rates are preferable. Important to consider is that false positive and false negative errors bear different costs to auditors. A false positive is defined as an error where a prediction model incorrectly identifies an error in an account balance. Whereas a false negative is defined as an error where a prediction model does not correctly identify an error in an account balance (Alpaydin 2014). While false positives would create more work for auditors, false negatives would be more detrimental to the audit firm. As an example, false positives would lead to auditors' investigating items that do not require investigation, but a false negative would fail to alert auditors of potential fraud, or material misstatement. Accordingly, it is important to investigate the error detection ability of prediction models that contain Twitter-based proxies.

A simulated experiment where errors are seeded into the dependent variable (i.e. overstating sales) is used to evaluate the error detection ability of the benchmark prediction models and prediction models with Twitter-based measures for the out-of-sample set. The error detection ability of the models is evaluated using different parameters. As quarterly financial statement information has been reviewed by auditors, false positives are tested by using quarterly financial information obtained from Compustat. Hence, a false positive is identified when a model detects an error when in reality, there is no error. In contrast, to test false negatives, errors are randomly seeded into sales, the dependent variable. The error rate seeded into the sales account balance is 4% of the actual balance. A false negative is identified when a model fails to flag the seeded error. Accounting errors related to sales are examined as AAERS (Accounting and Auditing Enforcement Releases) generally indicate that these are areas where fraud or errors are usually found.

The evaluation of error detection performance in an analytical model comprises two components (Kinney 1987; Hoitash et al. 2006; Kogan et al. 2014). First, a prediction interval for the predicted value is estimated and used as the acceptable threshold of variance. Second, a statistical investigation rule is applied to evaluate whether the value of the prediction falls within the acceptable threshold. If the value of the prediction falls outside the upper or lower limits of this threshold, then the observation would be identified as an error. The size of the prediction interval specifies the magnitude of tolerable error the auditor is willing to accept. In other words, it specifies the auditor's risk level, which is measured by $\alpha$. A smaller $\alpha$ will lead to a wider interval and fewer false positives but larger false negatives. A larger $\alpha$ will lead to a narrower interval and larger false positives but fewer false negatives. As a result, this study employs $\alpha = 0.33$ and $\alpha = 0.05$ to evaluate error detection ability under varying risk levels. Models that exhibit both lower false positive and lower false negative errors are considered to be more effective.

## 2.4 Results – Prediction Performance

The first research question investigates whether traditional and continuous substantive analytical models containing Twitter-based proxies for consumer interest, TCI, and satisfaction, TCS, produce more accurate revenue predictions than the benchmark models. The evaluation performance of models 5 and 1, 7 and 2, 9 and 3, and 11 and 4 for TCI and the evaluation performance of models 6 and 1, 8 and 2, 10 and 3, and 12 and 4 for TCS is examined. The MAPE is computed for each firm in the prediction period. The average MAPE is then computed to evaluate the prediction performance of the models with Twitter-based proxies with the benchmark models by industry. The Wilcoxon Sign-Rank test is used to evaluate whether the MAPE difference between the Twitter-based proxies and benchmark models is statistically different for each industry. Tables 4, 5, and 6 present the results for research question 1A. Tables 7, 8, and 9 present the result for research question 1B. Results are presented for the 24 industries contained in the sample.

***Traditional Substantive Analytical Models***

Table 4 presents the results of traditional analytical models with TCI, models 5, 7, 9 and 11, and traditional benchmark models 1, 2, 3, and 4. As indicated by Table 4, the models with TCI generate a smaller, or better, MAPE than the MAPE of the benchmark models for the majority of the industries that are examined. Models 5, 7, and 9 generate more accurate account predictions for 16 of the 24 industries, all of which have statistically significant differences. Interestingly, the results indicate that the simple model, model 5, generates account predictions that are just as superior as the account predictions generated by the models that incorporate more information, models 7, and 9. The prediction performance of model 11 is diluted as it generates more accurate predictions for 14 of the 24 industries, all of which have MAPE differences that are statistically significant.

The results in Table 5 illustrate the prediction performance of traditional analytical models with TCS, models, 6, 8, 10, and 12, and traditional benchmark models 1, 2, 3, and 4. The models with TCS generally produce more accurate account predictions, as indicated by the smaller MAPE, for the majority of the industries that are examined. Models 6, 8, 10, and 12 generate better predictions for 15, 14, 12 and 15 industries, respectively. The results of the nonparametric test suggest that the differences in MAPE are statistically significant for these models. Model 6, the simple model could be considered as the better model in this case as it is able to produce more accurate predictions for 15 of the 24 industries, while the prediction accuracy of models 8 and 10 appears to be diluted when GDP or AR information is included with TCS. The results from model 12 suggest that TCS is complemented by both external macroeconomic and contemporaneous firm-specific information as this is the model that incorporates lagged sales, TCS, GDP, and AR.

## Table 4: Prediction Performance of Traditional Substantive Analytical Models with TCI and without TCI (Models 5, 7, 9 and 11 and 1, 2, 3, and 4)

| | (1) | (5) | | | | (2) | (7) | | | | (3) | (9) | | | | (4) | (11) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Salest-12 | Salest-12+TweetCI | | | | Salest-12+GDPt-12 | Salest-12+TweetCI+GDP | | | | Salest-12+AR | Salest-12+AR+TweetCI | | | | Salest-12+AR+GDPt-12 | Salest12+AR+TweetCI+GD | | | |
| 2-Digit SIC | MAPE1 | MAPE5 | Difference | B/W | p-value | MAPE2 | MAPE7 | Difference | B/W | p-value | MAPE3 | MAPE9 | Difference | B/W | p-value | MAPE4 | MAPE11 | Difference | B/W | p-value |
| 20 | 0.0827 | 0.0729 | 0.0099 | B | 0.000 | 0.061 | 0.056 | 0.005 | B | 0.000 | 0.0531 | 0.0485 | 0.0046 | B | 0.001 | 0.044 | 0.041 | 0.003 | B | 0.000 |
| 21 | 0.0247 | 0.0211 | 0.0037 | B | 0.001 | 0.025 | 0.020 | 0.005 | B | 0.001 | 0.0203 | 0.0187 | 0.0016 | B | 0.001 | 0.020 | 0.019 | 0.001 | B | 0.001 |
| 23 | 0.0877 | 0.0703 | 0.0174 | B | 0.000 | 0.059 | 0.057 | 0.002 | B | 0.000 | 0.0664 | 0.0536 | 0.0129 | B | 0.000 | 0.044 | 0.045 | -0.001 | W | 0.077 |
| 28 | 0.0450 | 0.0468 | -0.0018 | W | 0.098 | 0.036 | 0.035 | 0.002 | B | 0.005 | 0.0311 | 0.0318 | -0.0007 | W | 0.014 | 0.030 | 0.031 | 0.000 | W | 0.224 |
| 29 | 0.1176 | 0.1307 | -0.0131 | W | 0.034 | 0.161 | 0.129 | 0.033 | B | 0.000 | 0.0776 | 0.0794 | -0.0019 | W | 0.034 | 0.074 | 0.080 | -0.006 | W | 0.034 |
| 30 | 0.0225 | 0.0224 | 0.0000 | NoDiff | 0.001 | 0.021 | 0.021 | 0.001 | B | 0.001 | 0.0234 | 0.0235 | -0.0001 | W | 0.001 | 0.022 | 0.022 | 0.001 | B | 0.001 |
| 31 | 0.0903 | 0.0968 | -0.0065 | W | 0.000 | 0.081 | 0.084 | -0.002 | W | 0.000 | 0.0838 | 0.0853 | -0.0015 | W | 0.000 | 0.063 | 0.065 | -0.002 | W | 0.034 |
| 35 | 0.0962 | 0.0593 | 0.0369 | B | 0.000 | 0.051 | 0.051 | 0.000 | NoDiff | 0.034 | 0.0721 | 0.0465 | 0.0256 | B | 0.034 | 0.044 | 0.044 | 0.000 | B | 0.034 |
| 36 | 0.0635 | 0.0584 | 0.0051 | B | 0.003 | 0.045 | 0.042 | 0.003 | B | 0.137 | 0.0318 | 0.0281 | 0.0038 | B | 0.003 | 0.042 | 0.041 | 0.001 | B | 0.420 |
| 37 | 0.1201 | 0.1033 | 0.0168 | B | 0.000 | 0.088 | 0.082 | 0.005 | B | 0.000 | 0.0873 | 0.0801 | 0.0072 | B | 0.015 | 0.072 | 0.071 | 0.001 | B | 0.005 |
| 39 | 0.0912 | 0.0686 | 0.0226 | B | 0.000 | 0.073 | 0.072 | 0.001 | B | 0.077 | 0.0946 | 0.0697 | 0.0249 | B | 0.000 | 0.073 | 0.075 | -0.002 | W | 0.000 |
| 42 | 0.0278 | 0.0234 | 0.0044 | B | 0.001 | 0.020 | 0.020 | 0.001 | B | 0.001 | 0.0243 | 0.0212 | 0.0031 | B | 0.001 | 0.022 | 0.021 | 0.002 | B | 0.001 |
| 44 | 0.0318 | 0.0242 | 0.0075 | B | 0.034 | 0.019 | 0.018 | 0.000 | B | 0.000 | 0.0219 | 0.0217 | 0.0002 | B | 0.034 | 0.019 | 0.019 | 0.000 | NoDiff | 0.034 |
| 45 | 0.0574 | 0.0559 | 0.0014 | B | 0.114 | 0.054 | 0.053 | 0.001 | B | 0.062 | 0.0480 | 0.0450 | 0.0030 | B | 0.000 | 0.048 | 0.045 | 0.003 | B | 0.000 |
| 47 | 0.0381 | 0.0392 | -0.0011 | W | 0.001 | 0.040 | 0.041 | -0.001 | W | 0.001 | 0.0441 | 0.0405 | 0.0036 | B | 0.001 | 0.038 | 0.039 | 0.000 | W | 0.001 |
| 48 | 0.0254 | 0.0243 | 0.0010 | B | 0.000 | 0.020 | 0.020 | -0.001 | W | 0.034 | 0.0200 | 0.0203 | -0.0002 | W | 0.034 | 0.020 | 0.020 | 0.000 | W | 0.034 |
| 53 | 0.0393 | 0.0407 | -0.0014 | W | 0.001 | 0.041 | 0.042 | -0.001 | W | 0.001 | 0.0384 | 0.0434 | -0.0050 | W | 0.001 | 0.044 | 0.042 | 0.002 | B | 0.001 |
| 55 | 0.0343 | 0.0277 | 0.0067 | B | 0.034 | 0.028 | 0.030 | -0.002 | W | 0.000 | 0.0556 | 0.0402 | 0.0154 | B | 0.034 | 0.039 | 0.039 | 0.000 | W | 0.034 |
| 57 | 0.0802 | 0.0800 | 0.0002 | B | 0.034 | 0.068 | 0.068 | 0.000 | W | 0.034 | 0.0808 | 0.0814 | -0.0005 | W | 0.034 | 0.070 | 0.070 | 0.000 | B | 0.034 |
| 58 | 0.0533 | 0.0502 | 0.0031 | B | 0.009 | 0.041 | 0.040 | 0.001 | B | 0.000 | 0.0457 | 0.0453 | 0.0004 | B | 0.000 | 0.039 | 0.038 | 0.001 | B | 0.000 |
| 59 | 0.0867 | 0.0860 | 0.0008 | B | 0.000 | 0.072 | 0.071 | 0.001 | B | 0.000 | 0.0659 | 0.0658 | 0.0001 | B | 0.077 | 0.065 | 0.064 | 0.001 | B | 0.000 |
| 70 | 0.1159 | 0.1058 | 0.0101 | B | 0.034 | 0.070 | 0.065 | 0.005 | B | 0.034 | 0.0488 | 0.0466 | 0.0021 | B | 0.034 | 0.047 | 0.045 | 0.001 | B | 0.034 |
| 73 | 0.0387 | 0.0376 | 0.0011 | B | 0.054 | 0.032 | 0.029 | 0.003 | B | 0.000 | 0.0407 | 0.0394 | 0.0013 | B | 0.003 | 0.034 | 0.029 | 0.005 | B | 0.000 |
| 75 | 0.0153 | 0.0153 | 0.0000 | NoDiff | 0.001 | 0.014 | 0.014 | 0.000 | B | 0.001 | 0.0131 | 0.0130 | 0.0000 | NoDiff | 0.001 | 0.011 | 0.010 | 0.001 | B | 0.001 |

## Table 5: Prediction Performance of Traditional Substantive Analytical Models with TCS and without TCS (Models 6, 8, 10 and 12 and 1, 2, 3, and 4)

| | (1) | (6) | | | | (2) | (8) | | | | (3) | (10) | | | | (4) | (12) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Salest-12 | Salest-12+TweetCS | | | | Salest-12+GDPt-12 | Salest-12+TweetCS+GD | | | | Salest-12+AR | Salest-12+AR+TweetCS | | | | Salest-12+AR+GDPt-12 | Salest-12+AR+TweetCS | | | |
| 2-Digit SIC | MAPE1 | MAPE6 | Difference | B/W | p-value | MAPE2 | MAPE8 | Difference | B/W | p-value | MAPE3 | MAPE10 | Difference | B/W | p-value | MAPE4 | MAPE12 | Difference | B/W | p-value |
| 20 | 0.0827 | 0.0760 | 0.0067 | B | 0.000 | 0.0613 | 0.0575 | 0.0038 | B | 0.000 | 0.0531 | 0.0490 | 0.0042 | B | 0.000 | 0.0442 | 0.0427 | 0.0015 | B | 0.000 |
| 21 | 0.0247 | 0.0237 | 0.0011 | B | 0.001 | 0.0252 | 0.0241 | 0.0011 | B | 0.001 | 0.0203 | 0.0190 | 0.0013 | B | 0.001 | 0.0204 | 0.0199 | 0.0005 | B | 0.001 |
| 23 | 0.0877 | 0.0765 | 0.0111 | B | 0.000 | 0.0587 | 0.0574 | 0.0013 | B | 0.000 | 0.0664 | 0.0537 | 0.0127 | B | 0.000 | 0.0441 | 0.0435 | 0.0006 | B | 0.000 |
| 28 | 0.0450 | 0.0422 | 0.0028 | B | 0.014 | 0.0363 | 0.0339 | 0.0023 | B | 0.000 | 0.0311 | 0.0321 | -0.0011 | W | 0.437 | 0.0302 | 0.0296 | 0.0006 | B | 0.000 |
| 29 | 0.1176 | 0.1184 | -0.0008 | W | 0.034 | 0.1613 | 0.1622 | -0.0009 | W | 0.000 | 0.0776 | 0.0789 | -0.0013 | W | 0.000 | 0.0738 | 0.0741 | -0.0003 | W | 0.000 |
| 30 | 0.0225 | 0.0224 | 0.0000 | NoDiff | 0.001 | 0.0214 | 0.0215 | 0.0000 | NoDiff | 0.001 | 0.0234 | 0.0234 | 0.0000 | NoDiff | 0.001 | 0.0224 | 0.0224 | 0.0000 | NoDiff | 0.001 |
| 31 | 0.0903 | 0.0913 | -0.0010 | W | 0.034 | 0.0813 | 0.0809 | 0.0004 | B | 0.034 | 0.0838 | 0.0841 | -0.0004 | W | 0.034 | 0.0630 | 0.0640 | -0.0011 | W | 0.034 |
| 35 | 0.0962 | 0.0942 | 0.0020 | B | 0.034 | 0.0512 | 0.0483 | 0.0029 | B | 0.034 | 0.0721 | 0.0694 | 0.0027 | B | 0.000 | 0.0442 | 0.0436 | 0.0006 | B | 0.034 |
| 36 | 0.0635 | 0.0480 | 0.0155 | B | 0.000 | 0.0449 | 0.0518 | -0.0068 | W | 0.000 | 0.0318 | 0.0441 | -0.0122 | W | 0.022 | 0.0422 | 0.0454 | -0.0032 | W | 0.022 |
| 37 | 0.1201 | 0.1075 | 0.0126 | B | 0.000 | 0.0877 | 0.0844 | 0.0033 | B | 0.883 | 0.0873 | 0.0800 | 0.0073 | B | 0.000 | 0.0721 | 0.0674 | 0.0047 | B | 0.000 |
| 39 | 0.0912 | 0.0778 | 0.0134 | B | 0.000 | 0.0729 | 0.0621 | 0.0108 | B | 0.000 | 0.0946 | 0.0805 | 0.0141 | B | 0.000 | 0.0734 | 0.0620 | 0.0114 | B | 0.000 |
| 42 | 0.0278 | 0.0269 | 0.0008 | B | 0.001 | 0.0202 | 0.0160 | 0.0041 | B | 0.001 | 0.0243 | 0.0248 | -0.0005 | W | 0.001 | 0.0224 | 0.0206 | 0.0018 | B | 0.001 |
| 44 | 0.0318 | 0.0299 | 0.0019 | B | 0.034 | 0.0186 | 0.0192 | -0.0006 | W | 0.000 | 0.0219 | 0.0225 | -0.0006 | W | 0.000 | 0.0187 | 0.0196 | -0.0009 | W | 0.000 |
| 45 | 0.0574 | 0.0537 | 0.0036 | B | 0.017 | 0.0542 | 0.0511 | 0.0031 | B | 0.000 | 0.0480 | 0.0478 | 0.0002 | B | 0.114 | 0.0477 | 0.0466 | 0.0011 | B | 0.001 |
| 47 | 0.0381 | 0.0398 | -0.0017 | W | 0.001 | 0.0405 | 0.0456 | -0.0051 | W | 0.001 | 0.0441 | 0.0423 | 0.0018 | B | 0.001 | 0.0382 | 0.0453 | -0.0071 | W | 0.001 |
| 48 | 0.0254 | 0.0351 | -0.0097 | W | 0.034 | 0.0196 | 0.0242 | -0.0046 | W | 0.034 | 0.0200 | 0.0250 | -0.0050 | W | 0.000 | 0.0199 | 0.0243 | -0.0043 | W | 0.000 |
| 53 | 0.0393 | 0.0408 | -0.0015 | W | 0.001 | 0.0411 | 0.0406 | 0.0005 | B | 0.001 | 0.0384 | 0.0400 | -0.0016 | W | 0.001 | 0.0439 | 0.0413 | 0.0027 | B | 0.001 |
| 55 | 0.0343 | 0.0356 | -0.0013 | W | 0.000 | 0.0282 | 0.0301 | -0.0019 | W | 0.000 | 0.0556 | 0.0495 | 0.0061 | B | 0.034 | 0.0392 | 0.0351 | 0.0040 | B | 0.034 |
| 57 | 0.0802 | 0.0719 | 0.0083 | B | 0.000 | 0.0683 | 0.0588 | 0.0095 | B | 0.034 | 0.0808 | 0.0735 | 0.0073 | B | 0.000 | 0.0697 | 0.0580 | 0.0117 | B | 0.034 |
| 58 | 0.0533 | 0.0503 | 0.0030 | B | 0.000 | 0.0406 | 0.0400 | 0.0006 | B | 0.326 | 0.0457 | 0.0430 | 0.0027 | B | 0.000 | 0.0392 | 0.0384 | 0.0008 | B | 0.027 |
| 59 | 0.0867 | 0.0979 | -0.0111 | W | 0.000 | 0.0718 | 0.0671 | 0.0047 | B | 0.000 | 0.0659 | 0.0662 | -0.0002 | W | 0.599 | 0.0646 | 0.0638 | 0.0008 | B | 0.077 |
| 70 | 0.1159 | 0.0598 | 0.0561 | B | 0.000 | 0.0703 | 0.0576 | 0.0127 | B | 0.000 | 0.0488 | 0.0460 | 0.0028 | B | 0.000 | 0.0469 | 0.0456 | 0.0013 | B | 0.000 |
| 73 | 0.0387 | 0.0347 | 0.0040 | B | 0.000 | 0.0322 | 0.0312 | 0.0010 | B | 0.000 | 0.0407 | 0.0355 | 0.0052 | B | 0.000 | 0.0339 | 0.0333 | 0.0006 | B | 0.437 |
| 75 | 0.0153 | 0.0168 | -0.0015 | W | 0.001 | 0.0142 | 0.0163 | -0.0021 | W | 0.001 | 0.0131 | 0.0154 | -0.0023 | W | 0.001 | 0.0112 | 0.0144 | -0.0032 | W | 0.001 |

Collectively, the results for RQ1A displayed in tables 4 and 5 suggest that both TCI and TCS contain incremental information that can enhance the prediction performance of traditional substantive analytical models. However, the results indicate that the simple model with TCI information generally provides higher prediction accuracy than the models with TCS information. Table 6 displays a summary of these findings. While both Twitter-based proxies have the potential to improve the prediction performance of traditional substantive analytical models, it is possible that model 5, which includes lagged sales and TCI, would be preferred by auditors over TCS models as this is the less complex model that generates superior account predictions for most of the industries that are examined.

**Table 6: Prediction Performance Summary of Traditional Substantive Analytical Models for 24 Industries**

| | Twitter Consumer Interest | | | | Twitter Consumer Satisfaction | | | |
|---|---|---|---|---|---|---|---|---|
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 16 of 24 | 16 of 24 | 16 of 24 | 14 of 24 | 15 of 24 | 14 of 24 | 12 of 24 | 15 of 24 |

***Continuous Substantive Analytical Models***

The results in Tables 7 and 8 present the results of continuous substantive analytical models, research question 1B. Table 7 compares models with TCI, models 5, 7, 9, and 11, to benchmark models 1, 2, 3, and 4. As illustrated in Table 7, model 5 produces a smaller MAPE and therefore generates better account predictions for 19 of the 24 industries. The predictions improve as more information is included. Models 7 and 11 produce marginally superior predictions for 21 and 22 of the 24 industries that are analyzed. However, when only contemporaneous financial information is included with TCI, model 9, the predictive power of the model is diluted as this model generated better predictions for 18 of the 24 industries. All presented differences in MAPE are statistically significant. Auditors may benefit from utilizing models 5, 7, or 11. However, the choice between these models would depend on the costs and benefits of obtaining additional information. Model 5 can produce superior account predictions for the majority of the industries, yet, model 7 and 11 produce marginally superior account predictions than model 5 but incorporate more information.

The results in Table 8 compare models with TCS, models 6, 8, 10, and 12, to benchmark models 1, 2, 3, and 4. The results in Table 8 indicate that models 6 and 10 produce superior account predictions for 14 of the 24 industries. The prediction performance improves as indicated by models 8 and 12 as they are able produce better account predictions for 20 or 22 industries, respectively. MAPE differences of the presented models are statistically significant. In the case of TCS, models 8 and 12 could be beneficial to auditors. Model 8 produces superior predictions for most industries while the prediction performance of model 12 is marginally superior than model 8 but would require the inclusion of additional information.

# Table 7: Prediction Performance of Continuous Substantive Analytical Models with TCI and without TCI (Models 5, 7, 9 and 11 and 1, 2, 3, and 4)

| | (1) | (5) | | | (2) | (7) | | | (3) | (9) | | | (4) | (11) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 | Saletst-1+TweetCI | | | Saletst-1+GDPt-1 | Saletst-1+TweetCI+GDPt-1 | | | Saletst-1+AR | Saletst-1+AR+TweetCI | | | Saletst-1+AR+GDPt-1 | Salest12+AR+TweetCI+GDPt-1 | | |
| 2-Digit SIC | MAPE1 | MAPE5 | Difference B/W | p-value | MAPE2 | MAPE7 | Difference B/W | p-value | MAPE3 | MAPE9 | Difference B/W | p-value | MAPE4 | MAPE11 | Difference B/W | p-value |
| 20 | 0.1015 | 0.0921 | 0.0094 B | 0.000 | 0.0855 | 0.0545 | 0.0310 B | 0.000 | 0.083 | 0.078 | 0.005 B | 0.000 | 0.079 | 0.040 | 0.039 B | 0.000 |
| 21 | 0.0577 | 0.0565 | 0.0012 B | 0.001 | 0.0492 | 0.0199 | 0.0292 B | 0.001 | 0.058 | 0.056 | 0.001 B | 0.001 | 0.047 | 0.019 | 0.028 B | 0.001 |
| 23 | 0.1439 | 0.1370 | 0.0069 B | 0.000 | 0.1271 | 0.0576 | 0.0695 B | 0.000 | 0.101 | 0.086 | 0.015 B | 0.000 | 0.091 | 0.045 | 0.046 B | 0.000 |
| 28 | 0.0735 | 0.0733 | 0.0001 B | 0.224 | 0.0547 | 0.0346 | 0.0200 B | 0.000 | 0.045 | 0.046 | -0.001 W | 0.043 | 0.046 | 0.030 | 0.016 B | 0.000 |
| 29 | 0.0578 | 0.0573 | 0.0005 B | 0.034 | 0.0825 | 0.1403 | -0.0578 W | 0.000 | 0.055 | 0.058 | -0.003 W | 0.000 | 0.055 | 0.079 | -0.024 W | 0.000 |
| 30 | 0.0498 | 0.0508 | -0.0010 W | 0.001 | 0.0349 | 0.0211 | 0.0137 B | 0.001 | 0.037 | 0.037 | 0.000 B | 0.001 | 0.035 | 0.022 | 0.013 B | 0.001 |
| 31 | 0.1686 | 0.1681 | 0.0005 B | 0.034 | 0.1411 | 0.0816 | 0.0595 B | 0.000 | 0.128 | 0.124 | 0.003 B | 0.034 | 0.107 | 0.064 | 0.042 B | 0.000 |
| 35 | 0.1193 | 0.1105 | 0.0088 B | 0.000 | 0.1057 | 0.0493 | 0.0564 B | 0.000 | 0.070 | 0.072 | -0.002 W | 0.000 | 0.066 | 0.043 | 0.022 B | 0.000 |
| 36 | 0.1020 | 0.0955 | 0.0064 B | 0.000 | 0.1090 | 0.0385 | 0.0705 B | 0.000 | 0.095 | 0.094 | 0.002 B | 0.137 | 0.104 | 0.039 | 0.065 B | 0.000 |
| 37 | 0.1122 | 0.1121 | 0.0001 B | 0.254 | 0.0879 | 0.0808 | 0.0071 B | 0.137 | 0.106 | 0.105 | 0.001 B | 0.841 | 0.080 | 0.071 | 0.009 B | 0.841 |
| 39 | 0.3305 | 0.3007 | 0.0298 B | 0.000 | 0.3411 | 0.0723 | 0.2688 B | 0.000 | 0.213 | 0.209 | 0.004 B | 0.077 | 0.180 | 0.075 | 0.105 B | 0.000 |
| 42 | 0.0754 | 0.0415 | 0.0339 B | 0.001 | 0.0560 | 0.0182 | 0.0378 B | 0.001 | 0.060 | 0.044 | 0.016 B | 0.001 | 0.053 | 0.020 | 0.033 B | 0.001 |
| 44 | 0.1499 | 0.1396 | 0.0103 B | 0.034 | 0.1525 | 0.0184 | 0.1341 B | 0.000 | 0.155 | 0.144 | 0.011 B | 0.000 | 0.152 | 0.018 | 0.134 B | 0.000 |
| 45 | 0.0745 | 0.0647 | 0.0098 B | 0.000 | 0.0646 | 0.0540 | 0.0106 B | 0.000 | 0.070 | 0.062 | 0.008 B | 0.000 | 0.062 | 0.045 | 0.016 B | 0.000 |
| 47 | 0.1753 | 0.1294 | 0.0459 B | 0.001 | 0.1318 | 0.0404 | 0.0914 B | 0.001 | 0.128 | 0.126 | 0.002 B | 0.001 | 0.125 | 0.040 | 0.084 B | 0.001 |
| 48 | 0.0305 | 0.0288 | 0.0018 B | 0.000 | 0.0301 | 0.0200 | 0.0101 B | 0.000 | 0.030 | 0.029 | 0.002 B | 0.000 | 0.030 | 0.020 | 0.010 B | 0.000 |
| 53 | 0.1947 | 0.1184 | 0.0764 B | 0.001 | 0.1801 | 0.0418 | 0.1383 B | 0.001 | 0.190 | 0.125 | 0.065 B | 0.001 | 0.181 | 0.043 | 0.138 B | 0.001 |
| 55 | 0.1088 | 0.0978 | 0.0110 B | 0.034 | 0.1212 | 0.0336 | 0.0875 B | 0.000 | 0.093 | 0.076 | 0.017 B | 0.000 | 0.069 | 0.041 | 0.028 B | 0.000 |
| 57 | 0.1206 | 0.1042 | 0.0163 B | 0.034 | 0.0986 | 0.0673 | 0.0313 B | 0.034 | 0.102 | 0.103 | -0.001 W | 0.034 | 0.106 | 0.068 | 0.038 B | 0.034 |
| 58 | 0.0687 | 0.0669 | 0.0017 B | 0.000 | 0.0555 | 0.0416 | 0.0139 B | 0.000 | 0.063 | 0.058 | 0.005 B | 0.000 | 0.055 | 0.039 | 0.016 B | 0.000 |
| 59 | 0.2546 | 0.2515 | 0.0031 B | 0.000 | 0.2451 | 0.0740 | 0.1710 B | 0.077 | 0.253 | 0.251 | 0.002 B | 0.000 | 0.243 | 0.063 | 0.180 B | 0.000 |
| 70 | 0.0530 | 0.0543 | -0.0013 W | 0.000 | 0.0498 | 0.0639 | -0.0141 W | 0.034 | 0.053 | 0.052 | 0.001 B | 0.034 | 0.049 | 0.045 | 0.004 B | 0.034 |
| 73 | 0.0849 | 0.0800 | 0.0049 B | 0.437 | 0.0732 | 0.0300 | 0.0431 B | 0.000 | 0.076 | 0.075 | 0.001 B | 0.003 | 0.071 | 0.031 | 0.040 B | 0.000 |
| 75 | 0.1525 | 0.0922 | 0.0602 B | 0.001 | 0.1469 | 0.0143 | 0.1326 B | 0.001 | 0.090 | 0.075 | 0.015 B | 0.001 | 0.092 | 0.011 | 0.081 B | 0.001 |

# Table 8: Prediction Performance of Continuous Substantive Analytical Models with TCS and without TCS (Models 6, 8, 10 and 12 and 1, 2, 3, and 4)

| | (1) | (6) | | | (2) | (8) | | | (3) | (10) | | | (4) | (12) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 | Saletst-1+TweetCS | | | Saletst-1+GDPt-1 | Saletst-1+TweetCS+GDPt-1 | | | Saletst-1+AR | Saletst-1+AR+TweetCS | | | Saletst-1+AR+GDPt-1 | Saletst-1+AR+TweetCS+GDPt-1 | | |
| 2-Digit SIC | MAPE1 | MAPE6 | Difference B/W | p-value | MAPE2 | MAPE8 | Difference B/W | p-value | MAPE3 | MAPE10 | Difference B/W | p-value | MAPE4 | MAPE12 | Difference B/W | p-value |
| 20 | 0.1015 | 0.0883 | 0.0132 B | 0.000 | 0.0855 | 0.0576 | 0.0280 B | 0.000 | 0.0828 | 0.0797 | 0.0031 B | 0.000 | 0.0792 | 0.0421 | 0.0372 B | 0.000 |
| 21 | 0.0577 | 0.0591 | -0.0014 W | 0.001 | 0.0492 | 0.0242 | 0.0249 B | 0.001 | 0.0576 | 0.0584 | -0.0008 W | 0.001 | 0.0472 | 0.0201 | 0.0271 B | 0.001 |
| 23 | 0.1439 | 0.1364 | 0.0075 B | 0.000 | 0.1271 | 0.0577 | 0.0694 B | 0.000 | 0.1013 | 0.0977 | 0.0036 B | 0.000 | 0.0908 | 0.0435 | 0.0473 B | 0.000 |
| 28 | 0.0735 | 0.0706 | 0.0028 B | 0.224 | 0.0547 | 0.0336 | 0.0211 B | 0.000 | 0.0455 | 0.0453 | 0.0001 B | 0.398 | 0.0459 | 0.0291 | 0.0168 B | 0.000 |
| 29 | 0.0578 | 0.0463 | 0.0116 B | 0.000 | 0.0825 | 0.1715 | -0.0889 W | 0.000 | 0.0551 | 0.0558 | -0.0006 W | 0.000 | 0.0553 | 0.0730 | -0.0178 W | 0.000 |
| 30 | 0.0498 | 0.0449 | 0.0049 B | 0.001 | 0.0349 | 0.0216 | 0.0132 B | 0.001 | 0.0372 | 0.0370 | 0.0002 B | 0.001 | 0.0351 | 0.0222 | 0.0129 B | 0.001 |
| 31 | 0.1686 | 0.1559 | 0.0127 B | 0.000 | 0.1411 | 0.0842 | 0.0570 B | 0.000 | 0.1276 | 0.1245 | 0.0031 B | 0.034 | 0.1067 | 0.0627 | 0.0440 B | 0.000 |
| 35 | 0.1193 | 0.1218 | -0.0025 W | 0.000 | 0.1057 | 0.0465 | 0.0592 B | 0.000 | 0.0695 | 0.0731 | -0.0036 W | 0.000 | 0.0655 | 0.0425 | 0.0230 B | 0.000 |
| 36 | 0.1020 | 0.1180 | -0.0161 W | 0.003 | 0.1090 | 0.0471 | 0.0619 B | 0.022 | 0.0954 | 0.1081 | -0.0126 W | 0.000 | 0.1036 | 0.0443 | 0.0592 B | 0.000 |
| 37 | 0.1122 | 0.1430 | -0.0308 W | 0.254 | 0.0879 | 0.0819 | 0.0060 B | 0.398 | 0.1058 | 0.0972 | 0.0086 B | 0.000 | 0.0798 | 0.0672 | 0.0126 B | 0.883 |
| 39 | 0.3305 | 0.3384 | -0.0080 W | 0.000 | 0.3411 | 0.0625 | 0.2786 B | 0.000 | 0.2132 | 0.2157 | -0.0025 W | 0.077 | 0.1797 | 0.0621 | 0.1176 B | 0.000 |
| 42 | 0.0754 | 0.0637 | 0.0117 B | 0.001 | 0.0560 | 0.0144 | 0.0415 B | 0.001 | 0.0605 | 0.0522 | 0.0083 B | 0.001 | 0.0532 | 0.0189 | 0.0343 B | 0.001 |
| 44 | 0.1499 | 0.1486 | 0.0013 B | 0.034 | 0.1525 | 0.0192 | 0.1332 B | 0.000 | 0.1548 | 0.1517 | 0.0030 B | 0.034 | 0.1523 | 0.0195 | 0.1328 B | 0.000 |
| 45 | 0.0745 | 0.0642 | 0.0103 B | 0.000 | 0.0646 | 0.0528 | 0.0118 B | 0.000 | 0.0697 | 0.0627 | 0.0070 B | 0.000 | 0.0616 | 0.0466 | 0.0149 B | 0.000 |
| 47 | 0.1753 | 0.1597 | 0.0156 B | 0.001 | 0.1318 | 0.0439 | 0.0879 B | 0.001 | 0.1278 | 0.1323 | -0.0045 W | 0.001 | 0.1249 | 0.0436 | 0.0813 B | 0.001 |
| 48 | 0.0305 | 0.0337 | -0.0032 W | 0.000 | 0.0301 | 0.0236 | 0.0065 B | 0.000 | 0.0304 | 0.0306 | -0.0002 W | 0.034 | 0.0299 | 0.0237 | 0.0062 B | 0.000 |
| 53 | 0.1947 | 0.1740 | 0.0207 B | 0.001 | 0.1801 | 0.0402 | 0.1399 B | 0.001 | 0.1899 | 0.1523 | 0.0375 B | 0.001 | 0.1810 | 0.0399 | 0.1411 B | 0.001 |
| 55 | 0.1088 | 0.1019 | 0.0070 B | 0.000 | 0.1212 | 0.0333 | 0.0879 B | 0.000 | 0.0931 | 0.0844 | 0.0087 B | 0.034 | 0.0692 | 0.0375 | 0.0317 B | 0.000 |
| 57 | 0.1206 | 0.1025 | 0.0181 B | 0.034 | 0.0986 | 0.0612 | 0.0374 B | 0.034 | 0.1020 | 0.0988 | 0.0031 B | 0.034 | 0.1063 | 0.0605 | 0.0458 B | 0.034 |
| 58 | 0.0687 | 0.0587 | 0.0100 B | 0.000 | 0.0555 | 0.0410 | 0.0145 B | 0.000 | 0.0626 | 0.0565 | 0.0062 B | 0.000 | 0.0550 | 0.0395 | 0.0155 B | 0.000 |
| 59 | 0.2546 | 0.2558 | -0.0012 W | 0.077 | 0.2451 | 0.0691 | 0.1760 B | 0.599 | 0.2532 | 0.2543 | -0.0011 W | 0.077 | 0.2430 | 0.0627 | 0.1803 B | 0.000 |
| 70 | 0.0530 | 0.0657 | -0.0127 W | 0.000 | 0.0498 | 0.0565 | -0.0067 W | 0.034 | 0.0532 | 0.0530 | 0.0003 B | 0.034 | 0.0491 | 0.0455 | 0.0036 B | 0.034 |
| 73 | 0.0849 | 0.0822 | 0.0027 B | 0.065 | 0.0732 | 0.0305 | 0.0427 B | 0.000 | 0.0762 | 0.0755 | 0.0007 B | 0.054 | 0.0714 | 0.0328 | 0.0387 B | 0.000 |
| 75 | 0.1525 | 0.1528 | -0.0004 W | 0.001 | 0.1469 | 0.0163 | 0.1306 B | 0.001 | 0.0901 | 0.1006 | -0.0105 W | 0.001 | 0.0919 | 0.0146 | 0.0773 B | 0.001 |

Taken together, the results in Tables 7 and 8 for RQ1B suggest that both Twitter proxies are useful in improving the prediction performance of continuous substantive analytical procedures. This is especially true when GDP information and TCI or TCS are included in the SAPs, as in models 7 and 8, or when GDP and AR information are incorporated along with TCI or TCS, as in models 11 and 12. In the case where the cost of obtaining additional information exceeds the benefits, model 5, with TCI, could be beneficial as it sacrifices the prediction performance of only 5 of the 24 industries while maintaining its simplicity.

Additionally, prediction performance of these models improved when compared to the prediction performance of traditional substantive analytical models suggesting that Twitter-based measures have more incremental predictive power for shorter time periods. Collectively, these results indicate that the timelier, or continuous substantive analytical models, that include information from social media are superior to traditional substantive analytical models that include this information. Table 9 displays a summary of the results for RQ1A and RQ1B.

**Table 9: Prediction Performance Summary of Traditional and Continuous Substantive Analytical Models for 24 Industries**

| | Twitter Consumer Interest | | | | Twitter Consumer Satisfaction | | | |
|---|---|---|---|---|---|---|---|---|
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 16 of 24 | 16 of 24 | 16 of 24 | 14 of 24 | 15 of 24 | 14 of 24 | 12 of 24 | 15 of 24 |
| Continuous - SAP | 19 of 24 | 21 of 24 | 18 of 24 | 22 of 24 | 14 of 24 | 20 of 24 | 14 of 24 | 22 of 24 |

## 2.5. Results – Error Detection Performance

The second research question examines whether traditional and continuous substantive analytical models that incorporate Twitter-based information of consumer interest and satisfaction are better able to detect errors than benchmark models. Simulated errors, 4% of the total of revenue, are randomly seeded into firm-month observations, and the procedure is repeated ten times (thus creating 10 datasets with seeded errors) to reduce bias in the results. Error detection performance is measured by identifying the number of observations that each of the models identified as an error.

Auditors are primarily concerned with high litigation costs associated with not identifying material misstatements, however, auditors also consider the costs of performing additional audit work for unusual items that do not translate to material misstatements as this has an impact on the budget. Hence, an effective model for error detection should produce relatively low false positive and false negative errors.

***Traditional Substantive Analytical Models***

Tables 10, 11, 12, and 13 display the results of error detection performance of traditional substantive analytical models with α 0.33[7] that have TCI and that do not have TCI information. Models 5 and 1, 7 and 2, 9 and 3, and 11 and 4 are compared in these tables. Collectively, the results indicate that the models generate superior error detection performance for false positives as the models with TCI produce lower false positive errors which comes at the cost of higher false negative errors. As documented in tables 10 and 11, models 5 and 7 generate lower false positive errors for 14 of the 24 industries that are analyzed, while models 9 and 11, shown on tables 12 and 13, generate lower false positive errors for only 11 of 24 industries. The error detection performance of models with TCI for false negatives is inferior to that of the benchmark models where the better performing

[7] Untabulated results of α 0.05 for all models, benchmark, TCI, and TCS, indicate the same pattern described above, though error detection performance marginally decreases.

models with TCI is able to produce at best, lower false negative errors for 6 of the 24 industries as indicated per tables 11 and 13, which display the error detection performance of models 7 and 11.

After comparing the error rates from tables 10, 11, 12, and 13, the model that leads to better error detection performance, in terms of achieving lower false positive and false negative errors, would be model 7, as it produces lower false positive and lower false negative errors for 5 of the 24 industries that are examined. Models 9 and 11 can achieve lower error rates for both types of errors for 4 industries and model 5 has lower error rates for 2 industries whereas benchmark models 3 and 4 are capable of achieving lower error percentages for 4 of the 24 industries.

Given that there is ambiguity as to the model that is more effective for error detection performance because one type of error is decreasing as the other type of error is increasing, it is necessary to analyze the ratio of the cost of errors for models that do not have Twitter information and for models with Twitter information. Analyzing the ratio of the costs of false positive and false negative errors can help to evaluate the tradeoff between the benchmark models and models with TCI (Hoitash et al. 2006). Accordingly, varying cost ratios that reflect the cost of identifying accounting errors when there are no accounting errors and the cost of not identifying accounting errors when there are accounting errors are examined.

Two cost ratios are evaluated, a cost ratio of 1:1, which assumes that false positives are as expensive as false negatives, and a cost ratio of 1:2, which assumes that false positives are half as expensive as false negatives[8]. Therefore, when the ratio of the total cost of errors is greater than 1 (i.e. sum of the costs of false positive and false negative errors for benchmark models/sum of the costs of false positive and false negative errors for models with TCI is greater than 1), it can be determined that the models with TCI are more effective than the benchmark models. As presented in tables 10, 11, 12, and 13, benchmark model 3, with AR, is more effective in detecting accounting errors as it is able to achieve better error detection performance for 16 industries, when the cost ratio is 1:1 or 1:2. For the models with TCI, model 7 is the better performing model as it is able to achieve better error detection performance for 12 industries, when the cost ratio is 1:1 or 1:2. Accordingly, traditional benchmark models outperform traditional models with TCI in terms of generating relatively low false positive and false negative error rates. Table 14 summarizes the better performing models with TCI.

[8] An additional cost ratio of 1:3 was analyzed in untabulated results, however, no significant differences across the 1:2 and 1:3 cost ratios were found. In general, under these varying cost ratios, the error detection performance did not change.

# Table 10: Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 5 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (5) | | | | | | | | | |
| | | Benchmark - Salest-12 | | Twitter - CI | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 37.11% | 21.88% | 35.10% | 19.67% | 2.01% | 2.21% | TCI | TCI | 1.08 | 1.09 | TCI | TCI |
| 21 | 12 | 37.05% | 30.00% | 34.84% | 30.00% | 2.21% | 0.00% | TCI | - | 1.03 | 1.02 | TCI* | TCI* |
| 23 | 36 | 44.44% | 19.00% | 41.48% | 19.50% | 2.96% | -0.50% | TCI | Benchmark | 1.04 | 1.02 | TCI* | TCI* |
| 28 | 72 | 37.13% | 1.43% | 40.85% | 9.64% | -3.72% | -8.21% | Benchmark | Benchmark | 0.76 | 0.66 | Benchmark | Benchmark |
| 29 | 24 | 43.00% | 13.33% | 45.26% | 0.00% | -2.27% | 13.33% | Benchmark | TCI | 1.24 | 1.54 | TCI* | TCI* |
| 30 | 12 | 34.44% | 5.00% | 33.27% | 5.00% | 1.17% | 0.00% | TCI | - | 1.03 | 1.03 | TCI* | TCI* |
| 31 | 24 | 36.02% | 25.00% | 40.20% | 25.00% | -4.18% | 0.00% | Benchmark | - | 0.94 | 0.95 | Benchmark* | Benchmark* |
| 35 | 24 | 44.01% | 0.00% | 34.89% | 11.67% | 9.12% | -11.67% | TCI | Benchmark | 0.95 | 0.76 | Benchmark* | Benchmark* |
| 36 | 48 | 33.72% | 16.00% | 30.38% | 17.33% | 3.34% | -1.33% | TCI | Benchmark | 1.04 | 1.01 | TCI* | TCI* |
| 37 | 84 | 44.16% | 13.69% | 41.71% | 19.11% | 2.45% | -5.42% | TCI | Benchmark | 0.95 | 0.90 | Benchmark* | Benchmark* |
| 39 | 36 | 30.37% | 34.00% | 26.16% | 40.50% | 4.21% | -6.50% | TCI | Benchmark | 0.97 | 0.92 | Benchmark* | Benchmark* |
| 42 | 12 | 13.97% | 0.00% | 7.50% | 0.00% | 6.47% | 0.00% | TCI | - | 1.86 | 1.86 | TCI* | TCI* |
| 44 | 24 | 16.22% | 20.00% | 7.10% | 23.33% | 9.12% | -3.33% | TCI | Benchmark | 1.19 | 1.05 | TCI* | TCI* |
| 45 | 96 | 44.79% | 13.56% | 43.95% | 15.33% | 0.85% | -1.78% | TCI | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 47 | 12 | 25.10% | 50.00% | 33.27% | 50.00% | -8.18% | 0.00% | Benchmark | - | 0.90 | 0.94 | Benchmark* | Benchmark* |
| 48 | 24 | 31.88% | 20.00% | 32.05% | 20.00% | -0.17% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 53 | 12 | 20.82% | 35.00% | 14.68% | 40.00% | 6.14% | -5.00% | TCI | Benchmark | 1.02 | 0.96 | TCI* | TCI* |
| 55 | 24 | 39.54% | 30.00% | 27.61% | 26.67% | 11.93% | 3.33% | TCI | TCI | 1.28 | 1.23 | TCI | TCI |
| 57 | 24 | 35.46% | 30.00% | 35.97% | 30.00% | -0.51% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 58 | 180 | 40.45% | 13.95% | 40.52% | 13.28% | -0.06% | 0.67% | Benchmark | TCI | 1.01 | 1.02 | TCI* | TCI* |
| 59 | 36 | 39.66% | 30.00% | 39.77% | 30.00% | -0.11% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 70 | 24 | 45.54% | 0.00% | 45.54% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 73 | 60 | 33.00% | 16.90% | 32.08% | 18.10% | 0.92% | -1.19% | TCI | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 11: Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 7 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (7) | | | | | | | | | |
| | | Benchmark - Salest-12 & GDPt-12 | | Twitter - CI & GDPt-12 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 35.51% | 22.41% | 34.36% | 23.35% | 1.16% | -0.94% | TCI | Benchmark | 1.00 | 0.99 | TCI* | TCI* |
| 21 | 12 | 36.37% | 35.00% | 32.87% | 30.00% | 3.50% | 5.00% | TCI | TCI | 1.14 | 1.15 | TCI | TCI |
| 23 | 36 | 37.37% | 18.50% | 37.60% | 23.50% | -0.23% | -5.00% | Benchmark | Benchmark | 0.91 | 0.88 | Benchmark | Benchmark |
| 28 | 72 | 34.51% | 14.64% | 35.86% | 15.71% | -1.35% | -1.07% | Benchmark | Benchmark | 0.95 | 0.95 | Benchmark | Benchmark |
| 29 | 24 | 47.62% | 0.00% | 46.20% | 10.00% | 1.42% | -10.00% | TCI | Benchmark | 0.85 | 0.72 | Benchmark* | Benchmark* |
| 30 | 12 | 37.09% | 5.00% | 36.73% | 5.00% | 0.36% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 31 | 24 | 28.31% | 5.00% | 31.40% | 11.67% | -3.10% | -6.67% | Benchmark | Benchmark | 0.77 | 0.70 | Benchmark | Benchmark |
| 35 | 24 | 34.71% | 10.00% | 34.71% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 36 | 48 | 27.50% | 14.00% | 32.80% | 27.62% | -5.30% | -13.62% | Benchmark | Benchmark | 0.69 | 0.63 | Benchmark | Benchmark |
| 37 | 84 | 43.75% | 18.64% | 40.30% | 21.83% | 3.44% | -3.19% | TCI | Benchmark | 1.00 | 0.96 | TCI* | TCI* |
| 39 | 36 | 24.64% | 38.50% | 24.47% | 40.50% | 0.17% | -2.00% | TCI | Benchmark | 0.97 | 0.96 | Benchmark* | Benchmark* |
| 42 | 12 | 13.97% | 0.00% | 6.67% | 0.00% | 7.31% | 0.00% | TCI | - | 2.10 | 2.10 | TCI* | TCI* |
| 44 | 24 | 0.00% | 31.67% | 0.00% | 31.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 45 | 96 | 43.66% | 13.17% | 43.07% | 17.68% | 0.59% | -4.51% | TCI | Benchmark | 0.94 | 0.89 | Benchmark* | Benchmark* |
| 47 | 12 | 25.10% | 50.00% | 25.10% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 48 | 24 | 24.01% | 6.67% | 25.26% | 6.67% | -1.25% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 53 | 12 | 25.62% | 30.00% | 8.33% | 25.00% | 17.29% | 5.00% | TCI | TCI | 1.67 | 1.47 | TCI | TCI |
| 55 | 24 | 32.49% | 30.00% | 25.15% | 26.67% | 7.34% | 3.33% | TCI | TCI | 1.21 | 1.18 | TCI | TCI |
| 57 | 24 | 24.90% | 25.00% | 28.25% | 16.67% | -3.35% | 8.33% | Benchmark | TCI | 1.11 | 1.22 | TCI* | TCI* |
| 58 | 180 | 39.60% | 21.10% | 39.02% | 17.08% | 0.58% | 4.03% | TCI | TCI | 1.08 | 1.12 | TCI | TCI |
| 59 | 36 | 36.77% | 27.50% | 33.14% | 27.50% | 3.63% | 0.00% | TCI | - | 1.06 | 1.04 | TCI* | TCI* |
| 70 | 24 | 46.54% | 3.33% | 43.29% | 0.00% | 3.25% | 3.33% | TCI | TCI | 1.15 | 1.23 | TCI | TCI |
| 73 | 60 | 28.10% | 19.05% | 22.18% | 19.05% | 5.92% | 0.00% | TCI | - | 1.14 | 1.10 | TCI* | TCI* |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 12: Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 9 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (9) | | | | | | | | | |
| | | Benchmark - Salest-12 & AR | | Twitter - CI & AR | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 36.54% | 25.61% | 33.03% | 23.66% | 3.51% | 1.95% | TCI | TCI | 1.10 | 1.09 | TCI | TCI |
| 21 | 12 | 35.98% | 30.00% | 37.45% | 25.00% | -1.48% | 5.00% | Benchmark | TCI | 1.06 | 1.10 | TCI* | TCI* |
| 23 | 36 | 38.42% | 31.00% | 33.97% | 30.00% | 4.45% | 1.00% | TCI | TCI | 1.09 | 1.07 | TCI | TCI |
| 28 | 72 | 31.58% | 12.14% | 33.12% | 12.14% | -1.54% | 0.00% | Benchmark | - | 0.97 | 0.97 | Benchmark* | Benchmark* |
| 29 | 24 | 44.01% | 3.33% | 48.95% | 13.33% | -4.94% | -10.00% | Benchmark | Benchmark | 0.76 | 0.67 | Benchmark | Benchmark |
| 30 | 12 | 39.10% | 15.00% | 38.13% | 15.00% | 0.96% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 31 | 24 | 39.20% | 16.67% | 39.88% | 16.67% | -0.68% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 35 | 24 | 36.20% | 3.33% | 30.76% | 13.33% | 5.43% | -10.00% | TCI | Benchmark | 0.90 | 0.75 | Benchmark* | Benchmark* |
| 36 | 48 | 24.90% | 25.33% | 33.22% | 36.38% | -8.32% | -11.05% | Benchmark | Benchmark | 0.72 | 0.71 | Benchmark | Benchmark |
| 37 | 84 | 42.41% | 15.17% | 40.68% | 22.08% | 1.73% | -6.91% | TCI | Benchmark | 0.92 | 0.86 | Benchmark* | Benchmark* |
| 39 | 36 | 30.66% | 32.50% | 24.47% | 40.50% | 6.20% | -8.00% | TCI | Benchmark | 0.97 | 0.91 | Benchmark* | Benchmark* |
| 42 | 12 | 12.56% | 15.00% | 13.27% | 15.00% | -0.71% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 44 | 24 | 3.91% | 26.67% | 3.91% | 26.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 45 | 96 | 43.91% | 20.46% | 41.56% | 19.29% | 2.34% | 1.17% | TCI | TCI | 1.06 | 1.06 | TCI | TCI |
| 47 | 12 | 25.10% | 50.00% | 32.87% | 50.00% | -7.77% | 0.00% | Benchmark | - | 0.91 | 0.94 | Benchmark* | Benchmark* |
| 48 | 24 | 27.00% | 6.67% | 27.49% | 6.67% | -0.49% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 53 | 12 | 20.82% | 35.00% | 14.68% | 40.00% | 6.14% | -5.00% | TCI | Benchmark | 1.02 | 0.96 | TCI* | TCI* |
| 55 | 24 | 42.50% | 10.00% | 39.71% | 18.33% | 2.79% | -8.33% | TCI | Benchmark | 0.90 | 0.82 | Benchmark* | Benchmark* |
| 57 | 24 | 38.85% | 33.33% | 38.85% | 33.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 58 | 180 | 39.07% | 16.66% | 39.91% | 17.05% | -0.84% | -0.39% | Benchmark | Benchmark | 0.98 | 0.98 | Benchmark | Benchmark |
| 59 | 36 | 30.35% | 19.00% | 29.76% | 19.00% | 0.59% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 70 | 24 | 45.40% | 13.33% | 44.15% | 6.67% | 1.25% | 6.67% | TCI | TCI | 1.16 | 1.25 | TCI | TCI |
| 73 | 60 | 33.80% | 22.32% | 34.35% | 23.21% | -0.54% | -0.89% | Benchmark | Benchmark | 0.98 | 0.97 | Benchmark | Benchmark |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 13: Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 11 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (11) | | | | | | | | | |
| | | Benchmark - Salest-12 & AR & GDPt-12 | | Twitter - CI & AR & GDPt-12 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 33.26% | 25.66% | 31.77% | 27.00% | 1.49% | -1.33% | TCI | Benchmark | 1.00 | 0.99 | TCI* | TCI* |
| 21 | 12 | 38.89% | 35.00% | 38.89% | 30.00% | 0.00% | 5.00% | - | TCI | 1.07 | 1.10 | TCI* | TCI* |
| 23 | 36 | 33.85% | 32.00% | 32.86% | 32.00% | 1.00% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 28 | 72 | 31.02% | 14.64% | 33.19% | 12.62% | -2.17% | 2.02% | Benchmark | TCI | 1.00 | 1.03 | Benchmark* | Benchmark* |
| 29 | 24 | 44.57% | 6.67% | 48.95% | 13.33% | -4.38% | -6.67% | Benchmark | Benchmark | 0.82 | 0.77 | Benchmark | Benchmark |
| 30 | 12 | 40.18% | 25.00% | 39.10% | 25.00% | 1.08% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 31 | 24 | 32.70% | 15.00% | 33.47% | 20.00% | -0.77% | -5.00% | Benchmark | Benchmark | 0.89 | 0.85 | Benchmark | Benchmark |
| 35 | 24 | 29.90% | 11.67% | 27.36% | 11.67% | 2.54% | 0.00% | TCI | - | 1.07 | 1.05 | TCI* | TCI* |
| 36 | 48 | 24.63% | 26.29% | 33.32% | 30.29% | -8.70% | -4.00% | Benchmark | Benchmark | 0.80 | 0.82 | Benchmark | Benchmark |
| 37 | 84 | 42.53% | 22.08% | 40.72% | 22.28% | 1.81% | -0.19% | TCI | Benchmark | 1.03 | 1.02 | TCI* | TCI* |
| 39 | 36 | 26.15% | 40.50% | 26.16% | 40.50% | -0.01% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 42 | 12 | 12.56% | 15.00% | 13.27% | 15.00% | -0.71% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 44 | 24 | 0.00% | 31.67% | 0.00% | 31.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 45 | 96 | 43.30% | 18.12% | 41.09% | 16.79% | 2.21% | 1.33% | TCI | TCI | 1.06 | 1.07 | TCI | TCI |
| 47 | 12 | 19.62% | 35.00% | 19.62% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 48 | 24 | 28.63% | 6.67% | 28.94% | 6.67% | -0.31% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 53 | 12 | 25.01% | 30.00% | 9.04% | 25.00% | 15.98% | 5.00% | TCI | TCI | 1.62 | 1.44 | TCI | TCI |
| 55 | 24 | 35.24% | 0.00% | 35.79% | 0.00% | -0.56% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 57 | 24 | 24.90% | 13.33% | 25.64% | 13.33% | -0.74% | 0.00% | Benchmark | - | 0.98 | 0.99 | Benchmark* | Benchmark* |
| 58 | 180 | 40.00% | 16.70% | 38.77% | 17.74% | 1.23% | -1.04% | TCI | Benchmark | 1.00 | 0.99 | TCI* | TCI* |
| 59 | 36 | 31.56% | 19.00% | 29.15% | 20.50% | 2.41% | -1.50% | TCI | Benchmark | 1.02 | 0.99 | TCI* | TCI* |
| 70 | 24 | 45.40% | 13.33% | 44.15% | 6.67% | 1.25% | 6.67% | TCI | TCI | 1.16 | 1.25 | TCI | TCI |
| 73 | 60 | 28.94% | 25.10% | 24.23% | 21.43% | 4.70% | 3.67% | TCI | TCI | 1.18 | 1.18 | TCI | TCI |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

**Table 14: Error Detection Performance Summary of Traditional Substantive Analytical Models with TCI for 24 Industries**

| Cost Ratio | Twitter Consumer Interest | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) |
| Traditional - SAP | 11 | 12 | 8 | 12 | 11 | 12 | 8 | 12 |

The error detection performance of traditional models with TCS is presented on tables 15, 16, 17, and 18. Models 6 and 1, 8 and 2, 10 and 3, and 12 and 4 are compared. A similar pattern is observed as with the models with TCI in that models with TCS lead to lower false positive errors at the cost of higher false negative errors. As presented in tables 15 and 17 models 6 and 10 generate lower false positive errors for 13 of the 24 industries that are analyzed, while models 8 and 12, displayed on tables 16 and 18, generate lower false positive errors for only 12 or 11 industries. The error detection performance of models with TCS for false negatives is inferior to that of the benchmark models. As documented on table 16, the better performing model with TCS is able to produce at best, lower false negative errors for 11 of the 24 industries. In terms of producing relatively low false positive and false negative errors, model 8 in table 16 is the better performer as it can achieve lower error rates for both types of errors for 5 industries, whereas benchmark model 2 is the better performer for 4 industries.

When evaluating the more effective model for error detection performance based on the varying ratio of costs of false positive and false negative errors, it can be determined that model 6 with TCS outperforms the other models since it leads to better error detection performance for 13 industries, when the cost ratio is 1:1 or 1:2. Benchmark models 2, 3, and 4 perform just as well as models 8, 10, and 12 with TCS as they lead to better detection performance for 12 industries when the cost ratio is 1:1 or 1:2. In summary, the traditional model with TCS, model 6, outperforms traditional benchmark models. Table 19 summarizes the better performing models with TCS.

# Table 15: Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 6 and 1)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (1) Benchmark - Salest-12 False Positive | (1) Benchmark - Salest-12 False Negative | (6) Twitter - CS False Positive | (6) Twitter - CS False Negative | Benchmark - CS Difference - FP | Benchmark - CS Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCS Total Cost | (1:2) Benchmark Total Cost /TCS Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 37.11% | 21.88% | 36.29% | 23.81% | 0.82% | -1.93% | TCS | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 21 | 12 | 37.05% | 30.00% | 36.96% | 10.00% | 0.09% | 20.00% | TCS | TCS | 1.43 | 1.70 | TCS | TCS |
| 23 | 36 | 44.44% | 19.00% | 41.59% | 15.00% | 2.85% | 4.00% | TCS | TCS | 1.12 | 1.15 | TCS | TCS |
| 28 | 72 | 37.13% | 1.43% | 39.10% | 8.57% | -1.97% | -7.14% | Benchmark | Benchmark | 0.81 | 0.71 | Benchmark | Benchmark |
| 29 | 24 | 43.00% | 13.33% | 43.00% | 13.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 30 | 12 | 34.44% | 5.00% | 33.27% | 5.00% | 1.17% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 31 | 24 | 36.02% | 25.00% | 39.37% | 16.67% | -3.34% | 8.33% | Benchmark | TCS | 1.09 | 1.18 | TCS* | TCS* |
| 35 | 24 | 44.01% | 0.00% | 46.47% | 0.00% | -2.46% | 0.00% | Benchmark | - | 0.95 | 0.95 | Benchmark* | Benchmark* |
| 36 | 48 | 33.72% | 16.00% | 27.50% | 18.29% | 6.22% | -2.29% | TCS | Benchmark | 1.09 | 1.03 | TCS* | TCS* |
| 37 | 84 | 44.16% | 13.69% | 42.40% | 18.14% | 1.76% | -4.44% | TCS | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 39 | 36 | 30.37% | 34.00% | 22.89% | 40.00% | 7.48% | -6.00% | TCS | Benchmark | 1.02 | 0.96 | TCS* | TCS* |
| 42 | 12 | 13.97% | 0.00% | 13.97% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 16.22% | 20.00% | 11.23% | 20.00% | 4.99% | 0.00% | TCS | - | 1.16 | 1.10 | TCS* | TCS* |
| 45 | 96 | 44.79% | 13.56% | 43.48% | 14.95% | 1.31% | -1.39% | TCS | Benchmark | 1.00 | 0.98 | Benchmark* | Benchmark* |
| 47 | 12 | 25.10% | 50.00% | 25.10% | 35.00% | 0.00% | 15.00% | - | TCS | 1.25 | 1.32 | TCS* | TCS* |
| 48 | 24 | 31.88% | 20.00% | 35.47% | 10.00% | -3.59% | 10.00% | Benchmark | TCS | 1.14 | 1.30 | TCS* | TCS* |
| 53 | 12 | 20.82% | 35.00% | 20.82% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 55 | 24 | 39.54% | 30.00% | 38.69% | 30.00% | 0.85% | 0.00% | TCS | - | 1.01 | 1.01 | TCS* | TCS* |
| 57 | 24 | 35.46% | 30.00% | 35.27% | 30.00% | 0.18% | 0.00% | TCS | - | 1.00 | 1.00 | TCS* | TCS* |
| 58 | 180 | 40.45% | 13.95% | 40.99% | 15.22% | -0.53% | -1.27% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 59 | 36 | 39.66% | 30.00% | 43.48% | 28.50% | -3.82% | 1.50% | Benchmark | TCS | 0.97 | 0.99 | Benchmark* | Benchmark* |
| 70 | 24 | 45.54% | 0.00% | 44.15% | 0.00% | 1.38% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 73 | 60 | 33.00% | 16.90% | 28.09% | 20.95% | 4.91% | -4.05% | TCS | Benchmark | 1.02 | 0.95 | TCS* | TCS* |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 16: Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 8 and 2)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (2) Benchmark - Salest-12 & GDPt-12 False Positive | (2) Benchmark - Salest-12 & GDPt-12 False Negative | (8) Twitter - CS & GDPt-12 False Positive | (8) Twitter - CS & GDPt-12 False Negative | Benchmark - CS Difference - FP | Benchmark - CS Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCS Total Cost | (1:2) Benchmark Total Cost /TCS Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 35.51% | 22.41% | 35.48% | 22.50% | 0.03% | -0.09% | TCS | Benchmark | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 21 | 12 | 36.37% | 35.00% | 38.49% | 35.00% | -2.12% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 23 | 36 | 37.37% | 18.50% | 34.26% | 22.50% | 3.11% | -4.00% | TCS | Benchmark | 0.98 | 0.94 | Benchmark* | Benchmark* |
| 28 | 72 | 34.51% | 14.64% | 34.64% | 14.64% | -0.13% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 29 | 24 | 47.62% | 0.00% | 48.84% | 0.00% | -1.22% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 30 | 12 | 37.09% | 5.00% | 36.73% | 15.00% | 0.36% | -10.00% | TCS | Benchmark | 0.81 | 0.71 | Benchmark* | Benchmark* |
| 31 | 24 | 28.31% | 5.00% | 32.70% | 11.67% | -4.39% | -6.67% | Benchmark | Benchmark | 0.75 | 0.68 | Benchmark | Benchmark |
| 35 | 24 | 34.71% | 10.00% | 34.89% | 11.67% | -0.18% | -1.67% | Benchmark | Benchmark | 0.96 | 0.94 | Benchmark | Benchmark |
| 36 | 48 | 27.50% | 14.00% | 29.89% | 22.00% | -2.39% | -8.00% | Benchmark | Benchmark | 0.80 | 0.75 | Benchmark | Benchmark |
| 37 | 84 | 43.75% | 18.64% | 42.96% | 21.64% | 0.79% | -3.00% | TCS | Benchmark | 0.97 | 0.94 | Benchmark* | Benchmark* |
| 39 | 36 | 24.64% | 38.50% | 19.69% | 36.50% | 4.95% | 2.00% | TCS | TCS | 1.12 | 1.10 | TCS | TCS |
| 42 | 12 | 13.97% | 0.00% | 6.67% | 0.00% | 7.31% | 0.00% | TCS | - | 2.10 | 2.10 | TCS* | TCS* |
| 44 | 24 | 0.00% | 31.67% | 0.00% | 30.00% | 0.00% | 1.67% | - | TCS | 1.06 | 1.06 | TCS* | TCS* |
| 45 | 96 | 43.66% | 13.17% | 42.93% | 13.90% | 0.73% | -0.73% | TCS | Benchmark | 1.00 | 0.99 | TCS* | TCS* |
| 47 | 12 | 25.10% | 50.00% | 29.88% | 25.00% | -4.78% | 25.00% | Benchmark | TCS | 1.37 | 1.57 | TCS* | TCS* |
| 48 | 24 | 24.01% | 6.67% | 23.53% | 6.67% | 0.49% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 53 | 12 | 25.62% | 30.00% | 29.88% | 30.00% | -4.26% | 0.00% | Benchmark | - | 0.93 | 0.95 | Benchmark* | Benchmark* |
| 55 | 24 | 32.49% | 30.00% | 33.61% | 33.33% | -1.13% | -3.33% | Benchmark | Benchmark | 0.93 | 0.92 | Benchmark | Benchmark |
| 57 | 24 | 24.90% | 25.00% | 24.63% | 23.33% | 0.27% | 1.67% | TCS | TCS | 1.04 | 1.05 | TCS | TCS |
| 58 | 180 | 39.60% | 21.10% | 39.09% | 16.95% | 0.51% | 4.15% | TCS | TCS | 1.08 | 1.12 | TCS | TCS |
| 59 | 36 | 36.77% | 27.50% | 34.77% | 20.00% | 2.00% | 7.50% | TCS | TCS | 1.17 | 1.23 | TCS | TCS |
| 70 | 24 | 46.54% | 3.33% | 46.59% | 0.00% | -0.05% | 3.33% | Benchmark | TCS | 1.07 | 1.14 | TCS* | TCS* |
| 73 | 60 | 28.10% | 19.05% | 27.43% | 17.38% | 0.68% | 1.67% | TCS | TCS | 1.05 | 1.06 | TCS | TCS |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 35.00% | 0.00% | 15.00% | - | TCS | 1.43 | 1.43 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 17: Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 10 and 3)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (3) Benchmark - Salest-12 & AR: False Positive | (3) Benchmark - Salest-12 & AR: False Negative | (10) Twitter - CS & AR: False Positive | (10) Twitter - CS & AR: False Negative | Benchmark - CS: Difference - FP | Benchmark - CS: Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCS Total Cost | (1:2) Benchmark Total Cost /TCS Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 36.54% | 25.61% | 34.55% | 30.20% | 1.98% | -4.59% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 21 | 12 | 35.98% | 30.00% | 38.49% | 5.00% | -2.52% | 25.00% | Benchmark | TCS | 1.52 | 1.98 | TCS* | TCS* |
| 23 | 36 | 38.42% | 31.00% | 33.46% | 31.00% | 4.96% | 0.00% | TCS | - | 1.08 | 1.05 | TCS* | TCS* |
| 28 | 72 | 31.58% | 12.14% | 34.77% | 10.71% | -3.18% | 1.43% | Benchmark | TCS | 0.96 | 0.99 | Benchmark* | Benchmark* |
| 29 | 24 | 44.01% | 3.33% | 45.26% | 3.33% | -1.25% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 30 | 12 | 39.10% | 15.00% | 37.09% | 25.00% | 2.01% | -10.00% | TCS | Benchmark | 0.87 | 0.79 | Benchmark* | Benchmark* |
| 31 | 24 | 39.20% | 16.67% | 38.54% | 6.67% | 0.66% | 10.00% | TCS | TCS | 1.24 | 1.40 | TCS | TCS |
| 35 | 24 | 36.20% | 3.33% | 35.25% | 0.00% | 0.95% | 3.33% | TCS | TCS | 1.12 | 1.22 | TCS | TCS |
| 36 | 48 | 24.90% | 25.33% | 27.14% | 31.00% | -2.24% | -5.67% | Benchmark | Benchmark | 0.86 | 0.85 | Benchmark | Benchmark |
| 37 | 84 | 42.41% | 15.17% | 41.31% | 19.89% | 1.10% | -4.72% | TCS | Benchmark | 0.94 | 0.90 | Benchmark* | Benchmark* |
| 39 | 36 | 30.66% | 32.50% | 22.51% | 40.00% | 8.16% | -7.50% | TCS | Benchmark | 1.01 | 0.93 | TCS* | TCS* |
| 42 | 12 | 12.56% | 15.00% | 12.56% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 3.91% | 26.67% | 3.91% | 23.33% | 0.00% | 3.33% | - | TCS | 1.12 | 1.13 | TCS* | TCS* |
| 45 | 96 | 43.91% | 20.46% | 43.30% | 19.58% | 0.61% | 0.89% | TCS | TCS | 1.02 | 1.03 | TCS | TCS |
| 47 | 12 | 25.10% | 50.00% | 29.88% | 30.00% | -4.78% | 20.00% | Benchmark | TCS | 1.25 | 1.39 | TCS* | TCS* |
| 48 | 24 | 27.00% | 6.67% | 25.05% | 6.67% | 1.95% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 53 | 12 | 20.82% | 35.00% | 25.01% | 35.00% | -4.19% | 0.00% | Benchmark | - | 0.93 | 0.96 | Benchmark* | Benchmark* |
| 55 | 24 | 42.50% | 10.00% | 41.45% | 16.67% | 1.06% | -6.67% | TCS | Benchmark | 0.90 | 0.84 | Benchmark* | Benchmark* |
| 57 | 24 | 38.85% | 33.33% | 36.95% | 30.00% | 1.90% | 3.33% | TCS | TCS | 1.08 | 1.09 | TCS | TCS |
| 58 | 180 | 39.07% | 16.66% | 38.29% | 20.00% | 0.77% | -3.34% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 59 | 36 | 30.35% | 19.00% | 34.37% | 15.00% | -4.02% | 4.00% | Benchmark | TCS | 1.00 | 1.06 | Benchmark* | Benchmark* |
| 70 | 24 | 45.40% | 13.33% | 46.60% | 6.67% | -1.19% | 6.67% | Benchmark | TCS | 1.10 | 1.20 | TCS* | TCS* |
| 73 | 60 | 33.80% | 22.32% | 28.28% | 24.70% | 5.53% | -2.38% | TCS | Benchmark | 1.06 | 1.01 | TCS* | TCS* |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 18: Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 12 and 4)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (4) Benchmark - Salest-12 & AR & GDPt-12: False Positive | (4) Benchmark - Salest-12 & AR & GDPt-12: False Negative | (12) Twitter - CS & AR & GDPt-12: False Positive | (12) Twitter - CS & AR & GDPt-12: False Negative | Benchmark - CS: Difference - FP | Benchmark - CS: Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCS Total Cost | (1:2) Benchmark Total Cost /TCS Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 33.26% | 25.66% | 33.16% | 26.49% | 0.10% | -0.83% | TCS | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 21 | 12 | 38.89% | 35.00% | 39.53% | 35.00% | -0.64% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 23 | 36 | 33.85% | 32.00% | 32.78% | 25.50% | 1.08% | 6.50% | TCS | TCS | 1.13 | 1.17 | TCS | TCS |
| 28 | 72 | 31.02% | 14.64% | 33.91% | 14.29% | -2.89% | 0.36% | Benchmark | TCS | 0.95 | 0.97 | Benchmark* | Benchmark* |
| 29 | 24 | 44.57% | 6.67% | 44.01% | 6.67% | 0.56% | 0.00% | TCS | - | 1.01 | 1.01 | TCS* | TCS* |
| 30 | 12 | 40.18% | 25.00% | 38.13% | 25.00% | 2.04% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 31 | 24 | 32.70% | 15.00% | 35.45% | 15.00% | -2.75% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 35 | 24 | 29.90% | 11.67% | 29.19% | 11.67% | 0.71% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 36 | 48 | 24.63% | 26.29% | 28.09% | 34.10% | -3.46% | -7.81% | Benchmark | Benchmark | 0.82 | 0.80 | Benchmark | Benchmark |
| 37 | 84 | 42.53% | 22.08% | 43.13% | 21.30% | -0.59% | 0.78% | Benchmark | TCS | 1.00 | 1.01 | TCS* | TCS* |
| 39 | 36 | 26.15% | 40.50% | 17.69% | 37.50% | 8.46% | 3.00% | TCS | TCS | 1.21 | 1.16 | TCS | TCS |
| 42 | 12 | 12.56% | 15.00% | 12.56% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 0.00% | 31.67% | 0.00% | 30.00% | 0.00% | 1.67% | - | TCS | 1.06 | 1.06 | TCS* | TCS* |
| 45 | 96 | 43.30% | 18.12% | 43.44% | 19.01% | -0.15% | -0.89% | Benchmark | Benchmark | 0.98 | 0.98 | Benchmark | Benchmark |
| 47 | 12 | 19.62% | 35.00% | 29.88% | 30.00% | -10.26% | 5.00% | Benchmark | TCS | 0.91 | 1.00 | Benchmark* | Benchmark* |
| 48 | 24 | 28.63% | 6.67% | 26.62% | 6.67% | 2.02% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 53 | 12 | 25.01% | 30.00% | 25.62% | 30.00% | -0.60% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 55 | 24 | 35.24% | 0.00% | 32.49% | 3.33% | 2.75% | -3.33% | TCS | Benchmark | 0.98 | 0.90 | Benchmark* | Benchmark* |
| 57 | 24 | 24.90% | 13.33% | 24.37% | 23.33% | 0.52% | -10.00% | TCS | Benchmark | 0.80 | 0.73 | Benchmark* | Benchmark* |
| 58 | 180 | 40.00% | 16.70% | 38.82% | 15.63% | 1.18% | 1.06% | TCS | TCS | 1.04 | 1.05 | TCS | TCS |
| 59 | 36 | 31.56% | 19.00% | 34.63% | 17.50% | -3.07% | 1.50% | Benchmark | TCS | 0.97 | 1.00 | Benchmark* | Benchmark* |
| 70 | 24 | 45.40% | 13.33% | 47.74% | 6.67% | -2.34% | 6.67% | Benchmark | TCS | 1.08 | 1.18 | TCS* | TCS* |
| 73 | 60 | 28.94% | 25.10% | 28.66% | 21.43% | 0.27% | 3.67% | TCS | TCS | 1.08 | 1.11 | TCS | TCS |
| 75 | 12 | 0.00% | 50.00% | 0.00% | 35.00% | 0.00% | 15.00% | - | TCS | 1.43 | 1.43 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

**Table 19: Error Detection Performance Summary of Traditional Substantive Analytical Models with TCS for 24 Industries**

| **Cost Ratio** | **Twitter Consumer Sentiment** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 13 | 12 | 12 | 12 | 13 | 12 | 12 | 12 |

***Continuous Substantive Analytical Models***

Tables 20, 21, 22, and 23 present the results of error detection performance for continuous analytical models that have TCI and that do not have TCI information. Models 5 and 1, 7 and 2, 9 and 3, and 11 and 4 are compared in these tables. As observed in the findings for traditional SAPs, the models with TCI are more effective in reducing the false positive error rate than in reducing the false negative error rate. Model 7 especially, is more effective in generating lower false positive errors for 16 industries, followed by model 11 which is more effective for 15 industries, while models 5, and 9 produce lower false positive errors for 14, and 11 industries. The error detection performance of models with TCI in relation to false negative error rate is inferior to that of the benchmark models as the more effective model with TCI, model 7, can lead to lower false negative errors for only 10 industries. With respect to maintaining relatively low error rates for both types of errors, benchmark model 1, is more effective as it can achieve lower false positive and false negative errors for 4 industries.

Model 7 is the better model once the criteria of the ratio of cost of errors is applied to unambiguously determine the overall benefits of continuous models with TCI and without TCI. Model 11 is superior to the other models as it leads to better detection performance for 14 industries when the cost ratio is 1:1 but only for 12 industries when the cost ratio is 1:2. The performance of model 11 is followed by that of model 7, which leads to superior performance for 13 industries when the cost ratio is 1:1 or 1:2. Consequently, the continuous model 7, is more effective than continuous benchmark models with respect to achieving lower false positive and false negative error rates under varying cost ratios. Table 24 summarizes the better performing models with TCI.

# Table 20: Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 5 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (5) | | | | | | | | | |
| | | Benchmark - Salest-1 | | Twitter - CI | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 44.72% | 18.01% | 43.99% | 17.37% | 0.73% | 0.64% | TCI | TCI | 1.02 | 1.03 | TCI | TCI |
| 21 | 12 | 47.86% | 5.00% | 50.00% | 0.00% | -2.14% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 23 | 36 | 45.09% | 2.50% | 47.11% | 10.00% | -2.02% | -7.50% | Benchmark | Benchmark | 0.83 | 0.75 | Benchmark | Benchmark |
| 28 | 72 | 43.97% | 11.07% | 45.00% | 13.21% | -1.03% | -2.14% | Benchmark | Benchmark | 0.95 | 0.93 | Benchmark | Benchmark |
| 29 | 24 | 45.67% | 16.67% | 41.94% | 28.33% | 3.73% | -11.67% | TCI | Benchmark | 0.89 | 0.80 | Benchmark* | Benchmark* |
| 30 | 12 | 50.00% | 0.00% | 47.86% | 5.00% | 2.14% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 31 | 24 | 42.40% | 6.67% | 43.87% | 6.67% | -1.47% | 0.00% | Benchmark | - | 0.97 | 0.97 | Benchmark* | Benchmark* |
| 35 | 24 | 42.84% | 15.00% | 42.85% | 16.67% | -0.01% | -1.67% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 36 | 48 | 43.58% | 16.57% | 42.39% | 21.90% | 1.18% | -5.33% | TCI | Benchmark | 0.94 | 0.89 | Benchmark* | Benchmark* |
| 37 | 84 | 45.15% | 15.44% | 43.63% | 17.94% | 1.52% | -2.50% | TCI | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 39 | 36 | 44.35% | 7.50% | 43.10% | 23.00% | 1.25% | -15.50% | TCI | Benchmark | 0.78 | 0.67 | Benchmark* | Benchmark* |
| 42 | 12 | 42.68% | 10.00% | 32.87% | 20.00% | 9.82% | -10.00% | TCI | Benchmark | 1.00 | 0.86 | Benchmark* | Benchmark* |
| 44 | 24 | 40.03% | 20.00% | 40.51% | 28.33% | -0.49% | -8.33% | Benchmark | Benchmark | 0.87 | 0.82 | Benchmark | Benchmark |
| 45 | 96 | 45.84% | 14.67% | 44.55% | 16.12% | 1.30% | -1.45% | TCI | Benchmark | 1.00 | 0.98 | Benchmark* | Benchmark* |
| 47 | 12 | 45.79% | 15.00% | 47.62% | 0.00% | -1.83% | 15.00% | Benchmark | TCI | 1.28 | 1.59 | TCI* | TCI* |
| 48 | 24 | 41.17% | 13.33% | 39.88% | 20.00% | 1.29% | -6.67% | TCI | Benchmark | 0.91 | 0.85 | Benchmark* | Benchmark* |
| 53 | 12 | 34.08% | 35.00% | 39.53% | 10.00% | -5.45% | 25.00% | Benchmark | TCI | 1.39 | 1.75 | TCI* | TCI* |
| 55 | 24 | 45.40% | 8.33% | 45.67% | 16.67% | -0.27% | -8.33% | Benchmark | Benchmark | 0.86 | 0.79 | Benchmark | Benchmark |
| 57 | 24 | 46.60% | 6.67% | 45.40% | 8.33% | 1.20% | -1.67% | TCI | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 58 | 180 | 45.09% | 13.83% | 44.76% | 14.81% | 0.33% | -0.97% | TCI | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 59 | 36 | 42.80% | 16.50% | 39.00% | 25.50% | 3.80% | -9.00% | TCI | Benchmark | 0.92 | 0.84 | Benchmark* | Benchmark* |
| 70 | 24 | 45.54% | 13.33% | 42.99% | 18.33% | 2.55% | -5.00% | TCI | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 73 | 60 | 37.50% | 27.68% | 38.13% | 27.08% | -0.64% | 0.60% | Benchmark | TCI | 1.00 | 1.01 | Benchmark* | TCI* |
| 75 | 12 | 48.10% | 10.00% | 33.68% | 30.00% | 14.42% | -20.00% | TCI | Benchmark | 0.91 | 0.73 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 21: Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 7 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (7) | | | | | | | | | |
| | | Benchmark - Salest-1 & GDPt-1 | | Twitter - CI & GDPt-1 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 43.27% | 17.98% | 41.72% | 21.66% | 1.56% | -3.67% | TCI | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 21 | 12 | 39.53% | 10.00% | 45.00% | 0.00% | -5.47% | 10.00% | Benchmark | TCI | 1.10 | 1.32 | TCI* | TCI* |
| 23 | 36 | 46.16% | 6.50% | 44.53% | 11.50% | 1.63% | -5.00% | TCI | Benchmark | 0.94 | 0.88 | Benchmark* | Benchmark* |
| 28 | 72 | 43.68% | 15.83% | 40.86% | 20.24% | 2.83% | -4.40% | TCI | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 29 | 24 | 49.07% | 6.67% | 42.09% | 30.00% | 6.98% | -23.33% | TCI | Benchmark | 0.77 | 0.61 | Benchmark* | Benchmark* |
| 30 | 12 | 36.73% | 20.00% | 39.53% | 10.00% | -2.80% | 10.00% | Benchmark | TCI | 1.15 | 1.29 | TCI* | TCI* |
| 31 | 24 | 40.85% | 6.67% | 39.22% | 6.67% | 1.63% | 0.00% | TCI | - | 1.04 | 1.03 | TCI* | TCI* |
| 35 | 24 | 44.15% | 13.33% | 42.55% | 10.00% | 1.60% | 3.33% | TCI | TCI | 1.09 | 1.13 | TCI | TCI |
| 36 | 48 | 44.23% | 16.00% | 41.56% | 22.95% | 2.67% | -6.95% | TCI | Benchmark | 0.93 | 0.87 | Benchmark* | Benchmark* |
| 37 | 84 | 42.88% | 19.55% | 43.96% | 16.19% | -1.08% | 3.36% | Benchmark | TCI | 1.04 | 1.07 | TCI* | TCI* |
| 39 | 36 | 46.86% | 2.50% | 42.70% | 14.00% | 4.16% | -11.50% | TCI | Benchmark | 0.87 | 0.73 | Benchmark* | Benchmark* |
| 42 | 12 | 40.18% | 20.00% | 32.87% | 20.00% | 7.31% | 0.00% | TCI | - | 1.14 | 1.10 | TCI* | TCI* |
| 44 | 24 | 44.29% | 16.67% | 45.54% | 13.33% | -1.24% | 3.33% | Benchmark | TCI | 1.04 | 1.08 | TCI* | TCI* |
| 45 | 96 | 45.00% | 19.02% | 44.89% | 16.73% | 0.10% | 2.29% | TCI | TCI | 1.04 | 1.06 | TCI | TCI |
| 47 | 12 | 45.52% | 10.00% | 47.62% | 0.00% | -2.10% | 10.00% | Benchmark | TCI | 1.17 | 1.38 | TCI* | TCI* |
| 48 | 24 | 42.70% | 13.33% | 39.87% | 18.33% | 2.83% | -5.00% | TCI | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 53 | 12 | 47.62% | 0.00% | 40.18% | 20.00% | 7.44% | -20.00% | TCI | Benchmark | 0.79 | 0.59 | Benchmark* | Benchmark* |
| 55 | 24 | 48.11% | 13.33% | 46.72% | 10.00% | 1.38% | 3.33% | TCI | TCI | 1.08 | 1.12 | TCI | TCI |
| 57 | 24 | 43.87% | 6.67% | 46.47% | 3.33% | -2.60% | 3.33% | Benchmark | TCI | 1.01 | 1.08 | TCI* | TCI* |
| 58 | 180 | 45.44% | 18.19% | 45.15% | 19.76% | 0.29% | -1.57% | TCI | Benchmark | 0.98 | 0.97 | Benchmark* | Benchmark* |
| 59 | 36 | 42.40% | 7.50% | 44.35% | 7.50% | -1.94% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 70 | 24 | 44.44% | 20.00% | 43.00% | 20.00% | 1.44% | 0.00% | TCI | - | 1.02 | 1.02 | TCI* | TCI* |
| 73 | 60 | 33.89% | 29.76% | 36.70% | 26.79% | -2.82% | 2.98% | Benchmark | TCI | 1.00 | 1.03 | TCI* | TCI* |
| 75 | 12 | 46.05% | 20.00% | 34.08% | 35.00% | 11.97% | -15.00% | TCI | Benchmark | 0.96 | 0.83 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 22: Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 9 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (9) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR | | Twitter - CI & AR | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 44.21% | 17.00% | 43.52% | 17.66% | 0.69% | -0.66% | TCI | Benchmark | 1.00 | 0.99 | TCI* | Benchmark* |
| 21 | 12 | 47.86% | 5.00% | 50.00% | 0.00% | -2.14% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 23 | 36 | 44.53% | 12.50% | 40.75% | 17.50% | 3.78% | -5.00% | TCI | Benchmark | 0.98 | 0.92 | Benchmark* | Benchmark* |
| 28 | 72 | 37.85% | 25.86% | 37.73% | 23.36% | 0.12% | 2.50% | TCI | TCI | 1.04 | 1.06 | TCI | TCI |
| 29 | 24 | 42.55% | 10.00% | 45.40% | 10.00% | -2.85% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 30 | 12 | 42.97% | 15.00% | 42.97% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 31 | 24 | 39.71% | 15.00% | 39.71% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 35 | 24 | 38.36% | 21.67% | 39.88% | 20.00% | -1.52% | 1.67% | Benchmark | TCI | 1.00 | 1.02 | TCI* | TCI* |
| 36 | 48 | 43.58% | 18.00% | 44.30% | 18.00% | -0.71% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 37 | 84 | 43.59% | 16.69% | 43.50% | 14.19% | 0.08% | 2.50% | TCI | TCI | 1.04 | 1.07 | TCI | TCI |
| 39 | 36 | 42.90% | 19.00% | 38.77% | 20.50% | 4.14% | -1.50% | TCI | Benchmark | 1.04 | 1.01 | TCI* | TCI* |
| 42 | 12 | 32.87% | 20.00% | 36.37% | 15.00% | -3.50% | 5.00% | Benchmark | TCI | 1.03 | 1.10 | TCI* | TCI* |
| 44 | 24 | 42.69% | 11.67% | 43.14% | 21.67% | -0.45% | -10.00% | Benchmark | Benchmark | 0.84 | 0.76 | Benchmark | Benchmark |
| 45 | 96 | 45.00% | 19.13% | 44.44% | 13.62% | 0.56% | 5.52% | TCI | TCI | 1.10 | 1.16 | TCI | TCI |
| 47 | 12 | 42.97% | 15.00% | 47.62% | 0.00% | -4.65% | 15.00% | Benchmark | TCI | 1.22 | 1.53 | TCI* | TCI* |
| 48 | 24 | 39.71% | 16.67% | 39.88% | 20.00% | -0.17% | -3.33% | Benchmark | Benchmark | 0.94 | 0.91 | Benchmark | Benchmark |
| 53 | 12 | 47.86% | 5.00% | 40.18% | 20.00% | 7.68% | -15.00% | TCI | Benchmark | 0.88 | 0.72 | Benchmark* | Benchmark* |
| 55 | 24 | 46.72% | 10.00% | 47.86% | 6.67% | -1.14% | 3.33% | Benchmark | TCI | 1.04 | 1.09 | TCI* | TCI* |
| 57 | 24 | 46.60% | 6.67% | 45.40% | 8.33% | 1.20% | -1.67% | TCI | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 58 | 180 | 46.24% | 10.96% | 45.34% | 15.42% | 0.89% | -4.46% | TCI | Benchmark | 0.94 | 0.89 | Benchmark* | Benchmark* |
| 59 | 36 | 43.00% | 20.50% | 43.87% | 18.00% | -0.87% | 2.50% | Benchmark | TCI | 1.03 | 1.05 | TCI* | TCI* |
| 70 | 24 | 47.98% | 10.00% | 44.29% | 15.00% | 3.70% | -5.00% | TCI | Benchmark | 0.98 | 0.92 | Benchmark* | Benchmark* |
| 73 | 60 | 40.54% | 21.85% | 40.60% | 23.51% | -0.07% | -1.67% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 75 | 12 | 40.82% | 30.00% | 29.42% | 30.00% | 11.40% | 0.00% | TCI | - | 1.19 | 1.13 | TCI* | TCI* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 23: Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 11 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (11) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR & GDPt-1 | | Twitter - CI & AR & GDPt-1 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 43.80% | 18.80% | 42.68% | 20.90% | 1.12% | -2.10% | TCI | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 21 | 12 | 42.39% | 5.00% | 36.01% | 10.00% | 6.38% | -5.00% | TCI | Benchmark | 1.03 | 0.94 | TCI* | Benchmark* |
| 23 | 36 | 45.27% | 7.50% | 45.27% | 7.50% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 28 | 72 | 40.64% | 26.19% | 39.50% | 24.76% | 1.14% | 1.43% | TCI | TCI | 1.04 | 1.04 | TCI | TCI |
| 29 | 24 | 41.47% | 18.33% | 43.87% | 6.67% | -2.40% | 11.67% | Benchmark | TCI | 1.18 | 1.37 | TCI* | TCI* |
| 30 | 12 | 36.73% | 20.00% | 36.73% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 31 | 24 | 42.40% | 6.67% | 40.85% | 6.67% | 1.55% | 0.00% | TCI | - | 1.03 | 1.03 | TCI* | TCI* |
| 35 | 24 | 38.19% | 20.00% | 36.95% | 28.33% | 1.24% | -8.33% | TCI | Benchmark | 0.89 | 0.84 | Benchmark* | Benchmark* |
| 36 | 48 | 42.33% | 23.33% | 40.78% | 23.33% | 1.55% | 0.00% | TCI | - | 1.02 | 1.02 | TCI* | TCI* |
| 37 | 84 | 43.46% | 13.89% | 44.56% | 10.97% | -1.10% | 2.92% | Benchmark | TCI | 1.03 | 1.07 | TCI* | TCI* |
| 39 | 36 | 33.33% | 32.50% | 35.67% | 29.50% | -2.34% | 3.00% | Benchmark | TCI | 1.01 | 1.04 | TCI* | TCI* |
| 42 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 45.67% | 16.67% | 46.85% | 13.33% | -1.18% | 3.33% | Benchmark | TCI | 1.04 | 1.07 | TCI* | TCI* |
| 45 | 96 | 44.62% | 18.28% | 43.55% | 18.34% | 1.07% | -0.06% | TCI | Benchmark | 1.02 | 1.01 | TCI* | TCI* |
| 47 | 12 | 45.52% | 10.00% | 47.62% | 0.00% | -2.10% | 10.00% | Benchmark | TCI | 1.17 | 1.38 | TCI* | TCI* |
| 48 | 24 | 41.31% | 15.00% | 39.87% | 18.33% | 1.44% | -3.33% | TCI | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 53 | 12 | 47.62% | 0.00% | 39.85% | 15.00% | 7.77% | -15.00% | TCI | Benchmark | 0.87 | 0.68 | Benchmark* | Benchmark* |
| 55 | 24 | 42.99% | 18.33% | 41.47% | 18.33% | 1.52% | 0.00% | TCI | - | 1.03 | 1.02 | TCI* | TCI* |
| 57 | 24 | 46.34% | 0.00% | 45.13% | 3.33% | 1.21% | -3.33% | TCI | Benchmark | 0.96 | 0.89 | Benchmark* | Benchmark* |
| 58 | 180 | 45.16% | 15.92% | 45.02% | 17.00% | 0.15% | -1.09% | TCI | Benchmark | 0.98 | 0.97 | Benchmark* | Benchmark* |
| 59 | 36 | 41.06% | 22.00% | 40.95% | 19.50% | 0.11% | 2.50% | TCI | TCI | 1.04 | 1.06 | TCI | TCI |
| 70 | 24 | 46.98% | 16.67% | 44.44% | 20.00% | 2.54% | -3.33% | TCI | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 73 | 60 | 38.13% | 27.08% | 41.17% | 22.14% | -3.04% | 4.94% | Benchmark | TCI | 1.03 | 1.08 | TCI* | TCI* |
| 75 | 12 | 43.55% | 25.00% | 25.10% | 35.00% | 18.46% | -10.00% | TCI | Benchmark | 1.14 | 0.98 | TCI* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

**Table 24: Error Detection Performance Summary of Continuous Substantive Analytical Models with TCI for 24 Industries**

| **Cost Ratio** | **Twitter Consumer Interest** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) |
| Continuous - SAP | 4 | 13 | 12 | 14 | 5 | 13 | 11 | 12 |

Tables 25, 26, 27, and 28 document the results of error detection performance for continuous models that have TCS and that do not have TCS information. Models 6 and 1, 8 and 2, 10 and 3, and 12 and 4 are compared in each table. In relation to error performance ability for false positives, model 6 can be more effective as it produces lower false positive errors for 15 industries in comparison to models 8, 10, and 12, which produce lower false positive errors for 12 and 14 industries. The error detection performance for false negatives is once again inferior to the performance of the benchmark models as the more effective model with TCS, model 8, can achieve lower false negative errors for only 10 industries. The model that can lead to lower error rates for both types of errors is model 8 as it produces lower error false positive and false negative error rates for 5 industries.

The results related to the cost ratio analysis suggests that model 8 outperforms the other models as it can achieve better detection performance for 13 industries when the cost ratio is 1:1 or 1:2. Accordingly, these findings suggest that model 8 can achieve superior error detection performance under varying cost ratios.  Table 29 summarizes the better performing models with TCS.

# Table 25: Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 6 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (6) | | | | | | | | | |
| | | Benchmark - Salest-1 | | Twitter - CS | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 44.72% | 18.01% | 44.00% | 17.85% | 0.73% | 0.16% | TCS | TCS | 1.01 | 1.01 | TCS | TCS |
| 21 | 12 | 47.86% | 5.00% | 41.14% | 35.00% | 6.72% | -30.00% | TCS | Benchmark | 0.69 | 0.52 | Benchmark* | Benchmark* |
| 23 | 36 | 45.09% | 2.50% | 45.99% | 2.50% | -0.90% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 28 | 72 | 43.97% | 11.07% | 43.63% | 13.69% | 0.34% | -2.62% | TCS | Benchmark | 0.96 | 0.93 | Benchmark* | Benchmark* |
| 29 | 24 | 45.67% | 16.67% | 45.67% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 30 | 12 | 50.00% | 0.00% | 48.10% | 10.00% | 1.90% | -10.00% | TCS | Benchmark | 0.86 | 0.73 | Benchmark* | Benchmark* |
| 31 | 24 | 42.40% | 6.67% | 39.22% | 6.67% | 3.18% | 0.00% | TCS | - | 1.07 | 1.06 | TCS* | TCS* |
| 35 | 24 | 42.84% | 15.00% | 42.84% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 36 | 48 | 43.58% | 16.57% | 44.22% | 14.95% | -0.64% | 1.62% | Benchmark | TCS | 1.02 | 1.03 | TCS* | TCS* |
| 37 | 84 | 45.15% | 15.44% | 45.04% | 12.44% | 0.12% | 3.00% | TCS | TCS | 1.05 | 1.09 | TCS | TCS |
| 39 | 36 | 44.35% | 7.50% | 42.60% | 11.50% | 1.75% | -4.00% | TCS | Benchmark | 0.96 | 0.90 | Benchmark* | Benchmark* |
| 42 | 12 | 42.68% | 10.00% | 32.87% | 20.00% | 9.82% | -10.00% | TCS | Benchmark | 1.00 | 0.86 | Benchmark* | Benchmark* |
| 44 | 24 | 40.03% | 20.00% | 41.31% | 15.00% | -1.29% | 5.00% | Benchmark | TCS | 1.07 | 1.12 | TCS* | TCS* |
| 45 | 96 | 45.84% | 14.67% | 43.91% | 17.46% | 1.94% | -2.80% | TCS | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 47 | 12 | 45.79% | 15.00% | 48.33% | 15.00% | -2.55% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 48 | 24 | 41.17% | 13.33% | 39.54% | 11.67% | 1.63% | 1.67% | TCS | TCS | 1.06 | 1.08 | TCS | TCS |
| 53 | 12 | 34.08% | 35.00% | 42.97% | 15.00% | -8.89% | 20.00% | Benchmark | TCS | 1.19 | 1.43 | TCS* | TCS* |
| 55 | 24 | 45.40% | 8.33% | 44.01% | 8.33% | 1.39% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 57 | 24 | 46.60% | 6.67% | 41.32% | 16.67% | 5.27% | -10.00% | TCS | Benchmark | 0.92 | 0.80 | Benchmark* | Benchmark* |
| 58 | 180 | 45.09% | 13.83% | 44.41% | 15.14% | 0.68% | -1.30% | TCS | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 59 | 36 | 42.80% | 16.50% | 42.80% | 15.50% | 0.00% | 1.00% | TCS | TCS | 1.02 | 1.03 | TCS | TCS |
| 70 | 24 | 45.54% | 13.33% | 44.29% | 16.67% | 1.24% | -3.33% | TCS | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 73 | 60 | 37.50% | 27.68% | 39.29% | 23.04% | -1.79% | 4.64% | Benchmark | TCS | 1.05 | 1.09 | TCS* | TCS* |
| 75 | 12 | 48.10% | 10.00% | 48.10% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 26: Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 8 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (8) | | | | | | | | | |
| | | Benchmark - Salest-1 & GDPt-1 | | Twitter - CS & GDPt-1 | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 43.27% | 17.98% | 43.88% | 15.18% | -0.60% | 2.80% | Benchmark | TCS | 1.04 | 1.07 | TCS* | TCS* |
| 21 | 12 | 39.53% | 10.00% | 39.21% | 5.00% | 0.32% | 5.00% | TCS | TCS | 1.12 | 1.21 | TCS | TCS |
| 23 | 36 | 46.16% | 6.50% | 44.25% | 5.00% | 1.91% | 1.50% | TCS | TCS | 1.07 | 1.09 | TCS | TCS |
| 28 | 72 | 43.68% | 15.83% | 41.90% | 20.07% | 1.79% | -4.24% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 29 | 24 | 49.07% | 6.67% | 49.07% | 6.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 30 | 12 | 36.73% | 20.00% | 40.50% | 25.00% | -3.77% | -5.00% | Benchmark | Benchmark | 0.87 | 0.85 | Benchmark | Benchmark |
| 31 | 24 | 40.85% | 6.67% | 40.85% | 6.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 35 | 24 | 44.15% | 13.33% | 41.17% | 13.33% | 2.99% | 0.00% | TCS | - | 1.05 | 1.04 | TCS* | TCS* |
| 36 | 48 | 44.23% | 16.00% | 44.86% | 14.67% | -0.63% | 1.33% | Benchmark | TCS | 1.01 | 1.03 | TCS* | TCS* |
| 37 | 84 | 42.88% | 19.55% | 44.40% | 16.97% | -1.53% | 2.58% | Benchmark | TCS | 1.02 | 1.05 | TCS* | TCS* |
| 39 | 36 | 46.86% | 2.50% | 47.78% | 5.00% | -0.92% | -2.50% | Benchmark | Benchmark | 0.94 | 0.90 | Benchmark | Benchmark |
| 42 | 12 | 40.18% | 20.00% | 37.45% | 30.00% | 2.72% | -10.00% | TCS | Benchmark | 0.89 | 0.82 | Benchmark* | Benchmark* |
| 44 | 24 | 44.29% | 16.67% | 44.44% | 20.00% | -0.14% | -3.33% | Benchmark | Benchmark | 0.95 | 0.92 | Benchmark | Benchmark |
| 45 | 96 | 45.00% | 19.02% | 44.30% | 18.86% | 0.70% | 0.16% | TCS | TCS | 1.01 | 1.01 | TCS | TCS |
| 47 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 48 | 24 | 42.70% | 13.33% | 41.01% | 10.00% | 1.69% | 3.33% | TCS | TCS | 1.10 | 1.14 | TCS | TCS |
| 53 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 55 | 24 | 48.11% | 13.33% | 46.85% | 13.33% | 1.25% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 57 | 24 | 43.87% | 6.67% | 42.40% | 6.67% | 1.47% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 58 | 180 | 45.44% | 18.19% | 44.24% | 15.14% | 1.20% | 3.05% | TCS | TCS | 1.07 | 1.10 | TCS | TCS |
| 59 | 36 | 42.40% | 7.50% | 45.36% | 10.00% | -2.96% | -2.50% | Benchmark | Benchmark | 0.90 | 0.88 | Benchmark | Benchmark |
| 70 | 24 | 44.44% | 20.00% | 43.14% | 21.67% | 1.30% | -1.67% | TCS | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 73 | 60 | 33.89% | 29.76% | 37.35% | 24.82% | -3.46% | 4.94% | Benchmark | TCS | 1.02 | 1.07 | TCS* | TCS* |
| 75 | 12 | 46.05% | 20.00% | 50.00% | 0.00% | -3.95% | 20.00% | Benchmark | TCS | 1.32 | 1.72 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 27: Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 10 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (10) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR | | Twitter - CS & AR | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 44.21% | 17.00% | 44.77% | 13.40% | -0.56% | 3.60% | Benchmark | TCS | 1.05 | 1.09 | TCS* | TCS* |
| 21 | 12 | 47.86% | 5.00% | 46.31% | 25.00% | 1.55% | -20.00% | TCS | Benchmark | 0.74 | 0.60 | Benchmark* | Benchmark* |
| 23 | 36 | 44.53% | 12.50% | 43.59% | 12.50% | 0.95% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 28 | 72 | 37.85% | 25.86% | 38.37% | 24.79% | -0.52% | 1.07% | Benchmark | TCS | 1.01 | 1.02 | TCS* | TCS* |
| 29 | 24 | 42.55% | 10.00% | 41.01% | 10.00% | 1.54% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 30 | 12 | 42.97% | 15.00% | 46.05% | 20.00% | -3.07% | -5.00% | Benchmark | Benchmark | 0.88 | 0.85 | Benchmark | Benchmark |
| 31 | 24 | 39.71% | 15.00% | 42.70% | 13.33% | -2.99% | 1.67% | Benchmark | TCS | 0.98 | 1.00 | Benchmark* | TCS* |
| 35 | 24 | 38.36% | 21.67% | 38.36% | 21.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 36 | 48 | 43.58% | 18.00% | 44.44% | 22.00% | -0.86% | -4.00% | Benchmark | Benchmark | 0.93 | 0.90 | Benchmark | Benchmark |
| 37 | 84 | 43.59% | 16.69% | 42.79% | 17.53% | 0.80% | -0.83% | TCS | Benchmark | 1.00 | 0.99 | Benchmark* | Benchmark* |
| 39 | 36 | 42.90% | 19.00% | 41.89% | 19.00% | 1.01% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 42 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 42.69% | 11.67% | 42.55% | 10.00% | 0.14% | 1.67% | TCS | TCS | 1.03 | 1.06 | TCS | TCS |
| 45 | 96 | 45.00% | 19.13% | 43.07% | 22.76% | 1.93% | -3.63% | TCS | Benchmark | 0.97 | 0.94 | Benchmark* | Benchmark* |
| 47 | 12 | 42.97% | 15.00% | 45.52% | 10.00% | -2.55% | 5.00% | Benchmark | TCS | 1.04 | 1.11 | TCS* | TCS* |
| 48 | 24 | 39.71% | 16.67% | 37.83% | 11.67% | 1.88% | 5.00% | TCS | TCS | 1.14 | 1.19 | TCS | TCS |
| 53 | 12 | 47.86% | 5.00% | 40.18% | 20.00% | 7.68% | -15.00% | TCS | Benchmark | 0.88 | 0.72 | Benchmark* | Benchmark* |
| 55 | 24 | 46.72% | 10.00% | 45.54% | 13.33% | 1.19% | -3.33% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 57 | 24 | 46.60% | 6.67% | 42.40% | 6.67% | 4.19% | 0.00% | TCS | - | 1.09 | 1.08 | TCS* | TCS* |
| 58 | 180 | 46.24% | 10.96% | 44.41% | 15.41% | 1.83% | -4.45% | TCS | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 59 | 36 | 43.00% | 20.50% | 42.30% | 27.00% | 0.70% | -6.50% | TCS | Benchmark | 0.92 | 0.87 | Benchmark* | Benchmark* |
| 70 | 24 | 47.98% | 10.00% | 45.54% | 13.33% | 2.45% | -3.33% | TCS | Benchmark | 0.98 | 0.94 | Benchmark* | Benchmark* |
| 73 | 60 | 40.54% | 21.85% | 42.29% | 20.24% | -1.75% | 1.61% | Benchmark | TCS | 1.00 | 1.02 | Benchmark* | TCS* |
| 75 | 12 | 40.82% | 30.00% | 45.79% | 15.00% | -4.97% | 15.00% | Benchmark | TCS | 1.17 | 1.33 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 28: Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 12 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (12) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR & GDPt-1 | | Twitter - CS & AR & GDPt-1 | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 43.80% | 18.80% | 44.42% | 16.58% | -0.62% | 2.21% | Benchmark | TCS | 1.03 | 1.05 | TCS* | TCS* |
| 21 | 12 | 42.39% | 5.00% | 39.21% | 5.00% | 3.18% | 0.00% | TCS | - | 1.07 | 1.06 | TCS* | TCS* |
| 23 | 36 | 45.27% | 7.50% | 42.80% | 16.50% | 2.47% | -9.00% | TCS | Benchmark | 0.89 | 0.80 | Benchmark* | Benchmark* |
| 28 | 72 | 40.64% | 26.19% | 40.43% | 22.62% | 0.21% | 3.57% | TCS | TCS | 1.06 | 1.09 | TCS | TCS |
| 29 | 24 | 41.47% | 18.33% | 41.47% | 18.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 30 | 12 | 36.73% | 20.00% | 40.50% | 25.00% | -3.77% | -5.00% | Benchmark | Benchmark | 0.87 | 0.85 | Benchmark | Benchmark |
| 31 | 24 | 42.40% | 6.67% | 41.00% | 8.33% | 1.40% | -1.67% | TCS | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 35 | 24 | 38.19% | 20.00% | 38.19% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 36 | 48 | 42.33% | 23.33% | 40.13% | 25.62% | 2.20% | -2.29% | TCS | Benchmark | 1.00 | 0.97 | Benchmark* | Benchmark* |
| 37 | 84 | 43.46% | 13.89% | 43.04% | 13.42% | 0.42% | 0.47% | TCS | TCS | 1.02 | 1.02 | TCS | TCS |
| 39 | 36 | 33.33% | 32.50% | 31.95% | 32.50% | 1.38% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 42 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 44 | 24 | 45.67% | 16.67% | 45.67% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 45 | 96 | 44.62% | 18.28% | 42.93% | 21.01% | 1.69% | -2.73% | TCS | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 47 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 48 | 24 | 41.31% | 15.00% | 41.01% | 10.00% | 0.30% | 5.00% | TCS | TCS | 1.10 | 1.17 | TCS | TCS |
| 53 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 55 | 24 | 42.99% | 18.33% | 42.99% | 18.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 57 | 24 | 46.34% | 0.00% | 42.40% | 6.67% | 3.94% | -6.67% | TCS | Benchmark | 0.94 | 0.83 | Benchmark* | Benchmark* |
| 58 | 180 | 45.16% | 15.92% | 43.88% | 16.22% | 1.29% | -0.31% | TCS | Benchmark | 1.02 | 1.01 | TCS* | TCS* |
| 59 | 36 | 41.06% | 22.00% | 39.11% | 27.00% | 1.95% | -5.00% | TCS | Benchmark | 0.95 | 0.91 | Benchmark* | Benchmark* |
| 70 | 24 | 46.98% | 16.67% | 44.44% | 20.00% | 2.54% | -3.33% | TCS | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 73 | 60 | 38.13% | 27.08% | 38.75% | 26.19% | -0.62% | 0.89% | Benchmark | TCS | 1.00 | 1.01 | TCS* | TCS* |
| 75 | 12 | 43.55% | 25.00% | 45.52% | 10.00% | -1.97% | 15.00% | Benchmark | TCS | 1.23 | 1.43 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

**Table 29: Error Detection Performance Summary of Continuous Substantive Analytical Models with TCS for 24 Industries**

| **Cost Ratio** | **Twitter Consumer Sentiment** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Continuous - SAP | 10 | 13 | 10 | 9 | 10 | 13 | 12 | 9 |

Taken together, the results of the error detection performance indicate that continuous SAPs, models 7 and 8 with TCI or TCS, are more effective. These findings indicate that auditors can benefit from incorporating TCI or TCS in SAPs that contain timelier information (one-month lag instead of twelve-months lag). Benchmark models would be more appropriate if auditors utilize prior year information in SAPs (i.e. traditional SAPs with no Twitter information). However, developing SAPs with prior year information would generally reduce the prediction ability of the benchmark models. Consequently, auditors could benefit from implementing the proposed continuous analytical models to fully exploit the benefits of SAPs, especially model 7, which contains timelier sales and GDP information and TCI, to experience improved prediction and error detection performance under varying cost ratios as this is the model that is superior for most of the industries in the sample.

To summarize, the findings of error detection performance parallel those of prediction performance as model 7 with TCI information can be more effective in generating superior predictions and at the same time, achieving superior error detection performance under varying cost ratios (model 11 can be effective for prediction and error detection performance, but only when false positives are equally expensive as false negatives, 1:1 cost ratio) for most industries. These findings indicate that TCI information has incremental value even when more traditional external information, such as lagged GDP, is added to the SAP.

## 2.6. Additional Analysis

***4-Digit SIC Analysis***

To provide a general view of prediction and error detection performance, this study presents the results for 88 firms that are aggregated by their 2-digit SIC industry codes. Thus, results for 24 industries are documented. Aggregating firms into their 4-digit SIC industry codes can provide a more specific view of prediction and error detection performance by industry. The caveat with this analysis however, is that 30 of the 88 firms that meet the sample criteria form 1 industry by themselves. When firms are separated by their 4-digit SIC industry code, the sample consists of 46 industries (30 industries that are made up by one firm and 16 industries that are made up by more than one firm). The appendix , Table 30 displays the descriptive statistics by 4-digit SIC industry code. Results by firms' 4-digit SIC industry codes are presented for comparability purposes.

As presented in Tables 31 and 32, the results indicate that the better model between TCI and TCS with respect to traditional SAPs is model 5, which is able to generate superior sales predictions for 29 of the 46 industries. The results for continuous SAPs are presented in Tables 33 and 34. With respect to continuous SAPs with TCI and TCS, the results indicate that models 11 and 12 generate superior sales predictions for 41 industries, followed by model 7, which is superior for 39 industries. These findings indicate that even after firms are aggregated by their 4-digit SIC industry code, the results do not substantially change.

Table 35 displays the comparative results by 4-digit and 2-digit SIC industry code. The results parallel those that are documented for the 2-digit SIC industry code analysis in that the prediction performance improves for the models that contain timelier information, especially for models 5, 7, 8, 11, and 12. In addition, when the relative performance of 4-digit and 2-digit SIC industry code is compared, the findings suggest that model 5 experiences a large decrease in prediction performance of 10% for traditional SAPs, and model 9 experiences a large increase in prediction performance of 10% for continuous SAPs, while the remaining models experience a slight decrease.

The results of error detection performance for traditional SAPs with TCI and TCS are presented in Tables 36, 37, 38, 39, 40, 41, 42, and 43. With respect to the model that achieves lower false positive and lower false negative error rates, the results suggest that model 11 is the better performer as it produces lower error rates for both types of errors for 7 of the industries, whereas the benchmark models are able to achieve at most better error detection performance for 6 of the industries. When the cost ratio is examined to determine the overall error detection performance of the models, the findings indicate that the benchmark models are the better performers than models with TCI or TCS. The results of error detection performance for continuous SAPs with TCI and TCS are displayed in Tables 44, 45, 46, 47, 48, 49, 50, and 51. In this case, the results indicate that benchmark model 1 is more effective as it can achieve lower error rates for 7 industries. For the cost ratio analysis however, the benchmark models are the better performers than models with TCI or TCS.

Taken together, the results of the 4-Digit SIC error detection performance analysis indicate that error detection performance is diluted, which could be attributed to the fact that about two thirds of the 4-Digit SIC industries are made up of one firm. Table 52 presents the comparative results by 4-digit and 2-digit SIC industry code. Benchmark models outperform models with TCI or TCS, for traditional SAPs, and models 7 and 8 outperform benchmark models, for continuous SAPs. When the relative differences between the 4-Digit SIC versus the 2-Digit SIC analysis are compared, the error detection performance of the models with TCI or TCS deteriorates as there are substantial decreases in the overall error detection performance of models with TCI or TCS for traditional SAPs or continuous SAPs. These decreases in error detection performance can be attributed to single firms that make up 30 of the 46 4-Digit SIC industries, more firms would be needed in order to obtain a more general view of the error detection performance by industry.

***Advertising Expense***

Twitter can be useful as a marketing mechanism (Burton and Soboleva 2011) when advertising is limited (Tang 2017). Therefore, it is expected that the predictive performance of models that have TCI or TCS would be inferior when compared to benchmark models that contain information related to advertising expenditures. To test this expectation, information related to advertising expense is extracted from the fundamentals annual Compustat database. As the advertising expense for many firms was not disclosed for 2012, this year is excluded from the analysis. Monthly advertising expense is interpolated for the period of 2013 to 2017 by calculating the ratio of monthly sales to annual sales. Subsequently, in order to estimate the monthly expense, the ratio of monthly sales to annual sales is multiplied by the annual advertising expense. The analysis for prediction and error detection performance is presented by 2-Digit SIC for continuous SAPs as these were deemed to be more effective than traditional SAPs.

Tables 53 and 54 document the results for models with TCI and TCS and benchmark models that include advertising expense as an explanatory variable. The results indicate that the prediction performance of models 5 and 9, with TCI, and 6 and 10, with TCS are inferior to benchmark models 1 and 3, which contain firm-specific information of lagged sales and lagged advertising expense, or lagged sales, AR, and lagged advertising expense. Accordingly, these results indicate that information from Twitter is not a substitute for firms' advertising initiatives, which parallels findings documented in prior research (Tang 2017).

Interestingly, models 7 and 11, with TCI, and 8 and 12, with TCS, are superior to benchmark models 2 and 4, which contain lagged sales, lagged GDP, and lagged advertising expense, or lagged sales, lagged GDP, AR, and lagged advertising expense. These findings indicate that although the predictive performance of models with TCI or TCS is limited for firms that incur advertising expenditures, Twitter information has the potential to complement models with macroeconomic information by producing better sales predictions than models that contain macroeconomic information and advertising information but no Twitter information. Overall, the results substantiate the evidence provided in prior research about the value of traditional as well as the value of nontraditional external information.

The results of error detection performance for models with TCI and TCS and benchmark models that contain advertising information are documented in Tables 55, 56, 57, 58, 59, 60, 61, and 62. In general, the results suggest that continuous SAPs that contain TCI or TCS generate lower false negative error rates, especially models 5 and 6, than benchmark models that contain advertising information. On the other hand, benchmark models, in particular models 1 and 2, produce lower false positive error rates than models with TCI or TCS. To evaluate the overall error detection performance of the models, the cost ratio of false positive to false negative errors is examined. Consequently, model 7 and model 8 are the more effective models as they experience improved error detection performance for 11 industries when the cost ratio is 1:1, or for 12 industries when the cost ratio is 1:2.

The presented results for prediction and error detection performance provide evidence that although Twitter information is not a perfect substitute for firms' advertising initiatives, it has the potential to provide superior prediction and error detection performance when incorporated in models that contain traditional external information, such as GDP.

## 2.7. Conclusion

Social media information has the potential to serve as an independent source of audit evidence. Social media information proxies of consumer interest and satisfaction are correlated with sales and could be used as indicators of firms' sales patterns. As a result, auditors could potentially utilize this information to enhance the power of traditional and continuous substantive analytical procedures.

For prediction performance, the results suggest that continuous SAPs with prior month sales, prior month GDP and TCI or TCS, or continuous SAPs with prior month sales, prior month GDP, AR and TCI or TCS, produce superior sales predictions than the benchmark models that do not incorporate TCI or TCS. However, for auditors to fully exploit the benefits of SAPs, it is also important to examine the error detection performance of the models. For error detection performance, the results indicate that continuous SAPs that contain prior month sales, prior month GDP and TCI or TCS outperform the benchmark models as they can achieve superior error detection performance under varying cost ratios. Accordingly, the more effective model for both prediction and error detection performance is the model with prior month sales, prior month GDP and TCI as it can produce superior prediction and error detection performance for most of the industries that are examined. Collectively, the findings in this paper offer valuable insights concerning the role of nontraditional, social media, information in a variety of analytical procedures. Thus, indicating that social media information could be utilized as a potential source of audit evidence.

This paper contributes to the auditing literature in analytical procedures by investigating the incremental contribution of information that is generated by third-parties on social media platforms. The evidence presented in this paper may be useful to audit researchers, practitioners, and standard-setters as they evaluate the relevance of new forms of external nonfinancial information in enhancing audit quality. This paper has a few limitations. First, monthly observations are estimated from quarterly observations. While this interpolation method has been applied in prior research, it is possible that the estimated monthly amounts do not entirely reflect actual recorded account balances. Therefore, future research could examine the value of social media information on a disaggregated dataset of actual recorded balances. Second, this study focuses on 24 business-to-consumer industries, as these are the industries that satisfied all sample requirements, to provide a general view of prediction and error detection performance at the industry level. Future research could incorporate more industries to this analysis and further assess the value of social media information. Third, only one social media platform is examined. Future research could expand this analysis by evaluating the incremental value of information from other social media platforms.

# Chapter 3. Redesigning the Audit Process: Towards Robotic Audit Process Automation

## 3.1. Introduction

Auditing, by and large, remains an artisanal process where audit firm manuals, testing templates, supportive software, and ad hoc judgments collectively generate an audit opinion (Moffitt, Vasarhelyi, and Rozario 2018). While useful, these audit tools do not fully exploit the benefits of technology in achieving process formalization and potentially higher quality audits. Public accounting firms are increasingly recognizing the potential impact of utilizing technology to perform better audits (Titera 2013; Appelbaum et al. 2017). Technology-based audit techniques such as deep learning for fraud risk assessments, and rules-based functions for substantive audit testing, have been studied by academics and audit professionals (Issa, Sun, and Vasarhelyi 2016), however, these techniques reflect the direct automation of manual audit tasks[1] and do not present a systematic approach for the integration of these elements into a well-orchestrated audit process. Conversely, emerging research in the area of RPA (robotic process automation automation) for auditing suggests that this technology can achieve near end-to-end process automation while at the same time shifting the responsibilities of auditors towards more value-added tasks[2].

RPA is being explored by the PCAOB and public accounting firms as an emerging technology that can fundamentally transform the financial reporting process and auditing (Cooper et al. 2018; Hamm 2018). RPA certainly has the potential to evolve auditing but

---

[1] These technology-based audit techniques consist of algorithms that can achieve the automation of tasks.
[2] Moffit et al. 2018 and Cooper et al. 2018 share the view that in addition to having the capability of automating a process, RPA can repurpose the role of auditors by automating repetitive and structured work that is time consuming.

by itself is not sufficient as it can conduct work the same way that human auditors do, i.e. RPA reflects the mere mechanization of audit tasks. Moffitt et al. (2018) indicate that with the advent of RPA, old audit processes should be redesigned. This notion is supported by existing literature in process redesign, which documents that leveraging technology to rethink a process can improve its efficiency and effectiveness (Hammer 1990; Davenport and Short 1990). Consequently, it is important to explore the impact of technological process reframing [3] using RPA. To explore the evolution of auditing as a production line, this essay aims to foresee the future of audit by introducing and applying a methodology for the redesign of the audit process using RPA.

While audit automation is not a new concept as it first emerged in the 1980s (e.g. Vasarhelyi 1984; Groomer and Murthy 1989), the rethinking of the audit process as a result of automation remains underexplored. The audit process can be described as a system that contains a series of elements that are interdependent and function as a whole (Von Bertalanffy 1968). In practice however, auditing is a labor-intensive process that is composed of elements that are often not integrated (Moffitt et al. 2018). As auditing evolves to closely integrate into a digital business environment, it is important to rethink the audit process by envisioning a systematic and well-orchestrated audit approach that is facilitated by technology, this process is referred to as RAPA (robotic audit process automation).

[3] Technological process reframing (TPR) is defined as the rethinking of the audit process in light of new technologies (Issa et al. 2016).

RAPA can be defined as a non-invasive and systematic method that can be used to integrate interacting audit tasks by leveraging technology[4]. RAPA has the potential to transform auditing by eliminating a number of manual tasks that are repetitive, structured, and time consuming. Accordingly, this paper proposes a framework for RAPA based on existing methodologies proposed in the RPA and process redesign literature. RPA is described as "a type of software that mimics the activity of a human being in carrying out a task within a process. It can do repetitive stuff more quickly, accurately, and tirelessly than humans, freeing them to do other tasks" (McKinsey 2016). With RPA, the audit process is mechanized but is left intact as RPA software replaces the work that human users perform. Process redesign on the other hand is defined as using technology to redesign existing business processes (Hammer 1990; Davenport and Short 1990). Process redesign focuses on improving process performance by streamlining processes (IBM 2017). Hence, it is important to explore the synergies of RPA and process redesign to obtain useful insights into the impact of these paradigms on audit quality.

To guide the presentation of the proposed framework and its implementation, this paper adapts a design science approach (Peffers et al. 2007). Design science research entails the following six classifications:

1) Problem identification and motivation,
2) Define objective of a solution,
3) Design and development of an artifact,
4) Demonstration of the solution,

[4] RPA does not require programming knowledge and can achieve end to end process automation through the presentation layer (Lacity, Willcocks, and Craig 2015; Moffitt et al. 2018). When applied to auditing, RPA can form the audit production line as envisioned by Issa et al. 2016.

5) Evaluation of the solution, and

6) Communication of the results

The framework proposed in this paper consists of six phases. Phase 1 consists of developing the vision and process objectives. Phase 2 consists of identifying an audit process suitable for automation. Phase 3 entails understanding the audit process to be automated. Phase 4 consists of designing and implementing an ADS (audit data standard)[5] to collect audit relevant attributes in a consistent format for testing. Phase 5 consists of designing and implementing audit apps. Finally, Phase 6 consists of feedback and evaluation.

The incorporation of the ADS as part of the process, the utilization of the Microsoft Access audit app that automatically executes tests on the full population of accounting records, and the utilization of the RPA audit app to connect these otherwise disintegrated process activities, denote the redesign of the audit process. The ADS facilitates a systematic approach for the collection and preparation of audit evidence. Audit queries that function as an audit app are preprogrammed into Microsoft Access; this audit app can automatically execute audit tests and test the full population of records in near real-time. The ADS and the Microsoft Access audit app are integrated with RPA to achieve near end-to-end audit process automation. Taken together, the redesigning of the audit process using RPA technology represent a substantial departure from traditional auditing methods.

To examine the viability of the framework, the loan testing audit sub-process of a public accounting firm is automated. The results of the framework implementation provide

[5] For more information on the Audit Data Standard initiative please refer to: https://www.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/auditdatastandards-gl-august2013.pdf

evidence that it is useful in guiding the application of RAPA. Specifically, four of the five audit tasks related to the loan testing audit sub-process were automated by applying process redesign using RPA. RPA software was utilized to facilitate the execution of audit evidence collection activities, the transfer of the standardized audit evidence to Microsoft Access, and the execution of preprogrammed Microsoft Access queries that automatically execute full population audit tests. Collectively, the output of the framework indicates that it can achieve the intended objective of near end-to-end audit process automation.

This study has two main contributions. First, it provides insights into how audit, as a process, can transform itself to achieve an audit production line by proposing a framework for RAPA that is founded on existing methodologies for RPA and business process redesign (Davenport and Short 1990; Attaran 2003; Moffit et al. 2018). Second, this paper provides guidance for the application of the framework by implementing it to the loan testing audit sub-process of a public accounting firm. While some research studies have examined the redesign of audit processes (e.g. Vasarhelyi and Halper 1991; Alles, Brennan, Kogan, and Vasarhelyi 2006; Issa and Kogan 2014), more research is needed to explore the potential of end-to-end process automation.

This study provides useful insights to academics, audit practitioners, standard-setters and regulators by expanding the literature on the impact of emerging technologies on auditing through a practical application of the proposed framework. Moreover, this paper can inform audit practitioners on the application of RPA to auditing and provide some clarity on the feasibility of RPA implementation to actual audit engagements and on how existing audit methodologies should be revised, or whether new methodologies should be created to reflect the use of RPA on auditing. Finally, this paper can help inform standard-setters and regulators on whether there is a need to revise standards.

The remaining sections of this paper are divided as follows. Section 2 presents a background on RPA and process redesign. Section 3 describes the proposed framework for RAPA. The implementation of the framework is described in Section 4. Lastly, Section 5 concludes the study and provides suggestions for future research.

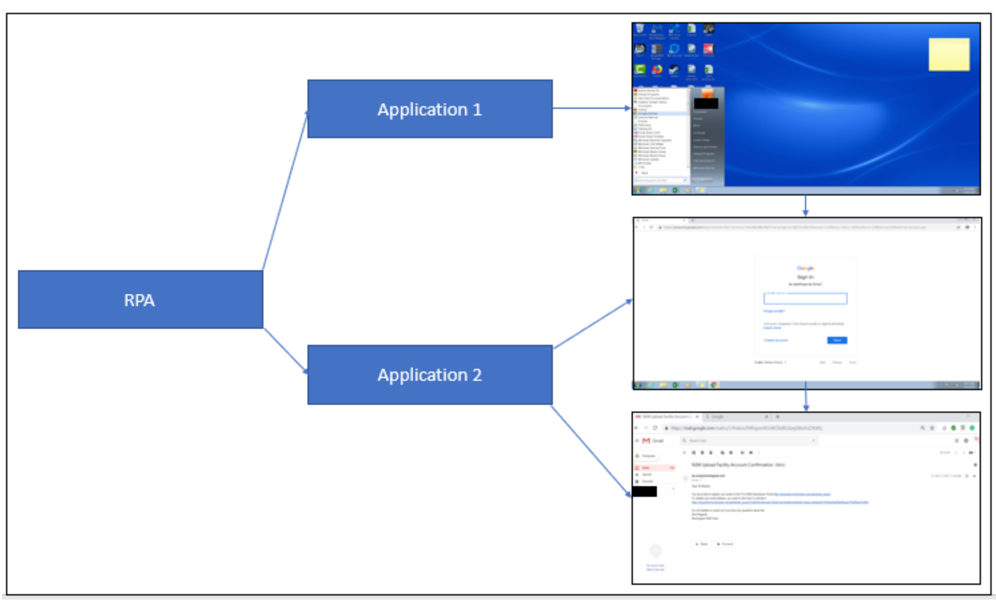
## 3.2. Background on RPA and Process Redesign

### 3.2.1. RPA

RPA is described as taking the robot out of the human by automating repetitive, manual, and structured tasks (McKinsey 2016). The IEEE Standards Association defines Robotic Process Automation (RPA) as: "a preconfigured software instance that uses business rules and predefined activity choreography to complete the autonomous execution of a combination of processes, activities, transactions, and tasks in one or more unrelated software systems to deliver a result or service with human exception management (IEEE Corporate Advisory Group, 2017)." These preconfigured software instances mimic the work that humans perform. It is important to note that RPA is a software robot and not a physical (hardware) robot that resides on top of the information technology infrastructure and that connects otherwise separated process activities.

RPA is different from other automation paradigms. First, RPA performs specific tasks in the same way that humans do. For example, to check email the RPA bot would follow the actions to click on the internet browser, type the email website, enter login credentials, and open email messages. Figure 7 illustrates these RPA tasks. Second, RPA is non-invasive as it represents a set of overlay software that resides on the presentation layer and connects disintegrated activities and software tools to form a smooth automated process (Lacity, Willcocks, and Craig 2015; Rozario and Vasarhelyi 2018b). This allows organizations to implement RPA without having to remodel the existing IT infrastructure. Third, RPA generally does not require programming skills as several of the RPA market leaders offer software that is user-friendly, consisting of drag and drop icons and recording options that facilitate the construction of a bot. Finally, when compared to regular, backend, automation RPA has demonstrated to maintain costs at a minimum and takes less time to implement (Lacity et al. 2015). Therefore, compared to regular automation, RPA is more flexible as it is capable of interacting with a variety of software applications, it is user-friendly compared to programming languages such as Python, R, or VBA, and is generally less expensive to implement.

**Figure 7: RPA Actions for Opening Email (Adapted and Modified from Lacity et al. 2015)**




RPA has been applied to various industries including the telecommunication, financial services, retail, manufacturing and the public accounting industry (Lacity et al.

2015; Willcocks and Lacity 2016; Seasongood 2016; Cooper et al. 2018). In the communication industry for example, RPA was implemented to process the swapping of telephone SIM cards and the pre-calculation of credit to a customer's account (Lacity et al. 2015). In the financial services industry most RPA implementation is occurring in the compliance domain, for KYC onboarding (Capgemini 2018). Finally, RPA in the retail and manufacturing industries has been adopted for the tracking of orders and returns (Seasongood 2016; Lin, Shih, Yang, and Kung 2018). The wide adoption of RPA across industries is not surprising as RPA implementation for the described use cases has demonstrated to improve process speed, decrease costs, and reduce the risk of error[6].

Although public accounting firms generally lag in technological innovation, Cooper et al. (2018) interview accounting professionals and document that RPA has gained momentum in the tax and advisory lines of service and that firms are beginning to experiment with this technology for assurance services. The challenge of applying RPA to auditing resides in the stringent regulations that publicly traded companies have to abide by. Nevertheless, the consensus among interviewees is that RPA would add value to auditing.

Moffitt et al. (2018) propose a framework for RPA implementation to auditing. The framework consists of seven stages, which can be divided into three major elements: 1) identifying the audit process to be automated and modularization of audit tasks, 2)

---

[6] In addition, a Deloitte RPA survey (2017) found that 53% of respondents from various organizations that were surveyed have started RPA projects and that by 2020, about 72% will embark on a RPA journey. Refer to: https://www2.deloitte.com/content/dam/Deloitte/bg/Documents/technology-media-telecommunications/Deloitte-us-cons-global-rpa-survey.pdf

developing an ADS, and 3) developing and testing audit apps. Huang (2019) applies RPA to the confirmation process to evaluate its feasibility. While the extant literature provides insights into RPA implementation to auditing, it does not fully address the redesigning of the audit process, which is necessary to fully exploit the benefits of technology (Alles et al. 2006; Davenport and Brain 2018).

To realize the maximum potential of RPA, it is necessary to rethink the audit process by shifting from manual audit tests to automated audit tests that enable full population testing near real-time (Vasarhelyi and Halper 1991; Alles et al. 2006; Chan and Vasarhelyi 2011; AICPA 2015; IAASB 2016, PCAOB 2017b; Byrnes 2015; Dai and Li 2016; Appelbaum et al. 2017). In addition, the inclusion of the ADS to the audit process could facilitate the execution of automated audit tests by structuring audit data in the same format (Li, Pawlicki, McQuilken, and Titera 2012). Collectively, automated audit tests that analyze the full population of accounting records in near real-time and the ADS represent a substantial departure from traditional audit methods.

### 3.2.2. Process Redesign

RPA automates the work that humans perform but preserves the flow of a process. Conversely, process redesign entails the transformation of a process. Existing literature in this domain indicates that a process should be redesigned to achieve process improvement and therefore enhance quality, reduce cost, and improve process speed (Gutierrez and Sastron 2018). Early articles by Hammer (1990) and Davenport and Short (1990), which propose ideas for process redesign, suggest that merely mechanizing a process does not address fundamental process deficiencies.

Changing the old rules of a process to new ones is consistent with the notion of process redesign. Hammer (1990) states that "[we should] use the power of modern information technology to radically redesign our business processes". In general, Hammer proposes a radical view to the redesign of a process suggesting that it is achieved when organizational structures and job designs are reshaped in a short period of time, an all or nothing approach. He describes case studies at Mutual Benefit Life and Ford to illustrate the importance of IT and organizational change in improving business performance. On the other hand, the non-radical view of process redesign is described as the specific function of reshaping a process using technology, rather than the reshaping of an organization (Davenport and Short 1990; Davenport 2015). Davenport and Short (1990) document the experiences of business process redesign at nineteen companies and propose a five-step method for process redesign. The five steps in the framework consist of 1) developing the business vision and process objective, 2) identifying the processes to be redesigned, 3) understanding and measuring existing processes, 4) identifying IT levers, and 5) designing and building a prototype of the new process.

The distinction between radical redesign and non-radical redesign can be made as the following: radical redesign is the transformation of a process, from the ground-up, using technology, whereas non-radical redesign is defined as a novel way of accomplishing the process in light of new technology (Mansar and Reijers 2007; Guitierrez and Sastron 2018). In addition, the information technology infrastructure, rather than the organization itself, is viewed as the primary enabler of non-radical process redesign (Davenport and Short 1990; Attaran 2003). Over the last three decades, radical and non-radical process redesign have had a significant impact on academia and practitioners. The literature in these domains have proposed a variety of frameworks and best practices.

Existing frameworks in both domains share similarities in that they generally prescribe the importance of understanding the process, envisioning and implementing the actions to the redesign of the process, and evaluating the implementation (Wastell, White and Kawalek 1994; Kettinger, Teng and Guha 1997; van Hee and Reijers 2000; Attaran 2003; Reijers and Mansar 2004). Frameworks that expand Hammer's theory of radical redesign also involve sociological aspects (e.g. Reijers and Mansar 2004), which is consistent with the notion that radical transformation includes changes to jobs and to the culture of an organization (Hammer 1990; Wastell et al. 1994). With respect to best practices, Mansar and Reijers (2007) conduct a literature review and find that the top 3 best practices are task elimination of unnecessary tasks, dividing large tasks into smaller tasks, and integrating technology to the process. Additionally, non-radical process redesign emphasizes the relationship between the process and IT where IT is viewed as the central precursor to process redesign (Davenport and Short 1990; Attaran 2003). Table 63 contrasts radical and non-radical process redesign.

**Table 63: Differences Between Radical and Non-Radical Process Redesign**

| | **Radical Process Redesign–- Hammer (1990)** | **Non-radical Process Redesign–- Davenport and Short (1990)** |
|---:|---|---|
| **Definition** | Defined as the transformation of a process, from the ground-up, using technology | Defined as the novel way of accomplishing the process in light of new technology |
| **Driver** | Technology and management are the primary drivers | Technology is the primary driver |
| **Level of improvement** | Revolutionary – dramatic improvements to the process and to the organization | Evolutionary – incremental improvements to the process |

| | | |
|---|---|---|
| | Process, organizational structures, and job designs are transformed | Process and some job design transformation |

Interestingly, research suggests that radical process redesign was not typically practiced and that it lost traction as use cases focused on "evolutionary" rather than "revolutionary" implementations (Davenport and Stoddard 1994; Jarvenpaa and Stoddard 1998). Consequently, because of the highly regulated nature of auditing, which may stifle audit innovation, non-radical process redesign can be more suitable than radical process redesign when envisioning the progressive transformation of the audit process as a result of technology. Essentially, non-radical process redesign would reflect the incremental improvements to the auditing process. As auditing evolves to closely integrate into a digital business environment, it is important to rethink the audit process by foreseeing auditing as a production line resulting from its redesign using RPA.

RPA and process redesign are similar in that these techniques utilize technology to improve processes and process outcomes, however what differentiates process redesign from RPA is the transformation of old process rules to new process rules. RPA mechanizes the work but does not fundamentally change the process. Alles et al. (2006) suggest that process redesign is a critical element of technology implementation in order to realize technological capabilities to the maximum. As a result, there is a natural synergy to be exploited from implementing process redesign and RPA.

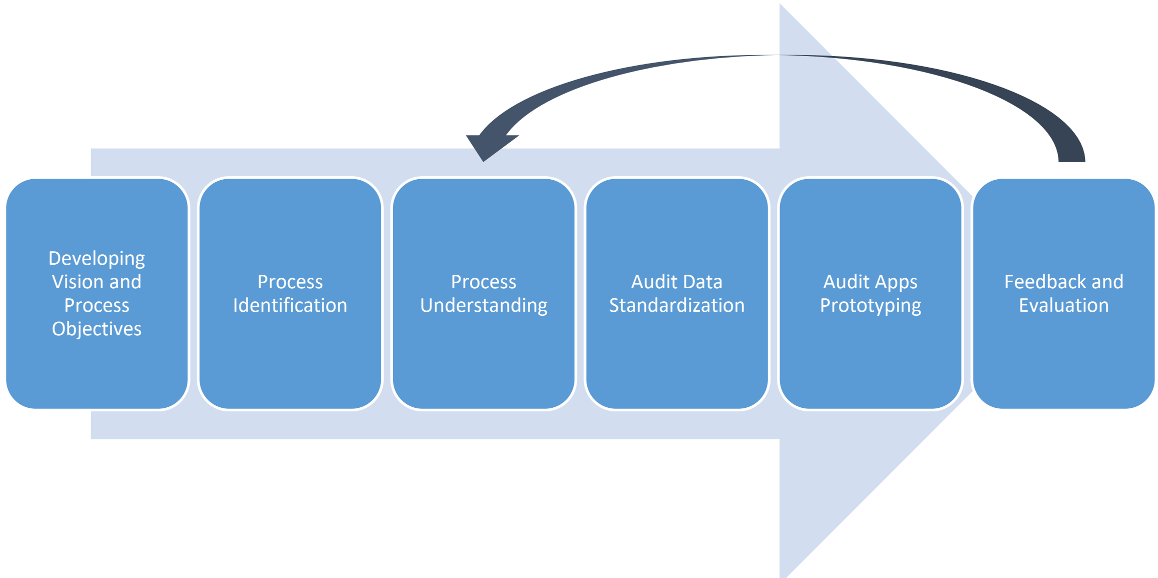
In auditing, shifting from sample-based to a full population testing audit approach, from manual evidence collection procedures to automated evidence collection procedures in a standardized format, i.e. the ADS, and from manual audit tests to automated audit tests, that can be executed near real-time, closely reflects the spirit of process redesign. Under the traditional audit approach, auditors manually collect audit evidence and select a sample of items that represent the population (AU-00350), manual audit procedures are then applied to each item in the sample. Items that do not conform to rules prescribed in audit procedures are then investigated and conclusions about the risk of material misstatement on the financial statements are derived on an annual basis. With RAPA, audits naturally shift to an audit by exception approach whereby evidence collection activities and audit tests are automated and full populations are tested in near real-time (Vasarhelyi and Halper 1991; Alles et al. 2006; Appelbaum et al. 2017), thus allowing auditors to estimate the risk of material misstatement more precisely.

Taken together, RPA and process redesign are necessary elements in rethinking how the audit process will evolve by using technology. The revival of process redesign as a result of RPA may be inevitable (Davenport 2015). Consequently, it is important for public accounting firms to explore the impact of both process redesign and RPA on auditing. RPA software robots would enable audit process efficiency by largely mechanizing parts of the audit process but does not expand the auditing paradigm to fully reap the benefits of automation using RPA. With process redesign, the auditing paradigm advances by shifting from anachronistic audit rules to new rules that closely reflect a digital business environment.

## 3.3. Robotic Audit Process Automation Framework

Facilitated by RPA and process predesign, RAPA has the potential to transform the way that audits are conducted. To provide guidance related to RAPA, a framework that expands the two process improvement paradigms is proposed:

**Figure 8: Robotic Audit Process Automation Framework**




As depicted in Figure 8, the six phases of the framework consist of 1) developing vision and process objectives, 2) process identification, 3) process understanding, 4) developing and implementing an audit data standard, 5) developing and implementing audit apps and 6) feedback and evaluation. The sections of the framework are described below:

### 3.3.1. Developing Vision and Process Objectives

Automation should be rationalized (Davenport and Short 1990; Kettinger et al. 1997; van Hee and Reijers 2000; Attaran 2003). Essentially, prior to embarking on the process automation journey, leaders of the process justify the reason for automation and generate objectives. Objectives for justifying process automation generally relate to 1) the need to reduce costs and the time that it takes to perform the task, 2) the need to improve process quality, and 3) the need to improve quality of work life; that is, the need to maintain employees motivated. This initial phase of process redesign emphasizes the reasoning for why it is necessary to automate.

### 3.3.2. Process Identification

Practically all processes can benefit from automation however, entities should aim for easy wins and target processes where the benefits of automation would exceed the costs. Consequently, public accounting firms can benefit from automating processes that consist of many rules-based tasks that are repetitive and that do not require audit judgment with RPA (Attaran 2003; Moffitt et al. 2018). Subject matter experts in public accounting firms can assist in the identification of a process where RAPA can add value. In addition, targeting low risk audit processes would be preferable as it can be less disruptive to internal and external stakeholders such as auditors and regulators (Rozario and Vasarhelyi 2018b).

### 3.3.3. Process Understanding

Upon identifying an audit process that would benefit from RPA, the RAPA team [7] within the audit firm would proceed to obtain an understanding of the process. Understanding a process from beginning to end can help identify problems within it so that they are not repeated, and can trigger discussions about how to improve the process to reflect new rules that closely parallel the digital business environment (Davenport 1990; Moffitt et al. 2018). By understanding the process, the RAPA team can envision the process areas that can be mechanized and the process areas that can be redesigned to achieve process improvement (van Hee and Reijers 2000; Attaran 2003).

In this phase of RAPA, procedures within an audit process can be segmented into micro audit procedures that can be interpreted by computer software (Alles et al. 2006; Moffitt et al. 2018). Through a brainstorming session, the RAPA team can come to the realization that not all audit procedures can be automated using RPA. Procedures could

[7] Moffitt et al. (2018) suggest that the public accounting firm, the RPA vendor, and a consulting firm can work together to implement RPA-enabled audit process automation. b

perhaps not be suitable for automation due to unstructured judgments that need to be made, or the lack of digitized information (Davenport and Short 1990; van Hee and Reijers 2000; Attaran 2003; Moffitt et al. 2018). The RAPA team can also brainstorm about the audit methods that can be transformed. In general, transformation of an audit process may comprise automating audit tests that examine complete population sets, however, another important aspect that audit firms should consider is the standardization of the data, which would enable the seamless execution of automated audit tests across several audit engagements.

### 3.3.4. Audit Data Standardization

Data is the new gold (The Economist 2017) and in auditing data is the foundation of an audit opinion. The audit opinion is derived from financial and non-financial audit evidence that underlies financial statements. As auditing moves towards using technology, a salient element of RAPA is the structuring of data into a consistent format. "Data should be in a structured format for the software program to successfully interpret the inputs (data attributes) of automated audit procedures. The reality, however, is that data that is collected as audit evidence come from different sources and in different labels, though the labels represent the same object" (Moffitt et al. 2018). Consequently, audit data standardization is necessary for RAPA to come to fruition.

The idea of standardizing audit relevant data is not new. The AICPA (American Institute of Certified Public Accountants) Assurance Services Executive Committee proposed the ADS (Audit Data Standard) in 2013, however, the ADS is increasingly gaining traction as audit firms launch audit automation initiatives. The creation of an ADS for a specific audit process can facilitate the deployment of automated audit tests by maintaining data in a structured format within a template (Zhang et al. 2012). Important to note is that the incorporation of an ADS within an audit process signifies a substantial departure from the traditional rules in an audit process. Under the traditional audit paradigm, audit evidence from company files is manually entered into audit workpapers.

Compiling audit relevant attributes from a myriad of company files into an ADS template can be a burdensome task for auditors and contrasts the notion of process redesign. The preparation of an ADS can be separated into micro audit procedures to open the source files, copy and paste information into the ADS template, and format the data in the template. RPA software robots can assist with the implementation of an audit process ADS. As a result, there are four components to the redesign of the audit process using RPA: 1) the inclusion of an ADS into an audit process, 2) automated audit tests that enable 3) full population testing, and 4) RPA software that connects process activities one, two, and three.

### 3.3.5. Audit Apps Prototyping

Audit apps are preprogramed audit procedures that execute formalized audit tests (Vasarhelyi, Warren, Teeter, and Titera 2014; Byrnes 2015; Dai and Li 2016). The RAPA team can preprogram the described micro audit procedures as audit apps to 1) collect audit evidence, 2) prepare the ADS, 3) import the ADS into audit software and 4) automatically execute audit tests. Audit apps may execute a single audit procedure, or a combination of audit procedures (Dai and Li 2016). A single audit procedure is performed by the audit app that automatically executes audit tests on a complete population of records, app number four, while RPA software is used to integrate these four procedures to achieve near end-to-end process automation. Hence, RPA software is the app that executes a combination of audit procedures, starting from audit evidence collection activities and finishing with the automated execution of audit tests. By redesigning the audit process using RPA, auditors can exploit the benefits of RAPA as it pertains to well-structured and defined audit tasks.

### 3.3.6. Feedback and Evaluation

Feedback and evaluation are vital to assessing the success of IT-enabled process redesign. Feedback is necessary to determine if there are any improvements that need to be made (van Hee and Reijers 2000). For example, the design of the audit apps are essentially prototypes that can be modified. The pilot implementation of the audit apps across a few audit engagements may indicate that the apps are not operating as intended (i.e. to achieve near end-to-end process automation). As a result, it is possible that several iterations of the apps would need to be designed to achieve the envisioned audit production line. Similarly, a method of evaluation is necessary to measure the value of the implementation (Attaran 2003).

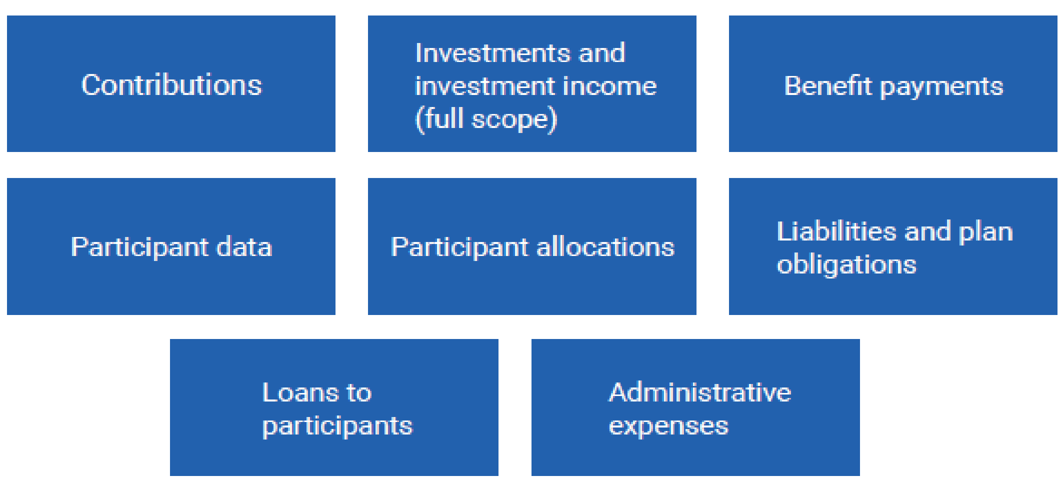
Evaluation can be measured in terms of efficiency and effectiveness (Kettinger et al. 1997; van Hee and Reijers 2000; Attaran 2003; Moffitt et al. 2018). Efficiency can be measured by comparing the cycle time under the RAPA approach (i.e. the total time RAPA spends conducting the work) to the cycle time under the traditional approach. Whereas effectiveness can be measured by comparing the quality of the product of the process under RAPA to that of the traditional approach. In auditing, effectiveness is measured by the incidence of material accounting anomalies that were discovered (Louwers et al. 2018). Another salient metric to determine the value of IT-enabled process redesign is cost. RAPA implementation should result in cost savings by reducing the time it takes to perform audit tasks and the number of accounting anomalies that remain undetected.

## 3.4. Application of Robotic Audit Process Automation Framework

A public accounting firm approached the research team to request for assistance in the implementation of RPA to automate their processes. The main motivation for the implementation of RPA was to envision a holistic approach for more effective and efficient audits as a result of disruptive technology. The EBP (employee benefit plan) audit process, which is a class of compliance audit, was targeted. This audit was deemed to be a qualified candidate for this research project as it consisted of sub-processes that contained several audit activities that were structured and repetitive. Moreover, the most recent survey of EBP audits indicates that approximately 40% of these audits contained deficiencies (U.S. Department of Labor 2015). Consequently, there is clearly a need to explore whether RPA can improve the quality of EBP audits.

One of the main objectives of an EBP audit is to "help protect the financial integrity of the employee benefit plan, which helps users determine whether the necessary funds will be available to pay retirement, health and other promised benefits to participants" (AICPA 2018). Figure 9 describes the sub-processes that are covered by EBP audits:

**Figure 9: EBP Audit Sub-Processes (Adapted from AICPA EBP Handbook)**



Four of the eight EBP audit sub-processes were selected as candidates for RAPA. These included: benefit payments, participant data, participant allocations, and loans to participants. In general, the audit activities within these sub-processes entailed the matching of data attributes, such as the date of hire of plan participants, contribution amount, and loan amount, from one source file to the others. As a result, it was determined that several of these audit activities could be converted into rules-based functions that would benefit from RAPA.

This paper documents the application of the proposed framework for RAPA to facilitate the redesign of one of these sub-processes, the loan testing audit sub-process[8]. This sub-process was selected to illustrate the feasibility of the framework as data for this audit area was readily available and in a machine-readable format.

### 3.4.1. Developing Vision and Process Objectives

The first phase of redesigning the audit process with RPA entails asking the question, why automate? The research team and the public accounting firm had several planning discussions about possibly automating audit sub-processes using RPA. The vision and objectives that emerged from these discussions was to apply RPA to achieve near end-to-end process automation. The motivation to move towards RPA-enabled audits was to help reduce the time that is spent performing tasks that do not require audit judgment, improve quality in the audit, and potentially encourage auditors to perform more meaningful work. In general, the firm expressed an interest in automating the EBP audit process using RPA since their auditors spent a significant number of hours performing

[8] Important to note is that while actual audit engagement data was used to understand the process and develop the RAPA prototype, this paper presents results that are based on simulated data.

audit tasks that primarily consisted of tasks that were rule-based. During these planning discussions, it was also noted that RPA technology could lead to enhanced audit quality by reducing the risk of human error and repurposing the responsibilities of their auditors to perform more value-added work, such as spending more time evaluating notable items.

### 3.4.2. Process Identification

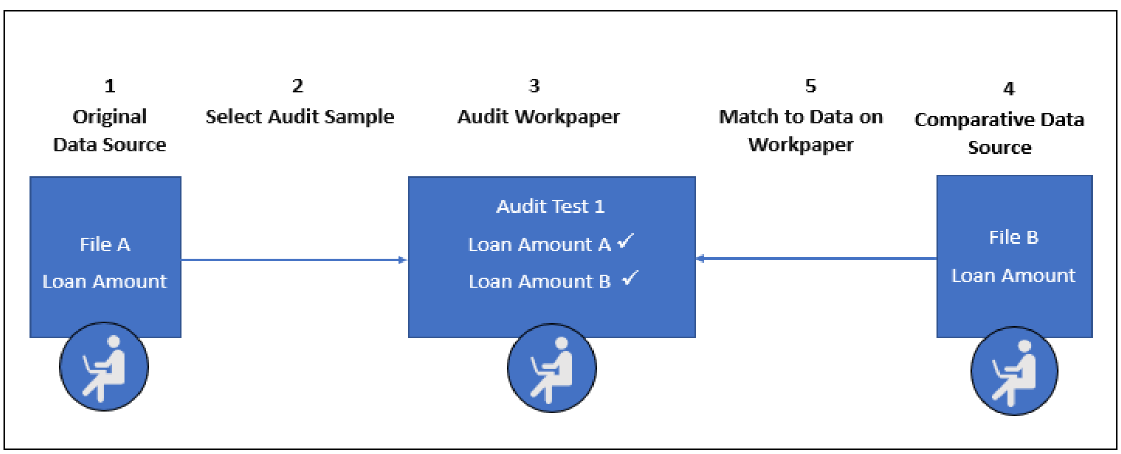
The loan testing audit sub-process at a public accounting firm was targeted as an area that would benefit from process automation using RPA. The audit objective of this process entails determining if employee loans are being administered in accordance with the company's employee loan agreement in order to ensure that they are properly presented in financial statements (AICPA 2018). The audit procedures within this sub-process required auditors to manually input information from company files into audit workpapers. Once entered into workpapers, auditors would manually complete a series of verification checks to provide assurance over the validity of employee loans. Specifically, the loan testing audit process consisted of the following five audit tests:

1) Ensure that the employee loan amount reported by the company matches the loan amount disbursed by the company
2) Ensure that the interest rate for the employee loan reported by the company matches the interest rate reported on the company's employee loan agreement
3) Ensure that the employee loan repayment amount from payroll matches the loan repayment amount per the loan amortization table
4) Ensure that the employee loan amount satisfies the minimum loan amount requirement as per the company's employee loan agreement

5) Ensure that the employee had the permissible number of loans as per the employee loan agreement

The execution of these audit tests within this audit sub-process do not require audit judgment as tests one to five represent rules-based audit activities that can be automated. In addition, the copying and pasting of information into an audit workpaper template is an activity that does not require analytical effort. Figure 10 presents an example of the five activities that are required to perform the first audit procedure. Under the traditional approach, the auditor is involved throughout the process where they first obtain the data that needs to be verified, select a sample, transfer it to the workpaper, obtain audit evidence to compare the original data source and match it. If the amounts match, the accounting record passes the test. These activities are performed for the remaining audit sample. As the activities within this audit procedure can be formalized into rules-based tasks, it would benefit from RAPA.

**Figure 10: Activities to Perform Loan Amount Match**



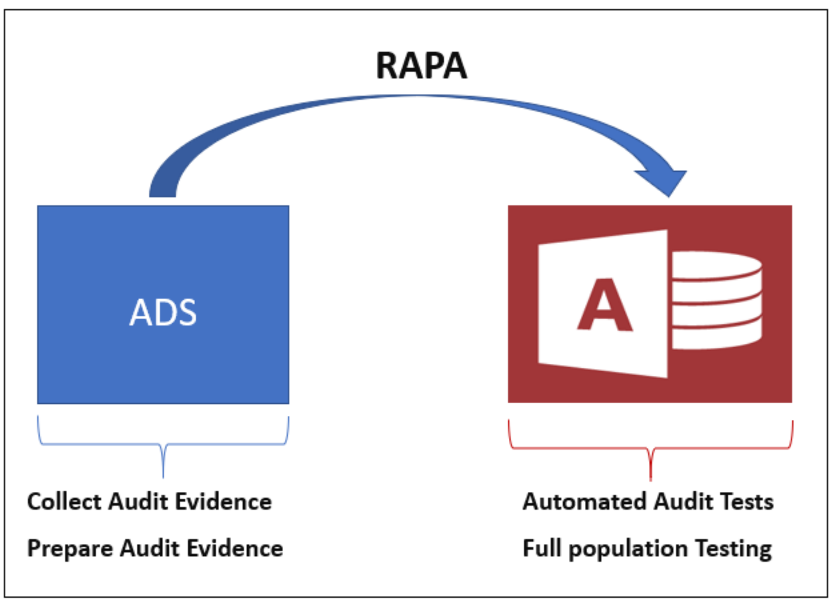
### 3.4.3. Process Understanding

To understand the loan testing audit sub-process the RAPA team (research team and the firm) reviewed prior year audit workpapers and audit evidence. Follow-up meetings were held with EBP audit subject matter experts to ensure understanding of these audit activities were correct. Understanding this sub-process from beginning to end helped the RAPA team identify activities that could be integrated with RPA. Procedures such as the collection of audit evidence and the automated execution of audit tests are disparate procedures that appeared feasible to integrate with RPA. The seamless automation of this process could help reduce the risk of human error in transposing evidence from company provided files into an audit workpaper and in identifying audit items that pass or fail audit tests efficiently. Moreover, the five audit tests outlined above could be divided into micro audit tests that can be preprogrammed into audit software, however, from discussions with the subject matter experts, the team concluded that automation could not be accomplished for audit step three in the process identification phase as the information on the payroll report was available via PDF file type and not in a file type that would be easily interpreted by a computer program. Attempts were made to convert the PDFs to machine readable files, albeit unsuccessfully. Accordingly, four of the five audit tests were selected for automation and full population testing. Taken together, the RAPA strategy developed by the team consisted of utilizing Microsoft Access as a prototype for automated audit queries and RPA to execute evidence collection activities, and audit queries.

### 3.4.4. Audit Data Standardization

RPA can facilitate the collection of audit evidence, however, merely importing audit evidence into an audit firm's testing template for one audit engagement only imitates the work that auditors perform for that specific audit engagement and does not exploit RPA to its full potential. Thus, training RPA software robots to collect and process audit evidence from a specific set of files and import it into an audit testing template limits RAPA to imitate the work of one audit engagement; whereas firms typically have numerous audit engagements. In practice, audit evidence files can come from different companies, and can contain different data labels, for the same data objects. Thus, it is essential to develop a systematic method, for evidence collection and processing activities to fully exploit the power of IT-enabled process redesign using RPA. Because RPA software robots have the capacity to collect and process information that is presented in a consistent format, which would include files that have the same labels that the RPA software was originally trained on, it is unavoidable to create a standard template for audit evidence. As a result, the overall RAPA strategy consisted of developing an ADS to maintain audit evidence in a consistent format and automating audit tests that perform full population testing. Figure 11 describes the overall RAPA strategy.

**Figure 11: RAPA Strategy**



The RAPA team developed the loan testing audit process ADS to achieve to goal of data standardization. The ADS for this audit process is reflected as an Excel workbook template with four components:

1) Link to Standard Field: Excel tab that represents the ADS dictionary, Figure 12. This tab contains the standard (generic) name of the audit data fields and maps it to the actual column names per company reports. This tab also contains the data type and the name of the reports the data was originally extracted from.

**Figure 12: ADS Dictionary**

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Standard Name | Column Name Per Report | Data Type | Report |
| 2 | Employee_ID | SSN | NUMERICA | Annual Loan Balance |
| 3 | Name | Participant Name | TEXT | Annual Loan Balance |
| 4 | Loan_Number | Loan ID | NUMERICA | Annual Loan Balance |
| 5 | Loan_Amount | Loan Amount | NUMERICA | Annual Loan Balance |
| 6 | Interest_Rate | Int Rate | Percentage | Annual Loan Balance |
| 7 | Date_Opened | Date Opened | DATE | Annual Loan Balance |
| 8 | Year_Opened | Date Opened2 | DATE | Annual Loan Balance |
| 9 | Employee_ID | SSN | NUMERICA | Check Register |
| 10 | Name | PAYEE | TEXT | Check Register |
| 11 | Loan_Amount_R2 | NET AMT | NUMERICA | Check Register |

2) Company reports: Excel tabs that represent direct data extracts from files that are collected by auditors. In this particular case, there are two Excel tabs, one for "Annual Loan Balance", and the other for "Check Register", Figures 13 and 14, respectively.

**Figure 13: Annual Loan Balance**

| | SSN | Participant Name | Loan ID | Int Rate | Date Oper | Loan Amo |
|---|---|---|---|---|---|---|
| 2 | XXX-XX-1234 | Farrah Stambaugh | LOAN 11 | 5 | 6132016 | 9199 |
| 3 | XXX-XX-1235 | Cecelia Kendra | LOAN 04 | 5 | 3302016 | 3739 |
| 4 | XXX-XX-1236 | Alba Moseley | LOAN 02 | 5 | 8182016 | 5160 |
| 5 | XXX-XX-1237 | Emil Stlouis | LOAN 03 | 5 | 1222016 | 8030 |
| 6 | XXX-XX-1238 | Taren Farrelly | LOAN 02 | 5 | 8082016 | 13202 |
| 7 | XXX-XX-1239 | Tiana Harstad | LOAN 03 | 5 | 8302016 | 8793 |
| 8 | XXX-XX-1240 | Bette Wildt | LOAN 02 | 5 | 12232016 | 10462 |
| 9 | XXX-XX-1241 | Gustavo Kocher | LOAN 03 | 5 | 5102016 | 10572 |
| 10 | XXX-XX-1242 | Latrina Pickel | LOAN 02 | 5 | 7202016 | 1412 |
| 11 | XXX-XX-1243 | Irena Wease | LOAN 03 | 5 | 6272016 | 14191 |
| 12 | XXX-XX-1244 | Aide Nuckles | LOAN 11 | 5 | 9062016 | 179 |
| 13 | XXX-XX-1245 | Ester Mullings | LOAN 04 | 5 | 4252016 | 1761 |
| 14 | XXX-XX-1246 | Russ Cushman | LOAN 02 | 5 | 2222016 | 14938 |
| 15 | XXX-XX-1247 | Allena Aldridge | LOAN 03 | 5 | 7192016 | 5426 |
| 16 | XXX-XX-1248 | Hermila Faw | LOAN 02 | 5 | 6272016 | 1579 |
| 17 | XXX-XX-1249 | Gerry Osby | LOAN 03 | 5 | 12292016 | 9025 |
| 18 | XXX-XX-1250 | Fernande Fuhr | LOAN 02 | 5 | 10102016 | 635 |
| 19 | XXX-XX-1251 | Maris Vicente | LOAN 03 | 5 | 12232016 | 13225 |
| 20 | XXX-XX-1252 | Natashia Maag | LOAN 02 | 5 | 2022016 | 14613 |
| 21 | XXX-XX-1253 | Odis Douglass | LOAN 03 | 5 | 12052016 | 7663 |
| 22 | XXX-XX-1254 | Letitia Gambrel | LOAN 11 | 5 | 6022016 | 7063 |

Link to Standard Field | Annual Loan Balance | Check Register | Lo

**Figure 14: Check Register**

| | SSN | PAYEE | AMOUNT |
|---|---|---|---|
| 2 | XXX-XX-1234 | Farrah Sta | 9199 |
| 3 | XXX-XX-1235 | Cecelia Ke | 3739 |
| 4 | XXX-XX-1236 | Alba Mose | 5160 |
| 5 | XXX-XX-1237 | Emil Stlou | 9999 |
| 6 | XXX-XX-1238 | Taren Farr | 13202 |
| 7 | XXX-XX-1239 | Tiana Hars | 8793 |
| 8 | XXX-XX-1240 | Bette Wild | 10462 |
| 9 | XXX-XX-1241 | Gustavo K | 10572 |
| 10 | XXX-XX-1242 | Latrina Pic | 1412 |
| 11 | XXX-XX-1243 | Irena Wea | 14191 |
| 12 | XXX-XX-1244 | Aide Nuck | 179 |
| 13 | XXX-XX-1245 | Ester Mull | 1761 |
| 14 | XXX-XX-1246 | Russ Cush | 14938 |
| 15 | XXX-XX-1247 | Allena Ald | 5426 |
| 16 | XXX-XX-1248 | Hermila Fa | 1579 |
| 17 | XXX-XX-1249 | Gerry Osb | 9025 |
| 18 | XXX-XX-1250 | Fernande | 6350 |
| 19 | XXX-XX-1251 | Maris Vice | 13225 |
| 20 | XXX-XX-1252 | Natashia M | 14613 |
| 21 | XXX-XX-1253 | Odis Doug | 7663 |
| 22 | XXX-XX-1254 | Letitia Gar | 7063 |

Link to Standard Field | Annual Loan Balance | Check Register

3) Loan Testing – ADS Data Prep: Excel tab that maps the original source of the data to the standard fields of the data, per the ADS dictionary, and the direct data extracts from company files, Figure 15. This template contains preprogrammed functions to preprocess the data. As an example, functions to extract the last four digits of the employees' social security number, which is used as the employee ID, and functions to trim the date to extract the year that the loan opened were programmed. Additionally, as the direct data extracts contain loan information for prior year and current year loans, the final step in this tab is to filter the "Date Opened" by the current year to reflect the employee loans that need to be verified for the current year under audit. There were 300 records that needed verification.

**Figure 15:ADS Data Preprocessing**

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Source | Annual Loan Balance | Annual Loan I | Annual Loan Bala | Annual Loan Bal | ance | Annual Loan Bal | Annual Loan Bal | Check Register | Check Register | Check Register |
| 2 | Standard Field | Employee ID | Name | Loan Number | Date Openec | Year Opened | Loan Amoun | Interest Rate | Employee_ID | Name | Loan_Amount_R25 |
| 9 | | 1234 | Farrah Stamb | LOAN 11 | 6132016 | 2016 | 9199 | 5.00 | 1240 | Bette Wildt | 10462.00 |
| 10 | | 1235 | Cecelia Kendr | LOAN 04 | 3302016 | 2016 | 3739 | 5.00 | 1241 | Gustavo Kocher | 10572.00 |
| 11 | | 1236 | Alba Moseley | LOAN 02 | 8182016 | 2016 | 5160 | 5.00 | 1242 | Latrina Pickel | 1412.00 |
| 12 | | 1237 | Emil Stlouis | LOAN 03 | 1222016 | 2016 | 8030 | 5.00 | 1243 | Irena Wease | 14191.00 |
| 13 | | 1238 | Taren Farrelly | LOAN 02 | 8082016 | 2016 | 13202 | 5.00 | 1244 | Aide Nuckles | 179.00 |
| 14 | | 1239 | Tiana Harstac | LOAN 03 | 8302016 | 2016 | 8793 | 5.00 | 1245 | Ester Mullings | 1761.00 |
| 15 | | 1240 | Bette Wildt | LOAN 02 | 12232016 | 2016 | 10462 | 5.00 | 1246 | Russ Cushman | 14938.00 |
| 16 | | 1241 | Gustavo Koch | LOAN 03 | 5102016 | 2016 | 10572 | 5.00 | 1247 | Allena Aldridge | 5426.00 |
| 17 | | 1242 | Latrina Pickel | LOAN 02 | 7202016 | 2016 | 1412 | 5.00 | 1248 | Hermila Faw | 1579.00 |
| 18 | | 1243 | Irena Wease | LOAN 03 | 6272016 | 2016 | 14191 | 5.00 | 1249 | Gerry Osby | 9025.00 |
| 19 | | 1244 | Aide Nuckles | LOAN 11 | 9062016 | 2016 | 179 | 5.00 | 1250 | Fernande Fuhr | 6350.00 |
| 20 | | 1245 | Ester Mulling | LOAN 04 | 4252016 | 2016 | 1761 | 5.00 | 1251 | Maris Vicente | 13225.00 |
| 21 | | 1246 | Russ Cushma | LOAN 02 | 2222016 | 2016 | 14938 | 5.00 | 1252 | Natashia Maag | 14613.00 |
| 22 | | 1247 | Allena Aldridg | LOAN 03 | 7192016 | 2016 | 5426 | 5.00 | 1253 | Odis Douglass | 7663.00 |
| 23 | | 1248 | Hermila Faw | LOAN 02 | 6272016 | 2016 | 1579 | 5.00 | 1254 | Letitia Gambrel | 7063.00 |
| 24 | | 1249 | Gerry Osby | LOAN 03 | 12292016 | 2016 | 9025 | 5.00 | 1255 | Diana Zager | 10892.00 |
| 25 | | 1250 | Fernande Fuh | LOAN 02 | 10102016 | 2016 | 635 | 5.00 | 1256 | Violeta Been | 8516.00 |
| 26 | | 1251 | Maris Vicente | LOAN 03 | 12232016 | 2016 | 13225 | 5.00 | 1257 | Asuncion Peskin | 10462.00 |
| 27 | | 1252 | Natashia Ma | LOAN 02 | 2022016 | 2016 | 14613 | 5.00 | 1258 | Versie Daniel | 12269.00 |
| 28 | | 1253 | Odis Douglas | LOAN 03 | 12052016 | 2016 | 7663 | 5.00 | 1259 | Yetta Cropp | 1739.00 |
| 29 | | 1254 | Letitia Gambr | LOAN 11 | 6022016 | 2016 | 7063 | 5.00 | 1260 | Margery Rain | 13917.00 |
| 30 | | 1255 | Diana Zager | LOAN 04 | 8292016 | 2016 | 10892 | 5.00 | 1261 | Caridad Finnie | 14678.00 |
| 31 | | 1256 | Violeta Been | LOAN 02 | 11252016 | 2016 | 8516 | 5.00 | 1262 | Hailey Lainez | 4598.00 |
| 32 | | 1257 | Asuncion Pes | LOAN 03 | 4252016 | 2016 | 10462 | 5.00 | 1263 | Dovie Garoutte | 2990.00 |

4) Loan Testing–- ADS Copy Paste: Excel tab that represents the integrated ADS structure for the loan testing audit sub-process, Figure 16. This tab is simply a copy and paste from the "Loan Testing – ADS Data Prep" tab. This template is subsequently imported into Microsoft Access

**Figure 16: Integrated ADS Structure**

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Company | Employee_ID | Name | Loan_Number | Date_Opened | Year_Opened | Loan_Amount | Interest_Rate | Loan_Amount_R25 |
| 2 | 1 | 1234 | Farrah Stambaugh | LOAN 11 | 6132016 | 2016 | 9199 | 5 | 9199 |
| 3 | 1 | 1235 | Cecelia Kendra | LOAN 04 | 3302016 | 2016 | 3739 | 5 | 3739 |
| 4 | 1 | 1236 | Alba Moseley | LOAN 02 | 8182016 | 2016 | 5160 | 5 | 5160 |
| 5 | 1 | 1237 | Emil Stlouis | LOAN 03 | 1222016 | 2016 | 8030 | 5 | 9999 |
| 6 | 1 | 1238 | Taren Farrelly | LOAN 02 | 8082016 | 2016 | 13202 | 5 | 13202 |
| 7 | 1 | 1239 | Tiana Harstad | LOAN 03 | 8302016 | 2016 | 8793 | 5 | 8793 |
| 8 | 1 | 1240 | Bette Wildt | LOAN 02 | 12232016 | 2016 | 10462 | 5 | 10462 |
| 9 | 1 | 1241 | Gustavo Kocher | LOAN 03 | 5102016 | 2016 | 10572 | 5 | 10572 |
| 10 | 1 | 1242 | Latrina Pickel | LOAN 02 | 7202016 | 2016 | 1412 | 5 | 1412 |
| 11 | 1 | 1243 | Irena Wease | LOAN 03 | 6272016 | 2016 | 14191 | 5 | 14191 |
| 12 | 1 | 1244 | Aide Nuckles | LOAN 11 | 9062016 | 2016 | 179 | 5 | 179 |
| 13 | 1 | 1245 | Ester Mullings | LOAN 04 | 4252016 | 2016 | 1761 | 5 | 1761 |
| 14 | 1 | 1246 | Russ Cushman | LOAN 02 | 2222016 | 2016 | 14938 | 5 | 14938 |
| 15 | 1 | 1247 | Allena Aldridge | LOAN 03 | 7192016 | 2016 | 5426 | 5 | 5426 |
| 16 | 1 | 1248 | Hermila Faw | LOAN 02 | 6272016 | 2016 | 1579 | 5 | 1579 |
| 17 | 1 | 1249 | Gerry Osby | LOAN 03 | 12292016 | 2016 | 9025 | 5 | 9025 |
| 18 | 1 | 1250 | Fernande Fuhr | LOAN 02 | 10102016 | 2016 | 635 | 5 | 6350 |
| 19 | 1 | 1251 | Maris Vicente | LOAN 03 | 12232016 | 2016 | 13225 | 5 | 13225 |
| 20 | 1 | 1252 | Natashia Maag | LOAN 02 | 2022016 | 2016 | 14613 | 5 | 14613 |
| 21 | 1 | 1253 | Odis Douglass | LOAN 03 | 12052016 | 2016 | 7663 | 5 | 7663 |
| 22 | 1 | 1254 | Letitia Gambrel | LOAN 11 | 6022016 | 2016 | 7063 | 5 | 7063 |

Loan Testing - ADS Copy Paste

Ready Average: 1383.5 Count: 301 Sum: 415050

### 3.4.5. Audit Apps Prototyping

The audit apps prototyping phase consists of two elements. The first element comprises the preprogramming of the four audit tests as Microsoft Access queries, the single audit app. The second element is the audit app that executes a combination of audit procedures and consists of the training of the UiPath[9] RPA software to 1) automatically extract information from several files, 2) import it into the Loan Testing ADS template, 3) select employee loans for the current year under audit, 4) transfer the information from the

[9] UiPath is one of the market leaders in the RPA industry. Please refer to https://www.uipath.com/platform.

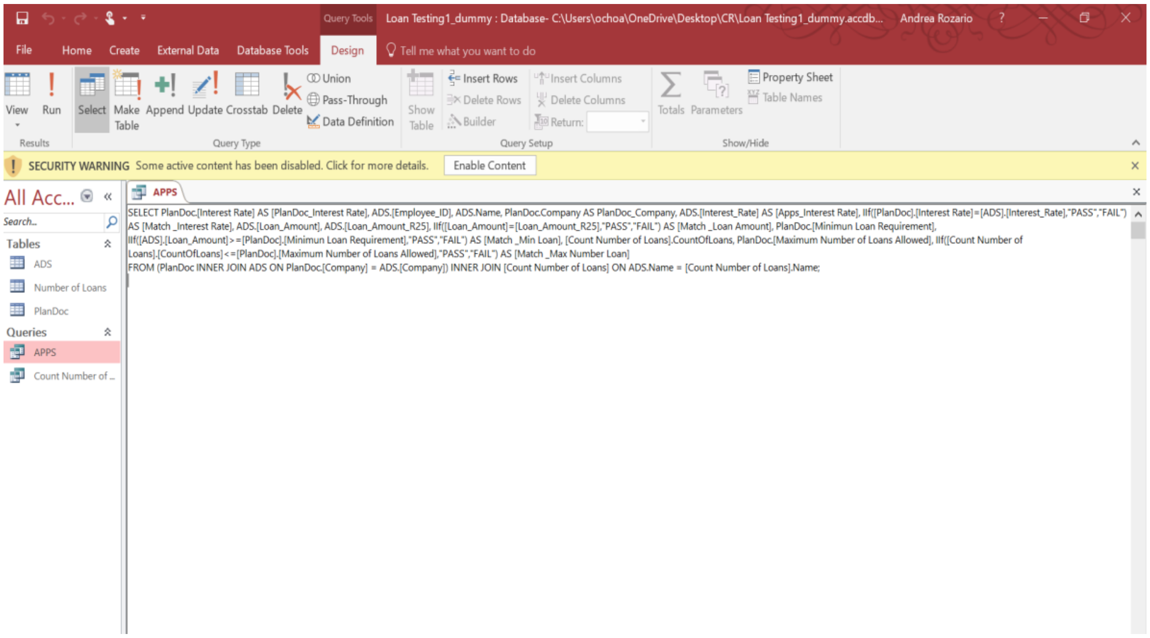
"Loan Testing – ADS Copy Paste" template into the "ADS" Microsoft Access table, 5) and execute the preprogrammed audit queries.

The first element within the audit apps prototyping section consists of the preprogramming of audit tests as audit queries in Microsoft Access. "IF-THEN" logic was used to program the following four audit test as Microsoft Access queries:

1) Ensure that the employee loan amount reported by the company matches the loan amount disbursed by the company
2) Ensure that the interest rate for the employee loan reported by the company matches the interest rate reported on the company's employee loan agreement
3) Ensure that the employee loan amount satisfies the minimum loan amount requirement as per the company's employee loan agreement
4) Ensure that the employee had the permissible number of loans as per the employee loan agreement

Figure 17 describes the functions that were entered to program the automated audit queries. Once audit data is loaded into the ADS Microsoft Access Table, RPA executes the "Run Query" function for audit queries to automatically execute the preprogrammed audit tests.

**Figure 17: Microsoft Access Audit Queries**




Figures 18, 19, and 20 illustrate the actions of the RPA audit app. RPA opens the company files, reads the information within the files and writes the information from the files into the “Annual Loan Balance”, and “Check Register” tabs within the ADS template. Figures 20 and 21 also depict the action to read the data in the “Loan Testing – ADS Data Prep” tab in order to filter for the current year employee loans.

**Figure 18: Open and Read company report**

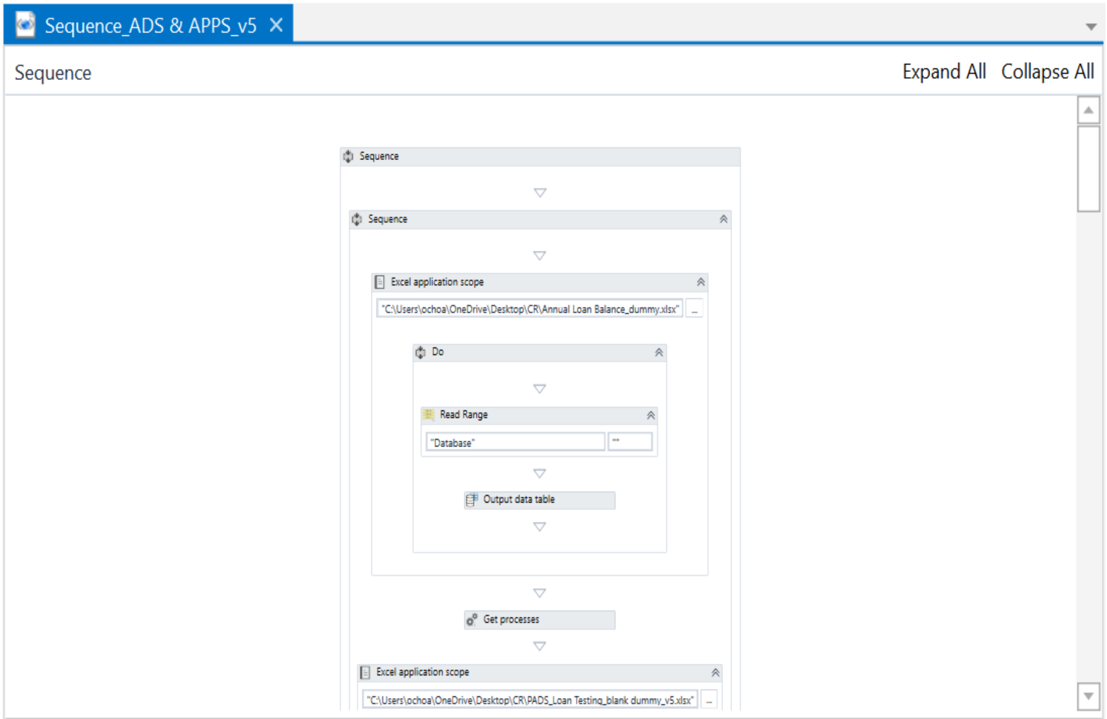



**Figure 19: Write company file and Read subsequent company report**

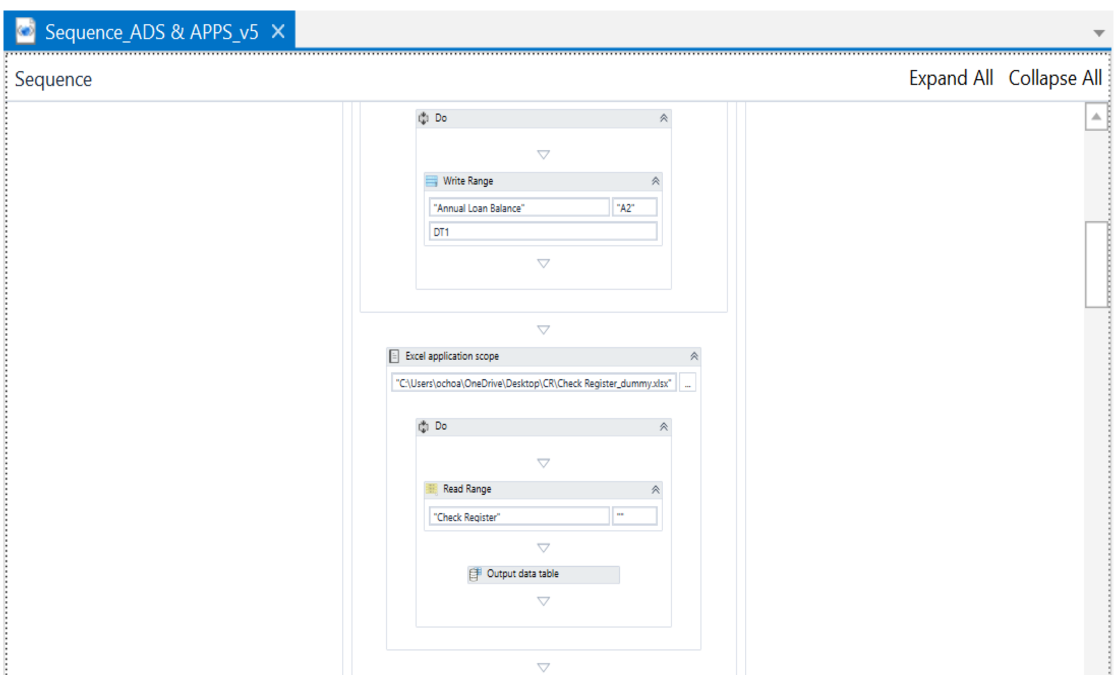
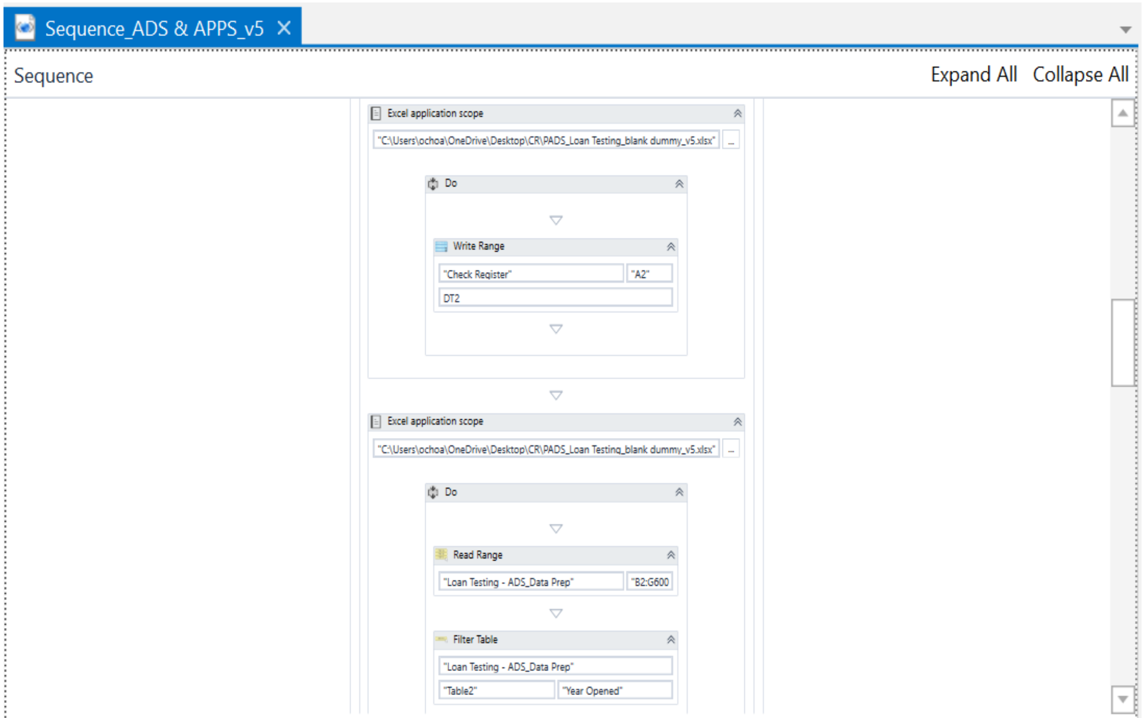


**Figure 20: Write company file and Read and Filter ADS information for current year**




Figures 21 and 22 present the actions of RPA to read information from the “Loan Testing – ADS Data Prep” tab and write it to the “Loan Testing – ADS Copy Paste” tab. Figure 19 also illustrates the RPA tasks to connect to the Microsoft Access application and read information from the “Loan Testing – ADS Copy Paste” tab.

**Figure 21: Filter ADS information for current year, cont'd, and Read ADS information**

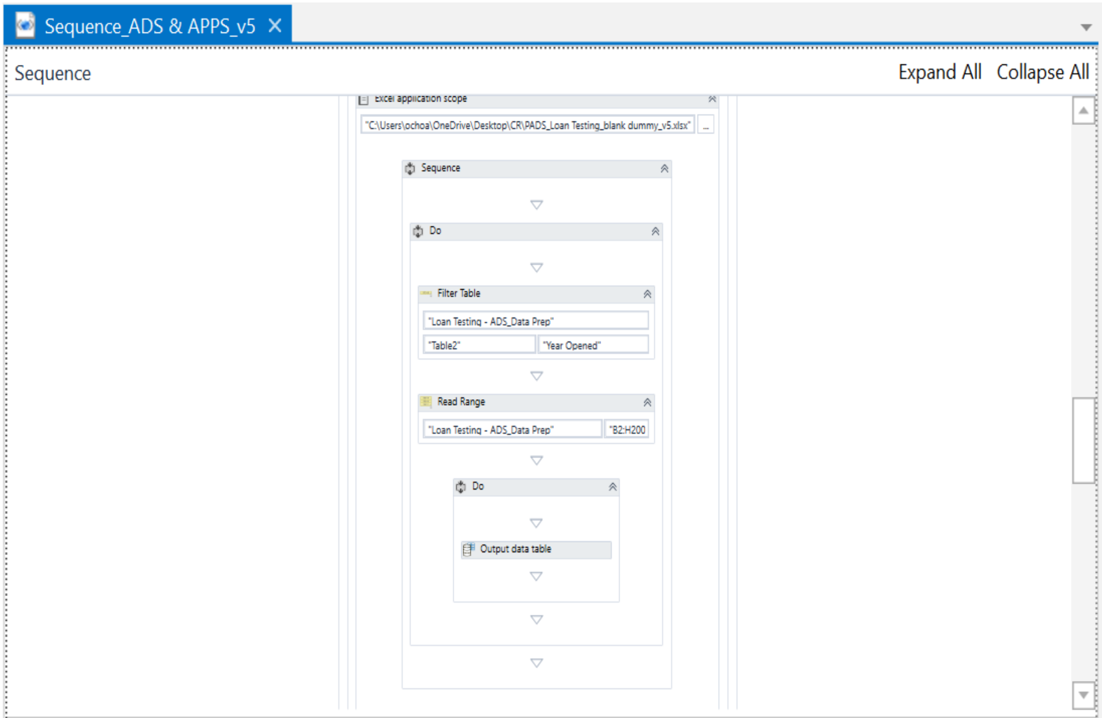



**Figure 22: Write ADS information to integrated ADS structure, Connect to Microsoft Access Program, and Read ADS information from integrated ADS structure**

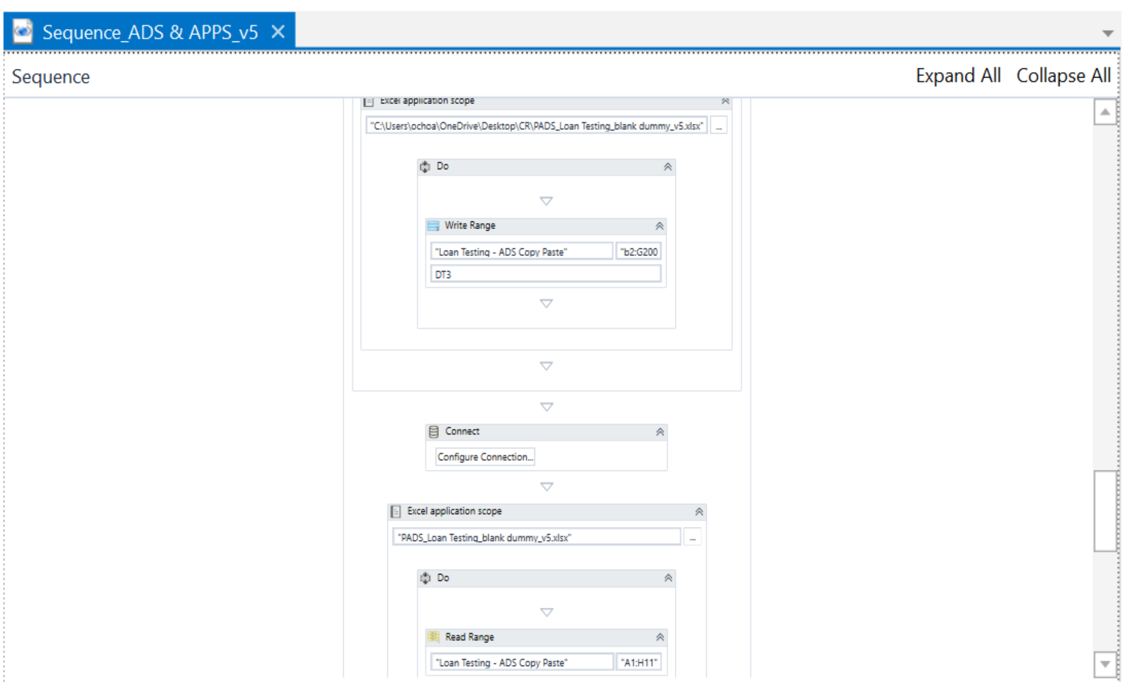
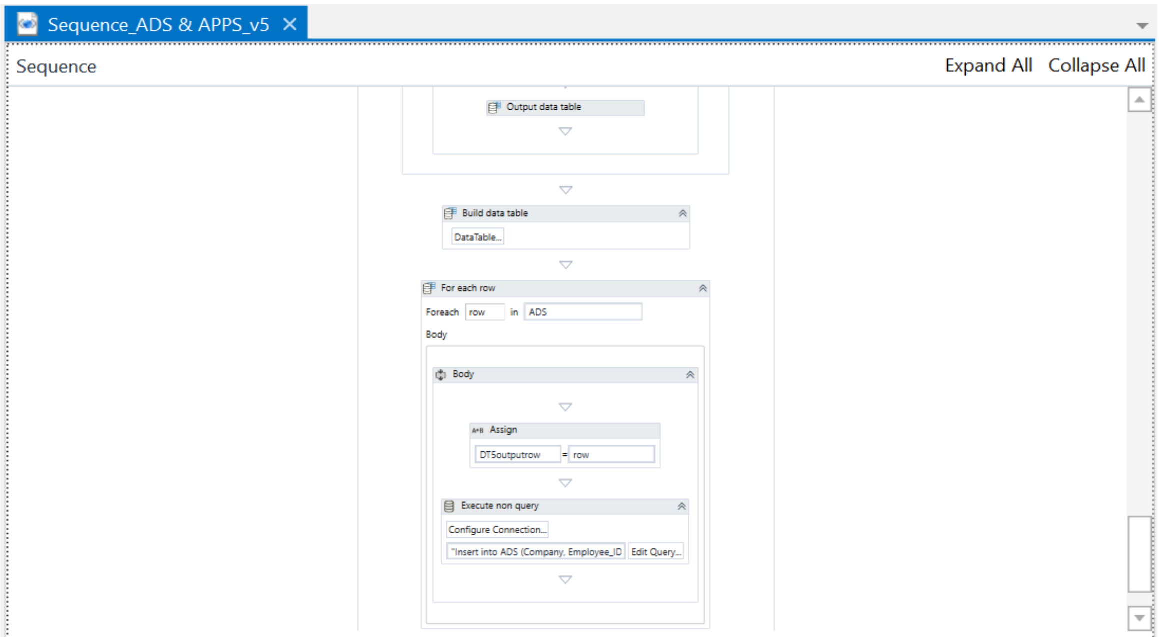


Figure 23 presents the RPA actions to write information from the integrated ADS structure to the “ADS” Microsoft Access table.

**Figure 23: Write ADS information to Microsoft Access Table**




Finally, Figures 24 to 27 illustrate RPA actions to automatically execute the preprogrammed audit queries. Specifically, the RPA opens the folder where the Microsoft Access database is located, opens the database and the ADS table within it, and the audit queries.

**Figure 24: Open folder to find Microsoft Access Database**

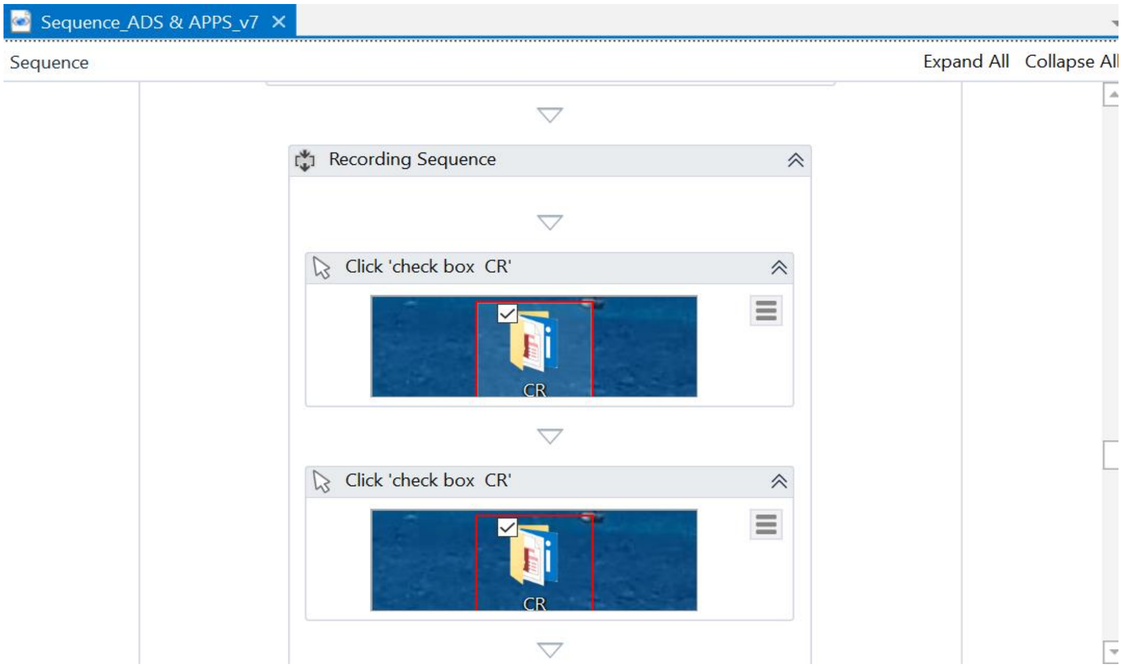



**Figure 25: Open Microsoft Access Database Open Microsoft Access Database**

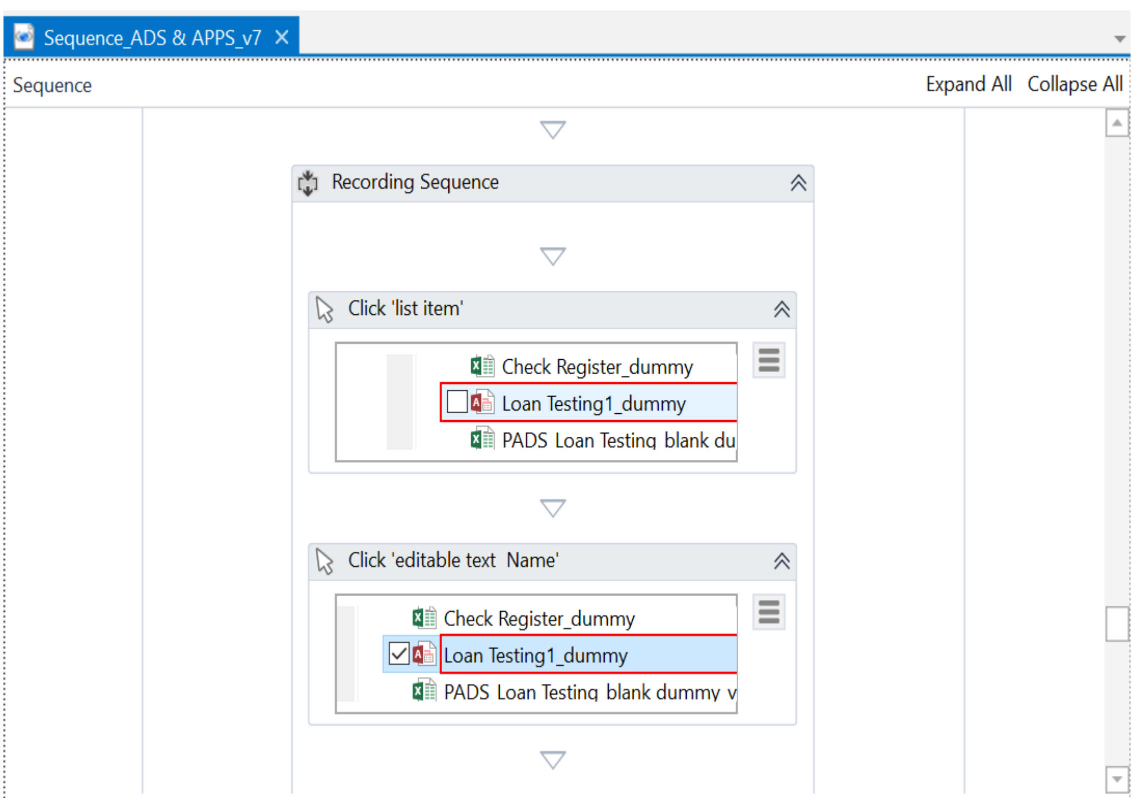


**Figure 26: Open Microsoft Access "ADS" Table**

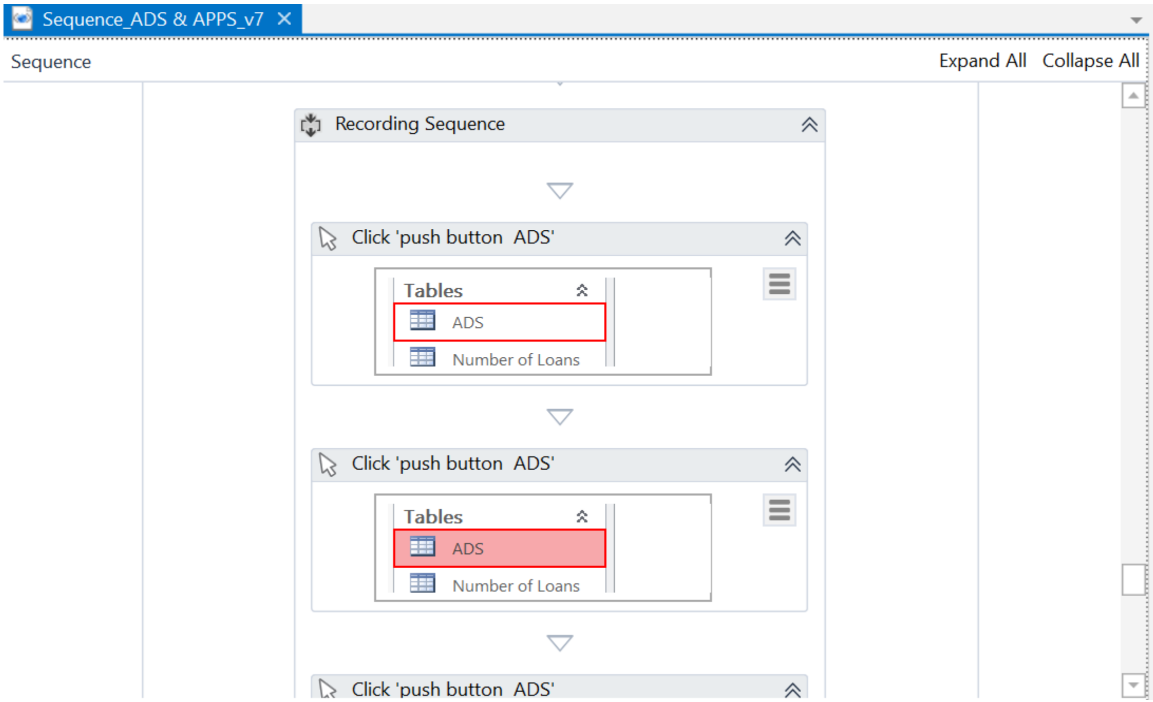
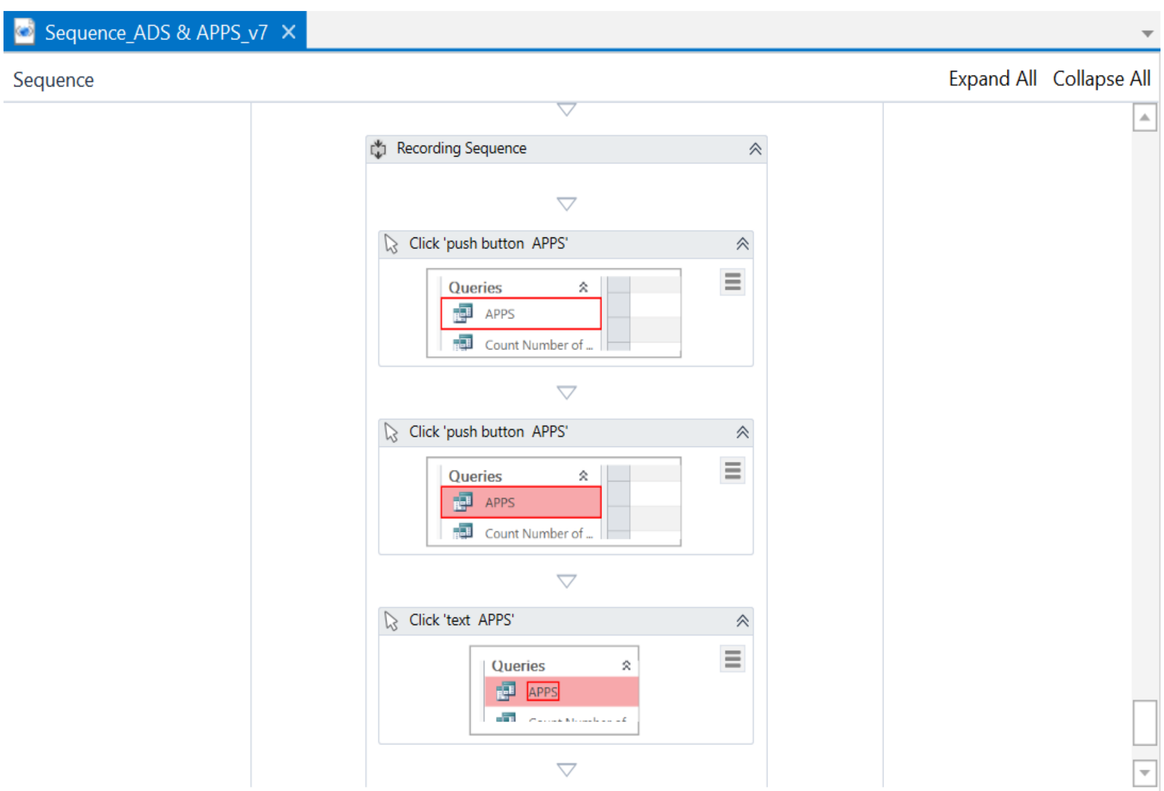
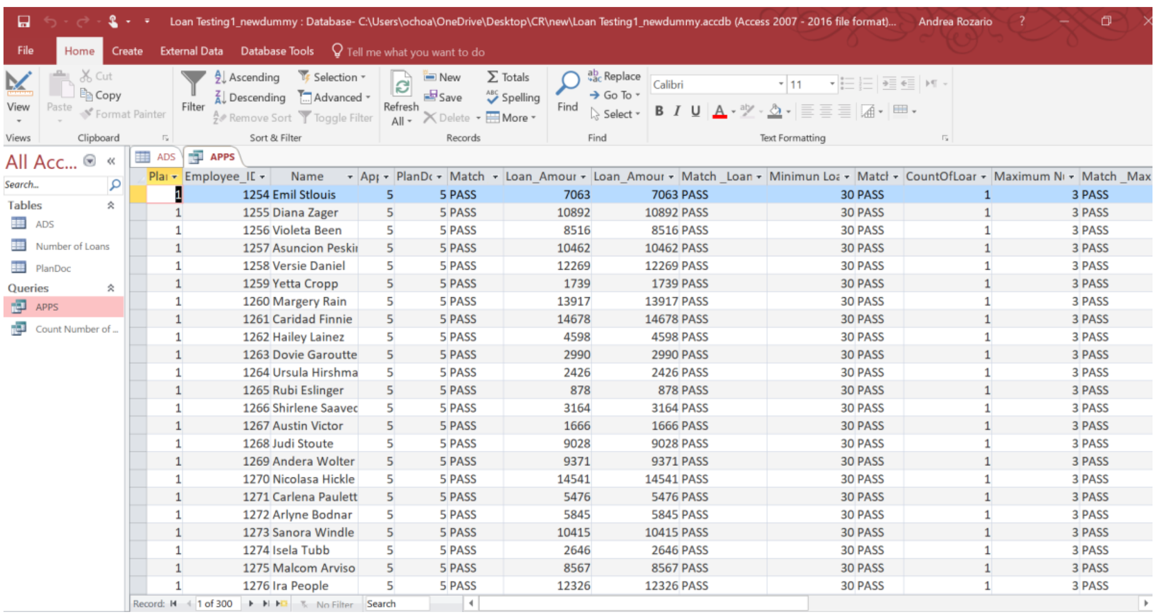



**Figure 27: Open Microsoft Access Queries to execute audit tests**



As illustrated above, RPA is an audit app that is capable of achieving near end-to-end process automation. In particular, for the sub-process presented in this paper RPA can automate audit procedures to 1) collect audit evidence, 2) prepare the ADS, 3) import it into the Microsoft Access program, and 4) execute automated audit tests. Figure 28, presents the results of the Microsoft Access audit apps that are executed by RPA. The results indicate that the population of employee loans passed the automated audit tests.

**Figure 28: Results of Microsoft Access Audit Apps**



| Plai | Employee_IC | Name | Apı | PlanDc | Match | Loan_Amour | Loan_Amour | Match _Loan | Minimun Loa | Matcl | CountOfLoar | Maximum Ni | Match _Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1254 | Emil Stlouis | 5 | 5 | PASS | 7063 | 7063 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1255 | Diana Zager | 5 | 5 | PASS | 10892 | 10892 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1256 | Violeta Been | 5 | 5 | PASS | 8516 | 8516 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1257 | Asuncion Peskii | 5 | 5 | PASS | 10462 | 10462 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1258 | Versie Daniel | 5 | 5 | PASS | 12269 | 12269 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1259 | Yetta Cropp | 5 | 5 | PASS | 1739 | 1739 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1260 | Margery Rain | 5 | 5 | PASS | 13917 | 13917 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1261 | Caridad Finnie | 5 | 5 | PASS | 14678 | 14678 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1262 | Hailey Lainez | 5 | 5 | PASS | 4598 | 4598 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1263 | Dovie Garoutte | 5 | 5 | PASS | 2990 | 2990 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1264 | Ursula Hirshma | 5 | 5 | PASS | 2426 | 2426 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1265 | Rubi Eslinger | 5 | 5 | PASS | 878 | 878 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1266 | Shirlene Saavec | 5 | 5 | PASS | 3164 | 3164 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1267 | Austin Victor | 5 | 5 | PASS | 1666 | 1666 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1268 | Judi Stoute | 5 | 5 | PASS | 9028 | 9028 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1269 | Andera Wolter | 5 | 5 | PASS | 9371 | 9371 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1270 | Nicolasa Hickle | 5 | 5 | PASS | 14541 | 14541 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1271 | Carlena Paulett | 5 | 5 | PASS | 5476 | 5476 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1272 | Arlyne Bodnar | 5 | 5 | PASS | 5845 | 5845 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1273 | Sanora Windle | 5 | 5 | PASS | 10415 | 10415 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1274 | Isela Tubb | 5 | 5 | PASS | 2646 | 2646 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1275 | Malcom Arviso | 5 | 5 | PASS | 8567 | 8567 | PASS | 30 | PASS | 1 | 3 | PASS |
| 1 | 1276 | Ira People | 5 | 5 | PASS | 12326 | 12326 | PASS | 30 | PASS | 1 | 3 | PASS |

The results of the UiPath RPA software and Microsoft Access queries provide evidence that the proposed RAPA framework can be used as a guide to technologically reframe the loan testing audit sub-process. Using this framework to automate a micro audit process, such as the one described in this research study, suggests that the framework could potentially be used to redesign other micro, or macro, audit processes using RPA. The results suggest that targeting the loan testing audit sub-process and redesigning it using

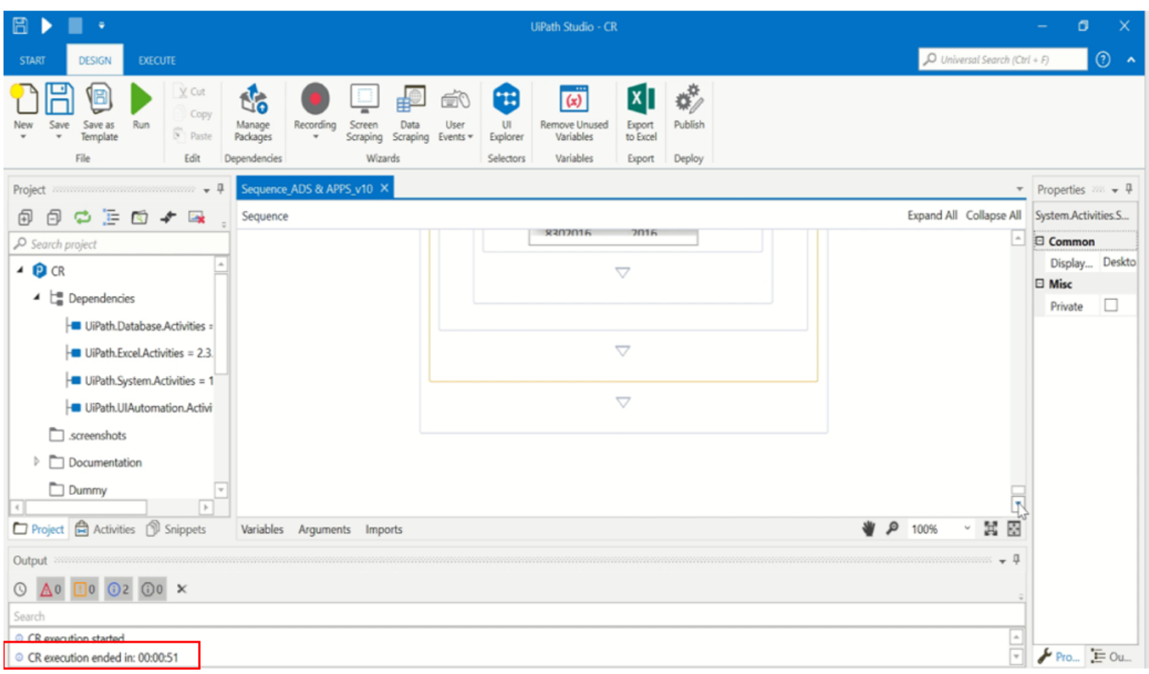
RPA can produce a well-orchestrated audit approach. A demonstration of the described near end-to-end process automation can be viewed in the following video:

RPA Video_Loan Testing_short.mp4

### 3.4.6. Feedback and Evaluation

Feedback and evaluation is a fundamental phase of the framework as it can help assess the value of RAPA. The public accounting firm is in the pilot implementation stages of RAPA and while its value is yet to be tabulated, preliminary assessments of efficiency and effectiveness can be conducted. For RAPA to be more efficient, it should conduct the work in less time than an auditor. Figure 29 illustrates that RAPA spends 51 seconds executing both evidence collection activities and audit tests for the 300 records that need to be verified.

**Figure 29: Execution Time of RAPA**

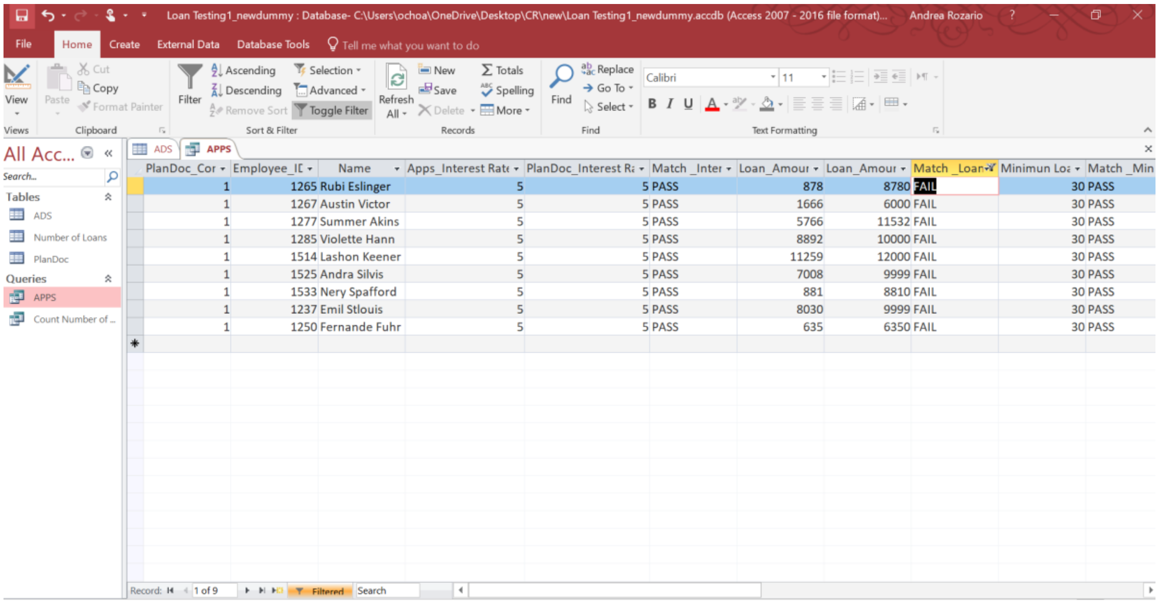


RAPA would clearly spend less time executing audit tasks than auditors but in addition, it could execute these tasks for a large number of audit engagements. For example, a conservative assessment would be that 30 minutes (1,800 seconds) of an auditors' time is allocated to perform loan testing for one audit engagement. Under the RAPA approach, loan testing can be performed for approximately 35 audit engagements (1,800 seconds divided by 51 seconds) in the same amount of time[10]. As a result, it can scale to support a large number of audit engagements.

To evaluate the effectiveness of RAPA, 9 errors were seeded into the data to overstate the loan amount that was disbursed. As loans reported in the "Annual Loan Balance" document (which is used to update the general ledger) are matched to the loans disbursed per the "Check Register" document, overstating loan amounts in the latter can result in the understatement of receivables that are due to the company from its employees. Consequently, loan receivables may not be included in the financial statements at the appropriate amount raising a concern for their valuation (Louwers et al. 2018). Figure 30 presents the output of RAPA and demonstrates that all the seeded errors were detected. Accordingly, RAPA could be used to detect accounting errors on the full population of accounting records near real-time and thus help auditors more precisely measure the risk of the inappropriate valuation of receivables. Importantly, as RAPA is scalable, it can detect accounting errors for more audit engagements in less time than it would take auditors if they were to manually do it.

[10] Even if 20 minutes, or 10 minutes, were allocated to the performance of tests within the loan-testing audit sub-process, the benefits from RAPA are undebatable as it could perform these tests for 23, or 11, audit engagements, respectively.

**Figure 30: Output of RAPA with Seeded Errors**




The illustrated measures have merit for demonstrating the potential cost savings that RAPA would produce as a result of decreased execution time and timelier error detection. Even if this approach is marginally less costly than the traditional audit approach, it can still lead to a good return on investment. In addition, by having the capability to detect accounting errors in near real-time for several audit engagements and by offering auditors the opportunity to focus on higher risk areas, RAPA can potentially lead to enhanced audit effectiveness.

## 3.5. Conclusion

More so than ever, technology is perceived as a disruptor in the auditing profession by challenging existing auditing methods that do not parallel a digital and real-time economy. Various technological audit tools have been studied and applied by academics and audit professionals, however, how auditors can leverage these tools to form a systematic audit process necessitates further consideration. To fill this research gap, this study proposed a framework to redesign an audit process using RPA to achieve near end-to-end process automation. With RPA and process redesign, auditors have the ability to seamlessly integrate evidence collection and processing activities and apply it to test the full population of records. Hence, the proposed framework can enable auditors to fully automate structured audit tasks.

The feasibility of the framework was examined by implementing it to the loan testing audit sub-process of a public accounting firm. This audit sub-process was selected as one of the key candidates for automation since it comprised several deterministic audit rules. Furthermore, from the understanding of the process, the RAPA team strategized to utilize RPA to facilitate the execution of audit evidence collection activities, the transfer of standardized audit evidence to Microsoft Access, and the execution of preprogrammed Microsoft Access queries. To proceed with the goal of automation, the RAPA team developed and implemented an ADS for the loan testing audit sub-process. Finally, RPA and Microsoft Access audit apps[11] were programmed and tests were conducted to evaluate the efficiency and effectiveness of RAPA. Collectively, the results provide evidence of the usefulness of the framework in applying RAPA to an audit process.

This study contributes to the literature by proposing and validating a framework for RAPA, to achieve near end-to-end audit process automation, however, it has a few limitations. First, the framework is applied to an audit sub-process to illustrate its applicability. The loan testing audit sub-process constitutes a minute portion of the holistic EBP audit process, which is composed of a series of other sub-processes. It would be

[11] The audit app supported by UiPath RPA software collects and prepares audit evidence, and executes automated audit tests while the audit app that contains preprogrammed audit tests was supported by Microsoft Access (other applications such as CaseWare IDEA, could be used to support automated audit tests).

interesting for future research to expand the application of this framework to an audit process, such as that of EBP audits, or even revenue audits. Second, the automation methods examined reflect simplistic functions that do not require auditor judgment, future research should investigate the implementation of this framework with audit tasks that are semi-structured or unstructured to further explore the benefits of RAPA with cognitive capabilities on the audit process. Third, while the measures for efficiency and effectiveness of RAPA demonstrate its untapped opportunity, a method of parallel implementation can provide further insight into the value of RAPA. For example, future research can design an experiment to investigate whether RAPA leads to enhanced auditor judgment and cost savings across various audit engagements. Finally, using technology to redesign an audit process can lead to the problem numerous notable items that are generated from full population tests, future research should consider expanding the proposed framework to include an approach for addressing these notable items using RPA, especially in a continuous auditing environment.

# Chapter 4. Reengineering the Audit with Blockchain and Smart Contracts

## 4.1. Introduction

Advances in technology have created a ‘real-time’ world in which economic transactions are processed electronically and immediately. Although businesses have adapted to this complex electronic world, the financial auditing paradigm remains in the status quo and continues to reflect a retrospective audit framework. Auditors continue to audit reactively and according to established archetypes, and yet they are still tasked with extending their confidence to the financial statements and protecting the public interest in regards to the financial statements in increasingly complicated and risky environments. It is not surprising that there is an expectation gap between the information auditors provide to financial statement users and the information expected by these users. This gap is the difference between what financial statement users demand, in terms of timely, relevant and reliable information and what they may receive as a result of the audit. Moreover, PCAOB inspection findings, in the areas of revenue for example, continue to illuminate an additional expectation gap between the procedures that auditors actually perform and the procedures they are required to perform in accordance to audit standards and regulations (PCAOB 2016). In both cases, the expectation gap represents a unique opportunity for improvement not only in audit quality, but also in the auditor’s ability to respond to client risks in a rapidly changing technological environment. In response to the challenges facing auditors, this paper proposes an external audit blockchain supported by smart audit procedures aimed at improving audit quality and meeting the information and performance demands of stakeholders.

Blockchain and smart contracts have great potential to improve business process quality. First coined in 2008 as the Bitcoin network, blockchain is essentially a distributed linear database that protects the integrity of its information with cryptography (Nakamoto 2008). Since blockchain provides a tamper-proof audit trail, which can be fused with smart contracts to autonomously execute tasks on behalf of human users (Szabo 1997; Kozlowski 2016), it is increasingly gaining popularity among business entities and audit firms. Various types of business entities are exploring the numerous blockchain applications that can improve efficiencies across the different components of the value chain. For example, blockchain applications for securing medical records, and supply chain provenance are among the many use cases of this technology[1].

Similarly, blockchain is gaining momentum in the public accounting industry. Although a relatively new technology, public accounting firms are finding it important to leverage blockchain technology to provide auditing and assurance services. For example, Deloitte was one of the first to successfully audit blockchain protocol[2], whereas PwC and EY have successfully developed auditing tools specifically for auditing blockchain transactions. Notably, PwC recently began to offer continuous auditing software to audit transactions on private business blockchains,[3] and EY developed the EY blockchain analyzer, which is capable of extracting transactions from multiple blockchain ledgers[4].

[1] See: https://medium.com/@matteozago/50-examples-of-how-blockchains-are-taking-over-the-world-4276bf488a4b
[2] See: https://www.ccn.com/big-four-giant-deloitte-completes-successful-blockchain-audit/
[3] See: https://www.pwc.com/us/en/about-us/new-ventures/pwc-blockchain-validation-solution.html
[4] See: https://www.ey.com/en_gl/news/2018/04/ey-announces-blockchain-audit-technology

Despite tremendous technological disruption in the last decade[5], the audit paradigm does not yet parallel the digital business world in adopting the use of new technologies as part of a change to methodology. Alles (2015) suggests that the audit clients' use of advanced technologies is likely to be the driver of adoption of such technologies by auditors. Blockchain and smart contracts technologies can potentially help the current audit framework to evolve by changing the way audit evidence is collected, analyzed and disseminated. By failing to take advantage of blockchain and smart contract technologies, the audit client's digital environment and associated risks will continue to outpace the effectiveness of the auditor's procedures. This will no doubt result in additional audit failures due to a declining trend in the auditor's ability to adequately assess risk and in overall reduction in audit quality. As a result, it is imperative to examine the extent to which blockchain and smart contracts can disrupt the financial statement auditing paradigm. This paper proposes a conceptual framework for an external audit blockchain in which smart contracts, referred to as "smart audit procedures" hereafter, can autonomously execute audit procedures and disclose audit procedures' results to participating users near real-time.

This paper extends the Dai and Vasarhelyi (2017) discussion on the possible applications of blockchain and smart contracts to transform auditing. However, several additional distinctive perspectives are also established. First, in this paper an external audit blockchain, which is supported by smart audit procedures, is proposed. The proposed blockchain can serve as an unified platform to enhance audit effectiveness and audit reporting. Second, this paper maps the characteristics of blockchain that can enhance audit evidence to the requirements of audit evidence described in PCAOB auditing standard

---

[5] See: http://usblogs.pwc.com/emerging-technology/rise-robotics-ai-infographic/

1105 (PCAOB 2010d). Furthermore, novel functions for the PCAOB, such as becoming the validators of smart audit procedures[6] and the reviewers of the results of these procedures near real-time, are proposed in an effort to support its evolving initiatives for improving audit quality (PCAOB 2018). Finally, this paper envisions the evolution of the financial audit paradigm by presenting a holistic audit framework that maps assertions to on and off-the-blockchain audit procedures and discusses issues related to the application of blockchain and smart contracts. Taken together, this paper offers useful insights into the potential use of blockchain and smart contracts by auditors to reduce the expectation gap dilemma.

Additionally, this paper uses design science research (DSR) methodology, the science of creating purposeful artifacts (Hevner, March, Park, and Ram 2004), as an overall guide for the discussion of the proposed audit methodology. Peffers et al. (2007) propose the following steps for design science research:

1) Problem identification and motivation,
2) Define objectives of a solution,
3) Design and development of an artifact,
4) Demonstration of the solution,
5) Evaluation of the solution, and
6) Communication of the results

It is noted that individual DSR studies are not expected to comprise all six activities (Peffers et al. 2007). Likewise, in this paper, the first three activities, comprising problem

[6] By becoming the validators of smart audit procedures, the PCAOB can verify that such audit procedures are designed to test specific account assertions as well as significant accounts, which can inherently improve the audit inspection process. Therefore, issues such as the performance of inadequate audit procedures could be mitigated proactively.

identification, defining the objectives of a solution and the design of an artifact, are discussed. As research progresses in the blockchain audit domain, it is expected that research papers concerning the development and evaluation of the design will be produced.

The remainder of this paper is organized as follows. Section 2 identifies the problem and discusses the motivation. Section 3 describes what a blockchain is, how it operates, and its characteristics. Section 4 and 5 define the objectives of the proposed framework. Specifically, Section 4 discusses how blockchain improves the reliability of financial and nonfinancial information and thus the reliability of potential audit evidence and Section 5 describes how blockchain audit evidence and smart audit procedures improve audit quality and narrow the expectation gap. The conceptual framework is presented in Section 6 and Section 7. Section 6 illustrates and describes interlinked blockchain ecosystems, which consist of a business blockchain and the proposed external audit blockchain. Section 7 describes the audit approach on the external audit blockchain. Section 8 presents a series of issues and opportunities for future research. Finally, the last section concludes the paper.

## 4.2. The Expectation Gap Dilemma in the Digital Era

The expectation gap dilemma that exists between both the auditor and the financial statements users and the auditor and the standard setters can be further broken down into an information gap and a performance gap (PCAOB 2016b). The information gap is the result of the information auditors provide to financial statement users versus the information financial statement users expect from auditors. The performance gap, on the other hand, arises from the disconnect between the audit procedures auditors are required to perform per audit standards and the audit procedures that are ultimately performed.

Collectively, these gaps call into question the usefulness of auditing, the audit opinion and the usefulness of audited financial statements.

The primary objective for a financial statement auditor is to provide the user of the financial statements with reasonable assurance that an entity's financial statements are free from material misstatement (Louwers et al. 2018). They lend their credibility and confidence to the financial statements, and with that, ensure that the public interest regarding the financial statements is protected. Users of the financial statements, like investors for example, do not have access to the internal systems and source documentation within an entity, yet they need to be able to rely on the information provided by the entity to make decisions and allocate their resources accordingly. Investors are demanding timely, relevant and reliable information, and they are starting to leverage unorthodox sources of information, such as real-time social media postings, to make financial decisions (Stein 2015). In the age of technological disruption and ever-increasing access to large amounts of information, understanding the information gap between what investors and financial statement users need and want from the auditor and what the auditor actually provides needs to be explored.

Moreover, it is the mission of the PCAOB to "oversee the audits of public companies in order to protect the interests of investors and further the public interest in the preparation of informative, accurate, and independent audit reports"[7], improve audit quality and, in some respects, ensure that auditors satisfy the demands of financial statement users. However, their inspection findings suggest that auditors are deficient in a

[7] See: https://pcaobus.org/About/History/Pages/default.aspx

variety of audit areas including the adequacy of substantive procedures performed, identification and testing of controls and the sufficiency of evidence to support the audit opinion (PCAOB 2016; PCAOB 2017). Therefore, a performance gap exists between the procedures auditors are expected to perform and the procedures that are actually performed and reviewed by PCAOB inspectors. Even if on a small scale, this performance gap undoubtedly illuminates the need for audit activities that can lead to higher audit quality.

It is critical for auditors to be proactive in understanding how disruptive technologies can evolve auditing to satisfy financial statement users demands and aid the PCAOB's continuous initiatives to improve overall audit quality (PCAOB 2017; PCAOB 2018). Some public accounting firms are adapting to business adoption of blockchain technology by exploring how to audit the blockchain protocol or by developing tools to audit the blockchain transactions of their clients[8], [9]. Accordingly, this paper proposes a way to enhance audit quality and meet different user demands in near real-time through the use of blockchain and smart contracts to ameliorate the expectation gap dilemma. In doing so, these technologies could evolve the way that financial statement audits are performed and delivered.

## 4.3. Blockchain for Auditing

Blockchain technology became increasingly popular primarily as a result of Bitcoin virtual currency (Nakamoto 2008). Blockchain is an open distributed ledger that enables users to transact directly with each other without the need of a trusted third party (Gruber 2013; Bryans 2014; Singh 2015). Bitcoin was one of the first publicly known applications

[8] See: https://www.pwc.com/us/en/about-us/new-ventures/pwc-blockchain-validation-solution.html
[9] See: https://www.ey.com/en_gl/news/2018/04/ey-announces-blockchain-audit-technology

of blockchain technology and functions as a secure peer-to-peer payment system. Using the Bitcoin network as an example of blockchain functionality, transactions would be sent directly by the payer to the payee and then broadcast on the Bitcoin network. These transactions are combined into blocks and validated by miners[10] about every 10 minutes through cryptography that combines a hash[11] of the transaction and the digital signature of the user. Upon validation of transactions, blocks are posted and time-stamped in the sequential blockchain ledger and are visible by all the nodes[12].

Key characteristics of the blockchain include decentralization, immutability and accountability. Decentralization is achieved by enabling various nodes (computers) to download the blockchain ledger, where every node has a local copy of the blockchain ledger and a universal view of the transactions. Due to the decentralized nature and the fact that each node has its own copy of the ledger, fraud on the blockchain would be unlikely to occur as participating nodes have access to view blockchain transactions as they are posted. In addition, any transactions that may appear fraudulent or in error would be corrected by appending a transaction adjustment to the blockchain.

Due to the cryptographic mechanism employed by the blockchain, immutability is also achieved. Once a block of transactions has been completed and added to the end of the blockchain, they cannot be reversed. This function prevents the problem of double spending coins from a digital wallet since the cryptography essentially prevents retroactive changes to the blockchain ledger. The cryptography and decentralization attributes provide

---

[10] Miners are nodes on the Bitcoin network that offer their computational resources to solve the hash function, once the hash function is solved, the transaction is posted to the blockchain ledger and visible by all participating nodes.
[11] Hash functions are used to encrypt and store data efficiently and securely. A hash function takes a string of characters, input x, and produces, a fixed-length, output y consisting of random numbers and characters.
[12] Note that a block is complete when all transactions within the block are validated.

the auditor with a tamper-resistant audit trail. Finally, accountability is achieved on the blockchain as the digital signature of the user binds him or her to the transaction enabling the auditor to verify the originator of the transaction. Collectively, these attributes make this technology appealing for accounting and assurance purposes as it provides a secure set of records, near real-time reporting, a robust audit trail and transparency.

More specifically from an auditing perspective, the blockchain can lower the risk of management override. For example, an ERP system by design contains the functionality for super user access where a designated database user can alter records on the database after they have been posted. Blockchain systems on the other hand, do not provide such functionality (Glaser 2017; Ibrahim 2017). In addition, blockchain securely stores records in such a way that the hash of current records contains information from the previous records, which would make it more difficult for users to alter transactions (Glaser 2017; Olsen, Borit, and Syed 2019). Taken together, blockchain can lower the risk of management override and therefore produce more reliable information.

Blockchain is most advantageous when applied to a trustless environment and while the audits of financial statements are a regulatory requirement that should help to maintain trust in the capital markets, audit failures, such as Enron, have raised the concern of whether auditing is meeting such objective. In an effort to restore confidence shortly after the Enron scandal, the PCAOB was created as a mechanism to oversee financial statement audits by ensuring that they are in compliance with standards. When complemented with oversight, audits have the potential to maintain trust in the capital markets, however, more recent audit failures, such as Lehman Brothers in 2008 and Wells Fargo in 2016 suggest that there is opportunity to further enhance confidence amongst investors and other financial statement users. Implementing audit activities on the blockchain could potentially enhance trust in the capital markets by enabling third party monitoring in a secure and timely environment (Alles et al. 2004). First, blockchain can mitigate the risk that audit workpapers will be manipulated[13]. Moreover, by offering audit inspectors the ability to monitor financial statement audits in near real-time, the inspection process can become more proactive and potentially result in more effective oversight activities as audit areas where firms may be deficient can be detected (and perhaps corrected) prior to the issuance of an audit report.

Therefore, combined with the increasing adoption of blockchain across a myriad of industries, the use of blockchain as a source of more reliable information and to securely enable third party monitoring in near real-time can be advantageous to auditing.

### 4.4. Blockchain can Improve the Reliability of External and Internal Audit Evidence

Ensuring that audit evidence collected for audit procedures is sufficient, relevant and reliable is paramount to auditors (PCAOB AS 1105 2010). Sufficiency is not likely to be a challenge in a blockchain environment as auditors would have the ability to test full populations of transactions that are directly extracted from the client's blockchain (for those transactions that are processed on the audit client's blockchain). As a result, auditors would shift their focus to the relevance and reliability requirements of audit evidence. The relevance of blockchain information will likely remain a matter of audit judgment (Brown-Liburd and Vasarhelyi 2015) as auditors would have to determine whether the evidence collected can satisfy audit objectives (e.g. collecting blockchain sales invoices[14] to verify

[13] During the Enron scandal, for example, Anderson auditors destroyed audit workpapers. Refer to: https://www.wsj.com/articles/SB1010695966620300040.

[14] https://gocardless.com/guides/invoicing/blockchain-and-e-billing/

the occurrence of sales transactions can satisfy the audit objective to verify that no fictitious transactions were recorded). With respect to reliability, the blockchain infrastructure has the potential to enhance the integrity of internal and external audit evidence.

In the current business environment, which primarily consists of centralized accounting ledgers, external audit evidence is generally considered more reliable than internal audit evidence as it is less likely for this information to be manipulated by management (PCAOB AS 1105 2010d). However, blockchain characteristics of decentralization, immutability and accountability can enhance the reliability of internal and external audit evidence as financial information, purchase orders[15], invoices and IoT[16] information can be stored on the secure and transparent blockchain ledger.

Since blockchain transactions require reconciliation by participating nodes, before they are posted to the ledger, completeness and accuracy checks are essentially performed proactively. Completeness and accuracy checks are also performed once transactions are posted as participating nodes have access to a universal view of blockchain transactions[17]. In addition, blockchain records are tamper-resistant due to the cryptographic mechanisms that are deployed. These records are protected by code and become irreversible as transaction hashes contain the information of the current transaction and the previous transaction. Finally, the originator of the record can be identified as the hash of the record

---

[15] See: https://www.sofocle.com/procure-pay-process-blockchain-way/

[16] IoT on the blockchain can help overcome security challenges. See: https://www.ibm.com/developerworks/cloud/library/cl-blockchain-for-cognitive-iot-apps-trs/cl-blockchain-for-cognitive-iot-apps-trs-pdf.pdf

[17] The consensus and decentralization infrastructure of the blockchain can help verify the accuracy and completeness of transactions. Validity checks of blockchain transactions is provided by transaction validators on the blockchain, once these validators reach consensus, transactions are posted to the blockchain. In addition, as every node has a local copy of the blockchain ledger they would have the ability to check whether their transactions have been posted, and whether they have been posted accurately.

also contains the user's digital signature. Taken together, blockchain attributes of decentralization, immutability and accountability help improve the integrity of internal and external data. Table 64 summarizes the challenges related to the veracity and variety of audit evidence that are mitigated by the blockchain.

**Table 64: Challenges of Gathering Audit Evidence that are Mitigated by the Blockchain**

| Challenges | Blockchain Attributes | Blockchain Benefits |
|---|---|---|
| **Traceable origins of sources (veracity)** | Decentralization<br><br>Immutability<br><br>Accountability | Data Integrity to improve the **reliability of audit evidence** |
| **Disaggregated data sources (variety)** | Decentralization | One distributed depository for financial and nonfinancial data to improve the accuracy and timeliness of audit procedures and obtain a deeper understanding of the client |

The use of financial and nonfinancial information as audit evidence is known to enhance the accuracy of audit procedures (Brazel, Jones, and Zimbelman 2009). However, the costs of preparing audit evidence from different sources has the potential to exceed its expected benefits as auditors may find it cumbersome to combine information from various sources (Appelbaum 2016). Blockchains have the ability to store audit evidence from a variety of sources, therefore helping to overcome the challenge of aggregating financial and nonfinancial information from internal or external sources.

The decentralized infrastructure of the blockchain also promotes the sharing of information in a more structured and similar format across companies or industries. In this manner, heterogeneous data is aggregated in near real-time into the blockchain distributed ledger and visible to participating users. On the blockchain, auditors would have access to all the reconciled financial transactions between, for example, the auditee (revenue) and its respective customers (payments), providing the auditors with one consistent version of economic transactions. Auditors could also benefit from using IoT information, such as locational data from GPS devices, or temperature data, stored on the blockchain to obtain a deeper understanding of the client's business and risk and to improve the accuracy of their estimates and valuations.

Although blockchain information has the potential to be more reliable than information from an ERP system, it is important for auditors to consider the risks that emerge in a blockchain environment. For example, the private keys of digital wallets can be stolen or lost, and there can be errors in smart contract code. Both of these risks could compromise the reliability of blockchain information.

## 4.5. Blockchain Audit Evidence and Smart Audit Procedures can Improve Audit Quality and Reporting

### 4.5.1 Blockchain Smart Contracts as Smart Audit Procedures

Ethereum, a Bitcoin competitor, has become prevalent in debates relating to the future of blockchain technology (Buterin 2013). The Ethereum blockchain network is a more general application of the Bitcoin blockchain network because it offers users the ability to create and execute a variety of smart contracts. Essentially, a smart contract is a software program that performs actions on behalf of the user based on pre-defined conditions (Szabo 1994). It refers to computer protocol that facilitates the process of engaging in contractual agreements including the enforcement, verification and performance of the terms of a contract (Szabo 1994), and within the blockchain, oversight authority of these contracts is distributed to the participating nodes (Dai and Vasarhelyi 2017). The emergence of blockchain technology has revitalized the concept of smart contracts, thus paving the way for the use of smart contracts as smart audit procedures.

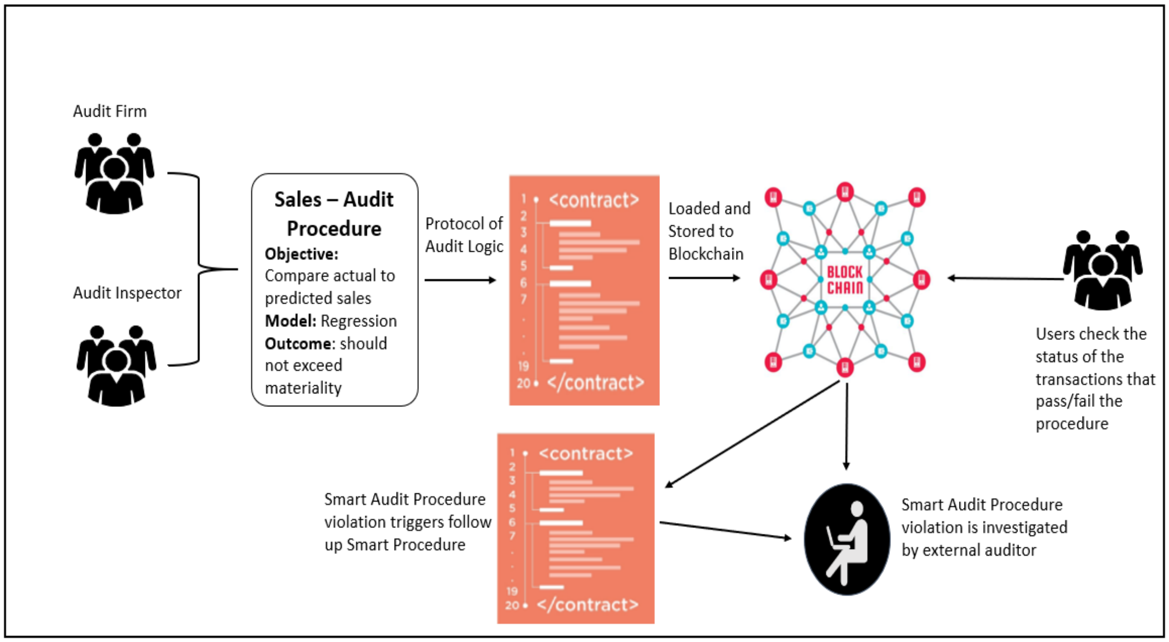
In addition to the variety of external and internal audit evidence from the client's blockchain having the potential to be a more reliable form of audit evidence, smart audit procedures on a blockchain that is operated by the external auditor and that leverage this audit evidence could also improve audit quality and reporting. Smart contracts on the blockchain are not restricted to legal agreements that become digitized, because they can be valuable in other contexts, such as auditing and accounting (Dai and Vasarhelyi 2017; Chou, Hwang, Wang, and Li 2019). In this way, smart contracts can become smart audit procedures. These smart audit procedures are essentially autonomous software programs that execute audit procedures based on pre-defined parameters on the blockchain. Smart audit procedures can mimic the function of software agents[18] that have the ability to analyze audit evidence on behalf of the auditor (Koslowski 2016). Rozario and Vasarhelyi (2018) propose that smart audit procedures can be preprogrammed as 'IF-THEN' rules and loaded to the blockchain set up by the external auditor.

Figure 31 describes an example of a smart analytical procedure to address the risk of material misstatement in sales. In the example, audit logic is translated to computer logic, which is the smart audit procedure, and loaded to the external audit blockchain that the audit firm has set up. The results of the smart audit procedures can then be verified by any blockchain users the audit firm has provided access to. Then, two methods could be

[18] Intelligent agents are computer programs that autonomously perform specific tasks on behalf of the human user (Nelson, Kogan, Srivastava, Vasarhelyi, and Lu 2000; Vasarhelyi and Hoitash 2005).

employed to manage the processing of notable items that are identified by smart audit procedures. First, the auditor could manually investigate the notable items. Alternatively, a follow-up smart audit procedure that prioritizes records that require further investigation could be pre-programmed.

**Figure 31: Smart audit procedure for sales (Adapted from Rozario and Vasarhelyi 2018)**




The aforementioned example describes a smart analytical procedure, however, simpler ones can be performed. For example, a smart control procedure can be described as a simplistic rules-based system that checks that the sales order, shipping documents and sales invoice match and take place in the correct order. If there are differences, like with traditional audit procedures, auditors would have to inquire with management and perform additional testing to ensure these differences do not indicate a material misstatement. In addition, follow-up smart audit procedures could also be pre-programmed to handle the notable items that are identified by other smart audit procedures.

### 4.5.2. Blockchain Audit Evidence and Smart Audit Procedures can Improve Audit Quality

Merged with blockchain technology, smart audit procedures have the potential to transform auditing. As discussed in the preceding section, blockchain mitigates challenges associated with gathering audit evidence and potentially improves the reliability of both internal and external audit evidence. For example, if a smart audit procedure processes unreliable information from the client's ERP system, the results could be misleading and potentially cause auditors to over or under estimate audit risk (Appelbaum 2016). However, the infrastructure of the blockchain, including decentralization, immutability and accountability characteristics, has the potential to substantially improve the reliability of financial and nonfinancial data, which could come from internal and external sources. Therefore, the variety of more reliable data on the blockchain could enhance the effectiveness of smart audit procedures by more accurately capturing the real risk of material misstatement.

In addition, blockchain audit evidence may enhance auditor judgment in a way that was not possible before. The auditor would now have access to a variety of immutable data that enhance the ability to assess risk in new and innovative ways. This variety of data (e.g. locational data from GPS devices, temperature sensors, weather data, etc.) can be connected to the blockchain, and new and reliable datasets can be generated. These datasets can enhance the accuracy of smart audit tests and auditors' understanding of the client's business environment. For example, for audit client's that deliver perishable food items to a customer's designated location, GPS locational data and temperature data on the blockchain can capture the exact time, date, place and temperature of these items providing the auditor with more visibility of the revenue process, including reasonableness of some of the client's estimates regarding spoilage, returns, etc. Incorporating these less traditional, nonfinancial predictors into a smart analytical model for sales could substantially improve the predictive power of the model (Yoon 2016) and provide new insights as to potential risks that may arise as a result of the client's business environment. For instance, temperature data on the blockchain can directly capture whether food items that are in transit are damaged. In this manner, nonfinancial information on the blockchain enables auditors to obtain deeper insights into risks that may lead to misstatements in revenue and other significant accounts such as inventory, accounts receivable and accounts payable.

Smart audit procedures that can automate manual and repetitive audit tasks that do not require audit judgment offer auditors the opportunity to focus resources on higher risk areas and thus improve audit quality. These high risks areas could include, but are not limited to, the analysis of notable items that are generated by smart audit procedures or the analysis of management's fair value assumptions.

#### 4.5.3. Blockchain Audit Evidence and Smart Audit Procedures can Improve Audit Reporting

The traditional audit paradigm is backward-looking and reflects a retrospective assurance model (Chan and Vasarhelyi 2011), because an opinion on the audited financial statements is issued at a point in time, several weeks after the occurrence of financial events. Therefore, the usefulness of annually audited financial statements in a modern world where financial statement users base their decisions on information that is available near real-time is questionable (Vasarhelyi and No 2017; IAASB 2016; Rozario and Vasarhelyi 2018).

With smart audit procedures, audits can naturally transition to a proactive audit model. Proactive audits have the potential to improve audit quality by detecting material misstatements at different points in time and by providing timelier and more transparent information to financial statement users (AICPA 2015). In addition, by executing smart audit procedures that use client blockchain information on the external audit blockchain, which is discussed in the following section, the reliability of audit evidence is preserved since management does not need to provide data to the auditor as they would directly extract information from the client's blockchain. This is important as audit procedures generally incorporate extracted information from the client's ERP system which could be subject to management manipulation[19].

Equally important, blockchain-based smart audit procedures can support the PCAOB's evolving initiatives in promoting audit quality (PCAOB 2017; PCAOB 2018) by enabling close monitoring of the audit firm's process near real-time. This idea of leveraging technology to "guard the guards" (i.e. auditing the auditors) is not a new concept. Alles, Kogan, and Vasarhelyi (2004) proposed a "black box log file", which could enable third party monitoring. This paper expands on this concept by proposing that blockchain and smart contract technologies could be implemented to securely store audit procedures that could be visible by relevant external parties that oversee financial statement audits, such as the PCAOB.

By closely monitoring the audit firms' process to address the risk of material misstatement and ultimately opine on financial statements, blockchain-based smart audit procedures offer the ability to perform more proactive audit inspections and potentially

[19] Although, external auditors generally engage IT auditors to test IT controls in order to ensure access to change management controls is restricted and ensure changes in the ERP system are valid, the risk of management making inappropriate changes still exists. Compared to an ERP system, this risk is lower on the blockchain since the ledger is distributed and relevant nodes would insure the data on the ledger is correct.

prevent audit failures. As suggested by Rozario and Vasarhelyi (2018) "with blockchain based smart audit procedures, both auditors and regulators have the opportunity to proactively address areas where audit firms have been deficient." Consequently, regulators can leverage blockchain-based smart audit procedures to improve the audit inspection process (PCAOB 2017).

In the following sections, interlinked blockchain ecosystems and the audit approach on the external audit blockchain ecosystem are described. The interlinked ecosystems consist of a business blockchain and a proposed external audit blockchain. Subsequently, the audit approach on the external audit blockchain ecosystem is proposed by describing a series of risks that relate to the revenue process and the potential smart audit procedures that could be used to address those risks.

## 4.6 Interlinked Blockchain Ecosystems

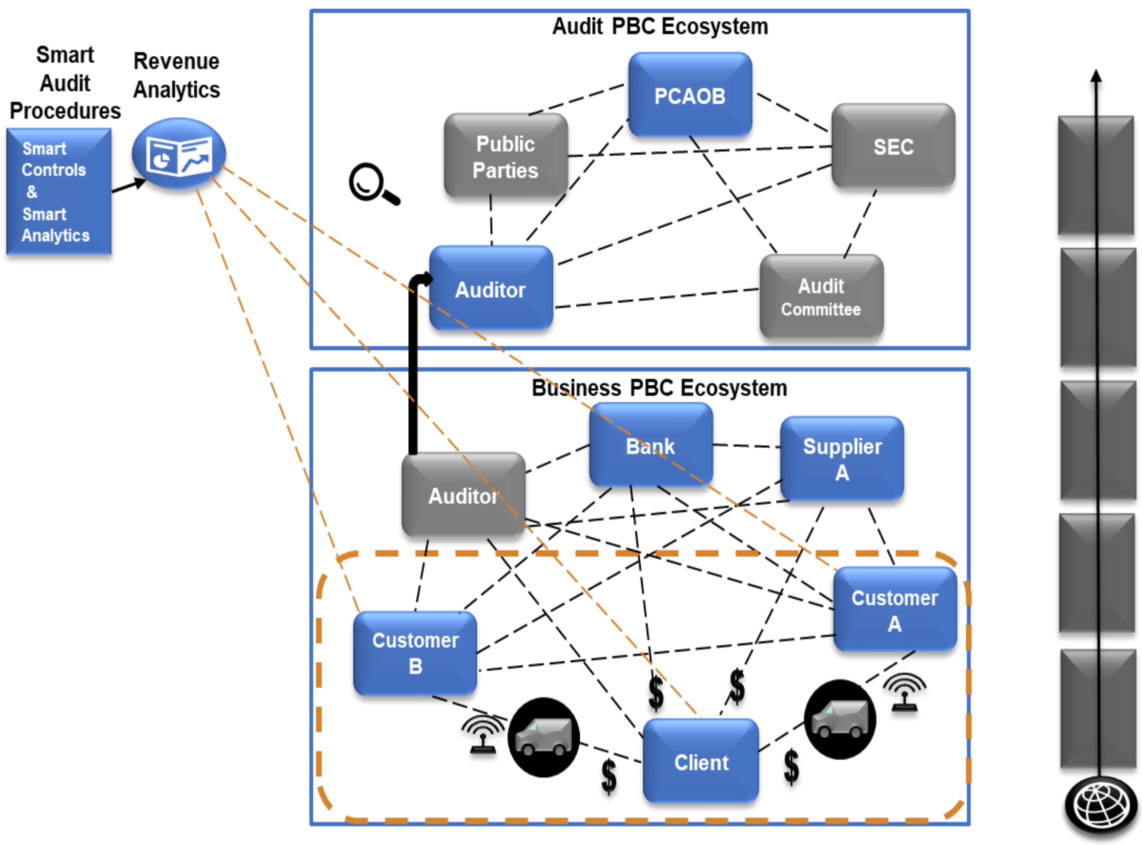
A business ecosystem is defined as "an economic community supported by a foundation of interacting organizations and individuals" (Moore 1996). Therefore, the byproduct of businesses shifting portions of their activities to the blockchain would be a blockchain ecosystem in which several business entities, such as Walmart and its suppliers [20], exploit the benefits of this technology. Similarly, as blockchain and smart contracts provide a unified and secure platform for the collection, analysis, and dissemination of audit evidence, it is plausible for auditing to evolve to be more closely aligned with external users' expectations.

---

[20] Walmart and its suppliers share relevant and reliable information about the quality of food items in a private and permissioned blockchain near real-time See: https://www.nytimes.com/2018/09/24/business/walmart-blockchain-lettuce.html

The evolving blockchain ecosystem is likely to be a multitude of interlinked blockchain ecosystems such as the ones described above. In this type of ecosystem, proactive audits are advantageous, because within it the auditor has the capability to view and extract a variety of reliable information from the client's blockchain without the need for laborious data standardization and then feed this data to pre-defined smart audit procedures that have been vetted by the PCAOB. Interlinked blockchain ecosystems are depicted in Figure 32. These ecosystems facilitate the seamless sharing of relevant and reliable information that is transmitted across active participants on numerous private and permissioned blockchains.

**Figure 32: Interlinked Blockchain Ecosystems**



### 4.6.1. Permissioned Business Blockchain Ecosystem

Private and permissioned blockchains may be appropriate in business and audit settings as they limit the amount of participants (Pilkington 2016). In addition, the responsibilities of these participants are defined in advance by the blockchain network administrator (Peters and Panayi 2015). This type of blockchain is useful as it helps preserve the confidentiality and security of information, and only a restricted amount of participants have access to information and the access is pre-determined. Certain participants may have the permission to send and/or receive transactions, others may have permission to only validate and post transactions, while others may have read-only access.

In the PBC (Private/Permissioned Blockchain) Business Ecosystem, the network administrator, which could be an employee of the audit client, would provide access to the client's blockchain. Accordingly, read and write access to customers A and B, supplier A and the bank can be restricted by the network administrator. The access that is granted to these participants would depend on the agreement between the client and these external parties[21].

Using Figure 32 for illustration, customer A could pay for purchased goods with digital currency and, along with the client, track the shipment of the goods with GPS and temperature sensors to verify the location, quality and time of shipment of the goods. Once the goods are delivered and payment has been satisfied, the audit client may use these funds to satisfy loan covenants with their bank and purchase raw materials from supplier A to

---

[21] While this type of blockchain can be useful in settings where confidentiality of the information is important, its major drawback is that it would be less decentralized than a public and permissionless blockchain, which would enable trust in a trustless environment. Therefore, this infrastructure is more beneficial when it is applied to a context where there is some trust amongst parties, yet trust needs to be enhanced (Glaser 2015; Gockel et al. 2018).

manufacture more inventory. In the blockchain, the aforementioned business events are visible by the audit client and its relevant participants in near real-time, as soon as transactions become posted to the blockchain. Additionally, the auditor can be a read-only node on the blockchain (PwC 2017) and have access to timely and reliable information as they are independent assurors of the audit client and do not directly engage in client business operations.

### 4.6.2. Permissioned External Audit Blockchain Ecosystem

There are benefits to having the external auditor be a node on the audit client's blockchain. The auditor, acting as an independent node on the client's blockchain, would have read-only access to the complete population of internal and external blockchain information, such as sales transactions, legal smart contracts, GPS data, various logs, etc. In addition, the client and auditor could reach the consensus to add the auditor as a node on the client's blockchain as sharing information on this platform would be less disruptive to the client than data extraction in a traditional auditing paradigm since less information would have to be extracted by management. Secondly, the auditor would benefit from being a node on the blockchain as the reliability of potential audit evidence is ensured by blockchain architecture.

There are also several reasons that should be considered in justifying the read-only access the auditor would have on the client's blockchain and why the smart audit procedures should take place on the external audit blockchain and not on the client's blockchain. First, the auditor is not an active participant engaging in transactions with the client's blockchain related participants, they are the independent verifiers of the client's assertions concerning financial statements (Louwers et al. 2018). Second, to preserve auditor independence on the blockchain, it would not be feasible for the auditor to perform smart audit procedures on the auditee's blockchain as it could be perceived as impairment of independence in appearance (Alles, Kogan, and Vasarhelyi 2002). Finally, read-only access is appropriate as it helps maintain the scalability of the private and permissioned blockchain by limiting the auditor to only have access to view and extract information.

Figure 32 also depicts the proposed independent external audit blockchain. Since the auditor is a node on the client's blockchain, they can extract audit relevant information, such as sales transactions, load it to their own blockchain and smart audit procedures could autonomously execute predetermined audit tests. The external audit blockchain ecosystem would consist of smart internal control tests, smart test of details and smart analytics that could enhance audit quality. These smart audit procedures could help detect material misstatements at different time intervals. After setting up their blockchain, the audit firm's IT team, in conjunction with the auditors, can design smart audit procedures and load those procedures to the external auditor blockchain. The PCAOB, as one of the active nodes on the auditor's blockchain, could vet these procedures as this is part of the consensus architecture of a blockchain. Auditors would then load smart audit procedures to the blockchain, but a different node on the blockchain, the PCAOB, could validate the smart audit procedures before they are activated.

One of the objectives of the PCAOB is to oversee audit firms to ensure audits are conducted in accordance with GAAS (generally accepted auditing standards). By including the PCAOB as a participating node on the auditor's blockchain, these regulators are able to provide oversight of the auditing firms prior to the firm's subsequent quality inspection. This can help improve the PCAOB's inspection process and communication between

PCAOB inspectors and audit firms. Currently, the PCAOB inspects large audit firms (firms that audit more than 100 public companies) on an annual basis and small audit firms on a triennial basis. These quality inspections occur after all audit testing has been concluded and the audit opinion has been issued. Hence, blockchain and smart contracts provide a way for the PCAOB to enhance the effectiveness of their inspection process by shifting it from a reactive to a more proactive inspection process that could detect potential audit deficiencies near real-time. We propose that the PCAOB would be appropriately qualified to verify the basic requirements of smart audit procedures, which could include verifying that smart audit procedures are designed to test significant financial statement accounts and specific assertions. Accordingly, the PCAOB would have the ability to validate the smart audit procedures that auditors post to their blockchain, view the results of smart audit procedures and also send inspection related information to auditors about their procedures. Therefore, issues previously discussed, like failure to appropriately test aspects of the audit or performing inadequate procedures, would be mitigated, as the PCAOB would be able to vet the appropriateness of the auditor's procedures prior to their execution. Potentially, to expedite this process, the PCAOB may develop its own smart audit procedures to determine the appropriateness of the auditor's procedures.

Secondary users of the external audit blockchain could also include other audit stakeholders like key investors, the SEC and the audit committee. The external audit blockchain would grant the auditors the ability to send relevant information to the appropriate parties. For example, analyses for revenue including checks that match the sales invoice, shipping, and sales order details and regression analysis that predicts future revenue or future customer churn could be executed by smart audit procedures on the external audit blockchain and sent to users of the financial statements to influence their investing decisions. The deployment of smart audit procedures on the blockchain can help improve audit quality by offering timelier information to financial statement users and assisting with the regulator's quality assessment process.

Smart audit procedures and their results would be visible by participating parties based on the following hierarchical structure.

***PCAOB***

First, smart audit procedures would be visible to the PCAOB for vetting and approval prior to execution, and their results would also subsequently be visible by PCAOB inspectors. The PCAOB could be given read and write access within the external audit blockchain to allow inspectors to view, comment on and make suggestions to the auditor's proposed smart audit procedures. However, it would not be feasible for the entire audit process to go on the blockchain. Since complex audit procedures such as the evaluation of the tax provision and fair value level 3 investments require a high degree of auditor judgment, they are likely to remain off-the-blockchain[22]. Thus, the PCAOB would likely adopt a hybrid approach to inspecting audit procedures for quality and collect documentation for audit procedures on and off-the-blockchain for their inspection process. Nevertheless, the visibility of smart audit procedures and their results on the tamper-proof external audit blockchain can facilitate the auditor's compliance with the PCAOB's

[22] The term off-the-blockchain refers to audit procedures that are not recorded on the decentralized audit blockchain depicted above. An example of on and off blockchain procedures is presented in Table 65.

requirements by preventing the use of inadequate audit procedures and thus help prevent audit failures.

***SEC, Audit Committee, Financial Statement Users***

Secondly, the SEC can be a node on the external audit blockchain and given read-only access to the results of smart audit procedures that have flagged potential fraud or restatement indicators, such information would be highly beneficial to SEC regulators as they perform risk assessment procedures to select potential companies for inspection. Also, having the audit committee as a node on the external audit blockchain would facilitate a more direct and timely method of communication from the auditors to the audit committee. Finally, the general public, including key investors, lenders and key suppliers, would have read-only access to view the results of smart audit procedures at the transactional level and to view the results of testing of the operating effectiveness of internal controls. This will provide them with access to more useful information than that provided by the aggregated nature of the traditional financial statements. Smart audit procedures have great potential to enhance the informational value that is provided by auditors. While a traditional audit opinion for financial statements may not yet be possible on the external audit blockchain, auditor certifications at the transaction and internal control level may provide timely and relevant information to various stakeholders (AICPA, 2015).

## 4.7. Continuous Audit and the Audit Approach on the External Audit Blockchain Ecosystem for Revenue

Vasarhelyi and Halper (1991) developed a Continuous Auditing Process System (CPAS) at AT&T Bell Labs that executed automated analyses, near real-time, on a complete population of records. The automated analyses comprised pre-defined benchmarks, based on auditor defined rules. Each time records exceeded the benchmarks they were flagged by the system and investigated by auditors. Since then, more Continuous Auditing (CA) applications have been developed and implemented by business entities (Kuenkaikaew and Vasarhelyi 2013).

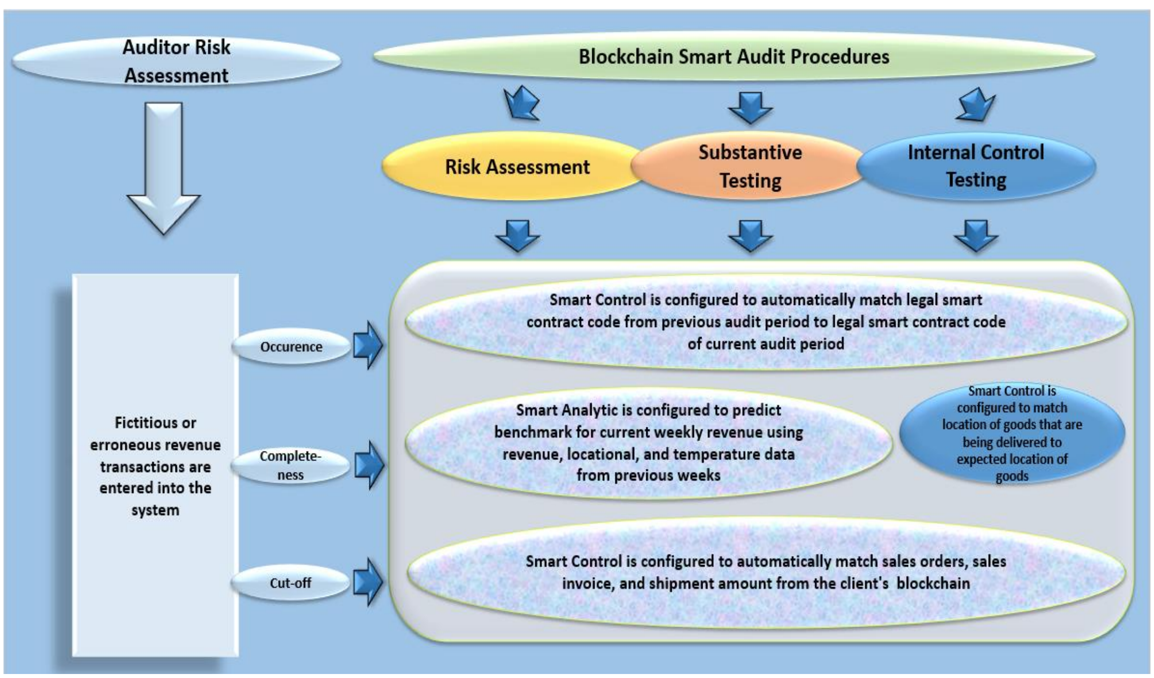
Despite the progressive adoption of this methodology by business organizations, adoption of CA by external auditors is practically non-existent. The major hurdle of CA adoption by external auditors resides within statutory requirements that mandate the auditor to be independent in appearance, which presents a conflict with CA methodologies that must be impounded on the client's computer system (Alles et al. 2002; Bumgarner and Vasarhelyi 2015). A viable solution to the independence problem for CA could be an audit data warehouse (Sigvaldason and Warren 2004). However, the aggregation of various sources of endogenous and exogenous data would remain a challenge. Moreover, for CA to become a reality in financial statement audits, a platform to securely execute automated audit procedures and to disseminate the results of those procedures would be required.

Blockchain can be an important facilitator for CA adoption by external auditors. By collecting a myriad of more reliable data from the client's blockchain and feeding such data to the independent external audit blockchain, where smart audit procedures execute audit tests, auditors insure the integrity of audit evidence and maintain independence while having the ability to provide near real-time assurance and reporting in a real-time economy. It is important to highlight that the blockchain provides a unique platform for both near real-time assurance and reporting, because it would facilitate the autonomous deployment of smart audit procedures.

**4.7.1. Responding to Revenue Risks with Blockchain Smart Audit Procedures**

Figure 33 provides an example of how smart audit procedures can work seamlessly to address multiple phases of the audit for the revenue cycle. This example illustrates how through the use of smart audit procedures executed on an external audit blockchain, the auditor's response to risk can become more proactive. As a result, the external audit blockchain ecosystem for revenue in Figure 33 depicts a variety of smart audit procedures that can be deployed to ensure that the risk of fictitious or erroneous revenue transactions is addressed and that the auditor's detection risk is reduced. The auditor would extract audit evidence from the client's blockchain and send this information to the hash of the smart audit procedure on the external audit blockchain. The smart audit procedure can then automatically perform the pre-defined audit test.

**Figure 33: Blockchain Smart Audit Procedures for Revenue**




To address the risk of fictitious or erroneous revenue transactions, four smart audit procedures could be applied: 1) a smart internal control test that matches the code of the legal smart contracts of the client could be set-up on the external audit blockchain; this procedure can be pre-programmed to match the code of the client's legal smart contracts in the current period under audit to the code of client's legal smart contracts from the prior audit period. By performing this procedure, the auditor can verify that the legal contracts of the client have not changed and if they have, the smart internal control test would flag any contract amendments. In addition, this procedure could also assist auditors in identifying new legal smart contracts that need to be audited. The described test could also serve as a triple-purpose audit procedure as it could assist with risk assessments, provide audit evidence for the existence and rights and obligations assertions over sales revenue and provide evidence of the operating effectiveness of internal controls.

The other smart audit procedures could be applied include: 2) a smart analytic that could execute a regression model based on pre-defined parameters including previous weeks' sales and IoT data that provides the location and temperature of the goods. This smart analytic can facilitate risk assessments and provide audit evidence about management's assertions of the revenue account balance; 3) in conjunction with this smart analytic, a smart internal control test could be automatically configured to use IoT data to match the location and temperature of the goods that are in transit to the expected location and temperature of the goods, as per the contract terms, thus providing audit evidence about the effectiveness of internal controls; and 4) finally, a three-way match smart internal control test could compare sales orders, sales invoices and shipment amounts, this control can serve as a triple-purpose audit procedure.

Collectively, these smart audit procedures would enable the auditor to assess the risk of material misstatement more accurately and in a timely manner. Therefore, as smart audit procedures increase audit efficiency, auditors would be able allocate more time to higher risk audit areas and areas requiring more complex auditor judgment while increasing the informational value they provide to various stakeholders.

### 4.7.2. Holistic Audit Framework for Revenue

The future audit with blockchain is likely to consist of both smart audit procedures on the blockchain and audit procedures off-the-blockchain. Both are needed to effectively conduct audits.

Even though an external audit blockchain ecosystem for revenue supports the automation and reporting of audit procedures, audit judgment will still remain salient. As a result, audit procedures that are unstructured due to a high level of subjectivity and complex judgments would remain outside of the blockchain. When testing using the blockchain, auditors will have to address notable items that smart audit procedures flag as not meeting pre-defined conditions. These items may require further investigation and perhaps paper-based audit evidence that would need additional documentation. The verification of legal paper-based company contracts and complex revenue estimates, such as when revenue is earned on a percentage of completion approach, are examples of additional procedures that may have to be manually verified and documented outside of the external auditor blockchain.

In addition, some accounting information will need to be verified outside of the blockchain environment due to the nature of the information. Information such as month-end adjusting journal entries, for example, would remain off-the-blockchain as these entries are generally related to company specific events (e.g. intercompany inventory transfers or consolidation adjustments). These type of entries do not pertain to routine business operations, and consequently may not satisfy the transaction validation criteria. Therefore, if accounting information exists outside of the blockchain it should be also be verified outside of it, since this technology can provide more reliable information only when there is a single version of transactions (O' Leary 2018).

Although it may not be reasonable to program smart audit procedures on the blockchain for highly subjective audit procedures and for information that exists outside of it, it is still very clear that an external audit blockchain ecosystem has great potential to enhance audit quality and audit reporting. Table 65 provides a holistic representation of an audit approach for revenue, which takes the above considerations into account. Particularly, significant audit risks that are generally important for manufacturing clients and those that carry inventory. Each risk described in the Table 65 is aligned with respective assertions (occurrence, completeness, cut-off) and relevant audit procedures for risk assessment, substantive testing and/or tests of controls. The last column in the table indicates whether the audit procedures would be executed on the external audit blockchain (i.e. smart audit procedures) or if they would remain in the traditional environment, off the external audit blockchain.

**Table 65: Example of Holistic Audit Approach for Revenue Adapted and Modified from Louwers et al. 2018**

| Risk | Assertions | Risk Assessment | Substantive Analytics | Tests of Controls | On BC? |
|---|---|---|---|---|---|
| **Fictitious or erroneous revenue transactions are entered into the system** | **Occurrence** | Cognitive analytics is used to read and analyze terms of pdf legal contracts, such as amount, approvals, contracting parties | | | No |
| | | Rules-based system is configured to automatically match the terms of legal contracts to the terms in legal smart contracts | | | No |
| | | Smart Control is configured to automatically match legal smart contract code from previous audit period to legal smart contract code of current audit period | | | Yes |
| | | Smart Analytic is configured to predict benchmark for current weekly revenue using revenue, locational, and temperature data from previous weeks | | Smart Control is configured to automatically match location and temperature of goods that are being delivered to expected location and temperature of goods | Yes |
| | | Smart Control is configured to automatically match revenue, invoice, and shipment amount from the client's blockchain | | | Yes |
| | | Not applicable | Not applicable | Smart Control is configured to automatically match the access level of customer node | Yes |
| | | Not applicable | Not applicable | Smart Control is configured to automatically match customer name per legal smart contract to customer name on active digital wallets | Yes |
| **Revenue transactions are not recorded in the correct period** | **Cut-off** | Not necessary, the record of the transaction and transaction event itself are triggered at the same time | | | Yes |
| | | Although not necessary to verify **cut-off** on BC, the following procedure, which is used to verify occurrence, can serve as a secondary test to verify the cut-off assertion:<br><br>Smart Control is configured to automatically match sales order, sales invoice, and shipment amount from the client's blockchain | | | Yes |
| **Revenue is not recorded** | **Complete-ness** | Not necessary, reconciliations occur as transactions are validated and then posted | | | Yes |
| | | Although not necessary to verify **completeness** on BC, the following procedure, which is used to verify occurrence, can serve as a secondary test to verify the completeness assertion:<br><br>Smart Control is configured to automatically match sales order, sales invoice, and shipment amount from the client's blockchain | | | Yes |
| **Revenue returns are not recognized** | **Occurrence** | Inspect and evaluate revenue return estimates | | | No |

For example, using Table 65, to address the risk of fictitious or erroneous revenue transactions, auditors would perform eight audit procedures, two of which would need to be performed off the external audit blockchain. Four of these procedures, which were described in Figure 33, could function as smart audit procedures on the blockchain. Moreover, two additional smart internal control tests could be designed to ensure that customers of the client maintain active digital wallets and have appropriate levels of access (e.g. access to send payment information but not to validate and post this information). For the risk that revenue transactions are not recorded in the correct period related to the cut-off assertion, additional audit procedures would not be necessary since, due to the inherent nature of blockchain technology, revenue is recorded at the same time the transaction occurs. Similarly, as reconciliations occur in near real-time by the auditee and its customer on the blockchain, ensuring the completeness of revenue transactions would also not be necessary. However, the three-way match that checks sales orders, sales invoices and shipping details can serve as a secondary test to verify cut-off and completeness. Finally, the evaluation of management revenue return estimates would be an audit procedure that remains off-the-blockchain. The smart audit procedures and off the blockchain audit procedures described within Table 65 are not all inclusive of a potential future hybrid audit approach. However instead, this table illustrates some of the significant risks and relevant audit procedures to describe the evolution of audit as audit clients and external auditors shift their business practices to the blockchain.

## 4.8. Issues and Future Research

The external audit blockchain proposed in this paper was supported by smart audit procedures that leveraged the benefits of blockchain technology under the assumption that blockchain would be widely but selectively adopted by businesses and that a single version of blockchain transactions is maintained (O'Leary 2018). Overall adoption of blockchain technology would in turn generate a higher demand for audit services on the blockchain (Alles 2015). However, there are a number of issues that should be considered for this concept to be realizable. These issues relate to the computational power, storage capabilities, cybersecurity risk, litigation risk, vulnerability of smart contracts, regulatory acceptance of the use of blockchain technology and the economics of the external audit blockchain.

***Computational power and Storage***

To ensure the integrity of the data, many blockchains such as Bitcoin and Ethereum, rely on cryptographic mechanisms that demand significant computational power. Permissioned blockchains have been proposed as a solution to ameliorate the challenge of computational complexity. However, when compared to traditional centralized databases, permissioned blockchain ledgers remain inefficient. Accordingly, the processing inefficiency of blockchain ledgers, in general, may hinder the wide adoption of this technology by businesses and subsequently by public accounting firms. In section 6 of this paper, an illustration of private and permissioned blockchain ecosystems for a business and a public accounting firm (with our primary focus being on the external auditor blockchain) were described, as use cases indicate that this type of blockchain provides the benefits of efficiency, transparency, and reliability in a secure environment [23] (PwC 2016; PwC 2017). Nevertheless, what remains unclear is the extent to which a private and permissioned

[23] In addition to PwC whitepapers, one of the major blockchain initiatives, hyperledger, concentrates on offering private and permissioned blockchain applications. See: https://www.hyperledger.org/.

blockchain is more beneficial than a traditional centralized database. Future research should examine the trade-off between the two and explore:

- What are the primary drivers for industry adoption of private and permissioned blockchains? Stated differently, why do users of private and permissioned blockchains opt out of using centralized databases?
- How to compare the benefits of public and permissionless blockchains, or private and permissioned blockchains to the benefits provided by a centralized database in relation to efficiency, transparency, and reliability?

Moreover, financial statement users could benefit from viewing the results of audit tests on a public and permissionless blockchain. As a result, future research could examine:

- What would be the intended and unintended consequences of disclosing audit information on a public and permissionless blockchain?
- Would the level of transparency of audit information on a public and permissionless blockchain, increase or decrease, compared to the level of transparency provided on a private and permissioned blockchain?

Blockchain by design is not equipped to store substantial volumes of data. Consequently, storing massive volumes of data could exacerbate the efficiency challenge of this technology. This is of paramount importance as the frequency to which IoT is placed on the blockchain increases[24]. Novel methods for solving the problem of data storage on

[24] IoT data on the blokchain. See: https://www.ibm.com/developerworks/cloud/library/cl-blockchain-for-cognitive-iot-apps-trs/cl-blockchain-for-cognitive-iot-apps-trs-pdf.pdf

blockchain including "decentralized storage"[25] and "blockweave"[26] have been proposed, however, the effectiveness of these methods is yet to be determined. Naturally, there is a need for future research to examine whether blockchain technology could be utilized to store big data. Specifically, future research could examine:

- Which methods can meet the demand to store big data on the blockchain?
- What latency is experienced on a blockchain system that stores big data?

### *Cybersecurity Risk*

Although the blockchain can be considered hack proof because it is a decentralized and immutable database, the risks of collusion and of private keys of digital wallets being stolen or lost emerge as cybersecurity threats. Collusion could occur when the majority of blockchain nodes control the blockchain network and retroactively alter transactions. In addition, blockchain users must be cognizant of securing access to the private keys of their digital wallets as they can be stolen (Gruber 2013). Hence, the possibility of collusion among blockchain users and the possibility of stealing private keys leads to the following research questions:

- How to design and implement a continuous monitoring method to reduce the risk of collusion on the blockchain?
- How to secure access of the private keys of blockchain users?

---

[25] Decentralized storage combines the characteristics of blockchain with techniques of sharding and swarming to meet the demand for the storage of massive amounts of information. See: https://dataconomy.com/2018/01/blockchain-data-storage-decentralized-future/.

[26] Blockweave offers low-cost storage on the blockchain. See: https://www.forbes.com/sites/shermanlee/2018/06/08/blockchain-is-critical-to-the-future-of-data-storage-heres-why/#3a423bea33e9 .

- With what frequency should blockchain information be backed up given that information is altered as a result of collusion, or given that information is stolen?

***Auditor Liability and Vulnerability of Smart Audit Procedures***

Providing more timely, relevant, reliable and transparent information can have adverse outcomes and increase auditors' litigation risk. Research in the critical audit matters area for example, suggests that providing additional information to financial statement users could increase auditors' liability (Gimbar, Hansen, and Ozlanski 2016). Hence, although the external audit blockchain supported by smart audit procedures could decrease the expectation gap between financial statement users and auditors, as it can provide more relevant, reliable and timely information to them, it is possible for auditor litigation risk to increase, a catch-22 situation. Research would be needed to examine what is the optimal balance between meeting financial statement users' demands while maintaining an acceptable level of audit litigation risk:

- Would auditor liability increase as a result of disclosing the results of audit tests at a disaggregate level to financial statement users?
- What is the optimal level of transparency and of litigation risk? In other words, how much transparency should be provided to financial statement users that rely on smart audit procedures while maintaining an acceptable level of audit litigation risk?

The vulnerability of smart audit procedures is another element to consider in relation to auditor liability. Smart audit procedures could be considered as the next generation of audit analytics (Rozario and Vasarhelyi 2018). Albeit, audit analytics are as effective as long as the code is free of error. Erroneous code in smart audit procedures could lead to inappropriate assessments about the risk of material misstatement as auditors would be relying on the output that is produced by these procedures. Consequently, future research could examine:

- What are the quality processes that public accounting firms should have in place to ensure smart audit procedures are free of error and thus reduce auditor liability?

***Regulatory Acceptance of an External Audit Blockchain***

The audit profession is currently experiencing a paradigm shift as a result of rapid technological advancements and regulators and standard-setters are taking notice. From the drafts for comment that the IAASB and the PCAOB have issued with relation to revising or creating new audit standards that recommend the use of more sophisticated audit analytics (IAASB 2016; PCAOB 2018), there is uncertainty around how audit standards should be updated or changed. As a result, the automated execution of audit analytics on the blockchain could create an entirely new set of challenges to regulators and standard-setters, which leads to the following research opportunities:

- How will the oversight model of financial statement audits be disrupted?
- How can auditors and regulators work in tandem to improve audit quality and inspection quality?
- What would the new standards that recommend the use of smart audit procedures on blockchain include?

In their latest strategic plan, the PCAOB stated that they are open to exploring the use of technology to automate their processes (PCAOB 2018). This paper proposed novel functions for the PCAOB, which would offer inspectors the ability to leverage blockchain and smart contract technologies to improve their inspection process and potentially prevent future audit failures. However, the feasibility of incorporating the PCAOB's inspection process into a blockchain environment should be empirically validated, which creates future opportunities for research:

- How to measure the benefits of adding the PCAOB as a validator of smart audit procedures and the reviewer of the results of audit tests?

  Do the benefits outweigh the costs?

***Economics of the External Audit Blockchain***

The burden of the costs for executing automated audit analytics on the external auditor blockchain will be on the audit firm as smart audit procedures would be impounded on the auditor's blockchain and not on the business blockchain. However, it would be of interest to investigate the cost dynamics of large public accounting firms versus mid-size public accounting firms. Large audit firms may be able to spread the cost of blockchain and smart contract technology implementation across their larger clients, however, for mid-size accounting firms, it is not clear how they would spread the cost. This premise leads to the following research questions:

- Would blockchain and smart contracts be developed in-house, or would it be outsourced?
- How would large and mid-size public accounting firms spread the cost of blockchain and smart contract implementation?

## 4.9. Conclusion

As blockchain and smart contracts are rapidly evolving business practices, the potential of these disruptive technologies in the external audit domain should not be neglected. This paper envisions the evolution of the financial statement audit paradigm by proposing an external audit blockchain supported by smart audit procedures in an attempt to foresee the benefits of this technology to auditing. The external audit blockchain benefits from the auditee's blockchain financial and nonfinancial information and has the potential to improve audit quality through the autonomous execution of audit procedures. Importantly, as smart audit procedures can autonomously disclose the results of audit procedures near real-time on the tamper-proof blockchain ledger, it is possible that these technologies could reduce the expectation gap between auditors, financial statement users and regulators (Rozario and Vasarhelyi 2018).

Although blockchain and smart audit procedures are important facilitators for evolving from a retroactive audit framework to a proactive audit framework to parallel a digital and near-the-event business environment, it is important to emphasize that the future audit framework is likely to comprise of on and off-the-blockchain audit procedures. Hence, while blockchain and smart audit procedures may radically evolve the way financial statement audits are performed and delivered, audit judgment is expected to remain a salient component of financial statement audits. Finally, while the benefits of blockchain and smart contracts to auditing were the primary focus of this paper, several issues and challenges to the adoption of these technologies, which lead to future research opportunities, were presented.

This research has a few limitations. First, existing audit client business risks were considered. It is likely that as these emerging technologies continue to mature, new risks that would have an effect on both the audit client and the audit firm would emerge. Future research could also investigate the new risks that emerge in a blockchain and smart contract ecosystem and their implications. Second, the scope of this paper largely focused on the revenue process as this process is generally a high risk area in audits. It would be interesting to expand this paper by addressing how the financial statement audit paradigm can evolve in other audit cycles beyond revenue. For example, how can auditors leverage drone related information logged into the blockchain to audit inventory? This paper also described an external audit blockchain and smart audit procedures that benefit from client blockchain information. Future research can expand the proposed conceptual framework by designing and implementing the described environments. Finally, it would be of interest to examine, in more detail, the challenges of applying blockchain technology to auditing.

# Chapter 5. Conclusion

The law of accelerating returns conjectures that technology evolves at an accelerated speed (Kurzweil 2004). Therefore, it is not surprising that companies are constantly evolving their business practices to keep pace with current technological developments. The auditing profession is by no means immune to technological disruption. Financial statement users have started to question the relevancy of the current audit model in a rapidly evolving business environment. Consequently, audit professionals, regulators, and academics have initiated discussions about how disruptive technologies can evolve auditing. Despite this ongoing debate, it is still unclear the extent to which technology can motivate the evolution of the current audit model and impact audit quality.

It is critical to examine the use of nontraditional sources of information and disruptive technologies on the audit process to ensure that auditing keeps pace with a rapidly changing business environment (and really, society!), and remains relevant. Accordingly, the purpose of this dissertation is to provide insights into the aforementioned debate by exploring the progressive evolution of auditing as a result of technological developments. This dissertation consisting of three essays on audit innovation explores 1) whether information from social media platforms can enhance traditional and continuous substantive analytical procedures; 2) the redesign of the audit process using RPA to achieve near end-to-end process automation; and 3) the impact of blockchain and smart contracts in evolving the way audits are performed and delivered.

## 5.1. Summary

The first essay examines whether information from social media platforms can enhance the effectiveness of traditional and continuous substantive analytical procedures. Specifically, this essay addresses whether Twitter-based proxies of consumer interest and consumer satisfaction improve the prediction performance and error detection performance of substantive analytical procedures. The PCAOB is increasingly recognizing that auditors use nontraditional sources of nonfinancial information to improve risk assessments (PCAOB 2017b), however, it is equally important to examine the potential of nontraditional nonfinancial information as audit evidence. Social media information about firms' brands or products that is generated by third parties can serve as a timely, and independent benchmark as it is less subject to manipulation by management and has been found to be correlated with sales (Tang 2017). The results suggest that the continuous substantive analytical models that incorporate TCI and GDP information experience improved prediction and error detection performance than the models that do not incorporate these measures suggesting that this information can complement traditional macroeconomic information and substitute contemporaneous firm-specific information such as accounts receivable. Taken together, the results provide evidence about the incremental value of social media information that is produced by third parties to analytical procedures.

The first essay indicates that social media information can be relevant to the performance of analytical procedures. In the second essay, a framework for redesigning the audit process using RPA is proposed and validated. Several research studies have proposed methodologies for automation in the audit (e.g. Alles et al. 2006; Issa and Kogan 2014), however, these methodologies often focus on automating specific audit steps rather than in forming a systematic audit production line (Issa et al. 2016). This essay conjectures that redesigning the audit process using RPA, which is referred to as RAPA, can enable near end-to-end process automation for audit processes that are labor-intensive and that do not require audit judgment. The proposed framework consists of 1) developing vision and process objectives, 2) identifying an audit process for automation, 3) understanding the process, 4) designing and implementing an ADS, 5) audit apps prototyping, and 6) feedback and evaluation. The loan testing audit sub-process of an accounting firm is selected to demonstrate the feasibility of the framework. The results of the framework validation indicate that it can be used to apply automation to an audit process. Specifically, the results suggest that RPA software can be used to achieve a systematic audit process and automate evidence collection activities and the execution of audit tests.

Similar to the second essay, the third essay of this dissertation explores the potential application of a different disruptive technology to auditing, blockchain and smart contracts. The current audit model emphasizes sampling, and a retroactive and binary audit opinion. Evolving auditing with blockchain and smart contracts can transform the way that audits are performed and delivered. This essay first proposes that a company's blockchain financial and nonfinancial records can satisfy the requirements of audit evidence as prescribed by the PCAOB (PCAOB 2010d, AS No. 1105). This is important as more reliable audit evidence can result in more reliable audit tests. Next, an external audit blockchain supported by smart audit procedures, which are autonomous audit procedures, is proposed. Smart audit procedures on the blockchain can autonomously execute audit procedures for the auditor as company information enters the external audit blockchain and the results of the audit procedures are disclosed to blockchain nodes near real-time. The external audit blockchain can serve as an important facilitator in the adoption of CA by external auditors as the integrity of audit evidence and auditor independence are preserved. Finally, novel functions for the PCAOB in a blockchain ecosystem and a holistic audit model consisting of on-the-blockchain and off-the-blockchain audit procedures are envisioned to provide guidance on the integration of blockchain and smart contracts in a risk-based audit.

## 5.2. Contributions

This dissertation contributes to knowledge in at least two ways. First, it explores the shift of auditing towards technological innovation with social media information, RPA, and blockchain smart contracts by proposing practical applications of these innovations. Second, it is among the first to discuss the impact of nontraditional audit evidence and disruptive technologies on audit quality.

The first study contributes to the literature on analytical procedures by analyzing a different form of nontraditional audit evidence (Yoon 2016), social media information that is generated by consumers, in a variety of analytical procedures. Prior research has found that social media information about consumer interest and satisfaction is associated with sales (Tang 2017), however, it is also important to evaluate the relevance of this information in the context of analytical procedures and assess whether it can improve the quality of the audit. This study provides useful insights to academics, audit practitioners, and regulators on the effectiveness of analytical procedures that incorporate a nontraditional source of audit evidence.

The second study offers useful insights to academics, audit practitioners, and regulators on the impact of disruptive technologies to the audit and its quality. Prior research has proposed a plethora of technological tools to automate audit procedures, however, how technology can be utilized to produce an audit production line that could result in improved audit quality requires more consideration. This essay proposes a framework for using RPA, referred to as RAPA (robotic audit process automation), and evaluates its feasibility by applying it to the sub-process of a public accounting firm. The application of the framework indicates that it can guide auditors in the implementation of RAPA to the audit process and that it can enhance audit efficiency and effectiveness. As a result, the evidence in this study can help auditors and regulators evaluate where in the audit process it may be beneficial to apply this novel technology and its impact on audit quality.

The last essay contributes to the emerging literature on blockchain for auditing, audit practice and standard setting by presenting the potential application of blockchain and smart contract technologies to transform audit procedures and reporting. Recent studies have explored the potential impact of these technologies by providing a general discussion of their application to accounting and auditing. This paper presents a more detailed discussion by linking these technologies to auditing standards, the audit process, and the responsibilities of regulators and by proposing ideas, such as the new responsibilities of regulators in a blockchain audit environment, to foresee the transformation of auditing and the regulatory environment in a digital economy. In addition, issues related to the application of these technologies are discussed. Auditors can benefit from the insights that are offered in this paper as they consider the expansion of blockchain auditing services to the audit of financial statements. Lastly, the arguments in this essay suggest that regulators, such as the PCAOB, may need to revamp their functions to support improvements in audit quality.

## 5.3. Limitations

One potential limitation of this dissertation is that it does not compare the current (traditional) audit approach to the proposed audit approach, in which the proposed innovations are utilized. An experimentation program, which evaluates the current to the new approach in parallel and in actual audit engagements, is essential to provide further insights into the trade-offs of the proposed applications and their usefulness (Rozario and Vasarhelyi 2018c). Another limitation exists with respect to the impact of the proposed applications on auditor's judgment, which could inherently impact audit quality. The described experimentation program can further assess the impact of these tools on auditor judgment. Future research should consider the case study method for the experimentation program where a team of experienced audit professionals, regulators, and academics work in tandem to analyze the benefits and challenges of the use of nontraditional and nonfinancial external information and disruptive technologies on auditing.

One limitation in the first essay exists in terms of the interpolated monthly financial information as quarterly financial information is extracted from Compustat to evaluate the effectiveness of analytical procedures that incorporate social media information on a disaggregated dataset. The interpolation technique has been used in prior research, but it has the limitation that it may not directly reflect the actual account balances of companies. Another limitation in this essay is that only one new source of nonfinancial external information is evaluated. Finally, the incremental contribution of social media information is limited to those companies in the business to consumer industries, therefore, these findings cannot be generalized to other industries.

The second essay proposes a framework for RAPA and applies it to a portion of an audit process to evaluate its feasibility. Although the insights from this essay provide an initial understanding of the impact of process redesign and RPA on the audit process, the holistic application of RPA to the audit process of several public accounting firms requires further consideration to be able to generalize the applicability of the framework. In addition, preliminary assessments of efficiency and effectiveness to measure the value of RAPA were conducted, however, the trade-offs of RAPA compared to the trade-offs of the traditional approach in a real audit engagement should be further evaluated.

The third study proposes an external audit blockchain supported by smart contracts as a possible way to narrow the expectation gap in the digital era. However, the new business and financial statement risks that emerge in a blockchain and smart contract ecosystem were discussed at a general level and require more in depth thinking. Moreover, the increasing adoption of blockchain technology by businesses and public accounting firms signals that this technology could be widely adopted and therefore replace traditional business and accounting practices, yet the interaction among different blockchain ecosystems which culminates with the disclosure of the results of smart audit procedures on the external audit blockchain should be empirically validated.

## 5.4. Future Research

For audit, innovation will drive quality (O’Donnell 2018). This dissertation is an initial attempt to understand the impact of nontraditional sources of information and disruptive technologies to the quality of audit services, which leads to several opportunities for future research. Future research can be conducted to evaluate the incremental informativeness of other forms of nontraditional audit evidence, such as sentiment from Facebook posts and the number of Google search queries about products or services, to analytical procedures as well as other procedures in the audit.

With respect to the disruptive technologies discussed in this dissertation, future research can explore the application of the RAPA framework to other processes in order to obtain further insight into its feasibility. For instance, future research can examine the application of RAPA for the risk assessment or concluding stages of the audit. In addition, as there is still much to be learned about blockchain and smart contracts and their implications to auditing, future research can explore the new risks that emerge in a blockchain and smart contract ecosystem and how blockchain adopters (business entities and public accounting firms) should respond to those risks. For example, future research can explore the new cybersecurity risks in the blockchain and the new audit services that should be designed and performed to mitigate those risks.

Finally, to fully assess the benefits and challenges of these novel tools for auditing, it is clear that an experimentation program is much needed (Rozario and Vasarhelyi 2018c). In specific, for auditing to advance as a profession and continue to remain relevant in the digital economy, it is vital to parallel test the proposed audit innovations to the traditional tools on real audit engagements. In addition to evaluating the performance of these tools, the experimentation program can facilitate the study of the impact of the proposed audit innovations on audit judgment.

It is difficult to exclude continuous auditing, full population testing, and an audit by exception approach as fundamental characteristics of the future financial statement audit. Therefore, the experimentation program should also examine the value of the proposed innovations under the umbrella of these characteristics.

This dissertation attempts to inform the audit community on the compelling issues of how to leverage nontraditional sources of information and disruptive technologies to advance auditing and audit quality. It is critical to examine the impact of technology on auditing to ensure that companies' digital business environments and associated risks do not outpace the relevance and effectiveness of audits. Accordingly, this dissertation contributes to the emerging literature on audit analytics by foreseeing the impact of technological innovations on the audit model and audit quality.

# REFERENCES

Allee, K.D., Baik, B. and Roh, Y., 2018. Detecting Financial Misreporting with Real Production Activity.

Allen, R. D., M. S. Beasley, and B. C. Branson. 1999. Improving analytical procedures: A case of using disaggregate multilocation data. *Auditing: A Journal of Practice & Theory* 18 (2):128-142.

Alles, M. G., Kogan, A., & Vasarhelyi, M. A. 2002. Feasibility and Economics of Continuous Assurance 1. In *Continuous Auditing: Theory and Application* (pp. 149-167). Emerald Publishing Limited.

Alles, M.G., Kogan, A. and Vasarhelyi, M.A., 2004. Restoring auditor credibility: tertiary monitoring and logging of continuous assurance systems. *International Journal of Accounting Information Systems*, *5*(2), pp.183-202.

Alles, M., Brennan, G., Kogan, A. and Vasarhelyi, M.A., 2006. Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems*, *7*(2), pp.137-161.

Alles, M. G. 2015. Drivers of the use and facilitators and obstacles of the evolution of Big Data by the audit profession. *Accounting Horizons* 29 (2): 439-449.

Alles, M., & Gray, G. L. 2016. Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors. *International Journal of Accounting Information Systems*, *22*, 44-59.

Alpaydin, E. (2014). *Introduction to machine learning*. The MIT Press.

American Institute of Certified Public Accountants (AICPA). 1988. Analytical Procedures. Statement of Auditing Standards No. 56. New York, NY: AICPA.

American Institute of Certified Public Accountants (AICPA). 2002. Consideration of Fraud in a Financial Statement Audit. Statement on Auditing Standards No. 99. New York, NY: AICPA.

American Institute of Certified Public Accountants (AICPA). 2012. Analytical Procedure. In *AU-C Section 520*. New York: American Institute of Certified Public Accountants.

American Institute of Certified Professional Accountants (AICPA). 2015. *Audit Analytics and Continuous Audit: Looking Toward the Future*, New York, NY.

American Institute of Certified Professional Accountants (AICPA). 2018. *Employee benefits plans – financial statement audits*, New York, NY.

Amir, E. and Lev, B., 1996. Value-relevance of nonfinancial information: The wireless communications industry. *Journal of accounting and economics*, *22*(1-3), pp.3-30.

Appelbaum, D. 2016. Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box as Reliable Evidence. *Journal of Emerging Technologies in Accounting*, *13*(1), 17-36.

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

Askitas, N. and Zimmermann, K.F., 2009. Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), pp.107-120.

Badertscher, B.A., Kim, J., Kinney, W. and Owens, E.L., 2017. Audit Procedures and Financial Statement Quality: The Positive Effects of Negative Assurance. Working Paper.

Bell, T., Marrs, F., Solomon, I., & Thomas, H. 1997. Auditing Organizations through a Strategic Lens: The KPMG business measurement process. *Montvale, NJ: KPMG Peat Marwick, LLP*.

Bollen, J., Mao, H., & Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8.

Brazel, J.F., Jones, K.L. and Zimbelman, M.F., 2009. Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research*, *47*(5), pp.1135-1166.

Brazel, J.F., Jones, K.L. and Prawitt, D.F., 2013. 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), pp.131-156.

Brown-Liburd, H., & Vasarhelyi, M. A. 2015. Big Data and audit evidence. *Journal of Emerging Technologies in Accounting*, *12*(1), 1-16.

Bryans, D. 2014. Bitcoin and money laundering: mining for an effective solution. *Ind. LJ*, *89*, 441.

Bumgarner, A., and Vasarhelyi, M. A. 2015. Continuous Auditing—A New View, Chapter 1 in AICPA, Audit Analytics and Continuous Audit: Looking Toward the Future, *American Institute Of Certified Public Accountants.* New York, NY 2015

Burton, S., & Soboleva, A. 2011. Interactive or reactive? Marketing with Twitter. *Journal of Consumer Marketing*, *28*(7), 491-499.

Buterin, V. 2013. Ethereum white paper.

Byrnes, P.E., 2015. *Developing automated applications for clustering and outlier detection: data mining implications for auditing practice* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Capgemini. 2018. Why RPA Constitutes a Success Factor for Financial Institutions KYC Digitization. Retrieved from: https://www.capgemini.com/de-de/2018/02/why-rpa-constitutes-a-success-factor-for-financial-institutions-kyc-digitization/

Chan, D. Y., & Vasarhelyi, M. A. 2011. Innovation and practice of continuous auditing. *International Journal of Accounting Information Systems*, *12*(2), 152-160.

Chartered Professional Accountants of Canada. 2016. Audit Data Analytics Alert, Audit Data Analytics, Keeping Up with the Pace of Change.

Chen, Y. and Leitch, R.A., 1998. The error detection of structural analytical procedures: A simulation study. *Auditing*, *17*(2), p.36.

Chen, Y. and Leitch, R.A., 1999. An analysis of the relative power characteristics of analytical procedures. *Auditing: A Journal of Practice & Theory*, *18*(2), pp.35-69.

Chiu, T. and Jans, M.J., 2017. Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness.

Choi, H. and Varian, H., 2012. Predicting the present with Google Trends. *Economic Record*, *88*(s1), pp.2-9.

Chou, C., Hwang, R., Wang, T., and Li, C. 2019. Using Smart Contract to Establish Decentralized Accounting Standards: An Example of Revenue Recognition. *Working paper. California State University, Monterey Bay.*

Cohen, J. R., G. Krishnamoorthy, and A. M. Wright. 2000. Evidence on the effect of financial and nonfinancial trends on analytical review. *Auditing: A Journal of Practice & Theory* 19 (1):27-48.

Cooper, L. A., Holderness, D.K., Sorensen, T., and Wood, D.A. 2018. Robotic Process Automation in Public Accounting. Working Paper.

Curtis, M. B., and E. A. Payne. 2008. An examination of contextual factors and individual characteristics affecting technology implementation decisions in auditing. International Journal of Accounting Information Systems 9(2): 104-121.

Da, Z., J. Engelberg, and P. Gao. 2011a. In search of fundamentals. *Working paper*.

Da, Z., Engelberg, J., & Gao, P. 2011b. In search of attention. *The Journal of Finance*, *66*(5), 1461-1499.

Dai, J. and Li, Q., 2016. Designing Audit Apps for Armchair Auditors to Analyze Government Procurement Contracts. *Journal of Emerging Technologies in Accounting*, *13*(2), pp.71-88.

Dai, J., & Vasarhelyi, M. A. 2017. Towards Blockchain-based Accounting and Assurance. *Journal of Information Systems*.

Davenport, T.H. and Short, J.E., 1990. The new industrial engineering: information technology and business process redesign.

Davenport, T.H. and Stoddard, D.B., 1994. Reengineering: business change of mythic proportions?. *MIS quarterly*, pp.121-127.

Davenport, T.H. 2015. Process Automation and the Rebirth of Reengineering. Retrieved from:https://blogs.wsj.com/cio/2015/07/08/process-automation-and-the-rebirth-of-reengineering/

Davenport, H. and Brain, D. 2018. Before Automating Your Company's Processes, Find Ways to Improve Them. Harvard Business Review. Retrieved from: https://hbr.org/2018/06/before-automating-your-companys-processes-find-ways-to-improve-them

Du, R. Y., & Kamakura, W. A. 2012. Quantitative trendspotting. *Journal of Marketing Research*, *49*(4), 514-536.

Dzeng, S. C. 1994. A comparison of analytical procedure expectation models using both aggregate and disaggregate data. *AUDITING: A Journal of Practice & Theory*, *13*(2), 1.

Ettredge, M., Gerdes, J., & Karuga, G. 2005. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87-92.

Gayo-Avello, D. 2013. A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, *31*(6), 649-679.

Gimbar, C., Hansen, B. and Ozlanski, M.E., 2016. The effects of critical audit matter paragraphs and accounting standard precision on auditor liability. *The Accounting Review*, *91*(6), pp.1629-1646.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012.

Glaser, F., 2017. Pervasive decentralisation of digital infrastructures: a framework for blockchain enabled system and use case analysis.

Gockel, B., Acar, T., and Forster, M. 2018. Blockchain in Logistics. *DHL Trend Research*.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. 2010. Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences*, *107*(41), 17486-17490.

Groomer, S., & Murthy, U. 1989. Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems*, *3*(2), 53–69.

Gruber, S. 2013. Trust, Identity and Disclosure: Are Bitcoin Exchanges the Next Virtual Havens for Money Laundering and Tax Evasion. *Quinnipiac L. Rev.*, *32*, 135.

Gutierrez, M. and Sastron, F., 2018. Framework for a New Wave of IT-Enabled Business Process Transformations. In *Closing the Gap Between Practice and Research in Industrial Engineering* (pp. 351-358). Springer, Cham.

Hamm, K. 2018. Mexican Mangos, Diamonds, Cargo Shipping Containers, Oh My! What Auditors Need to Know about Blockchain and Other Emerging Technologies: A Regulator's Perspective. Retrievedfrom: http://raw.rutgers.edu/docs/wcars/43wcars/Sides/Mexican%20Mangos%20Diamonds%20Shipping%20Containers%20Oh%20My.11.02.18%20Hamm_.pdf

Hammer, M., 1990. Reengineering work: don't automate, obliterate. *Harvard business review*, *68*(4), pp.104-112.

Hevner, A., March, S. T., Park, J. and Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75 – 105.

Hirst, D.E. and Koonce, L., 1996. Audit analytical procedures: A field investigation. *Contemporary Accounting Research*, *13*(2), pp.457-486.

Hoitash, R., A. Kogan, and M. Vasarhelyi. 2006. Peer-based approach for analytical procedures. *Auditing: A Journal of Practice and Theory* 25 (2): 53–84.

Huang, F., 2019. *Three essays on emerging technologies in accounting* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

IBM. 2017. How is RPA different from other enterprise automation tools such as BPM/ODM? Retrieved from: https://www.ibm.com/blogs/insights-on-business/gbs-strategy/rpa-different-enterprise-automation-tools-bpmodm/

Ibrahim, A., 2017. *APPLICABILITY OF BLOCKCHAIN TECHNOLOGY IN TELECOMMUNICATIONS SERVICE MANAGEMENT* (Doctoral dissertation, Leiden University).

IEEE Corporate Advisory Group. 2017. IEEE Guide for Terms and Concepts in Intelligent Process Automation. New York City: IEEE.

International Auditing and Assurance Standards Board (IAASB). 2016. Exploring the Growing Use of Technology in the Audit, with a Focus on Data Analytics. New York, NY: IFAC.

Ittner, C.D. and Larcker, D.F., 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of accounting research*, *36*, pp.1-35.

Issa, H., 2013. *Exceptional exceptions* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Issa, H. and Kogan, A., 2014. A predictive ordered logistic regression model as a tool for quality review of control risk assessments. *Journal of Information Systems*, *28*(2), pp.209-229.

Issa, H., Sun, T. and Vasarhelyi, M.A., 2016. Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), pp.1-20.

Jarvenpaa, S.L. and Stoddard, D.B., 1998. Business process redesign: Radical and evolutionary change. *Journal of Business Research*, *41*(1), pp.15-27.

Ji, X., Chun, S. A., & Geller, J. 2013. Monitoring public health concerns using twitter sentiment classifications. In *Healthcare Informatics (ICHI), 2013 IEEE International Conference on* (pp. 335-344). IEEE.

Kinney, W. R. 1978. ARIMA and regression in analytical review: An empirical test. *The Accounting Review, 53*(1), 48-60.

Kinney, W.R., 1987. ATTENTION-DIRECTING ANALYTICAL REVIEW USING ACCOUNTING RATIOS-A CASE-STUDY. *AUDITING-A JOURNAL OF PRACTICE & THEORY*, *6*(2), pp.59-73.

Kogan, A., Alles, M.G., Vasarhelyi, M.A. and Wu, J., 2014. Design and evaluation of a continuous data level auditing system. *Auditing: A Journal of Practice & Theory*, *33*(4), pp.221-245.

Kozlowski, S. 2016. A Vision of an ENHanced ANalytic Constituent Environment: ENHANCE. Doctoral dissertation, Rutgers, The State University of New Jersey.

Kuenkaikaew, S., & Vasarhelyi, M. A. 2013. The Predictive Audit Framework. *International Journal of Digital Accounting Research*. Vol. 13: 37-71.

Kurzweil, R., 2004. The law of accelerating returns. In *Alan Turing: Life and legacy of a great thinker* (pp. 381-416). Springer, Berlin, Heidelberg.

Lacity, M., Willcocks, L., & Craig, A. 2015. Robotic Process Automation at Telefonica 02. The Outsourcing Unit, pp. 1-19.

Lee, L.F., Hutton, A.P. and Shu, S., 2015. The role of social media in the capital market: evidence from consumer product recalls. *Journal of Accounting Research*, *53*(2), pp.367-404.

Lev, B., 1980. On the use of index models in analytical reviews by auditors. *Journal of Accounting Research*, pp.524-550.

Lev, B., & Gu, F. 2016. *The end of accounting and the path forward for investors and managers*. John Wiley & Sons.

Limam Mansar, S. and Reijers, H.A., 2007. Best practices in business process redesign: use and impact. *Business Process Management Journal*, *13*(2), pp.193-213.

Likefolio. 2018. Who are you? Retrieved from: https://home.likefolio.com/

Lin, S., Shih, L., Yang, D., Kung, J. 2018. Apply RPA (Robotic Process Automation) in Semiconductor Smart Manufacturing. e-Manufacturing & Design Collaboration Symposium (eMDC), Hsinchu, Taiwan, 2018, pp. 1-3.

Louwers, T. J., Ramsay, R. J., Sinason, D. H., Strawser, J. R., & Thibodeau, J. C. 2018. *Auditing and assurance services*. New York, NY: McGraw-Hill/Irwin.

Mainelli, M., & Smith, M. 2015. Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers (aka blockchain technology). *The Journal of Financial Perspectives*, *3*(3), 38-69.

McKinsey. 2016. The next acronym you need to know about: RPA (robotic process automation). Retrieved from: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-next-acronym-you-need-to-know-about-rpa

Meza, Miguel Angel Minutti. 2011. Using Peer Firms to Examine whether Auditor Industry Specialization Improves Audit Quality and to Enhance Expectation Models for Analytical Audit Procedures. *Diss. University of Toronto*.

Moffit, K., Rozario, A. M., Vasarhelyi, M. A. 2018. Robotic Process Automation for Auditing. Forthcoming in “Journal of Emerging Technologies in Accounting”.

Moon, D. 2016. “Monitoring Reputation Risk Using KRIs: Using Twitter data to Detect a Sign of Reputation Damage in Real Time”. Chapter 4 of the dissertation titled *Continuous Risk Monitoring and Assessment: CRMA*, Rutgers University, Newark N.J.

Moore, J. 1996. The Death of Competition: Leadership & Strategy. *The Age of Business Ecosystems, HarperCollins, New York*.

Murthy, D. 2015. Twitter and elections: are tweets, predictive, reactive, or a form of buzz?. *Information, Communication & Society*, *18*(7), 816-831.

Nakamoto, S. 2008. *Bitcoin: A peer-to-peer electronic cash system*. Working paper.

Nelson, K. M., Kogan, A., Srivastava, R. P., Vasarhelyi, M. A., & Lu, H. 2000. Virtual auditing agents: the EDGAR Agent challenge. Decision Support Systems, 28(3), 241-253.

No, W. G., & Vasarhelyi, M. A. 2017. Cybersecurity and Continuous Assurance. *Journal of Emerging Technologies in Accounting*, *14*(1), 1-12.

O' Donnell. 2018. Innovation: Keeping up with the rising bar of quality. Retrieved from: https://www.youtube.com/watch?v=9izk5BPhdgo.

O'Leary, D.E., 2018. Open Information Enterprise Transactions: Business Intelligence and Wash and Spoof Transactions in Blockchain and Social Commerce. *Intelligent Systems in Accounting, Finance and Management*, *25*(3), pp.148-158.

Olsen, P., Borit, M. and Syed, S., 2019. Applications, limitations, costs, and benefits related to the use of blockchain technology in the food industry. *Nofima rapportserie*.

Paniagua, J., & Sapena, J. 2014. Business performance and social media: Love or hate? *Business Horizons, 57*, 719-728. doi:10.1016/j.bushor.2014.07.005

Peters, G. W., and E. Panayi. 2015. Understanding Modern Banking Ledgers through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money. Working paper. Cornell University.

Pilkington, M. 2016. Blockchain Technology: Principles and Applications. Research Handbook on Digital Transformations, edited by F. Xavier Olleros and Majlinda Zhegu. Edward Elgar.

Pricewaterhousecoopers. (PwC). 2016. The argument for private blockchains. Retrieved from: http://usblogs.pwc.com/emerging-technology/the-argument-for-private-blockchains/

Pricewaterhousecoopers. (PwC). 2017. PwC Blockchain Validation Solution. https://www.pwc.com/us/en/about-us/new-ventures/pwc-blockchain-validation-solution.html

Public Company Accounting Oversight Board (PCAOB). 2007. Observations on Auditors' Implementation of PCAOB Standards Relating to Auditors' Responsibilities.

Public Company Accounting Oversight Board (PCAOB). 2010a. *Substantive Analytical Procedures.* Auditing Standard (AS) No. 2305. Washington D.C.: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2010.b *Identifying and Assessing Risks of Material Misstatement.* Auditing Standard (AS) No. 2110. Washington D.C.: PCAOB

Public Company Accounting Oversight Board (PCAOB). 2010c. *Evaluating Audit Results.* Auditing Standard (AS) No. 2810. Washington, D.C.: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2010d. Audit Evidence. PCAOB Auditing Standard No. 1105. Washington, DC: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2016a. Staff Inspection Brief. Preview of Observations from 2015 Inspections of Auditors of Issuers. Washington D.C.: PCAOB

Public Company Accounting Oversight Board (PCAOB). 2016b. Audit Expectations Gap: A Framework for Regulatory Analysis. Washington, D.C: PCAOB. Retrieved from:https://pcaobus.org/News/Speech/Pages/Franzel-speech-Institute-12-13-16.aspx

Public Company Accounting Oversight Board (PCAOB). 2016c. Areas of Focus for Global Audit Regulators: Washington, DC: PCAOB. Retrieved from: https://pcaobus.org/News/Speech/Pages/Harris-global-audit-Institute-12-13-14.aspx

Public Company Accounting Oversight Board (PCAOB). 2017a. The State of Audit Quality and Regulatory Approaches to Achieving High Quality Audits. Retrieved from: https://pcaobus.org/News/Speech/Pages/Franzel-state-audit-quality-regulatory-approaches-achieving-high-quality-audits-12-7-17.aspx

Public Company Accounting Oversight Board (PCAOB). 2017b. Technology and the Audit of Today and Tomorrow. Retrieved from: https://pcaobus.org/News/Speech/Pages/Harris-statement-PCAOB-AAA-4-20-17.aspx

Public Company Accounting Oversight Board (PCAOB). 2018. Strategic Plan 2018-2022 Draft for Comment. Retrieved from: https://pcaobus.org/About/Administration/Documents/Strategic%20Plans/Draft-2018-2022.pdf

Reijers, H.A. and Mansar, S.L., 2005. Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega*, *33*(4), pp.283-306.

Rozario, A. M., & Vasarhelyi, M. A. 2018a. Auditing with smart contracts. International Journal of Digital Accounting Research.

Rozario A. M., Vasarhelyi M. A. 2018b. How Robotic Process Automation Is Transforming Accounting and Auditing. The CPA Journal.

Rozario A.M., Vasarhelyi M. A. 2018c. Comments on the PCAOB's Strategic Plan for 2018-2022. Retrieved from:

https://pcaobus.org/About/Administration/StrategicPlanComments/12_Rutgers%20Carlab.pdf

Seasongood, S. 2016. NOT JUST FOR THE ASSEMBLY LINE: A Case for Robotics in Accounting and Finance. Financial Executive, 3231-39.

Securities and Exchange Commission (SEC). 2015. Accountants and Capital Markets in an Era of Digital Disruption: Remarks to the Institute of Chartered Accountants in England and Wales and British American Business. Retrieved from: https://www.sec.gov/news/speech/remarks-inst-chartered-acctnts.html

Sigvaldason, T., and Warren, J.D., 2004. *Solving the Software Architecture Riddle to Deliver Enterprise-wide Continuous Financial Process Monitoring and "Auditing."* Financial Market Solutions, LLC.

Singh, K. 2015. The New Wild West: Preventing Money Laundering in the Bitcoin Network. *Nw. J. Tech. & Intell. Prop.*, *13*, iii.

Stephens-Davidowitz, S. and Pinker, S., 2017. *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*.

Stieglitz, S., & Dang-Xuan, L. 2013. Emotions and information diffusion in social media sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248. doi:10.2753/MIS0742-1222290408.

Sun, T., 2018. Deep learning applications in audit decision making (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Szabo, N. 1994. Smart contracts. *Retrieved from: http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html*

Tang, V.W., 2017. Wisdom of Crowds: Cross-sectional Variation in the Informativeness of Third-Party-Generated Product Information on Twitter. *Journal of Accounting Research*.

Tapscott, D. and Tapscott, A. 2016. *Blockchain revolution*. 1st ed. New York: Portfolio-Penguin.

The Economist. 2017. The world's most valuable resource is no longer oil, but data. Retrieved from: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Titera, W. R. 2013. Updating audit standard—Enabling audit data analysis. *Journal of Information Systems*, *27*(1), 325-331.

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

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm*, *10*(1), 178-185.

U.S. Department of Labor. 2015. Assessing the Quality of Employee Benefit Plan Audits.

van Hee, K.M. and Reijers, H.A., 2000. Using formal analysis techniques in business process redesign. In *Business process management* (pp. 142-160). Springer, Berlin, Heidelberg.

Vasarhelyi, M.A., 1984. Automation and changes in the audit process. *Auditing: A journal*.

Vasarhelyi, M. A., & Halper, F. B. 1991. The continuous audit of online systems. In *Auditing: A Journal of Practice and Theory*.

Vasarhelyi, M. A., & Hoitash, R. 2005. Intelligent Software Agents In Accounting: an evolving scenario. *M. Vasarhelyi, A. Kogan (a cura di), The Evolving Paradigms of Artificial Intelligence and Expert Systems: An International View*, *6*.

Vasarhelyi, M. A., J. D. Warren Jr, R. A. Teeter, and W. R. Titera. 2014. Embracing the Automated Audit. Journal of Accountancy. Available at: http://www.journalofaccountancy.com/issues/2014/apr/automated-audits-20127039.html

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. 2015. Big Data in accounting: An overview. *Accounting Horizons*, *29*(2), 381-396.

Vaziri, A. 2016. Smart BoL – Reducing Contractual Enforcement Costs. Retrieved from: https://www.chainofthings.com/news/2016/11/23/smart-bol-reducing-contractual-enforcement-costs

Von Bertalanffy, L., 1968. General system theory. *New York*, *41973*(1968), p.40.

Wheeler, S. and Pany, K., 1990. Assessing the performance of analytical procedures: A best case scenario. *Accounting Review*, pp.557-577.

Wild, J. J. 1987. The prediction performance of a structural model of accounting numbers. *Journal of Accounting Research*, *25*(1), 139-160.

Willcocks, L.P. and Lacity, M., 2016. *Service automation robots and the future of work*. SB Publishing.

Yin, C. 2018. Privacy-Preserving Information Sharing within and Audit Firm. Dissertation, Rutgers University, Newark N.J.

Yoon, K., Hoogduin, L., & Zhang, L. 2015. Big data as complementary audit evidence. *Accounting Horizons*, *29*(2), 431-438.

Yoon, K. 2016. “Big Data as Audit Evidence: Utilizing Weather Indicators.” Chapter 3 of the dissertation titled *Three Essays on Unorthodox Audit Evidence*, Rutgers University, Newark N.J.

Zhang, L., Pawlicki, A., McQuilken, D., and Titera, W. 2012. The AICPA Assurance Services Executive Committee Emerging Assurance Technologies Task Force: The Audit Data Standards (ADS) Initiative. Journal of Information Systems, 26(1), 199-205.

# APPENDIX – Chapter 2

## Table 30: 4-Digit SIC - Descriptive Statistics - Financial Information for Final Sample, from 2012-2017

**Descriptive Statistics - Financial Information - Firm-Quarter Observations from 2012-2017**

| 4-Digit SIC Code | Industry Name | Number of Firm-Quarter Observations | Revenue | Accounts Receivable |
|---|---|---|---|---|
| 2000 | Food and Kindred Products | 48 | 4176.59 | 2672.12 |
| 2033 | Canned, Fruits, Vegetable Preserves, Jams & Jellies | 24 | 259.91 | 81.12 |
| 2040 | Grain Mill Products | 48 | 3845.96 | 1505.65 |
| 2080 | Beverages | 24 | 16165.42 | 7408.54 |
| 2082 | Malt Beverages | 48 | 1369.65 | 734.27 |
| 2086 | Bottled & Canned Soft Drinks & Carbonated Waters | 72 | 4386.50 | 1889.14 |
| 2090 | Miscellaneous Food Preparations & Kindred Products | 24 | 1080.11 | 437.20 |
| 2111 | Cigarettes | 24 | 4616.67 | 161.38 |
| 2300 | Apparel and Other Finished Prods of Fabrics & Similar Materia | 48 | 1160.38 | 494.00 |
| 2320 | Men's & Boys' Furnishings, Work Clothing, & Allied Garments | 24 | 1764.53 | 594.95 |
| 2840 | Soap, Detergents, Cleaning Preparations, Perfumes, Cosmetics | 24 | 18343.21 | 5608.46 |
| 2842 | Specialty Cleaning, Polishing and Sanitation Preparations | 24 | 1432.67 | 525.79 |
| 2844 | Perfumes, Cosmetics & Other Toilet Preparations | 72 | 2455.73 | 1120.70 |
| 2890 | Miscellaneous Chemical Products | 24 | 93.60 | 61.22 |
| 2911 | Petroleum Refining | 48 | 29763.38 | 10859.92 |
| 3021 | Rubber & Plastics Footwear | 24 | 7498.13 | 3441.46 |
| 3100 | Leather and Leather Products | 24 | 1182.97 | 237.74 |
| 3140 | Footwear - except Rubber | 24 | 696.59 | 368.74 |
| 3540 | Metalworking Machinery & Equipment | 24 | 2820.91 | 1689.10 |
| 3577 | Computer Peripheral Equipment, NEC | 24 | 540.81 | 261.30 |
| 3630 | Household Appliances | 24 | 4986.71 | 2487.79 |
| 3663 | Radio & TV Broadcasting & Communications Equipment | 48 | 25977.13 | 11675.50 |
| 3674 | Semiconductors & Related Devices | 24 | 14142.58 | 4326.13 |
| 3711 | Motor Vehicles & Passenger Car Bodies | 144 | 34647.45 | 41809.17 |
| 3751 | Motorcycles, Bicycles & Parts | 24 | 1472.83 | 2383.89 |
| 3942 | Dolls & Stuffed Toys | 24 | 1458.91 | 1112.19 |
| 3944 | Games, Toys & Children's Vehicles - except Dolls & | 24 | 1130.23 | 962.71 |
| 3949 | Sporting & Athletic Goods, NEC | 24 | 221.98 | 185.14 |
| 4210 | Trucking & Courier Services - except Air | 24 | 14705.75 | 6303.17 |
| 4400 | Water Transportation | 48 | 3034.01 | 423.68 |
| 4512 | Air Transportation, Scheduled | 168 | 5236.72 | 884.61 |
| 4513 | Air Courier Services | 24 | 12548.88 | 5982.29 |
| 4700 | Transportation Services | 24 | 1669.62 | 1090.61 |
| 4812 | Radiotelephone Communications | 24 | 1005.86 | 579.76 |
| 4832 | Radio Broadcasting Stations | 24 | 1099.78 | 217.34 |
| 5399 | Retail - Miscellaneous General Merchandise Stores | 24 | 28794.04 | 1269.33 |
| 5500 | Retail - Auto Dealers & Gasoline Stations | 24 | 3674.47 | 95.03 |
| 5531 | Retail - Auto & Home Supply Stores | 24 | 2134.21 | 504.07 |
| 5700 | Retail - Home Furniture, Furnishings & Equipment Stores | 24 | 451.87 | 27.29 |
| 5731 | Retail - Radio, TV & Consumer Electronics Stores | 24 | 9988.00 | 1267.29 |
| 5812 | Retail - Eating Places | 360 | 1401.46 | 220.35 |
| 5912 | Retail - Drug Stores and Proprietary Stores | 72 | 22690.37 | 5330.04 |
| 7011 | Hotels & Motels | 48 | 2475.56 | 800.65 |
| 7370 | Services - Computer Programming, Data Processing, Etc. | 48 | 9485.26 | 5604.75 |
| 7372 | Services - Prepackaged Software | 48 | 7962.84 | 5216.05 |
| 7510 | Services - Auto Rental & Leasing - except Drivers | 24 | 772.17 | 401.16 |

## Table 31: 4-Digit SIC - Prediction Performance of Traditional Substantive Analytical Models with TCI and without TCI (Models 5, 7, 9 and 11 and 1, 2, 3, and 4)

| | (1) | (5) | | | | (2) | (7) | | | | (3) | (9) | | | | (4) | (11) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Salest-12 | Salest-12+TweetCI | | | | Salest-12+GDPt-12 | Salest-12+TweetCI+GD | | | | Salest-12+AR | Salest-12+AR+TweetCI | | | | Salest-12+AR+GDPt-12 | Salest12+AR+TweetCI+ | | | |
| 4-Digit SIC | MAPE1 | MAPE5 | Differenc | B/W | p-value | MAPE2 | MAPE7 | Differenc | B/W | p-value | MAPE3 | MAPE9 | Differenc | B/W | p-value | MAPE4 | MAPE11 | Differenc | B/W | p-value |
| 2000 | 0.0618 | 0.0492 | 0.0126 | B | 0.0000 | 0.058 | 0.058 | 0.0002 | B | 0.0340 | 0.0370 | 0.0367 | 0.0003 | B | 0.0340 | 0.053 | 0.054 | -0.001 | W | 0.000 |
| 2033 | 0.0429 | 0.0463 | -0.0034 | W | 0.0005 | 0.030 | 0.025 | 0.0053 | B | 0.0005 | 0.0384 | 0.0375 | 0.0009 | B | 0.0005 | 0.033 | 0.024 | 0.009 | B | 0.001 |
| 2040 | 0.0448 | 0.0263 | 0.0185 | B | 0.0000 | 0.030 | 0.028 | 0.0012 | B | 0.0000 | 0.0511 | 0.0253 | 0.0259 | B | 0.0000 | 0.026 | 0.025 | 0.001 | B | 0.000 |
| 2080 | 0.0172 | 0.0198 | -0.0027 | W | 0.0005 | 0.018 | 0.018 | -0.0001 | W | 0.0005 | 0.0173 | 0.0198 | -0.0025 | W | 0.0005 | 0.017 | 0.018 | 0.000 | W | 0.001 |
| 2082 | 0.2499 | 0.2199 | 0.0300 | B | 0.0340 | 0.179 | 0.159 | 0.0200 | B | 0.0340 | 0.1238 | 0.1200 | 0.0037 | B | 0.0000 | 0.090 | 0.085 | 0.005 | B | 0.000 |
| 2086 | 0.0553 | 0.0549 | 0.0004 | B | 0.0000 | 0.034 | 0.028 | 0.0059 | B | 0.0000 | 0.0403 | 0.0416 | -0.0013 | W | 0.0000 | 0.034 | 0.028 | 0.006 | B | 0.077 |
| 2090 | 0.0535 | 0.0527 | 0.0008 | B | 0.0005 | 0.053 | 0.053 | 0.0000 | No Differ | 0.0005 | 0.0370 | 0.0358 | 0.0011 | B | 0.0005 | 0.038 | 0.037 | 0.001 | B | 0.0005 |
| 2111 | 0.0247 | 0.0211 | 0.0037 | B | 0.0005 | 0.025 | 0.020 | 0.0049 | B | 0.0005 | 0.0203 | 0.0187 | 0.0016 | B | 0.0005 | 0.020 | 0.019 | 0.0014 | B | 0.0005 |
| 2300 | 0.0726 | 0.0654 | 0.0072 | B | 0.0000 | 0.058 | 0.055 | 0.0027 | B | 0.0000 | 0.0397 | 0.0377 | 0.0020 | B | 0.0340 | 0.036 | 0.038 | -0.0017 | W | 0.0340 |
| 2320 | 0.1178 | 0.0799 | 0.0378 | B | 0.0005 | 0.061 | 0.061 | 0.0000 | No Differ | 0.0005 | 0.1198 | 0.0853 | 0.0345 | B | 0.0005 | 0.060 | 0.061 | -0.0004 | W | 0.0005 |
| 2840 | 0.0302 | 0.0407 | -0.0105 | W | 0.0005 | 0.050 | 0.051 | -0.0008 | W | 0.0005 | 0.0198 | 0.0204 | -0.0006 | W | 0.0005 | 0.023 | 0.022 | 0.0007 | B | 0.0005 |
| 2842 | 0.0176 | 0.0138 | 0.0038 | B | 0.0005 | 0.015 | 0.014 | 0.0010 | B | 0.0005 | 0.0155 | 0.0141 | 0.0014 | B | 0.0005 | 0.014 | 0.014 | -0.0001 | W | 0.0005 |
| 2844 | 0.0652 | 0.0658 | -0.0006 | W | 0.0772 | 0.044 | 0.040 | 0.0034 | B | 0.0000 | 0.0422 | 0.0437 | -0.0015 | W | 0.0772 | 0.041 | 0.042 | -0.0009 | W | 0.5993 |
| 2890 | 0.0269 | 0.0289 | -0.0020 | W | 0.0005 | 0.021 | 0.021 | -0.0001 | W | 0.0005 | 0.0245 | 0.0253 | -0.0008 | W | 0.0005 | 0.022 | 0.023 | -0.0002 | W | 0.0005 |
| 2911 | 0.1176 | 0.1307 | -0.0131 | W | 0.0340 | 0.161 | 0.129 | 0.0328 | B | 0.0000 | 0.0776 | 0.0794 | -0.0019 | W | 0.0340 | 0.074 | 0.080 | -0.0063 | W | 0.0340 |
| 3021 | 0.0225 | 0.0224 | 0.0000 | No Differ | 0.0005 | 0.021 | 0.021 | 0.0007 | B | 0.0005 | 0.0234 | 0.0235 | -0.0001 | W | 0.0005 | 0.022 | 0.022 | 0.0006 | B | 0.0005 |
| 3100 | 0.1405 | 0.1478 | -0.0073 | W | 0.0005 | 0.140 | 0.144 | -0.0043 | W | 0.0005 | 0.1210 | 0.1239 | -0.0029 | W | 0.0005 | 0.093 | 0.097 | -0.0043 | W | 0.0005 |
| 3140 | 0.0402 | 0.0458 | -0.0056 | W | 0.0005 | 0.022 | 0.023 | -0.0004 | W | 0.0005 | 0.0466 | 0.0467 | -0.0001 | W | 0.0005 | 0.033 | 0.033 | 0.0002 | B | 0.0005 |
| 3540 | 0.0704 | 0.0639 | 0.0065 | B | 0.0005 | 0.055 | 0.055 | -0.0005 | W | 0.0005 | 0.0442 | 0.0451 | -0.0009 | W | 0.0005 | 0.041 | 0.041 | -0.0003 | W | 0.0005 |
| 3577 | 0.1220 | 0.0547 | 0.0673 | B | 0.0005 | 0.048 | 0.047 | 0.0005 | B | 0.0005 | 0.0999 | 0.0479 | 0.0520 | B | 0.0005 | 0.047 | 0.047 | 0.0009 | B | 0.0005 |
| 3630 | 0.0152 | 0.0157 | -0.0005 | W | 0.0005 | 0.016 | 0.026 | -0.0092 | W | 0.0005 | 0.0173 | 0.0243 | -0.0070 | W | 0.0005 | 0.016 | 0.025 | -0.0086 | W | 0.0005 |
| 3663 | 0.1002 | 0.0942 | 0.0060 | B | 0.0340 | 0.073 | 0.062 | 0.0109 | B | 0.0000 | 0.0462 | 0.0346 | 0.0115 | B | 0.0000 | 0.068 | 0.062 | 0.0061 | B | 0.0340 |
| 3674 | 0.0385 | 0.0297 | 0.0088 | B | 0.0005 | 0.017 | 0.018 | -0.0003 | W | 0.0005 | 0.0178 | 0.0188 | -0.0010 | W | 0.0005 | 0.017 | 0.017 | -0.0001 | W | 0.0005 |
| 3711 | 0.1199 | 0.1097 | 0.0102 | B | 0.0000 | 0.088 | 0.085 | 0.0030 | B | 0.0000 | 0.0827 | 0.0826 | 0.0000 | No Differ | 0.6852 | 0.070 | 0.070 | -0.0008 | W | 0.0000 |
| 3751 | 0.12105 | 0.06513 | 0.0559 | B | 0.0005 | 0.08508 | 0.06654 | 0.0185 | B | 0.0005 | 0.1151 | 0.0647 | 0.0504 | B | 0.0005 | 0.08729 | 0.07423 | 0.0131 | B | 0.0005 |
| 3942 | 0.08959 | 0.07609 | 0.0135 | B | 0.0005 | 0.07685 | 0.0766 | 0.0002 | B | 0.0005 | 0.08865 | 0.07392 | 0.0147 | B | 0.0005 | 0.07239 | 0.07395 | -0.0016 | W | 0.0005 |
| 3944 | 0.04401 | 0.04332 | 0.0007 | B | 0.0005 | 0.05708 | 0.05845 | -0.0014 | W | 0.0005 | 0.04044 | 0.04034 | 0.0001 | B | 0.0005 | 0.06038 | 0.06115 | -0.0008 | W | 0.0005 |
| 3949 | 0.14007 | 0.08644 | 0.0536 | B | 0.0005 | 0.08471 | 0.08147 | 0.0032 | B | 0.0005 | 0.15485 | 0.09497 | 0.0599 | B | 0.0005 | 0.08755 | 0.09005 | -0.0025 | W | 0.0005 |
| 4210 | 0.02775 | 0.02339 | 0.0044 | B | 0.0005 | 0.02016 | 0.01959 | 0.0006 | B | 0.0005 | 0.02429 | 0.02116 | 0.0031 | B | 0.0005 | 0.02239 | 0.02088 | 0.0015 | B | 0.0005 |
| 4400 | 0.03178 | 0.02423 | 0.0075 | B | 0.0340 | 0.01861 | 0.01842 | 0.0002 | B | 0.0000 | 0.0219 | 0.02169 | 0.0002 | B | 0.0340 | 0.01868 | 0.01865 | 0.0000 | No Differ | 0.0340 |
| 4512 | 0.05349 | 0.05323 | 0.0003 | B | 0.8827 | 0.0533 | 0.05323 | 0.0001 | B | 0.8827 | 0.04711 | 0.04428 | 0.0028 | B | 0.0021 | 0.04682 | 0.04434 | 0.0025 | B | 0.0000 |
| 4513 | 0.08444 | 0.07462 | 0.0098 | B | 0.0005 | 0.06068 | 0.05539 | 0.0053 | B | 0.0005 | 0.05453 | 0.05017 | 0.0044 | B | 0.0005 | 0.05349 | 0.04915 | 0.0043 | B | 0.0005 |
| 4700 | 0.03815 | 0.03921 | -0.0011 | W | 0.0005 | 0.04049 | 0.04122 | -0.0007 | W | 0.0005 | 0.0441 | 0.04052 | 0.0036 | B | 0.0005 | 0.03817 | 0.03859 | -0.0004 | W | 0.0005 |
| 4812 | 0.04384 | 0.04234 | 0.0015 | B | 0.0005 | 0.03226 | 0.03398 | -0.0017 | W | 0.0005 | 0.03345 | 0.03444 | -0.0010 | W | 0.0005 | 0.0332 | 0.03433 | -0.0011 | W | 0.0005 |
| 4832 | 0.0069 | 0.00634 | 0.0006 | B | 0.0005 | 0.00698 | 0.00645 | 0.0005 | B | 0.0005 | 0.00661 | 0.00609 | 0.0005 | B | 0.0005 | 0.00665 | 0.00615 | 0.0005 | B | 0.0005 |
| 5399 | 0.03932 | 0.04069 | -0.0014 | W | 0.0005 | 0.04108 | 0.04202 | -0.0009 | W | 0.0005 | 0.03836 | 0.04339 | -0.0050 | W | 0.0005 | 0.04393 | 0.04209 | 0.0018 | B | 0.0005 |
| 5500 | 0.02247 | 0.02607 | -0.0036 | W | 0.0005 | 0.01969 | 0.02293 | -0.0032 | W | 0.0005 | 0.02597 | 0.02783 | -0.0019 | W | 0.0005 | 0.02478 | 0.02541 | -0.0006 | W | 0.0005 |
| 5531 | 0.04622 | 0.02924 | 0.0170 | B | 0.0005 | 0.03668 | 0.03679 | -0.0001 | W | 0.0005 | 0.08522 | 0.05256 | 0.0327 | B | 0.0005 | 0.05356 | 0.05309 | 0.0005 | B | 0.0005 |
| 5700 | 0.03958 | 0.03907 | 0.0005 | B | 0.0005 | 0.01759 | 0.01533 | 0.0023 | B | 0.0005 | 0.03969 | 0.03963 | 0.0001 | B | 0.0005 | 0.01755 | 0.01595 | 0.0016 | B | 0.0005 |
| 5731 | 0.12076 | 0.12084 | -0.0001 | W | 0.0005 | 0.11901 | 0.1215 | -0.0025 | W | 0.0005 | 0.12192 | 0.12307 | -0.0012 | W | 0.0005 | 0.12187 | 0.12333 | -0.0015 | W | 0.0005 |
| 5812 | 0.06066 | 0.05908 | 0.0016 | B | 0.0086 | 0.0406 | 0.03976 | 0.0008 | B | 0.0002 | 0.06312 | 0.06017 | 0.0029 | B | 0.0000 | 0.03921 | 0.0379 | 0.0013 | B | 0.0000 |
| 5912 | 0.08672 | 0.08597 | 0.0008 | B | 0.0000 | 0.07177 | 0.07104 | 0.0007 | B | 0.0000 | 0.06594 | 0.0658 | 0.0001 | B | 0.0772 | 0.06457 | 0.06361 | 0.0010 | B | 0.0000 |
| 7011 | 0.11588 | 0.10578 | 0.0101 | B | 0.0340 | 0.07026 | 0.06546 | 0.0048 | B | 0.0340 | 0.04876 | 0.04661 | 0.0021 | B | 0.0340 | 0.04691 | 0.04549 | 0.0014 | B | 0.0340 |
| 7370 | 0.03235 | 0.03064 | 0.0017 | B | 0.0000 | 0.02477 | 0.02101 | 0.0038 | B | 0.0000 | 0.03519 | 0.03403 | 0.0012 | B | 0.0000 | 0.02576 | 0.0207 | 0.0051 | B | 0.0000 |
| 7372 | 0.0429 | 0.04216 | 0.0007 | B | 0.5993 | 0.03719 | 0.03408 | 0.0031 | B | 0.0000 | 0.04432 | 0.043 | 0.0013 | B | 0.0772 | 0.03931 | 0.03492 | 0.0044 | B | 0.0000 |
| 7510 | 0.01532 | 0.01532 | 0.0000 | No Differ | 0.0005 | 0.0142 | 0.01385 | 0.0004 | B | 0.0005 | 0.01306 | 0.01302 | 0.0000 | No Differ | 0.0005 | 0.01124 | 0.00996 | 0.0013 | B | 0.0005 |

## Table 32: 4-Digit SIC - Prediction Performance of Traditional Substantive Analytical Models with TCS and without TCS (Models 6, 8, 10 and 12 and 1, 2, 3, and 4)

| 4-Digit SIC | (1) Salest-12 MAPE1 | (6) Salest-12+TweetCS MAPE6 | Differenc | B/W | p-value | (2) Salest-12+GDPt-12 MAPE2 | (8) Salest-12+TweetCS+G MAPE8 | Differenc | B/W | p-value | (3) Salest-12+AR MAPE3 | (10) Salest-12+AR+TweetC MAPE10 | Differenc | B/W | p-value | (4) Salest-12+AR+GDPt-12 MAPE4 | (12) Salest-12+AR+TweetC MAPE12 | Differenc | B/W | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2000 | 0.0618 | 0.0609 | 0.0009 | B | 0.0000 | 0.0580 | 0.0574 | 0.0007 | B | 0.0000 | 0.0370 | 0.0399 | -0.0029 | W | 0.0340 | 0.0533 | 0.0523 | 0.0010 | B | 0.0340 |
| 2033 | 0.0429 | 0.0489 | -0.0060 | W | 0.0005 | 0.0301 | 0.0361 | -0.0060 | W | 0.0005 | 0.0384 | 0.0347 | 0.0037 | B | 0.0005 | 0.0331 | 0.0313 | 0.0018 | B | 0.0005 |
| 2040 | 0.0448 | 0.0320 | 0.0127 | B | 0.0000 | 0.0295 | 0.0290 | 0.0006 | B | 0.0340 | 0.0511 | 0.0367 | 0.0144 | B | 0.0000 | 0.0260 | 0.0254 | 0.0006 | B | 0.0340 |
| 2080 | 0.0172 | 0.0153 | 0.0019 | B | 0.0005 | 0.0177 | 0.0198 | -0.0021 | W | 0.0005 | 0.0173 | 0.0143 | 0.0030 | B | 0.0005 | 0.0173 | 0.0190 | -0.0018 | W | 0.0005 |
| 2082 | 0.2499 | 0.2460 | 0.0039 | B | 0.0000 | 0.1786 | 0.1564 | 0.0221 | B | 0.0000 | 0.1238 | 0.1228 | 0.0009 | B | 0.0340 | 0.0901 | 0.0868 | 0.0032 | B | 0.0000 |
| 2086 | 0.0553 | 0.0387 | 0.0167 | B | 0.0772 | 0.0343 | 0.0317 | 0.0026 | B | 0.0000 | 0.0403 | 0.0340 | 0.0063 | B | 0.0772 | 0.0344 | 0.0317 | 0.0026 | B | 0.0000 |
| 2090 | 0.0535 | 0.0534 | 0.0002 | B | 0.0005 | 0.0530 | 0.0533 | -0.0003 | W | 0.0005 | 0.0370 | 0.0377 | -0.0007 | W | 0.0005 | 0.0380 | 0.0380 | 0.0000 | No Differ | 0.0005 |
| 2111 | 0.0247 | 0.0237 | 0.0011 | B | 0.0005 | 0.0252 | 0.0241 | 0.0011 | B | 0.0005 | 0.0203 | 0.0190 | 0.0013 | B | 0.0005 | 0.0204 | 0.0199 | 0.0005 | B | 0.0005 |
| 2300 | 0.0726 | 0.0715 | 0.0011 | B | 0.0340 | 0.0578 | 0.0564 | 0.0014 | B | 0.0340 | 0.0397 | 0.0373 | 0.0024 | B | 0.0000 | 0.0360 | 0.0353 | 0.0007 | B | 0.0000 |
| 2320 | 0.1178 | 0.0864 | 0.0313 | B | 0.0005 | 0.0607 | 0.0595 | 0.0012 | B | 0.0005 | 0.1198 | 0.0864 | 0.0334 | B | 0.0005 | 0.0603 | 0.0599 | 0.0004 | B | 0.0005 |
| 2840 | 0.0302 | 0.0290 | 0.0012 | B | 0.0005 | 0.0502 | 0.0484 | 0.0018 | B | 0.0005 | 0.0198 | 0.0189 | 0.0009 | B | 0.0005 | 0.0226 | 0.0218 | 0.0008 | B | 0.0005 |
| 2842 | 0.0176 | 0.0174 | 0.0002 | B | 0.0005 | 0.0154 | 0.0155 | -0.0001 | W | 0.0005 | 0.0155 | 0.0155 | 0.0000 | No Differ | 0.0005 | 0.0143 | 0.0146 | -0.0002 | W | 0.0005 |
| 2844 | 0.0652 | 0.0607 | 0.0045 | B | 0.5993 | 0.0437 | 0.0396 | 0.0041 | B | 0.0000 | 0.0422 | 0.0447 | -0.0025 | W | 0.0000 | 0.0406 | 0.0396 | 0.0010 | B | 0.0000 |
| 2890 | 0.0269 | 0.0248 | 0.0020 | B | 0.0005 | 0.0209 | 0.0209 | 0.0000 | W | 0.0005 | 0.0245 | 0.0243 | 0.0002 | B | 0.0005 | 0.0224 | 0.0224 | 0.0001 | B | 0.0005 |
| 2911 | 0.1176 | 0.1184 | -0.0008 | W | 0.0340 | 0.1613 | 0.1622 | -0.0009 | W | 0.0000 | 0.0776 | 0.0789 | -0.0013 | W | 0.0000 | 0.0738 | 0.0741 | -0.0003 | W | 0.0000 |
| 3021 | 0.0225 | 0.0224 | 0.0000 | No Differ | 0.0005 | 0.0214 | 0.0215 | 0.0000 | No Differ | 0.0005 | 0.0234 | 0.0234 | 0.0000 | No Differ | 0.0005 | 0.0224 | 0.0224 | 0.0000 | No Differ | 0.0005 |
| 3100 | 0.1405 | 0.1404 | 0.0001 | B | 0.0005 | 0.1402 | 0.1393 | 0.0010 | B | 0.0005 | 0.1210 | 0.1243 | -0.0033 | W | 0.0005 | 0.0928 | 0.0925 | 0.0004 | B | 0.0005 |
| 3140 | 0.0402 | 0.0422 | -0.0020 | W | 0.0005 | 0.0224 | 0.0226 | -0.0002 | W | 0.0005 | 0.0466 | 0.0440 | 0.0026 | B | 0.0005 | 0.0331 | 0.0356 | -0.0025 | W | 0.0005 |
| 3540 | 0.0704 | 0.0704 | 0.0000 | No Differ | 0.0005 | 0.0548 | 0.0466 | 0.0082 | B | 0.0005 | 0.0442 | 0.0419 | 0.0023 | B | 0.0005 | 0.0410 | 0.0367 | 0.0043 | B | 0.0005 |
| 3577 | 0.1220 | 0.1181 | 0.0039 | B | 0.0005 | 0.0477 | 0.0501 | -0.0024 | W | 0.0005 | 0.0999 | 0.0968 | 0.0031 | B | 0.0005 | 0.0474 | 0.0505 | -0.0032 | W | 0.0005 |
| 3630 | 0.0152 | 0.0152 | 0.0000 | No Differ | 0.0005 | 0.0164 | 0.0194 | -0.0030 | W | 0.0005 | 0.0173 | 0.0174 | -0.0001 | W | 0.0005 | 0.0160 | 0.0168 | -0.0009 | W | 0.0005 |
| 3663 | 0.1002 | 0.0713 | 0.0289 | B | 0.0000 | 0.0730 | 0.0847 | -0.0117 | W | 0.0000 | 0.0462 | 0.0695 | -0.0234 | W | 0.0340 | 0.0679 | 0.0735 | -0.0056 | W | 0.0340 |
| 3674 | 0.0385 | 0.0341 | 0.0044 | B | 0.0005 | 0.0174 | 0.0184 | -0.0010 | W | 0.0005 | 0.0178 | 0.0199 | -0.0021 | W | 0.0005 | 0.0171 | 0.0176 | -0.0006 | W | 0.0005 |
| 3711 | 0.1199 | 0.1136 | 0.0063 | B | 0.0002 | 0.0881 | 0.0879 | 0.0002 | B | 0.0137 | 0.0827 | 0.0816 | 0.0010 | B | 0.0002 | 0.0695 | 0.0685 | 0.0011 | B | 0.0137 |
| 3751 | 0.12105 | 0.07062 | 0.0504 | B | 0.0005 | 0.08508 | 0.0632 | 0.0219 | B | 0.0005 | 0.1151 | 0.07063 | 0.0445 | B | 0.0005 | 0.08729 | 0.06103 | 0.0263 | B | 0.0005 |
| 3942 | 0.08959 | 0.0499 | 0.0397 | B | 0.0005 | 0.07685 | 0.05499 | 0.0219 | B | 0.0005 | 0.08865 | 0.04952 | 0.0391 | B | 0.0005 | 0.07239 | 0.05346 | 0.0189 | B | 0.0005 |
| 3944 | 0.04401 | 0.04914 | -0.0051 | W | 0.0005 | 0.05708 | 0.0506 | 0.0065 | B | 0.0005 | 0.04044 | 0.04759 | -0.0071 | W | 0.0005 | 0.06038 | 0.05529 | 0.0051 | B | 0.0005 |
| 3949 | 0.14007 | 0.1345 | 0.0056 | B | 0.0005 | 0.08471 | 0.08075 | 0.0040 | B | 0.0005 | 0.15485 | 0.14444 | 0.0104 | B | 0.0005 | 0.08755 | 0.07724 | 0.0103 | B | 0.0005 |
| 4210 | 0.02775 | 0.02691 | 0.0008 | B | 0.0005 | 0.02016 | 0.01605 | 0.0041 | B | 0.0005 | 0.02429 | 0.02477 | -0.0005 | W | 0.0005 | 0.02239 | 0.02062 | 0.0018 | B | 0.0005 |
| 4400 | 0.03178 | 0.02987 | 0.0019 | B | 0.0340 | 0.01861 | 0.01916 | -0.0006 | W | 0.0000 | 0.0219 | 0.02252 | -0.0006 | W | 0.0000 | 0.01868 | 0.01956 | -0.0009 | W | 0.0000 |
| 4512 | 0.05349 | 0.04977 | 0.0037 | B | 0.2543 | 0.0533 | 0.05 | 0.0033 | B | 0.0000 | 0.04711 | 0.04686 | 0.0003 | B | 0.0744 | 0.04682 | 0.04555 | 0.0013 | B | 0.0000 |
| 4513 | 0.08444 | 0.08139 | 0.0031 | B | 0.0005 | 0.06068 | 0.05921 | 0.0015 | B | 0.0005 | 0.05453 | 0.0547 | -0.0002 | W | 0.0005 | 0.05349 | 0.0537 | -0.0002 | W | 0.0005 |
| 4700 | 0.03815 | 0.0398 | -0.0017 | W | 0.0005 | 0.04049 | 0.04555 | -0.0051 | W | 0.0005 | 0.0441 | 0.04233 | 0.0018 | B | 0.0005 | 0.03817 | 0.0453 | -0.0071 | W | 0.0005 |
| 4812 | 0.04384 | 0.06322 | -0.0194 | W | 0.0005 | 0.03226 | 0.04155 | -0.0093 | W | 0.0005 | 0.03345 | 0.04341 | -0.0100 | W | 0.0005 | 0.0332 | 0.04185 | -0.0086 | W | 0.0005 |
| 4832 | 0.0069 | 0.00689 | 0.0000 | No Differ | 0.0005 | 0.00698 | 0.00694 | 0.0000 | No Differ | 0.0005 | 0.00661 | 0.00667 | -0.0001 | W | 0.0005 | 0.00665 | 0.00668 | 0.0000 | No Differ | 0.0005 |
| 5399 | 0.03932 | 0.0408 | -0.0015 | W | 0.0005 | 0.04108 | 0.04063 | 0.0005 | B | 0.0005 | 0.03836 | 0.03995 | -0.0016 | W | 0.0005 | 0.04393 | 0.04125 | 0.0027 | B | 0.0005 |
| 5500 | 0.02247 | 0.02328 | -0.0008 | W | 0.0005 | 0.01969 | 0.02122 | -0.0015 | W | 0.0005 | 0.02597 | 0.02629 | -0.0003 | W | 0.0005 | 0.02478 | 0.02485 | -0.0001 | W | 0.0005 |
| 5531 | 0.04622 | 0.048 | -0.0018 | W | 0.0005 | 0.03668 | 0.03905 | -0.0024 | W | 0.0005 | 0.08522 | 0.07263 | 0.0126 | B | 0.0005 | 0.05356 | 0.04542 | 0.0081 | B | 0.0005 |
| 5700 | 0.03958 | 0.02836 | 0.0112 | B | 0.0005 | 0.01759 | 0.01981 | -0.0022 | W | 0.0005 | 0.03969 | 0.0284 | 0.0113 | B | 0.0005 | 0.01755 | 0.01834 | -0.0008 | W | 0.0005 |
| 5731 | 0.12076 | 0.11545 | 0.0053 | B | 0.0005 | 0.11901 | 0.09773 | 0.0213 | B | 0.0005 | 0.12192 | 0.11862 | 0.0033 | B | 0.0005 | 0.12187 | 0.09766 | 0.0242 | B | 0.0005 |
| 5812 | 0.06066 | 0.06216 | -0.0015 | W | 0.0000 | 0.0406 | 0.03997 | 0.0006 | B | 0.3261 | 0.06312 | 0.05876 | 0.0044 | B | 0.0000 | 0.03921 | 0.03841 | 0.0008 | B | 0.0266 |
| 5912 | 0.08672 | 0.09786 | -0.0111 | W | 0.0000 | 0.07177 | 0.06709 | 0.0047 | B | 0.0000 | 0.06594 | 0.06617 | -0.0002 | W | 0.5993 | 0.06457 | 0.06376 | 0.0008 | B | 0.0772 |
| 7011 | 0.11588 | 0.05975 | 0.0561 | B | 0.0000 | 0.07026 | 0.05761 | 0.0127 | B | 0.0000 | 0.04876 | 0.04596 | 0.0028 | B | 0.0000 | 0.04691 | 0.04562 | 0.0013 | B | 0.0000 |
| 7370 | 0.03235 | 0.02697 | 0.0054 | B | 0.0000 | 0.02477 | 0.0238 | 0.0010 | B | 0.0000 | 0.03519 | 0.02955 | 0.0056 | B | 0.0000 | 0.02576 | 0.02684 | -0.0011 | W | 0.0000 |
| 7372 | 0.0429 | 0.03987 | 0.0030 | B | 0.0000 | 0.03719 | 0.03611 | 0.0011 | B | 0.0000 | 0.04432 | 0.03949 | 0.0048 | B | 0.5993 | 0.03931 | 0.03764 | 0.0017 | B | 0.0000 |
| 7510 | 0.01532 | 0.01681 | -0.0015 | W | 0.0005 | 0.0142 | 0.01631 | -0.0021 | W | 0.0005 | 0.01306 | 0.01537 | -0.0023 | W | 0.0005 | 0.01124 | 0.01441 | -0.0032 | W | 0.0005 |

# Table 33: 4-Digit SIC - Prediction Performance of Continuous Substantive Analytical Models with TCI and without TCI (Models 5, 7, 9 and 11 and 1, 2, 3, and 4)

| | (1) | (5) | | | | (2) | (7) | | | | (3) | (9) | | | | (4) | (11) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 | Saletst-1+TweetCI | | | | Saletst-1+GDPt-1 | Saletst-1+TweetCI+GDPt-1 | | | | Saletst-1+AR | Saletst-1+AR+TweetCI | | | | Saletst-1+AR+GDPt-1 | Salest12+AR+TweetCI+GDPt-1 | | | |
| 4-Digit SIC | MAPE1 | MAPE5 | Difference | B/W | p-value | MAPE2 | MAPE7 | Difference | B/W | p-value | MAPE3 | MAPE9 | Difference | B/W | p-value | MAPE4 | MAPE11 | Difference | B/W | p-value |
| 2000 | 0.0840 | 0.0613 | 0.0227 | B | 0.0000 | 0.0665 | 0.0592 | 0.0073 | B | 0.0000 | 0.0547 | 0.0544 | 0.0003 | B | 0.0340 | 0.0618 | 0.0516 | 0.0102 | B | 0.0340 |
| 2033 | 0.1089 | 0.1094 | -0.0005 | W | 0.0005 | 0.0800 | 0.0229 | 0.0571 | B | 0.0005 | 0.0850 | 0.0823 | 0.0027 | B | 0.0005 | 0.0746 | 0.0246 | 0.0500 | B | 0.0005 |
| 2040 | 0.0554 | 0.0494 | 0.0060 | B | 0.0000 | 0.0417 | 0.0277 | 0.0140 | B | 0.0000 | 0.0612 | 0.0525 | 0.0087 | B | 0.0340 | 0.0461 | 0.0252 | 0.0208 | B | 0.0000 |
| 2080 | 0.1423 | 0.1365 | 0.0059 | B | 0.0005 | 0.1418 | 0.0171 | 0.1247 | B | 0.0005 | 0.1520 | 0.1497 | 0.0023 | B | 0.0005 | 0.1540 | 0.0169 | 0.1371 | B | 0.0005 |
| 2082 | 0.1399 | 0.1171 | 0.0229 | B | 0.0000 | 0.1313 | 0.1536 | -0.0223 | W | 0.0340 | 0.1133 | 0.0985 | 0.0148 | B | 0.0000 | 0.1020 | 0.0821 | 0.0198 | B | 0.0000 |
| 2086 | 0.0980 | 0.0989 | -0.0009 | W | 0.0772 | 0.0771 | 0.0268 | 0.0503 | B | 0.0000 | 0.0753 | 0.0740 | 0.0014 | B | 0.0000 | 0.0716 | 0.0269 | 0.0447 | B | 0.0000 |
| 2090 | 0.1139 | 0.1076 | 0.0063 | B | 0.0005 | 0.0944 | 0.0528 | 0.0416 | B | 0.0005 | 0.0725 | 0.0704 | 0.0021 | B | 0.0005 | 0.0881 | 0.0374 | 0.0507 | B | 0.0005 |
| 2111 | 0.0577 | 0.0565 | 0.0012 | B | 0.0005 | 0.0492 | 0.0199 | 0.0292 | B | 0.0005 | 0.0576 | 0.0565 | 0.0011 | B | 0.0005 | 0.0472 | 0.0190 | 0.0282 | B | 0.0005 |
| 2300 | 0.1409 | 0.1408 | 0.0001 | B | 0.0340 | 0.1229 | 0.0560 | 0.0668 | B | 0.0000 | 0.0750 | 0.0680 | 0.0070 | B | 0.0000 | 0.0806 | 0.0366 | 0.0440 | B | 0.0000 |
| 2320 | 0.1499 | 0.1295 | 0.0204 | B | 0.0005 | 0.1357 | 0.0608 | 0.0748 | B | 0.0005 | 0.1539 | 0.1219 | 0.0320 | B | 0.0005 | 0.1111 | 0.0605 | 0.0506 | B | 0.0005 |
| 2840 | 0.0444 | 0.0445 | -0.0001 | W | 0.0005 | 0.0538 | 0.0481 | 0.0057 | B | 0.0005 | 0.0222 | 0.0221 | 0.0000 | B | 0.0005 | 0.0228 | 0.0210 | 0.0019 | B | 0.0005 |
| 2842 | 0.0745 | 0.0715 | 0.0030 | B | 0.0005 | 0.0693 | 0.0144 | 0.0549 | B | 0.0005 | 0.0719 | 0.0706 | 0.0013 | B | 0.0005 | 0.0717 | 0.0144 | 0.0573 | B | 0.0005 |
| 2844 | 0.0985 | 0.0987 | -0.0002 | W | 0.5993 | 0.0620 | 0.0415 | 0.0205 | B | 0.0000 | 0.0533 | 0.0552 | -0.0019 | W | 0.0000 | 0.0542 | 0.0413 | 0.0128 | B | 0.0000 |
| 2890 | 0.0264 | 0.0281 | -0.0017 | W | 0.0005 | 0.0187 | 0.0206 | -0.0019 | W | 0.0005 | 0.0190 | 0.0193 | -0.0003 | W | 0.0005 | 0.0182 | 0.0226 | -0.0044 | W | 0.0005 |
| 2911 | 0.0578 | 0.0573 | 0.0005 | B | 0.0340 | 0.0825 | 0.1403 | -0.0578 | W | 0.0000 | 0.0551 | 0.0581 | -0.0029 | W | 0.0000 | 0.0553 | 0.0791 | -0.0239 | W | 0.0000 |
| 3021 | 0.0498 | 0.0508 | -0.0010 | W | 0.0005 | 0.0349 | 0.0211 | 0.0137 | B | 0.0005 | 0.0372 | 0.0371 | 0.0001 | B | 0.0005 | 0.0351 | 0.0216 | 0.0135 | B | 0.0005 |
| 3100 | 0.1636 | 0.1670 | -0.0034 | W | 0.0005 | 0.1640 | 0.1444 | 0.0196 | B | 0.0005 | 0.1626 | 0.1629 | -0.0003 | W | 0.0005 | 0.1363 | 0.0945 | 0.0418 | B | 0.0005 |
| 3140 | 0.1736 | 0.1691 | 0.0044 | B | 0.0005 | 0.1183 | 0.0188 | 0.0995 | B | 0.0005 | 0.0925 | 0.0860 | 0.0065 | B | 0.0005 | 0.0771 | 0.0345 | 0.0426 | B | 0.0005 |
| 3540 | 0.0819 | 0.0811 | 0.0008 | B | 0.0005 | 0.0650 | 0.0536 | 0.0114 | B | 0.0005 | 0.0717 | 0.0739 | -0.0021 | W | 0.0005 | 0.0597 | 0.0404 | 0.0193 | B | 0.0005 |
| 3577 | 0.1567 | 0.1399 | 0.0168 | B | 0.0005 | 0.1463 | 0.0450 | 0.1013 | B | 0.0005 | 0.0673 | 0.0702 | -0.0029 | W | 0.0005 | 0.0714 | 0.0465 | 0.0248 | B | 0.0005 |
| 3630 | 0.0800 | 0.0750 | 0.0050 | B | 0.0005 | 0.0512 | 0.0280 | 0.0232 | B | 0.0005 | 0.0551 | 0.0556 | -0.0005 | W | 0.0005 | 0.0517 | 0.0268 | 0.0249 | B | 0.0005 |
| 3663 | 0.1295 | 0.1232 | 0.0063 | B | 0.0340 | 0.1674 | 0.0539 | 0.1135 | B | 0.0000 | 0.1411 | 0.1374 | 0.0037 | B | 0.0340 | 0.1586 | 0.0555 | 0.1030 | B | 0.0340 |
| 3674 | 0.0689 | 0.0606 | 0.0082 | B | 0.0005 | 0.0500 | 0.0181 | 0.0320 | B | 0.0005 | 0.0445 | 0.0437 | 0.0008 | B | 0.0005 | 0.0454 | 0.0180 | 0.0274 | B | 0.0005 |
| 3711 | 0.0911 | 0.0877 | 0.0034 | B | 0.0000 | 0.0652 | 0.0833 | -0.0181 | W | 0.0000 | 0.0831 | 0.0838 | -0.0008 | W | 0.0037 | 0.0629 | 0.0704 | -0.0075 | W | 0.0037 |
| 3751 | 0.2389 | 0.2585 | -0.0196 | W | 0.0005 | 0.2240 | 0.0658 | 0.1582 | B | 0.0005 | 0.2423 | 0.2309 | 0.0114 | B | 0.0005 | 0.1812 | 0.0737 | 0.1075 | B | 0.0005 |
| 3942 | 0.4645 | 0.4330 | 0.0315 | B | 0.0005 | 0.4506 | 0.0770 | 0.3736 | B | 0.0005 | 0.2670 | 0.3071 | -0.0400 | W | 0.0005 | 0.2955 | 0.0738 | 0.2218 | B | 0.0005 |
| 3944 | 0.3263 | 0.3152 | 0.0110 | B | 0.0005 | 0.3727 | 0.0593 | 0.3133 | B | 0.0005 | 0.2120 | 0.2075 | 0.0045 | B | 0.0005 | 0.1511 | 0.0615 | 0.0896 | B | 0.0005 |
| 3949 | 0.2007 | 0.1538 | 0.0469 | B | 0.0005 | 0.1999 | 0.0805 | 0.1194 | B | 0.0005 | 0.1606 | 0.1121 | 0.0486 | B | 0.0005 | 0.0926 | 0.0889 | 0.0036 | B | 0.0005 |
| 4210 | 0.0754 | 0.0415 | 0.0339 | B | 0.0005 | 0.0560 | 0.0182 | 0.0378 | B | 0.0005 | 0.0605 | 0.0442 | 0.0163 | B | 0.0005 | 0.0532 | 0.0201 | 0.0331 | B | 0.0005 |
| 4400 | 0.1499 | 0.1396 | 0.0103 | B | 0.0340 | 0.1525 | 0.0184 | 0.1341 | B | 0.0000 | 0.1548 | 0.1438 | 0.0110 | B | 0.0000 | 0.1523 | 0.0185 | 0.1338 | B | 0.0000 |
| 4512 | 0.0790 | 0.0681 | 0.0109 | B | 0.0000 | 0.0682 | 0.0538 | 0.0143 | B | 0.0000 | 0.0739 | 0.0652 | 0.0088 | B | 0.0000 | 0.0648 | 0.0448 | 0.0200 | B | 0.0000 |
| 4513 | 0.0428 | 0.0408 | 0.0020 | B | 0.0005 | 0.0395 | 0.0554 | -0.0159 | W | 0.0005 | 0.0397 | 0.0407 | -0.0010 | W | 0.0005 | 0.0392 | 0.0492 | -0.0100 | W | 0.0005 |
| 4700 | 0.1753 | 0.1294 | 0.0459 | B | 0.0005 | 0.1318 | 0.0404 | 0.0914 | B | 0.0005 | 0.1278 | 0.1261 | 0.0017 | B | 0.0005 | 0.1249 | 0.0405 | 0.0844 | B | 0.0005 |
| 4812 | 0.0424 | 0.0417 | 0.0006 | B | 0.0005 | 0.0422 | 0.0341 | 0.0081 | B | 0.0005 | 0.0420 | 0.0417 | 0.0004 | B | 0.0005 | 0.0422 | 0.0343 | 0.0079 | B | 0.0005 |
| 4832 | 0.0187 | 0.0158 | 0.0029 | B | 0.0005 | 0.0180 | 0.0058 | 0.0122 | B | 0.0005 | 0.0188 | 0.0159 | 0.0030 | B | 0.0005 | 0.0176 | 0.0058 | 0.0118 | B | 0.0005 |
| 5399 | 0.1947 | 0.1184 | 0.0764 | B | 0.0005 | 0.1801 | 0.0418 | 0.1383 | B | 0.0005 | 0.1899 | 0.1250 | 0.0649 | B | 0.0005 | 0.1810 | 0.0434 | 0.1376 | B | 0.0005 |
| 5500 | 0.0688 | 0.0426 | 0.0262 | B | 0.0005 | 0.0590 | 0.0208 | 0.0382 | B | 0.0005 | 0.0609 | 0.0427 | 0.0181 | B | 0.0005 | 0.0591 | 0.0232 | 0.0359 | B | 0.0005 |
| 5531 | 0.1488 | 0.1529 | -0.0041 | W | 0.0005 | 0.1833 | 0.0465 | 0.1368 | B | 0.0005 | 0.1253 | 0.1096 | 0.0156 | B | 0.0005 | 0.0794 | 0.0584 | 0.0210 | B | 0.0005 |
| 5700 | 0.0989 | 0.0989 | 0.0000 | No Differe | 0.0005 | 0.1129 | 0.0128 | 0.1001 | B | 0.0005 | 0.0994 | 0.0990 | 0.0004 | B | 0.0005 | 0.1142 | 0.0125 | 0.1017 | B | 0.0005 |
| 5731 | 0.1423 | 0.1096 | 0.0327 | B | 0.0005 | 0.0844 | 0.1218 | -0.0374 | W | 0.0005 | 0.1045 | 0.1076 | -0.0031 | W | 0.0005 | 0.0984 | 0.1234 | -0.0250 | W | 0.0005 |
| 5812 | 0.0687 | 0.0669 | 0.0017 | B | 0.0000 | 0.0555 | 0.0416 | 0.0139 | B | 0.0000 | 0.0626 | 0.0579 | 0.0048 | B | 0.0000 | 0.0550 | 0.0386 | 0.0164 | B | 0.0000 |
| 5912 | 0.2546 | 0.2515 | 0.0031 | B | 0.0000 | 0.2451 | 0.0740 | 0.1710 | B | 0.0772 | 0.2532 | 0.2511 | 0.0021 | B | 0.0000 | 0.2430 | 0.0634 | 0.1796 | B | 0.0000 |
| 7011 | 0.0530 | 0.0543 | -0.0013 | W | 0.0000 | 0.0498 | 0.0639 | -0.0141 | W | 0.0340 | 0.0532 | 0.0519 | 0.0013 | B | 0.0340 | 0.0491 | 0.0453 | 0.0037 | B | 0.0340 |
| 7370 | 0.1199 | 0.1065 | 0.0134 | B | 0.0340 | 0.1010 | 0.0275 | 0.0735 | B | 0.0000 | 0.0900 | 0.0881 | 0.0019 | B | 0.0000 | 0.0905 | 0.0274 | 0.0631 | B | 0.0000 |
| 7372 | 0.0616 | 0.0623 | -0.0007 | W | 0.0772 | 0.0546 | 0.0317 | 0.0229 | B | 0.0000 | 0.0670 | 0.0666 | 0.0004 | B | 0.5993 | 0.0587 | 0.0333 | 0.0254 | B | 0.0000 |
| 7510 | 0.1525 | 0.0922 | 0.0602 | B | 0.0005 | 0.1469 | 0.0143 | 0.1326 | B | 0.0005 | 0.0901 | 0.0754 | 0.0147 | B | 0.0005 | 0.0919 | 0.0107 | 0.0812 | B | 0.0005 |

# Table 34: 4-Digit SIC - Prediction Performance of Continuous Substantive Analytical Models with TCS and without TCS (Models 6, 8, 10 and 12 and 1, 2, 3, and 4)

| | (1) | (6) | | | | (2) | (8) | | | | (3) | (10) | | | | (4) | (12) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 | Saletst-1+TweetCS | | | | Saletst-1+GDPt-1 | Saletst-1+TweetCS+GDPt-1 | | | | Saletst-1+AR | Saletst-1+AR+TweetCS | | | | Saletst-1+AR+GDPt-1 | Saletst-1+AR+TweetCS+GDPt-1 | | | |
| 4-Digit SIC | MAPE1 | MAPE6 | Difference | B/W | p-value | MAPE2 | MAPE8 | Difference | B/W | p-value | MAPE3 | MAPE10 | Difference | B/W | p-value | MAPE4 | MAPE12 | Difference | B/W | p-value |
| 2000 | 0.0840 | 0.0633 | 0.0207 | B | 0.0000 | 0.0665 | 0.0592 | 0.0072 | B | 0.0000 | 0.0547 | 0.0565 | -0.0018 | W | 0.0340 | 0.0618 | 0.0492 | 0.0126 | B | 0.0000 |
| 2033 | 0.1089 | 0.1009 | 0.0080 | B | 0.0005 | 0.0800 | 0.0360 | 0.0441 | B | 0.0005 | 0.0850 | 0.0813 | 0.0036 | B | 0.0005 | 0.0746 | 0.0310 | 0.0437 | B | 0.0005 |
| 2040 | 0.0554 | 0.0464 | 0.0090 | B | 0.0340 | 0.0417 | 0.0296 | 0.0121 | B | 0.0340 | 0.0612 | 0.0508 | 0.0104 | B | 0.0000 | 0.0461 | 0.0271 | 0.0190 | B | 0.0000 |
| 2080 | 0.1423 | 0.1517 | -0.0094 | W | 0.0005 | 0.1418 | 0.0189 | 0.1228 | B | 0.0005 | 0.1520 | 0.1437 | 0.0083 | B | 0.0005 | 0.1540 | 0.0181 | 0.1358 | B | 0.0005 |
| 2082 | 0.1399 | 0.1238 | 0.0161 | B | 0.0000 | 0.1313 | 0.1561 | -0.0248 | W | 0.0340 | 0.1133 | 0.1085 | 0.0048 | B | 0.0000 | 0.1020 | 0.0858 | 0.0161 | B | 0.0000 |
| 2086 | 0.0980 | 0.0785 | 0.0196 | B | 0.0000 | 0.0771 | 0.0310 | 0.0461 | B | 0.0000 | 0.0753 | 0.0723 | 0.0031 | B | 0.0772 | 0.0716 | 0.0311 | 0.0405 | B | 0.0000 |
| 2090 | 0.1139 | 0.1048 | 0.0091 | B | 0.0005 | 0.0944 | 0.0529 | 0.0415 | B | 0.0005 | 0.0725 | 0.0831 | -0.0105 | W | 0.0005 | 0.0881 | 0.0383 | 0.0498 | B | 0.0005 |
| 2111 | 0.0577 | 0.0591 | -0.0014 | W | 0.0005 | 0.0492 | 0.0242 | 0.0249 | B | 0.0005 | 0.0576 | 0.0584 | -0.0008 | W | 0.0005 | 0.0472 | 0.0201 | 0.0271 | B | 0.0005 |
| 2300 | 0.1409 | 0.1325 | 0.0084 | B | 0.0000 | 0.1229 | 0.0568 | 0.0660 | B | 0.0000 | 0.0750 | 0.0730 | 0.0019 | B | 0.0340 | 0.0806 | 0.0355 | 0.0452 | B | 0.0000 |
| 2320 | 0.1499 | 0.1443 | 0.0057 | B | 0.0005 | 0.1357 | 0.0596 | 0.0761 | B | 0.0005 | 0.1539 | 0.1470 | 0.0069 | B | 0.0005 | 0.1111 | 0.0597 | 0.0514 | B | 0.0005 |
| 2840 | 0.0444 | 0.0467 | -0.0023 | W | 0.0005 | 0.0538 | 0.0453 | 0.0084 | B | 0.0005 | 0.0222 | 0.0238 | -0.0016 | W | 0.0005 | 0.0228 | 0.0204 | 0.0024 | B | 0.0005 |
| 2842 | 0.0745 | 0.0752 | -0.0007 | W | 0.0005 | 0.0693 | 0.0155 | 0.0538 | B | 0.0005 | 0.0719 | 0.0710 | 0.0009 | B | 0.0005 | 0.0717 | 0.0145 | 0.0572 | B | 0.0005 |
| 2844 | 0.0985 | 0.0933 | 0.0051 | B | 0.0772 | 0.0620 | 0.0399 | 0.0221 | B | 0.0000 | 0.0533 | 0.0528 | 0.0004 | B | 0.0772 | 0.0542 | 0.0392 | 0.0150 | B | 0.0000 |
| 2890 | 0.0264 | 0.0218 | 0.0046 | B | 0.0005 | 0.0187 | 0.0206 | -0.0019 | W | 0.0005 | 0.0190 | 0.0187 | 0.0002 | B | 0.0005 | 0.0182 | 0.0223 | -0.0040 | W | 0.0005 |
| 2911 | 0.0578 | 0.0463 | 0.0116 | B | 0.0000 | 0.0825 | 0.1715 | -0.0889 | W | 0.0000 | 0.0551 | 0.0558 | -0.0006 | W | 0.0000 | 0.0553 | 0.0730 | -0.0178 | W | 0.0000 |
| 3021 | 0.0498 | 0.0449 | 0.0049 | B | 0.0005 | 0.0349 | 0.0216 | 0.0132 | B | 0.0005 | 0.0372 | 0.0370 | 0.0002 | B | 0.0005 | 0.0351 | 0.0222 | 0.0129 | B | 0.0005 |
| 3100 | 0.1636 | 0.1624 | 0.0012 | B | 0.0005 | 0.1640 | 0.1390 | 0.0250 | B | 0.0005 | 0.1626 | 0.1543 | 0.0083 | B | 0.0005 | 0.1363 | 0.0904 | 0.0458 | B | 0.0005 |
| 3140 | 0.1736 | 0.1494 | 0.0242 | B | 0.0005 | 0.1183 | 0.0294 | 0.0889 | B | 0.0005 | 0.0925 | 0.0947 | -0.0022 | W | 0.0005 | 0.0771 | 0.0350 | 0.0421 | B | 0.0005 |
| 3540 | 0.0819 | 0.0847 | -0.0028 | W | 0.0005 | 0.0650 | 0.0448 | 0.0202 | B | 0.0005 | 0.0717 | 0.0744 | -0.0026 | W | 0.0005 | 0.0597 | 0.0362 | 0.0235 | B | 0.0005 |
| 3577 | 0.1567 | 0.1590 | -0.0023 | W | 0.0005 | 0.1463 | 0.0482 | 0.0981 | B | 0.0005 | 0.0673 | 0.0719 | -0.0045 | W | 0.0005 | 0.0714 | 0.0488 | 0.0226 | B | 0.0005 |
| 3630 | 0.0800 | 0.0777 | 0.0023 | B | 0.0005 | 0.0512 | 0.0221 | 0.0291 | B | 0.0005 | 0.0551 | 0.0551 | 0.0000 | No Differe | 0.0005 | 0.0517 | 0.0179 | 0.0338 | B | 0.0005 |
| 3663 | 0.1295 | 0.1668 | -0.0373 | W | 0.0000 | 0.1674 | 0.0739 | 0.0935 | B | 0.0340 | 0.1411 | 0.1663 | -0.0252 | W | 0.0000 | 0.1586 | 0.0704 | 0.0881 | B | 0.0340 |
| 3674 | 0.0689 | 0.0610 | 0.0078 | B | 0.0005 | 0.0500 | 0.0184 | 0.0317 | B | 0.0005 | 0.0445 | 0.0445 | 0.0000 | No Differe | 0.0005 | 0.0454 | 0.0186 | 0.0269 | B | 0.0005 |
| 3711 | 0.0911 | 0.1299 | -0.0389 | W | 0.6852 | 0.0652 | 0.0852 | -0.0200 | W | 0.0002 | 0.0831 | 0.0818 | 0.0013 | B | 0.0000 | 0.0629 | 0.0685 | -0.0056 | W | 0.0137 |
| 3751 | 0.2389 | 0.2214 | 0.0175 | B | 0.0005 | 0.2240 | 0.0617 | 0.1623 | B | 0.0005 | 0.2423 | 0.1897 | 0.0526 | B | 0.0005 | 0.1812 | 0.0595 | 0.1217 | B | 0.0005 |
| 3942 | 0.4645 | 0.4730 | -0.0085 | W | 0.0005 | 0.4506 | 0.0561 | 0.3945 | B | 0.0005 | 0.2670 | 0.2806 | -0.0135 | W | 0.0005 | 0.2955 | 0.0545 | 0.2410 | B | 0.0005 |
| 3944 | 0.3263 | 0.3317 | -0.0054 | W | 0.0005 | 0.3727 | 0.0508 | 0.3219 | B | 0.0005 | 0.2120 | 0.2155 | -0.0035 | W | 0.0005 | 0.1511 | 0.0555 | 0.0956 | B | 0.0005 |
| 3949 | 0.2007 | 0.2106 | -0.0099 | W | 0.0005 | 0.1999 | 0.0805 | 0.1194 | B | 0.0005 | 0.1606 | 0.1509 | 0.0097 | B | 0.0005 | 0.0926 | 0.0763 | 0.0162 | B | 0.0005 |
| 4210 | 0.0754 | 0.0637 | 0.0117 | B | 0.0005 | 0.0560 | 0.0144 | 0.0415 | B | 0.0005 | 0.0605 | 0.0522 | 0.0083 | B | 0.0005 | 0.0532 | 0.0189 | 0.0343 | B | 0.0005 |
| 4400 | 0.1499 | 0.1486 | 0.0013 | B | 0.0340 | 0.1525 | 0.0192 | 0.1332 | B | 0.0000 | 0.1548 | 0.1517 | 0.0030 | B | 0.0340 | 0.1523 | 0.0195 | 0.1328 | B | 0.0000 |
| 4512 | 0.0790 | 0.0676 | 0.0114 | B | 0.0000 | 0.0682 | 0.0518 | 0.0164 | B | 0.0000 | 0.0739 | 0.0660 | 0.0079 | B | 0.0000 | 0.0648 | 0.0456 | 0.0191 | B | 0.0000 |
| 4513 | 0.0428 | 0.0403 | 0.0024 | B | 0.0005 | 0.0395 | 0.0597 | -0.0202 | W | 0.0005 | 0.0397 | 0.0393 | 0.0005 | B | 0.0005 | 0.0392 | 0.0538 | -0.0146 | W | 0.0005 |
| 4700 | 0.1753 | 0.1597 | 0.0156 | B | 0.0005 | 0.1318 | 0.0439 | 0.0879 | B | 0.0005 | 0.1278 | 0.1323 | -0.0045 | W | 0.0005 | 0.1249 | 0.0436 | 0.0813 | B | 0.0005 |
| 4812 | 0.0424 | 0.0477 | -0.0053 | W | 0.0005 | 0.0422 | 0.0411 | 0.0011 | B | 0.0005 | 0.0420 | 0.0429 | -0.0009 | W | 0.0005 | 0.0422 | 0.0412 | 0.0011 | B | 0.0005 |
| 4832 | 0.0187 | 0.0197 | -0.0011 | W | 0.0005 | 0.0180 | 0.0062 | 0.0118 | B | 0.0005 | 0.0188 | 0.0184 | 0.0005 | B | 0.0005 | 0.0176 | 0.0062 | 0.0114 | B | 0.0005 |
| 5399 | 0.1947 | 0.1740 | 0.0207 | B | 0.0005 | 0.1801 | 0.0402 | 0.1399 | B | 0.0005 | 0.1899 | 0.1523 | 0.0375 | B | 0.0005 | 0.1810 | 0.0399 | 0.1411 | B | 0.0005 |
| 5500 | 0.0688 | 0.0669 | 0.0019 | B | 0.0005 | 0.0590 | 0.0201 | 0.0389 | B | 0.0005 | 0.0609 | 0.0620 | -0.0012 | W | 0.0005 | 0.0591 | 0.0228 | 0.0363 | B | 0.0005 |
| 5531 | 0.1488 | 0.1368 | 0.0120 | B | 0.0005 | 0.1833 | 0.0465 | 0.1368 | B | 0.0005 | 0.1253 | 0.1067 | 0.0186 | B | 0.0005 | 0.0794 | 0.0523 | 0.0271 | B | 0.0005 |
| 5700 | 0.0989 | 0.1097 | -0.0108 | W | 0.0005 | 0.1129 | 0.0215 | 0.0914 | B | 0.0005 | 0.0994 | 0.1135 | -0.0141 | W | 0.0005 | 0.1142 | 0.0211 | 0.0931 | B | 0.0005 |
| 5731 | 0.1423 | 0.0953 | 0.0470 | B | 0.0005 | 0.0844 | 0.1009 | -0.0165 | W | 0.0005 | 0.1045 | 0.0842 | 0.0203 | B | 0.0005 | 0.0984 | 0.0998 | -0.0015 | W | 0.0005 |
| 5812 | 0.0687 | 0.0587 | 0.0100 | B | 0.0000 | 0.0555 | 0.0410 | 0.0145 | B | 0.0000 | 0.0626 | 0.0565 | 0.0062 | B | 0.0000 | 0.0550 | 0.0395 | 0.0155 | B | 0.0000 |
| 5912 | 0.2546 | 0.2558 | -0.0012 | W | 0.0772 | 0.2451 | 0.0691 | 0.1760 | B | 0.5993 | 0.2532 | 0.2543 | -0.0011 | W | 0.0772 | 0.2430 | 0.0627 | 0.1803 | B | 0.0000 |
| 7011 | 0.0530 | 0.0657 | -0.0127 | W | 0.0000 | 0.0498 | 0.0565 | -0.0067 | W | 0.0340 | 0.0532 | 0.0530 | 0.0003 | B | 0.0340 | 0.0491 | 0.0455 | 0.0036 | B | 0.0340 |
| 7370 | 0.1199 | 0.1092 | 0.0108 | B | 0.0000 | 0.1010 | 0.0270 | 0.0740 | B | 0.0000 | 0.0900 | 0.0924 | -0.0024 | W | 0.0000 | 0.0905 | 0.0296 | 0.0609 | B | 0.0000 |
| 7372 | 0.0616 | 0.0642 | -0.0026 | W | 0.0000 | 0.0546 | 0.0328 | 0.0219 | B | 0.0000 | 0.0670 | 0.0642 | 0.0028 | B | 0.5993 | 0.0587 | 0.0349 | 0.0239 | B | 0.0000 |
| 7510 | 0.1525 | 0.1528 | -0.0004 | W | 0.0005 | 0.1469 | 0.0163 | 0.1306 | B | 0.0005 | 0.0901 | 0.1006 | -0.0105 | W | 0.0005 | 0.0919 | 0.0146 | 0.0773 | B | 0.0005 |

**Table 35: 4-Digit SIC vs. 2-Digit SIC - Prediction Performance Summary of Traditional and Continuous Substantive Analytical Models**

| **4 - Digit SIC** | **Twitter Consumer Interest** | | | | **Twitter Consumer** | | | |
|---|---|---|---|---|---|---|---|---|
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 29 of 46 | 28 of 46 | 28 of 46 | 25 of 46 | 28 of 46 | 25 of 46 | 25 of 46 | 27 of 46 |
| Continuous - SAP | 34 of 46 | 39 of 46 | 34 of 46 | 41 of 46 | 28 of 46 | 38 of 46 | 25 of 46 | 41 of 46 |
| **2 - Digit SIC** | **Twitter Consumer Interest** | | | | **Twitter Consumer** | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 16 of 24 | 16 of 24 | 16 of 24 | 14 of 24 | 15 of 24 | 14 of 24 | 12 of 24 | 15 of 24 |
| Continuous - SAP | 19 of 24 | 21 of 24 | 18 of 24 | 22 of 24 | 14 of 24 | 20 of 24 | 14 of 24 | 22 of 24 |

| **4 - Digit SIC** | **Twitter Consumer Interest** | | | | **Twitter Consumer** | | | |
|---|---|---|---|---|---|---|---|---|
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 57% | 61% | 61% | 54% | 61% | 54% | 54% | 59% |
| Continuous - SAP | 74% | 85% | 85% | 89% | 61% | 83% | 54% | 89% |
| **2 - Digit SIC** | **Twitter Consumer Interest** | | | | **Twitter Consumer** | | | |
| Traditional - SAP | 67% | 67% | 67% | 58% | 63% | 58% | 50% | 63% |
| Continuous - SAP | 79% | 88% | 75% | 92% | 58% | 83% | 58% | 92% |
| **Relative Difference 4 - Digit SIC vs. 2 - Digit SIC** | | | | | | | | |
| **Traditional - SAP % change** | -10% | -6% | -6% | -4% | -2% | -4% | 4% | -4% |
| **Continuous - SAP % change** | -5% | -3% | 10% | -3% | 3% | -1% | -4% | -3% |

# Table 36: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 5 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (5) | | | | | | | | | |
| | | Benchmark - Salest-12 | | Twitter - CI | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 41.77% | 26.67% | 41.01% | 15.00% | 0.76% | 11.67% | TCI | TCI | 1.22 | 1.34 | TCI | TCI |
| 2033 | 12 | 34.08% | 20.00% | 34.80% | 20.00% | -0.72% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 2040 | 24 | 47.86% | 6.67% | 36.74% | 10.00% | 11.12% | -3.33% | TCI | Benchmark | 1.17 | 1.08 | TCI* | TCI* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 38.71% | 16.67% | 40.38% | 6.67% | -1.66% | 10.00% | Benchmark | TCI | 1.18 | 1.34 | TCI* | TCI* |
| 2086 | 36 | 36.02% | 10.00% | 36.26% | 14.00% | -0.24% | -4.00% | Benchmark | Benchmark | 0.92 | 0.87 | Benchmark | Benchmark |
| 2090 | 12 | 25.10% | 30.00% | 25.10% | 30.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 37.05% | 30.00% | 34.84% | 30.00% | 2.21% | 0.00% | TCI | - | 1.03 | 1.02 | TCI* | TCI* |
| 2300 | 24 | 41.17% | 21.67% | 44.01% | 21.67% | -2.85% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 2320 | 12 | 50.00% | 5.00% | 35.65% | 5.00% | 14.35% | 0.00% | TCI | - | 1.35 | 1.31 | TCI* | TCI* |
| 2840 | 12 | 39.85% | 0.00% | 47.86% | 5.00% | -8.00% | -5.00% | Benchmark | Benchmark | 0.75 | 0.69 | Benchmark | Benchmark |
| 2842 | 12 | 12.56% | 0.00% | 12.56% | 25.00% | 0.00% | -25.00% | - | Benchmark | 0.33 | 0.20 | Benchmark* | Benchmark* |
| 2844 | 36 | 41.16% | 2.50% | 42.70% | 2.50% | -1.54% | 0.00% | Benchmark | - | 0.97 | 0.97 | Benchmark* | Benchmark* |
| 2890 | 12 | 38.78% | 0.00% | 45.79% | 0.00% | -7.01% | 0.00% | Benchmark | - | 0.85 | 0.85 | Benchmark* | Benchmark* |
| 2911 | 24 | 43.00% | 13.33% | 45.26% | 0.00% | -2.27% | 13.33% | Benchmark | TCI | 1.24 | 1.54 | TCI* | TCI* |
| 3021 | 12 | 34.44% | 5.00% | 33.27% | 5.00% | 1.17% | 0.00% | TCI | - | 1.03 | 1.03 | TCI* | TCI* |
| 3100 | 12 | 41.75% | 20.00% | 42.97% | 20.00% | -1.22% | 0.00% | Benchmark | - | 0.98 | 0.99 | Benchmark* | Benchmark* |
| 3140 | 12 | 28.96% | 20.00% | 37.09% | 20.00% | -8.13% | 0.00% | Benchmark | - | 0.86 | 0.89 | Benchmark* | Benchmark* |
| 3540 | 12 | 42.95% | 0.00% | 45.26% | 5.00% | -2.32% | -5.00% | Benchmark | Benchmark | 0.85 | 0.78 | Benchmark | Benchmark |
| 3577 | 12 | 45.00% | 0.00% | 19.62% | 15.00% | 25.38% | -15.00% | TCI | Benchmark | 1.30 | 0.91 | TCI* | Benchmark* |
| 3630 | 12 | 8.08% | 0.00% | 0.00% | 0.00% | 8.08% | 0.00% | TCI | - | 0.00 | 0.00 | Benchmark* | Benchmark* |
| 3663 | 24 | 36.96% | 23.33% | 35.29% | 23.33% | 1.66% | 0.00% | TCI | - | 1.03 | 1.02 | TCI* | TCI* |
| 3674 | 12 | 43.55% | 5.00% | 39.53% | 10.00% | 4.02% | -5.00% | TCI | Benchmark | 0.98 | 0.90 | Benchmark* | Benchmark* |
| 3711 | 72 | 44.39% | 9.05% | 44.21% | 11.90% | 0.19% | -2.86% | TCI | Benchmark | 0.95 | 0.92 | Benchmark* | Benchmark* |
| 3751 | 12 | 42.68% | 25.00% | 20.22% | 40.00% | 22.46% | -15.00% | TCI | Benchmark | 1.12 | 0.92 | TCI* | Benchmark* |
| 3942 | 12 | 28.50% | 30.00% | 18.41% | 45.00% | 10.09% | -15.00% | TCI | Benchmark | 0.92 | 0.82 | Benchmark* | Benchmark* |
| 3944 | 12 | 8.33% | 50.00% | 8.33% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 45.00% | 0.00% | 42.68% | 10.00% | 2.32% | -10.00% | TCI | Benchmark | 0.85 | 0.72 | Benchmark* | Benchmark* |
| 4210 | 12 | 13.97% | 0.00% | 7.50% | 0.00% | 6.47% | 0.00% | TCI | - | 1.86 | 1.86 | TCI* | TCI* |
| 4400 | 24 | 16.22% | 20.00% | 7.10% | 23.33% | 9.12% | -3.33% | TCI | Benchmark | 1.19 | 1.05 | TCI* | TCI* |
| 4512 | 84 | 44.32% | 13.19% | 43.30% | 15.83% | 1.02% | -2.64% | TCI | Benchmark | 0.97 | 0.94 | Benchmark* | Benchmark* |
| 4513 | 12 | 47.86% | 10.00% | 48.10% | 5.00% | -0.24% | 5.00% | Benchmark | TCI | 1.09 | 1.17 | TCI* | TCI* |
| 4700 | 12 | 25.10% | 50.00% | 33.27% | 50.00% | -8.18% | 0.00% | Benchmark | - | 0.90 | 0.94 | Benchmark* | Benchmark* |
| 4812 | 12 | 45.26% | 30.00% | 44.97% | 30.00% | 0.29% | 0.00% | TCI | - | 1.00 | 1.00 | TCI* | TCI* |
| 4832 | 12 | 9.74% | 0.00% | 10.95% | 0.00% | -1.21% | 0.00% | Benchmark | - | 0.89 | 0.89 | Benchmark* | Benchmark* |
| 5399 | 12 | 20.82% | 35.00% | 14.68% | 40.00% | 6.14% | -5.00% | TCI | Benchmark | 1.02 | 0.96 | TCI* | Benchmark* |
| 5500 | 12 | 43.24% | 0.00% | 40.18% | 0.00% | 3.06% | 0.00% | TCI | - | 1.08 | 1.08 | TCI* | TCI* |
| 5531 | 12 | 35.29% | 45.00% | 8.33% | 40.00% | 26.96% | 5.00% | TCI | TCI | 1.66 | 1.42 | TCI | TCI |
| 5700 | 12 | 30.33% | 45.00% | 30.74% | 45.00% | -0.40% | 0.00% | Benchmark | - | 0.99 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 39.85% | 0.00% | 40.43% | 0.00% | -0.58% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 5812 | 180 | 40.45% | 13.95% | 40.52% | 13.28% | -0.06% | 0.67% | Benchmark | TCI | 1.01 | 1.02 | TCI* | TCI* |
| 5912 | 36 | 39.66% | 30.00% | 39.77% | 30.00% | -0.11% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 7011 | 24 | 45.54% | 0.00% | 45.54% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 7370 | 24 | 18.76% | 16.67% | 18.76% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 7372 | 36 | 39.99% | 15.00% | 38.76% | 17.50% | 1.23% | -2.50% | TCI | Benchmark | 0.98 | 0.95 | Benchmark* | Benchmark* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 37: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 7 and 2)

Error Detection Ability - Alpha = 0.33

| 4-Digit SIC | Number of Observations | (2) Benchmark - Salest-12 & GDPt-12: False Positive | (2) Benchmark - Salest-12 & GDPt-12: False Negative | (7) Twitter - CI & GDPt-12: False Positive | (7) Twitter - CI & GDPt-12: False Negative | Benchmark - CI: Difference - FP | Benchmark - CI: Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCI Total Cost | (1:2) Benchmark Total Cost /TCI Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2000 | 24 | 45.54% | 16.67% | 44.57% | 16.67% | 0.97% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 2033 | 12 | 36.01% | 15.00% | 24.05% | 15.00% | 11.97% | 0.00% | TCI | - | 1.31 | 1.22 | TCI* | TCI* |
| 2040 | 24 | 34.66% | 3.33% | 35.46% | 3.33% | -0.80% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.67% | 10.00% | 45.80% | 10.00% | -0.13% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2086 | 36 | 31.49% | 27.00% | 30.44% | 31.00% | 1.05% | -4.00% | TCI | Benchmark | 0.95 | 0.92 | Benchmark* | Benchmark* |
| 2090 | 12 | 19.62% | 30.00% | 19.62% | 30.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 36.37% | 35.00% | 32.87% | 30.00% | 3.50% | 5.00% | TCI | TCI | 1.14 | 1.15 | TCI | TCI |
| 2300 | 24 | 36.77% | 16.67% | 37.99% | 25.00% | -1.21% | -8.33% | Benchmark | Benchmark | 0.85 | 0.80 | Benchmark | Benchmark |
| 2320 | 12 | 38.45% | 15.00% | 36.73% | 15.00% | 1.72% | 0.00% | TCI | - | 1.03 | 1.03 | TCI* | TCI* |
| 2840 | 12 | 42.97% | 0.00% | 50.00% | 0.00% | -7.03% | 0.00% | Benchmark | - | 0.86 | 0.86 | Benchmark* | Benchmark* |
| 2842 | 12 | 8.33% | 20.00% | 15.38% | 25.00% | -7.05% | -5.00% | Benchmark | Benchmark | 0.70 | 0.74 | Benchmark | Benchmark |
| 2844 | 36 | 36.87% | 15.00% | 34.77% | 15.00% | 2.10% | 0.00% | TCI | - | 1.04 | 1.03 | TCI* | TCI* |
| 2890 | 12 | 35.92% | 5.00% | 36.37% | 5.00% | -0.45% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 2911 | 24 | 47.62% | 0.00% | 46.20% | 10.00% | 1.42% | -10.00% | TCI | Benchmark | 0.85 | 0.72 | Benchmark* | Benchmark* |
| 3021 | 12 | 37.09% | 5.00% | 36.73% | 5.00% | 0.36% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 3100 | 12 | 39.21% | 5.00% | 42.39% | 15.00% | -3.18% | -10.00% | Benchmark | Benchmark | 0.77 | 0.68 | Benchmark | Benchmark |
| 3140 | 12 | 12.56% | 5.00% | 14.98% | 5.00% | -2.42% | 0.00% | Benchmark | - | 0.88 | 0.90 | Benchmark* | Benchmark* |
| 3540 | 12 | 45.00% | 5.00% | 45.00% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3577 | 12 | 19.62% | 10.00% | 19.62% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 8.33% | 5.00% | 26.01% | 40.00% | -17.68% | -35.00% | Benchmark | Benchmark | 0.20 | 0.17 | Benchmark | Benchmark |
| 3663 | 24 | 35.27% | 16.67% | 35.28% | 20.00% | -0.01% | -3.33% | Benchmark | Benchmark | 0.94 | 0.91 | Benchmark | Benchmark |
| 3674 | 12 | 25.10% | 5.00% | 33.68% | 10.00% | -8.58% | -5.00% | Benchmark | Benchmark | 0.69 | 0.65 | Benchmark | Benchmark |
| 3711 | 72 | 45.54% | 13.21% | 42.46% | 16.07% | 3.09% | -2.86% | TCI | Benchmark | 1.00 | 0.96 | TCI* | Benchmark* |
| 3751 | 12 | 29.88% | 35.00% | 22.92% | 40.00% | 6.96% | -5.00% | TCI | Benchmark | 1.03 | 0.97 | TCI* | Benchmark* |
| 3942 | 12 | 16.39% | 45.00% | 18.41% | 45.00% | -2.01% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 3944 | 12 | 15.38% | 50.00% | 15.38% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 37.45% | 0.00% | 36.01% | 10.00% | 1.44% | -10.00% | TCI | Benchmark | 0.81 | 0.67 | Benchmark* | Benchmark* |
| 4210 | 12 | 13.97% | 0.00% | 6.67% | 0.00% | 7.31% | 0.00% | TCI | - | 2.10 | 2.10 | TCI* | TCI* |
| 4400 | 24 | 0.00% | 31.67% | 0.00% | 31.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4512 | 84 | 43.00% | 11.67% | 42.27% | 16.61% | 0.73% | -4.94% | TCI | Benchmark | 0.93 | 0.88 | Benchmark* | Benchmark* |
| 4513 | 12 | 47.86% | 15.00% | 48.10% | 15.00% | -0.24% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 25.10% | 50.00% | 25.10% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4812 | 12 | 33.62% | 10.00% | 35.46% | 10.00% | -1.84% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 4832 | 12 | 10.95% | 0.00% | 10.95% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5399 | 12 | 25.62% | 30.00% | 8.33% | 25.00% | 17.29% | 5.00% | TCI | TCI | 1.67 | 1.47 | TCI | TCI |
| 5500 | 12 | 37.09% | 0.00% | 33.27% | 0.00% | 3.82% | 0.00% | TCI | - | 1.11 | 1.11 | TCI* | TCI* |
| 5531 | 12 | 27.13% | 45.00% | 14.68% | 40.00% | 12.45% | 5.00% | TCI | TCI | 1.32 | 1.24 | TCI | TCI |
| 5700 | 12 | 0.00% | 25.00% | 10.90% | 10.00% | -10.90% | 15.00% | Benchmark | TCI | 1.20 | 1.62 | TCI* | TCI* |
| 5731 | 12 | 39.85% | 15.00% | 39.85% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5812 | 180 | 39.60% | 21.10% | 39.02% | 17.08% | 0.58% | 4.03% | TCI | TCI | 1.08 | 1.12 | TCI | TCI |
| 5912 | 36 | 36.77% | 27.50% | 33.14% | 27.50% | 3.63% | 0.00% | TCI | - | 1.06 | 1.04 | TCI* | TCI* |
| 7011 | 24 | 46.54% | 3.33% | 43.29% | 0.00% | 3.25% | 3.33% | TCI | TCI | 1.15 | 1.23 | TCI | TCI |
| 7370 | 24 | 10.90% | 23.33% | 3.04% | 23.33% | 7.86% | 0.00% | TCI | - | 1.30 | 1.16 | TCI* | TCI* |
| 7372 | 36 | 36.28% | 12.50% | 31.21% | 12.50% | 5.07% | 0.00% | TCI | - | 1.12 | 1.09 | TCI* | TCI* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 38: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 9 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (9) | | | | | | | | | |
| | | Benchmark - Salest-12 & AR | | Twitter - CI & AR | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 37.12% | 10.00% | 38.53% | 13.33% | -1.41% | -3.33% | Benchmark | Benchmark | 0.91 | 0.88 | Benchmark* | Benchmark* |
| 2033 | 12 | 37.77% | 30.00% | 32.41% | 30.00% | 5.36% | 0.00% | TCI | - | 1.09 | 1.06 | TCI* | TCI* |
| 2040 | 24 | 48.95% | 28.33% | 32.64% | 6.67% | 16.31% | 21.67% | TCI | TCI | 1.97 | 2.30 | TCI | TCI |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 44.15% | 16.67% | 42.84% | 16.67% | 1.31% | 0.00% | TCI | - | 1.02 | 1.02 | TCI* | TCI* |
| 2086 | 36 | 31.38% | 16.50% | 32.78% | 20.50% | -1.40% | -4.00% | Benchmark | Benchmark | 0.90 | 0.87 | Benchmark | Benchmark |
| 2090 | 12 | 19.62% | 35.00% | 19.62% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 35.98% | 30.00% | 37.45% | 25.00% | -1.48% | 5.00% | Benchmark | TCI | 1.06 | 1.10 | TCI* | TCI* |
| 2300 | 24 | 30.33% | 38.33% | 26.82% | 38.33% | 3.51% | 0.00% | TCI | - | 1.05 | 1.03 | TCI* | TCI* |
| 2320 | 12 | 50.00% | 5.00% | 44.71% | 0.00% | 5.29% | 5.00% | TCI | TCI | 1.23 | 1.34 | TCI | TCI |
| 2840 | 12 | 25.62% | 5.00% | 33.68% | 0.00% | -8.06% | 5.00% | Benchmark | TCI | 0.91 | 1.06 | Benchmark* | TCI* |
| 2842 | 12 | 17.70% | 20.00% | 12.56% | 25.00% | 5.14% | -5.00% | TCI | Benchmark | 1.00 | 0.92 | TCI* | Benchmark* |
| 2844 | 36 | 33.16% | 10.00% | 33.98% | 7.50% | -0.81% | 2.50% | Benchmark | TCI | 1.04 | 1.09 | TCI* | TCI* |
| 2890 | 12 | 41.72% | 0.00% | 43.53% | 5.00% | -1.81% | -5.00% | Benchmark | Benchmark | 0.86 | 0.78 | Benchmark | Benchmark |
| 2911 | 24 | 44.01% | 3.33% | 48.95% | 13.33% | -4.94% | -10.00% | Benchmark | Benchmark | 0.76 | 0.67 | Benchmark | Benchmark |
| 3021 | 12 | 39.10% | 15.00% | 38.13% | 15.00% | 0.96% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 3100 | 12 | 42.39% | 5.00% | 41.75% | 5.00% | 0.64% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 3140 | 12 | 35.56% | 20.00% | 37.81% | 20.00% | -2.25% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 3540 | 12 | 32.87% | 5.00% | 39.85% | 10.00% | -6.99% | -5.00% | Benchmark | Benchmark | 0.76 | 0.72 | Benchmark | Benchmark |
| 3577 | 12 | 39.17% | 0.00% | 18.21% | 15.00% | 20.97% | -15.00% | TCI | Benchmark | 1.18 | 0.81 | TCI* | Benchmark* |
| 3630 | 12 | 14.58% | 5.00% | 39.53% | 30.00% | -24.95% | -25.00% | Benchmark | Benchmark | 0.28 | 0.25 | Benchmark | Benchmark |
| 3663 | 24 | 29.25% | 33.33% | 31.46% | 41.67% | -2.20% | -8.33% | Benchmark | Benchmark | 0.86 | 0.84 | Benchmark | Benchmark |
| 3674 | 12 | 24.57% | 10.00% | 29.42% | 10.00% | -4.85% | 0.00% | Benchmark | - | 0.88 | 0.90 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.80% | 11.55% | 43.10% | 14.52% | -0.30% | -2.98% | Benchmark | Benchmark | 0.94 | 0.91 | Benchmark* | Benchmark* |
| 3751 | 12 | 39.85% | 25.00% | 20.22% | 45.00% | 19.63% | -20.00% | TCI | Benchmark | 0.99 | 0.82 | Benchmark* | Benchmark* |
| 3942 | 12 | 28.50% | 25.00% | 18.41% | 45.00% | 10.09% | -20.00% | TCI | Benchmark | 0.84 | 0.72 | Benchmark* | Benchmark* |
| 3944 | 12 | 8.33% | 50.00% | 8.33% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 45.52% | 0.00% | 39.53% | 10.00% | 5.99% | -10.00% | TCI | Benchmark | 0.92 | 0.76 | Benchmark* | Benchmark* |
| 4210 | 12 | 12.56% | 15.00% | 13.27% | 15.00% | -0.71% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 4400 | 24 | 3.91% | 26.67% | 3.91% | 26.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4512 | 84 | 43.67% | 20.25% | 40.90% | 18.83% | 2.76% | 1.41% | TCI | TCI | 1.07 | 1.07 | TCI | TCI |
| 4513 | 12 | 45.52% | 15.00% | 45.79% | 15.00% | -0.26% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 25.10% | 50.00% | 32.87% | 50.00% | -7.77% | 0.00% | Benchmark | - | 0.91 | 0.94 | Benchmark* | Benchmark* |
| 4812 | 12 | 39.41% | 10.00% | 39.41% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4832 | 12 | 7.62% | 0.00% | 9.29% | 0.00% | -1.67% | 0.00% | Benchmark | - | 0.82 | 0.82 | Benchmark* | Benchmark* |
| 5399 | 12 | 20.82% | 35.00% | 14.68% | 40.00% | 6.14% | -5.00% | TCI | Benchmark | 1.02 | 0.96 | TCI* | Benchmark* |
| 5500 | 12 | 44.29% | 0.00% | 44.66% | 5.00% | -0.37% | -5.00% | Benchmark | Benchmark | 0.89 | 0.81 | Benchmark | Benchmark |
| 5531 | 12 | 40.50% | 15.00% | 33.68% | 25.00% | 6.82% | -10.00% | TCI | Benchmark | 0.95 | 0.84 | Benchmark* | Benchmark* |
| 5700 | 12 | 37.77% | 45.00% | 37.77% | 45.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 39.85% | 10.00% | 39.85% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5812 | 180 | 39.07% | 16.66% | 39.91% | 17.05% | -0.84% | -0.39% | Benchmark | Benchmark | 0.98 | 0.98 | Benchmark | Benchmark |
| 5912 | 36 | 30.35% | 19.00% | 29.76% | 19.00% | 0.59% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 7011 | 24 | 45.40% | 13.33% | 44.15% | 6.67% | 1.25% | 6.67% | TCI | TCI | 1.16 | 1.25 | TCI | TCI |
| 7370 | 24 | 23.33% | 13.33% | 20.85% | 16.67% | 2.48% | -3.33% | TCI | Benchmark | 0.98 | 0.92 | Benchmark* | Benchmark* |
| 7372 | 36 | 39.33% | 25.50% | 41.04% | 24.50% | -1.71% | 1.00% | Benchmark | TCI | 0.99 | 1.00 | Benchmark* | TCI* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 39: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCI and without TCI (Models 11 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (11) | | | | | | | | | |
| | | Benchmark - Salest-12 & AR & GDPt-12 | | Twitter - CI & AR & GDPt-12 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 41.63% | 16.67% | 40.19% | 16.67% | 1.44% | 0.00% | TCI | - | 1.03 | 1.02 | TCI* | TCI* |
| 2033 | 12 | 36.01% | 30.00% | 24.05% | 30.00% | 11.97% | 0.00% | TCI | - | 1.22 | 1.14 | TCI* | TCI* |
| 2040 | 24 | 34.49% | 0.00% | 34.28% | 0.00% | 0.21% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 41.32% | 20.00% | 42.51% | 23.33% | -1.19% | -3.33% | Benchmark | Benchmark | 0.93 | 0.91 | Benchmark | Benchmark |
| 2086 | 36 | 29.88% | 27.00% | 27.88% | 31.00% | 2.00% | -4.00% | TCI | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 2090 | 12 | 19.62% | 35.00% | 18.91% | 35.00% | 0.71% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 2111 | 12 | 38.89% | 35.00% | 38.89% | 30.00% | 0.00% | 5.00% | - | TCI | 1.07 | 1.10 | TCI* | TCI* |
| 2300 | 24 | 27.36% | 35.00% | 24.71% | 35.00% | 2.65% | 0.00% | TCI | - | 1.04 | 1.03 | TCI* | TCI* |
| 2320 | 12 | 43.84% | 20.00% | 44.71% | 20.00% | -0.87% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 2840 | 12 | 25.62% | 0.00% | 36.73% | 0.00% | -11.11% | 0.00% | Benchmark | - | 0.70 | 0.70 | Benchmark* | Benchmark* |
| 2842 | 12 | 9.74% | 20.00% | 15.38% | 25.00% | -5.64% | -5.00% | Benchmark | Benchmark | 0.74 | 0.76 | Benchmark | Benchmark |
| 2844 | 36 | 35.15% | 16.50% | 33.99% | 11.50% | 1.17% | 5.00% | TCI | TCI | 1.14 | 1.20 | TCI | TCI |
| 2890 | 12 | 38.09% | 0.00% | 40.07% | 0.00% | -1.97% | 0.00% | Benchmark | - | 0.95 | 0.95 | Benchmark* | Benchmark* |
| 2911 | 24 | 44.57% | 6.67% | 48.95% | 13.33% | -4.38% | -6.67% | Benchmark | Benchmark | 0.82 | 0.77 | Benchmark | Benchmark |
| 3021 | 12 | 40.18% | 25.00% | 39.10% | 25.00% | 1.08% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 3100 | 12 | 32.87% | 20.00% | 32.87% | 25.00% | 0.00% | -5.00% | - | Benchmark | 0.91 | 0.88 | Benchmark* | Benchmark* |
| 3140 | 12 | 32.46% | 5.00% | 33.95% | 5.00% | -1.48% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 3540 | 12 | 37.81% | 5.00% | 37.09% | 5.00% | 0.72% | 0.00% | TCI | - | 1.02 | 1.02 | TCI* | TCI* |
| 3577 | 12 | 19.62% | 15.00% | 13.97% | 15.00% | 5.64% | 0.00% | TCI | - | 1.19 | 1.13 | TCI* | TCI* |
| 3630 | 12 | 15.18% | 15.00% | 39.53% | 30.00% | -24.35% | -15.00% | Benchmark | Benchmark | 0.43 | 0.45 | Benchmark | Benchmark |
| 3663 | 24 | 27.85% | 33.33% | 31.67% | 33.33% | -3.82% | 0.00% | Benchmark | - | 0.94 | 0.96 | Benchmark* | Benchmark* |
| 3674 | 12 | 25.95% | 10.00% | 29.42% | 10.00% | -3.47% | 0.00% | Benchmark | - | 0.91 | 0.93 | Benchmark* | Benchmark* |
| 3711 | 72 | 44.21% | 18.64% | 42.20% | 17.57% | 2.01% | 1.07% | TCI | TCI | 1.05 | 1.05 | TCI | TCI |
| 3751 | 12 | 29.88% | 30.00% | 29.88% | 35.00% | 0.00% | -5.00% | - | Benchmark | 0.92 | 0.90 | Benchmark* | Benchmark* |
| 3942 | 12 | 18.41% | 45.00% | 18.41% | 45.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3944 | 12 | 15.38% | 50.00% | 15.38% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 39.53% | 10.00% | 39.53% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4210 | 12 | 12.56% | 15.00% | 13.27% | 15.00% | -0.71% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 4400 | 24 | 0.00% | 31.67% | 0.00% | 31.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4512 | 84 | 42.96% | 17.67% | 40.35% | 15.92% | 2.61% | 1.75% | TCI | TCI | 1.08 | 1.08 | TCI | TCI |
| 4513 | 12 | 45.52% | 15.00% | 45.79% | 15.00% | -0.26% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 19.62% | 35.00% | 19.62% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4812 | 12 | 38.95% | 10.00% | 38.09% | 10.00% | 0.87% | 0.00% | TCI | - | 1.02 | 1.01 | TCI* | TCI* |
| 4832 | 12 | 13.97% | 0.00% | 16.49% | 0.00% | -2.52% | 0.00% | Benchmark | - | 0.85 | 0.85 | Benchmark* | Benchmark* |
| 5399 | 12 | 25.01% | 30.00% | 9.04% | 25.00% | 15.98% | 5.00% | TCI | TCI | 1.62 | 1.44 | TCI | TCI |
| 5500 | 12 | 40.46% | 0.00% | 41.37% | 0.00% | -0.90% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 5531 | 12 | 28.96% | 0.00% | 28.96% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5700 | 12 | 0.00% | 10.00% | 2.50% | 10.00% | -2.50% | 0.00% | Benchmark | - | 0.80 | 0.89 | Benchmark* | Benchmark* |
| 5731 | 12 | 39.85% | 10.00% | 39.85% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5812 | 180 | 40.00% | 16.70% | 38.77% | 17.74% | 1.23% | -1.04% | TCI | Benchmark | 1.00 | 0.99 | TCI* | Benchmark* |
| 5912 | 36 | 31.56% | 19.00% | 29.15% | 20.50% | 2.41% | -1.50% | TCI | Benchmark | 1.02 | 0.99 | TCI* | Benchmark* |
| 7011 | 24 | 45.40% | 13.33% | 44.15% | 6.67% | 1.25% | 6.67% | TCI | TCI | 1.16 | 1.25 | TCI | TCI |
| 7370 | 24 | 14.71% | 31.67% | 6.74% | 26.67% | 7.97% | 5.00% | TCI | TCI | 1.39 | 1.30 | TCI | TCI |
| 7372 | 36 | 36.03% | 16.50% | 32.64% | 15.00% | 3.39% | 1.50% | TCI | TCI | 1.10 | 1.10 | TCI | TCI |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 40: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 6 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (6) | | | | | | | | | |
| | | Benchmark - Salest-12 | | Twitter - CS | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 41.77% | 26.67% | 44.29% | 18.33% | -2.53% | 8.33% | Benchmark | TCS | 1.09 | 1.17 | TCS* | TCS* |
| 2033 | 12 | 34.08% | 20.00% | 37.45% | 5.00% | -3.37% | 15.00% | Benchmark | TCS | 1.27 | 1.56 | TCS* | TCS* |
| 2040 | 24 | 47.86% | 6.67% | 41.41% | 18.33% | 6.45% | -11.67% | TCS | Benchmark | 0.91 | 0.78 | Benchmark* | Benchmark* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 30.00% | 0.00% | -5.00% | - | Benchmark | 0.83 | 0.83 | Benchmark* | Benchmark* |
| 2082 | 24 | 38.71% | 16.67% | 38.71% | 10.00% | 0.00% | 6.67% | - | TCS | 1.14 | 1.23 | TCS* | TCS* |
| 2086 | 36 | 36.02% | 10.00% | 35.13% | 23.50% | 0.90% | -13.50% | TCS | Benchmark | 0.79 | 0.68 | Benchmark* | Benchmark* |
| 2090 | 12 | 25.10% | 30.00% | 25.10% | 30.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 37.05% | 30.00% | 36.96% | 10.00% | 0.09% | 20.00% | TCS | TCS | 1.43 | 1.70 | TCS | TCS |
| 2300 | 24 | 41.17% | 21.67% | 41.32% | 16.67% | -0.16% | 5.00% | Benchmark | TCS | 1.08 | 1.13 | TCS* | TCS* |
| 2320 | 12 | 50.00% | 5.00% | 42.11% | 5.00% | 7.89% | 0.00% | TCS | - | 1.17 | 1.15 | TCS* | TCS* |
| 2840 | 12 | 39.85% | 0.00% | 39.85% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 12.56% | 0.00% | 12.56% | 20.00% | 0.00% | -20.00% | - | Benchmark | 0.39 | 0.24 | Benchmark* | Benchmark* |
| 2844 | 36 | 41.16% | 2.50% | 43.46% | 5.00% | -2.29% | -2.50% | Benchmark | Benchmark | 0.90 | 0.86 | Benchmark | Benchmark |
| 2890 | 12 | 38.78% | 0.00% | 42.39% | 0.00% | -3.62% | 0.00% | Benchmark | - | 0.91 | 0.91 | Benchmark* | Benchmark* |
| 2911 | 24 | 43.00% | 13.33% | 43.00% | 13.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3021 | 12 | 34.44% | 5.00% | 33.27% | 5.00% | 1.17% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 3100 | 12 | 41.75% | 20.00% | 40.39% | 20.00% | 1.36% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 3140 | 12 | 28.96% | 20.00% | 38.17% | 5.00% | -9.21% | 15.00% | Benchmark | TCS | 1.13 | 1.43 | TCS* | TCS* |
| 3540 | 12 | 42.95% | 0.00% | 47.86% | 0.00% | -4.91% | 0.00% | Benchmark | - | 0.90 | 0.90 | Benchmark* | Benchmark* |
| 3577 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 8.08% | 0.00% | 9.74% | 0.00% | -1.67% | 0.00% | Benchmark | - | 0.83 | 0.83 | Benchmark* | Benchmark* |
| 3663 | 24 | 36.96% | 23.33% | 29.70% | 28.33% | 7.26% | -5.00% | TCS | Benchmark | 1.04 | 0.97 | TCS* | Benchmark* |
| 3674 | 12 | 43.55% | 5.00% | 36.01% | 5.00% | 7.54% | 0.00% | TCS | - | 1.18 | 1.16 | TCS* | TCS* |
| 3711 | 72 | 44.39% | 9.05% | 44.49% | 10.12% | -0.09% | -1.07% | Benchmark | Benchmark | 0.98 | 0.97 | Benchmark | Benchmark |
| 3751 | 12 | 42.68% | 25.00% | 25.49% | 40.00% | 17.20% | -15.00% | TCS | Benchmark | 1.03 | 0.88 | TCS* | Benchmark* |
| 3942 | 12 | 28.50% | 30.00% | 6.67% | 50.00% | 21.83% | -20.00% | TCS | Benchmark | 1.03 | 0.83 | TCS* | Benchmark* |
| 3944 | 12 | 8.33% | 50.00% | 0.00% | 50.00% | 8.33% | 0.00% | TCS | - | 1.17 | 1.08 | TCS* | TCS* |
| 3949 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4210 | 12 | 13.97% | 0.00% | 13.97% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 16.22% | 20.00% | 11.23% | 20.00% | 4.99% | 0.00% | TCS | - | 1.16 | 1.10 | TCS* | TCS* |
| 4512 | 84 | 44.32% | 13.19% | 42.79% | 14.86% | 1.53% | -1.67% | TCS | Benchmark | 1.00 | 0.98 | Benchmark* | Benchmark* |
| 4513 | 12 | 47.86% | 10.00% | 47.86% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 25.10% | 50.00% | 25.10% | 35.00% | 0.00% | 15.00% | - | TCS | 1.25 | 1.32 | TCS* | TCS* |
| 4812 | 12 | 45.26% | 30.00% | 49.76% | 15.00% | -4.50% | 15.00% | Benchmark | TCS | 1.16 | 1.32 | TCS* | TCS* |
| 4832 | 12 | 9.74% | 0.00% | 9.74% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5399 | 12 | 20.82% | 35.00% | 20.82% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5500 | 12 | 43.24% | 0.00% | 40.46% | 0.00% | 2.77% | 0.00% | TCS | - | 1.07 | 1.07 | TCS* | TCS* |
| 5531 | 12 | 35.29% | 45.00% | 36.73% | 45.00% | -1.44% | 0.00% | Benchmark | - | 0.98 | 0.99 | Benchmark* | Benchmark* |
| 5700 | 12 | 30.33% | 45.00% | 19.62% | 45.00% | 10.72% | 0.00% | TCS | - | 1.17 | 1.10 | TCS* | TCS* |
| 5731 | 12 | 39.85% | 0.00% | 45.79% | 0.00% | -5.93% | 0.00% | Benchmark | - | 0.87 | 0.87 | Benchmark* | Benchmark* |
| 5812 | 180 | 40.45% | 13.95% | 40.99% | 15.22% | -0.53% | -1.27% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 5912 | 36 | 39.66% | 30.00% | 43.48% | 28.50% | -3.82% | 1.50% | Benchmark | TCS | 0.97 | 0.99 | Benchmark* | Benchmark* |
| 7011 | 24 | 45.54% | 0.00% | 44.15% | 0.00% | 1.38% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 7370 | 24 | 18.76% | 16.67% | 11.24% | 16.67% | 7.53% | 0.00% | TCS | - | 1.27 | 1.17 | TCS* | TCS* |
| 7372 | 36 | 39.99% | 15.00% | 36.14% | 22.50% | 3.86% | -7.50% | TCS | Benchmark | 0.94 | 0.86 | Benchmark* | Benchmark* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 41: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 8 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (8) | | | | | | | | | | |
| | | Benchmark - Salest-12 & GDPt-12 | | Twitter - CS & GDPt-12 | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 45.54% | 16.67% | 45.24% | 16.67% | 0.30% | 0.00% | TCS | - | 1.00 | 1.00 | TCS* | TCS* |
| 2033 | 12 | 36.01% | 15.00% | 40.18% | 20.00% | -4.16% | -5.00% | Benchmark | Benchmark | 0.85 | 0.82 | Benchmark | Benchmark |
| 2040 | 24 | 34.66% | 3.33% | 35.46% | 0.00% | -0.80% | 3.33% | Benchmark | TCS | 1.07 | 1.17 | TCS* | TCS* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 30.00% | 0.00% | -5.00% | - | Benchmark | 0.83 | 0.83 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.67% | 10.00% | 43.14% | 6.67% | 2.53% | 3.33% | TCS | TCS | 1.12 | 1.16 | TCS | TCS |
| 2086 | 36 | 31.49% | 27.00% | 31.92% | 27.00% | -0.43% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 2090 | 12 | 19.62% | 30.00% | 19.62% | 30.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 36.37% | 35.00% | 38.49% | 35.00% | -2.12% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 2300 | 24 | 36.77% | 16.67% | 32.92% | 23.33% | 3.86% | -6.67% | TCS | Benchmark | 0.95 | 0.88 | Benchmark* | Benchmark* |
| 2320 | 12 | 38.45% | 15.00% | 36.73% | 15.00% | 1.72% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 2840 | 12 | 42.97% | 0.00% | 42.97% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 8.33% | 20.00% | 8.33% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2844 | 36 | 36.87% | 15.00% | 37.13% | 15.00% | -0.26% | 0.00% | Benchmark | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2890 | 12 | 35.92% | 5.00% | 35.92% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2911 | 24 | 47.62% | 0.00% | 48.84% | 0.00% | -1.22% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 3021 | 12 | 37.09% | 5.00% | 36.73% | 15.00% | 0.36% | -10.00% | TCS | Benchmark | 0.81 | 0.71 | Benchmark* | Benchmark* |
| 3100 | 12 | 39.21% | 5.00% | 42.39% | 15.00% | -3.18% | -10.00% | Benchmark | Benchmark | 0.77 | 0.68 | Benchmark | Benchmark |
| 3140 | 12 | 12.56% | 5.00% | 19.01% | 5.00% | -6.45% | 0.00% | Benchmark | - | 0.73 | 0.78 | Benchmark* | Benchmark* |
| 3540 | 12 | 45.00% | 5.00% | 42.39% | 5.00% | 2.61% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 3577 | 12 | 19.62% | 10.00% | 25.10% | 15.00% | -5.48% | -5.00% | Benchmark | Benchmark | 0.74 | 0.72 | Benchmark | Benchmark |
| 3630 | 12 | 8.33% | 5.00% | 23.89% | 25.00% | -15.55% | -20.00% | Benchmark | Benchmark | 0.27 | 0.25 | Benchmark | Benchmark |
| 3663 | 24 | 35.27% | 16.67% | 33.70% | 16.67% | 1.57% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 3674 | 12 | 25.10% | 5.00% | 26.86% | 10.00% | -1.77% | -5.00% | Benchmark | Benchmark | 0.82 | 0.75 | Benchmark | Benchmark |
| 3711 | 72 | 45.54% | 13.21% | 45.50% | 14.64% | 0.04% | -1.43% | TCS | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 3751 | 12 | 29.88% | 35.00% | 20.82% | 45.00% | 9.05% | -10.00% | TCS | Benchmark | 0.99 | 0.90 | Benchmark* | Benchmark* |
| 3942 | 12 | 16.39% | 45.00% | 6.67% | 35.00% | 9.73% | 10.00% | TCS | TCS | 1.47 | 1.39 | TCS | TCS |
| 3944 | 12 | 15.38% | 50.00% | 15.38% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 37.45% | 0.00% | 32.46% | 5.00% | 4.99% | -5.00% | TCS | Benchmark | 1.00 | 0.88 | Benchmark* | Benchmark* |
| 4210 | 12 | 13.97% | 0.00% | 6.67% | 0.00% | 7.31% | 0.00% | TCS | - | 2.10 | 2.10 | TCS* | TCS* |
| 4400 | 24 | 0.00% | 31.67% | 0.00% | 30.00% | 0.00% | 1.67% | - | TCS | 1.06 | 1.06 | TCS* | TCS* |
| 4512 | 84 | 43.00% | 11.67% | 42.15% | 12.44% | 0.86% | -0.78% | TCS | Benchmark | 1.00 | 0.99 | TCS* | Benchmark* |
| 4513 | 12 | 47.86% | 15.00% | 47.86% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 25.10% | 50.00% | 29.88% | 25.00% | -4.78% | 25.00% | Benchmark | TCS | 1.37 | 1.57 | TCS* | TCS* |
| 4812 | 12 | 33.62% | 10.00% | 32.87% | 10.00% | 0.75% | 0.00% | TCS | - | 1.02 | 1.01 | TCS* | TCS* |
| 4832 | 12 | 10.95% | 0.00% | 10.95% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5399 | 12 | 25.62% | 30.00% | 29.88% | 30.00% | -4.26% | 0.00% | Benchmark | - | 0.93 | 0.95 | Benchmark* | Benchmark* |
| 5500 | 12 | 37.09% | 0.00% | 37.09% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5531 | 12 | 27.13% | 45.00% | 29.64% | 50.00% | -2.51% | -5.00% | Benchmark | Benchmark | 0.91 | 0.90 | Benchmark | Benchmark |
| 5700 | 12 | 0.00% | 25.00% | 6.67% | 25.00% | -6.67% | 0.00% | Benchmark | - | 0.79 | 0.88 | Benchmark* | Benchmark* |
| 5731 | 12 | 39.85% | 15.00% | 36.73% | 15.00% | 3.12% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 5812 | 180 | 39.60% | 21.10% | 39.09% | 16.95% | 0.51% | 4.15% | TCS | TCS | 1.08 | 1.12 | TCS | TCS |
| 5912 | 36 | 36.77% | 27.50% | 34.77% | 20.00% | 2.00% | 7.50% | TCS | TCS | 1.17 | 1.23 | TCS | TCS |
| 7011 | 24 | 46.54% | 3.33% | 46.59% | 0.00% | -0.05% | 3.33% | Benchmark | TCS | 1.07 | 1.14 | TCS* | TCS* |
| 7370 | 24 | 10.90% | 23.33% | 5.83% | 23.33% | 5.07% | 0.00% | TCS | - | 1.17 | 1.10 | TCS* | TCS* |
| 7372 | 36 | 36.28% | 12.50% | 37.00% | 10.00% | -0.71% | 2.50% | Benchmark | TCS | 1.04 | 1.08 | TCS* | TCS* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 35.00% | 0.00% | 15.00% | - | TCS | 1.43 | 1.43 | TCS* | TCS* |

# Table 42: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 10 and 3)

Error Detection Ability - Alpha = 0.33

| | | (3) | | (10) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Benchmark - Salest-12 & AR | | Twitter - CS & AR | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 37.12% | 10.00% | 36.40% | 25.00% | 0.72% | -15.00% | TCS | Benchmark | 0.77 | 0.66 | Benchmark* | Benchmark* |
| 2033 | 12 | 37.77% | 30.00% | 30.33% | 20.00% | 7.44% | 10.00% | TCS | TCS | 1.35 | 1.39 | TCS | TCS |
| 2040 | 24 | 48.95% | 28.33% | 44.29% | 25.00% | 4.67% | 3.33% | TCS | TCS | 1.12 | 1.12 | TCS | TCS |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 30.00% | 0.00% | -5.00% | - | Benchmark | 0.83 | 0.83 | Benchmark* | Benchmark* |
| 2082 | 24 | 44.15% | 16.67% | 43.29% | 20.00% | 0.87% | -3.33% | TCS | Benchmark | 0.96 | 0.93 | Benchmark* | Benchmark* |
| 2086 | 36 | 31.38% | 16.50% | 31.66% | 27.00% | -0.28% | -10.50% | Benchmark | Benchmark | 0.82 | 0.75 | Benchmark | Benchmark |
| 2090 | 12 | 19.62% | 35.00% | 19.62% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 35.98% | 30.00% | 38.49% | 5.00% | -2.52% | 25.00% | Benchmark | TCS | 1.52 | 1.98 | TCS* | TCS* |
| 2300 | 24 | 30.33% | 38.33% | 28.07% | 38.33% | 2.26% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 2320 | 12 | 50.00% | 5.00% | 42.11% | 5.00% | 7.89% | 0.00% | TCS | - | 1.17 | 1.15 | TCS* | TCS* |
| 2840 | 12 | 25.62% | 5.00% | 25.62% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 17.70% | 20.00% | 14.68% | 20.00% | 3.02% | 0.00% | TCS | - | 1.09 | 1.06 | TCS* | TCS* |
| 2844 | 36 | 33.16% | 10.00% | 39.19% | 7.50% | -6.03% | 2.50% | Benchmark | TCS | 0.92 | 0.98 | Benchmark* | Benchmark* |
| 2890 | 12 | 41.72% | 0.00% | 42.62% | 0.00% | -0.90% | 0.00% | Benchmark | - | 0.98 | 0.98 | Benchmark* | Benchmark* |
| 2911 | 24 | 44.01% | 3.33% | 45.26% | 3.33% | -1.25% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 3021 | 12 | 39.10% | 15.00% | 37.09% | 25.00% | 2.01% | -10.00% | TCS | Benchmark | 0.87 | 0.79 | Benchmark* | Benchmark* |
| 3100 | 12 | 42.39% | 5.00% | 42.39% | 10.00% | 0.00% | -5.00% | - | Benchmark | 0.90 | 0.84 | Benchmark* | Benchmark* |
| 3140 | 12 | 35.56% | 20.00% | 34.08% | 0.00% | 1.48% | 20.00% | TCS | TCS | 1.63 | 2.22 | TCS | TCS |
| 3540 | 12 | 32.87% | 5.00% | 30.23% | 0.00% | 2.64% | 5.00% | TCS | TCS | 1.25 | 1.42 | TCS | TCS |
| 3577 | 12 | 39.17% | 0.00% | 39.53% | 0.00% | -0.36% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 3630 | 12 | 14.58% | 5.00% | 18.41% | 30.00% | -3.83% | -25.00% | Benchmark | Benchmark | 0.40 | 0.31 | Benchmark | Benchmark |
| 3663 | 24 | 29.25% | 33.33% | 29.70% | 35.00% | -0.44% | -1.67% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 3674 | 12 | 24.57% | 10.00% | 29.42% | 10.00% | -4.85% | 0.00% | Benchmark | - | 0.88 | 0.90 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.80% | 11.55% | 43.39% | 12.62% | -0.59% | -1.07% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark | Benchmark |
| 3751 | 12 | 39.85% | 25.00% | 24.57% | 40.00% | 15.28% | -15.00% | TCS | Benchmark | 1.00 | 0.86 | TCS* | Benchmark* |
| 3942 | 12 | 28.50% | 25.00% | 6.67% | 50.00% | 21.83% | -25.00% | TCS | Benchmark | 0.94 | 0.74 | Benchmark* | Benchmark* |
| 3944 | 12 | 8.33% | 50.00% | 5.00% | 50.00% | 3.33% | 0.00% | TCS | - | 1.06 | 1.03 | TCS* | TCS* |
| 3949 | 12 | 45.52% | 0.00% | 42.68% | 0.00% | 2.84% | 0.00% | TCS | - | 1.07 | 1.07 | TCS* | TCS* |
| 4210 | 12 | 12.56% | 15.00% | 12.56% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 3.91% | 26.67% | 3.91% | 23.33% | 0.00% | 3.33% | - | TCS | 1.12 | 1.13 | TCS* | TCS* |
| 4512 | 84 | 43.67% | 20.25% | 42.96% | 19.28% | 0.71% | 0.97% | TCS | TCS | 1.03 | 1.03 | TCS | TCS |
| 4513 | 12 | 45.52% | 15.00% | 45.52% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 25.10% | 50.00% | 29.88% | 30.00% | -4.78% | 20.00% | Benchmark | TCS | 1.25 | 1.39 | TCS* | TCS* |
| 4812 | 12 | 39.41% | 10.00% | 36.73% | 10.00% | 2.68% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 4832 | 12 | 7.62% | 0.00% | 7.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5399 | 12 | 20.82% | 35.00% | 25.01% | 35.00% | -4.19% | 0.00% | Benchmark | - | 0.93 | 0.96 | Benchmark* | Benchmark* |
| 5500 | 12 | 44.29% | 0.00% | 43.24% | 0.00% | 1.05% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 5531 | 12 | 40.50% | 15.00% | 39.49% | 25.00% | 1.00% | -10.00% | TCS | Benchmark | 0.86 | 0.79 | Benchmark* | Benchmark* |
| 5700 | 12 | 37.77% | 45.00% | 19.62% | 45.00% | 18.15% | 0.00% | TCS | - | 1.28 | 1.17 | TCS* | TCS* |
| 5731 | 12 | 39.85% | 10.00% | 48.10% | 0.00% | -8.24% | 10.00% | Benchmark | TCS | 1.04 | 1.24 | TCS* | TCS* |
| 5812 | 180 | 39.07% | 16.66% | 38.29% | 20.00% | 0.77% | -3.34% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 5912 | 36 | 30.35% | 19.00% | 34.37% | 15.00% | -4.02% | 4.00% | Benchmark | TCS | 1.00 | 1.06 | Benchmark* | TCS* |
| 7011 | 24 | 45.40% | 13.33% | 46.60% | 6.67% | -1.19% | 6.67% | Benchmark | TCS | 1.10 | 1.20 | TCS* | TCS* |
| 7370 | 24 | 23.33% | 13.33% | 16.64% | 20.00% | 6.69% | -6.67% | TCS | Benchmark | 1.00 | 0.88 | TCS* | Benchmark* |
| 7372 | 36 | 39.33% | 25.50% | 34.37% | 25.50% | 4.95% | 0.00% | TCS | - | 1.08 | 1.06 | TCS* | TCS* |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 43: 4-Digit SIC - Error Detection Performance for Traditional Substantive Analytical Models with TCS and without TCS (Models 12 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (12) | | | | | | | | | |
| | | Benchmark - Salest-12 & AR & GDPt-12 | | Twitter - CS & AR & GDPt-12 | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 41.63% | 16.67% | 41.63% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2033 | 12 | 36.01% | 30.00% | 36.37% | 25.00% | -0.36% | 5.00% | Benchmark | TCS | 1.08 | 1.11 | TCS* | TCS* |
| 2040 | 24 | 34.49% | 0.00% | 33.30% | 3.33% | 1.19% | -3.33% | TCS | Benchmark | 0.94 | 0.86 | Benchmark* | Benchmark* |
| 2080 | 12 | 0.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 41.32% | 20.00% | 42.52% | 20.00% | -1.20% | 0.00% | Benchmark | - | 0.98 | 0.99 | Benchmark* | Benchmark* |
| 2086 | 36 | 29.88% | 27.00% | 29.01% | 29.50% | 0.87% | -2.50% | TCS | Benchmark | 0.97 | 0.95 | Benchmark* | Benchmark* |
| 2090 | 12 | 19.62% | 35.00% | 19.62% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 38.89% | 35.00% | 39.53% | 35.00% | -0.64% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 2300 | 24 | 27.36% | 35.00% | 24.88% | 26.67% | 2.48% | 8.33% | TCS | TCS | 1.21 | 1.24 | TCS | TCS |
| 2320 | 12 | 43.84% | 20.00% | 44.42% | 15.00% | -0.58% | 5.00% | Benchmark | TCS | 1.07 | 1.13 | TCS* | TCS* |
| 2840 | 12 | 25.62% | 0.00% | 35.92% | 5.00% | -10.30% | -5.00% | Benchmark | Benchmark | 0.63 | 0.56 | Benchmark | Benchmark |
| 2842 | 12 | 9.74% | 20.00% | 9.74% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2844 | 36 | 35.15% | 16.50% | 37.48% | 14.00% | -2.32% | 2.50% | Benchmark | TCS | 1.00 | 1.04 | TCS* | TCS* |
| 2890 | 12 | 38.09% | 0.00% | 37.73% | 0.00% | 0.36% | 0.00% | TCS | - | 1.01 | 1.01 | TCS* | TCS* |
| 2911 | 24 | 44.57% | 6.67% | 44.01% | 6.67% | 0.56% | 0.00% | TCS | - | 1.01 | 1.01 | TCS* | TCS* |
| 3021 | 12 | 40.18% | 25.00% | 38.13% | 25.00% | 2.04% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 3100 | 12 | 32.87% | 20.00% | 35.92% | 20.00% | -3.06% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3140 | 12 | 32.46% | 5.00% | 34.84% | 5.00% | -2.38% | 0.00% | Benchmark | - | 0.94 | 0.95 | Benchmark* | Benchmark* |
| 3540 | 12 | 37.81% | 5.00% | 33.27% | 5.00% | 4.54% | 0.00% | TCS | - | 1.12 | 1.10 | TCS* | TCS* |
| 3577 | 12 | 19.62% | 15.00% | 24.49% | 15.00% | -4.88% | 0.00% | Benchmark | - | 0.88 | 0.91 | Benchmark* | Benchmark* |
| 3630 | 12 | 15.18% | 15.00% | 18.41% | 30.00% | -3.22% | -15.00% | Benchmark | Benchmark | 0.62 | 0.58 | Benchmark | Benchmark |
| 3663 | 24 | 27.85% | 33.33% | 31.46% | 38.33% | -3.61% | -5.00% | Benchmark | Benchmark | 0.88 | 0.87 | Benchmark | Benchmark |
| 3674 | 12 | 25.95% | 10.00% | 29.42% | 10.00% | -3.47% | 0.00% | Benchmark | - | 0.91 | 0.93 | Benchmark* | Benchmark* |
| 3711 | 72 | 44.21% | 18.64% | 45.68% | 15.12% | -1.47% | 3.52% | Benchmark | TCS | 1.03 | 1.07 | TCS* | TCS* |
| 3751 | 12 | 29.88% | 30.00% | 20.82% | 40.00% | 9.05% | -10.00% | TCS | Benchmark | 0.98 | 0.89 | Benchmark* | Benchmark* |
| 3942 | 12 | 18.41% | 45.00% | 6.67% | 35.00% | 11.74% | 10.00% | TCS | TCS | 1.52 | 1.41 | TCS | TCS |
| 3944 | 12 | 15.38% | 50.00% | 15.38% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 39.53% | 10.00% | 28.04% | 10.00% | 11.49% | 0.00% | TCS | - | 1.30 | 1.24 | TCS* | TCS* |
| 4210 | 12 | 12.56% | 15.00% | 12.56% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 0.00% | 31.67% | 0.00% | 30.00% | 0.00% | 1.67% | - | TCS | 1.06 | 1.06 | TCS* | TCS* |
| 4512 | 84 | 42.96% | 17.67% | 43.13% | 18.64% | -0.17% | -0.97% | Benchmark | Benchmark | 0.98 | 0.97 | Benchmark | Benchmark |
| 4513 | 12 | 45.52% | 15.00% | 45.52% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 19.62% | 35.00% | 29.88% | 30.00% | -10.26% | 5.00% | Benchmark | TCS | 0.91 | 1.00 | Benchmark* | Benchmark* |
| 4812 | 12 | 38.95% | 10.00% | 36.33% | 10.00% | 2.63% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 4832 | 12 | 13.97% | 0.00% | 13.27% | 0.00% | 0.71% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 5399 | 12 | 25.01% | 30.00% | 25.62% | 30.00% | -0.60% | 0.00% | Benchmark | - | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 5500 | 12 | 40.46% | 0.00% | 37.09% | 0.00% | 3.37% | 0.00% | TCS | - | 1.09 | 1.09 | TCS* | TCS* |
| 5531 | 12 | 28.96% | 0.00% | 27.13% | 5.00% | 1.83% | -5.00% | TCS | Benchmark | 0.90 | 0.78 | Benchmark* | Benchmark* |
| 5700 | 12 | 0.00% | 10.00% | 7.50% | 25.00% | -7.50% | -15.00% | Benchmark | Benchmark | 0.31 | 0.35 | Benchmark | Benchmark |
| 5731 | 12 | 39.85% | 10.00% | 36.01% | 15.00% | 3.84% | -5.00% | TCS | Benchmark | 0.98 | 0.91 | Benchmark* | Benchmark* |
| 5812 | 180 | 40.00% | 16.70% | 38.82% | 15.63% | 1.18% | 1.06% | TCS | TCS | 1.04 | 1.05 | TCS | TCS |
| 5912 | 36 | 31.56% | 19.00% | 34.63% | 17.50% | -3.07% | 1.50% | Benchmark | TCS | 0.97 | 1.00 | Benchmark* | Benchmark* |
| 7011 | 24 | 45.40% | 13.33% | 47.74% | 6.67% | -2.34% | 6.67% | Benchmark | TCS | 1.08 | 1.18 | TCS* | TCS* |
| 7370 | 24 | 14.71% | 31.67% | 14.71% | 31.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 7372 | 36 | 36.03% | 16.50% | 35.66% | 7.50% | 0.37% | 9.00% | TCS | TCS | 1.22 | 1.36 | TCS | TCS |
| 7510 | 12 | 0.00% | 50.00% | 0.00% | 35.00% | 0.00% | 15.00% | - | TCS | 1.43 | 1.43 | TCS* | TCS* |

# Table 44: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 5 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (5) | | | | | | | | | |
| | | Benchmark - Salest-1 | | Twitter - CI | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 46.85% | 13.33% | 45.53% | 11.67% | 1.32% | 1.67% | TCI | TCI | 1.05 | 1.07 | TCI | TCI |
| 2033 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2040 | 24 | 45.40% | 10.00% | 44.29% | 16.67% | 1.11% | -6.67% | TCI | Benchmark | 0.91 | 0.84 | Benchmark* | Benchmark* |
| 2080 | 12 | 39.85% | 15.00% | 39.85% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.80% | 20.00% | 46.72% | 10.00% | -0.92% | 10.00% | Benchmark | TCI | 1.16 | 1.29 | TCI* | TCI* |
| 2086 | 36 | 43.68% | 15.00% | 44.53% | 12.50% | -0.85% | 2.50% | Benchmark | TCI | 1.03 | 1.06 | TCI* | TCI* |
| 2090 | 12 | 43.26% | 20.00% | 33.27% | 25.00% | 9.99% | -5.00% | TCI | Benchmark | 1.09 | 1.00 | TCI* | Benchmark* |
| 2111 | 12 | 47.86% | 5.00% | 50.00% | 0.00% | -2.14% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 2300 | 24 | 45.13% | 3.33% | 46.85% | 13.33% | -1.72% | -10.00% | Benchmark | Benchmark | 0.81 | 0.70 | Benchmark | Benchmark |
| 2320 | 12 | 45.00% | 0.00% | 47.62% | 0.00% | -2.62% | 0.00% | Benchmark | - | 0.94 | 0.94 | Benchmark* | Benchmark* |
| 2840 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2844 | 36 | 44.44% | 10.00% | 44.62% | 14.00% | -0.18% | -4.00% | Benchmark | Benchmark | 0.93 | 0.89 | Benchmark | Benchmark |
| 2890 | 12 | 39.53% | 10.00% | 45.52% | 10.00% | -5.99% | 0.00% | Benchmark | - | 0.89 | 0.91 | Benchmark* | Benchmark* |
| 2911 | 24 | 45.67% | 16.67% | 41.94% | 28.33% | 3.73% | -11.67% | TCI | Benchmark | 0.89 | 0.80 | Benchmark* | Benchmark* |
| 3021 | 12 | 50.00% | 0.00% | 47.86% | 5.00% | 2.14% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 3100 | 12 | 39.21% | 5.00% | 39.21% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3140 | 12 | 45.26% | 5.00% | 47.86% | 5.00% | -2.60% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3540 | 12 | 43.26% | 20.00% | 43.26% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3577 | 12 | 42.39% | 5.00% | 42.39% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 42.68% | 10.00% | 45.52% | 10.00% | -2.84% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3663 | 24 | 41.63% | 21.67% | 41.63% | 21.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3674 | 12 | 47.86% | 5.00% | 40.50% | 25.00% | 7.36% | -20.00% | TCI | Benchmark | 0.81 | 0.64 | Benchmark* | Benchmark* |
| 3711 | 72 | 45.13% | 16.31% | 42.90% | 20.24% | 2.23% | -3.93% | TCI | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 3751 | 12 | 45.26% | 5.00% | 47.62% | 0.00% | -2.36% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 3942 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3944 | 12 | 42.68% | 10.00% | 46.05% | 20.00% | -3.36% | -10.00% | Benchmark | Benchmark | 0.80 | 0.73 | Benchmark | Benchmark |
| 3949 | 12 | 45.26% | 5.00% | 37.45% | 30.00% | 7.81% | -25.00% | TCI | Benchmark | 0.75 | 0.57 | Benchmark* | Benchmark* |
| 4210 | 12 | 42.68% | 10.00% | 32.87% | 20.00% | 9.82% | -10.00% | TCI | Benchmark | 1.00 | 0.86 | Benchmark* | Benchmark* |
| 4400 | 24 | 40.03% | 20.00% | 40.51% | 28.33% | -0.49% | -8.33% | Benchmark | Benchmark | 0.87 | 0.82 | Benchmark | Benchmark |
| 4512 | 84 | 45.85% | 13.69% | 45.11% | 13.58% | 0.74% | 0.11% | TCI | TCI | 1.01 | 1.01 | TCI | TCI |
| 4513 | 12 | 45.79% | 15.00% | 40.18% | 20.00% | 5.61% | -5.00% | TCI | Benchmark | 1.01 | 0.95 | TCI* | Benchmark* |
| 4700 | 12 | 45.79% | 15.00% | 47.62% | 0.00% | -1.83% | 15.00% | Benchmark | TCI | 1.28 | 1.59 | TCI* | TCI* |
| 4812 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4832 | 12 | 42.11% | 0.00% | 39.53% | 10.00% | 2.57% | -10.00% | TCI | Benchmark | 0.85 | 0.71 | Benchmark* | Benchmark* |
| 5399 | 12 | 34.08% | 35.00% | 39.53% | 10.00% | -5.45% | 25.00% | Benchmark | TCI | 1.39 | 1.75 | TCI* | TCI* |
| 5500 | 12 | 45.52% | 10.00% | 45.79% | 15.00% | -0.26% | -5.00% | Benchmark | Benchmark | 0.91 | 0.86 | Benchmark | Benchmark |
| 5531 | 12 | 45.26% | 5.00% | 45.52% | 10.00% | -0.26% | -5.00% | Benchmark | Benchmark | 0.91 | 0.84 | Benchmark | Benchmark |
| 5700 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 50.00% | 0.00% | 47.86% | 5.00% | 2.14% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5812 | 180 | 45.09% | 13.83% | 44.76% | 14.81% | 0.33% | -0.97% | TCI | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 5912 | 36 | 42.80% | 16.50% | 39.00% | 25.50% | 3.80% | -9.00% | TCI | Benchmark | 0.92 | 0.84 | Benchmark* | Benchmark* |
| 7011 | 24 | 45.54% | 13.33% | 42.99% | 18.33% | 2.55% | -5.00% | TCI | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 7370 | 24 | 40.68% | 31.67% | 40.37% | 28.33% | 0.31% | 3.33% | TCI | TCI | 1.05 | 1.07 | TCI | TCI |
| 7372 | 36 | 35.16% | 22.50% | 36.53% | 24.00% | -1.37% | -1.50% | Benchmark | Benchmark | 0.95 | 0.95 | Benchmark | Benchmark |
| 7510 | 12 | 48.10% | 10.00% | 33.68% | 30.00% | 14.42% | -20.00% | TCI | Benchmark | 0.91 | 0.73 | Benchmark* | Benchmark* |

# Table 45: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 7 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (7) | | | | | | | | | |
| | | Benchmark - Salest-1 & GDPt-1 | | Twitter - CI & GDPt-1 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 45.80% | 18.33% | 43.43% | 26.67% | 2.37% | -8.33% | TCI | Benchmark | 0.91 | 0.85 | Benchmark* | Benchmark* |
| 2033 | 12 | 45.26% | 5.00% | 42.68% | 10.00% | 2.58% | -5.00% | TCI | Benchmark | 0.95 | 0.88 | Benchmark* | Benchmark* |
| 2040 | 24 | 44.15% | 13.33% | 38.18% | 18.33% | 5.97% | -5.00% | TCI | Benchmark | 1.02 | 0.95 | TCI* | Benchmark* |
| 2080 | 12 | 39.85% | 15.00% | 36.73% | 20.00% | 3.12% | -5.00% | TCI | Benchmark | 0.97 | 0.91 | Benchmark* | Benchmark* |
| 2082 | 24 | 46.97% | 15.00% | 45.79% | 16.67% | 1.18% | -1.67% | TCI | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 2086 | 36 | 40.74% | 15.50% | 41.69% | 14.00% | -0.95% | 1.50% | Benchmark | TCI | 1.01 | 1.03 | TCI* | TCI* |
| 2090 | 12 | 35.65% | 5.00% | 39.53% | 10.00% | -3.88% | -5.00% | Benchmark | Benchmark | 0.82 | 0.77 | Benchmark | Benchmark |
| 2111 | 12 | 39.53% | 10.00% | 45.00% | 0.00% | -5.47% | 10.00% | Benchmark | TCI | 1.10 | 1.32 | TCI* | TCI* |
| 2300 | 24 | 45.40% | 8.33% | 42.84% | 15.00% | 2.56% | -6.67% | TCI | Benchmark | 0.93 | 0.85 | Benchmark* | Benchmark* |
| 2320 | 12 | 47.62% | 0.00% | 47.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 45.00% | 0.00% | 42.11% | 0.00% | 2.89% | 0.00% | TCI | - | 1.07 | 1.07 | TCI* | TCI* |
| 2844 | 36 | 42.10% | 24.00% | 38.19% | 30.50% | 3.91% | -6.50% | TCI | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 2890 | 12 | 39.53% | 10.00% | 36.01% | 10.00% | 3.52% | 0.00% | TCI | - | 1.08 | 1.06 | TCI* | TCI* |
| 2911 | 24 | 49.07% | 6.67% | 42.09% | 30.00% | 6.98% | -23.33% | TCI | Benchmark | 0.77 | 0.61 | Benchmark* | Benchmark* |
| 3021 | 12 | 36.73% | 20.00% | 39.53% | 10.00% | -2.80% | 10.00% | Benchmark | TCI | 1.15 | 1.29 | TCI* | TCI* |
| 3100 | 12 | 35.65% | 5.00% | 35.65% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3140 | 12 | 45.26% | 5.00% | 42.39% | 5.00% | 2.87% | 0.00% | TCI | - | 1.06 | 1.05 | TCI* | TCI* |
| 3540 | 12 | 45.52% | 10.00% | 42.68% | 10.00% | 2.84% | 0.00% | TCI | - | 1.05 | 1.05 | TCI* | TCI* |
| 3577 | 12 | 42.68% | 10.00% | 42.39% | 5.00% | 0.29% | 5.00% | TCI | TCI | 1.11 | 1.20 | TCI | TCI |
| 3630 | 12 | 33.27% | 25.00% | 33.27% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3663 | 24 | 46.60% | 6.67% | 44.15% | 13.33% | 2.44% | -6.67% | TCI | Benchmark | 0.93 | 0.85 | Benchmark* | Benchmark* |
| 3674 | 12 | 48.10% | 10.00% | 43.26% | 20.00% | 4.83% | -10.00% | TCI | Benchmark | 0.92 | 0.82 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.90% | 19.83% | 43.30% | 18.33% | -0.39% | 1.50% | Benchmark | TCI | 1.02 | 1.03 | TCI* | TCI* |
| 3751 | 12 | 42.68% | 10.00% | 47.62% | 0.00% | -4.93% | 10.00% | Benchmark | TCI | 1.11 | 1.32 | TCI* | TCI* |
| 3942 | 12 | 45.00% | 0.00% | 42.39% | 5.00% | 2.61% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 3944 | 12 | 50.00% | 0.00% | 47.62% | 0.00% | 2.38% | 0.00% | TCI | - | 1.05 | 1.05 | TCI* | TCI* |
| 3949 | 12 | 45.26% | 5.00% | 37.09% | 25.00% | 8.17% | -20.00% | TCI | Benchmark | 0.81 | 0.63 | Benchmark* | Benchmark* |
| 4210 | 12 | 40.18% | 20.00% | 32.87% | 20.00% | 7.31% | 0.00% | TCI | - | 1.14 | 1.10 | TCI* | TCI* |
| 4400 | 24 | 44.29% | 16.67% | 45.54% | 13.33% | -1.24% | 3.33% | Benchmark | TCI | 1.04 | 1.08 | TCI* | TCI* |
| 4512 | 84 | 45.54% | 15.58% | 44.76% | 15.92% | 0.78% | -0.34% | TCI | Benchmark | 1.01 | 1.00 | TCI* | TCI* |
| 4513 | 12 | 40.82% | 30.00% | 45.79% | 15.00% | -4.97% | 15.00% | Benchmark | TCI | 1.17 | 1.33 | TCI* | TCI* |
| 4700 | 12 | 45.52% | 10.00% | 47.62% | 0.00% | -2.10% | 10.00% | Benchmark | TCI | 1.17 | 1.38 | TCI* | TCI* |
| 4812 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4832 | 12 | 45.00% | 0.00% | 39.53% | 10.00% | 5.47% | -10.00% | TCI | Benchmark | 0.91 | 0.76 | Benchmark* | Benchmark* |
| 5399 | 12 | 47.62% | 0.00% | 40.18% | 20.00% | 7.44% | -20.00% | TCI | Benchmark | 0.79 | 0.59 | Benchmark* | Benchmark* |
| 5500 | 12 | 48.10% | 10.00% | 45.79% | 15.00% | 2.31% | -5.00% | TCI | Benchmark | 0.96 | 0.90 | Benchmark* | Benchmark* |
| 5531 | 12 | 48.10% | 10.00% | 47.62% | 0.00% | 0.48% | 10.00% | TCI | TCI | 1.22 | 1.43 | TCI | TCI |
| 5700 | 12 | 42.11% | 0.00% | 45.00% | 0.00% | -2.89% | 0.00% | Benchmark | - | 0.94 | 0.94 | Benchmark* | Benchmark* |
| 5731 | 12 | 45.52% | 10.00% | 47.86% | 5.00% | -2.33% | 5.00% | Benchmark | TCI | 1.05 | 1.13 | TCI* | TCI* |
| 5812 | 180 | 45.44% | 18.19% | 45.15% | 19.76% | 0.29% | -1.57% | TCI | Benchmark | 0.98 | 0.97 | Benchmark* | Benchmark* |
| 5912 | 36 | 42.40% | 7.50% | 44.35% | 7.50% | -1.94% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 7011 | 24 | 44.44% | 20.00% | 43.00% | 20.00% | 1.44% | 0.00% | TCI | - | 1.02 | 1.02 | TCI* | TCI* |
| 7370 | 24 | 38.54% | 26.67% | 41.32% | 16.67% | -2.78% | 10.00% | Benchmark | TCI | 1.12 | 1.23 | TCI* | TCI* |
| 7372 | 36 | 30.37% | 30.00% | 33.19% | 30.00% | -2.82% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 7510 | 12 | 46.05% | 20.00% | 34.08% | 35.00% | 11.97% | -15.00% | TCI | Benchmark | 0.96 | 0.83 | Benchmark* | Benchmark* |

# Table 46: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 9 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (9) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR | | Twitter - CI & AR | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 42.98% | 16.67% | 47.86% | 6.67% | -4.88% | 10.00% | Benchmark | TCI | 1.09 | 1.25 | TCI* | TCI* |
| 2033 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2040 | 24 | 46.72% | 10.00% | 41.64% | 23.33% | 5.09% | -13.33% | TCI | Benchmark | 0.87 | 0.76 | Benchmark* | Benchmark* |
| 2080 | 12 | 42.39% | 5.00% | 42.39% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 46.72% | 10.00% | 42.70% | 13.33% | 4.03% | -3.33% | TCI | Benchmark | 1.01 | 0.96 | TCI* | Benchmark* |
| 2086 | 36 | 43.49% | 10.00% | 43.49% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2090 | 12 | 37.81% | 35.00% | 37.81% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 47.86% | 5.00% | 50.00% | 0.00% | -2.14% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 2300 | 24 | 42.85% | 16.67% | 38.37% | 23.33% | 4.48% | -6.67% | TCI | Benchmark | 0.96 | 0.90 | Benchmark* | Benchmark* |
| 2320 | 12 | 47.62% | 0.00% | 45.00% | 0.00% | 2.62% | 0.00% | TCI | - | 1.06 | 1.06 | TCI* | TCI* |
| 2840 | 12 | 29.88% | 35.00% | 29.88% | 35.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 50.00% | 0.00% | 47.62% | 0.00% | 2.38% | 0.00% | TCI | - | 1.05 | 1.05 | TCI* | TCI* |
| 2844 | 36 | 35.67% | 29.50% | 35.41% | 24.50% | 0.26% | 5.00% | TCI | TCI | 1.09 | 1.12 | TCI | TCI |
| 2890 | 12 | 36.01% | 10.00% | 39.53% | 10.00% | -3.52% | 0.00% | Benchmark | - | 0.93 | 0.94 | Benchmark* | Benchmark* |
| 2911 | 24 | 42.55% | 10.00% | 45.40% | 10.00% | -2.85% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3021 | 12 | 42.97% | 15.00% | 42.97% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3100 | 12 | 36.37% | 15.00% | 39.85% | 15.00% | -3.48% | 0.00% | Benchmark | - | 0.94 | 0.95 | Benchmark* | Benchmark* |
| 3140 | 12 | 42.68% | 10.00% | 39.53% | 10.00% | 3.15% | 0.00% | TCI | - | 1.06 | 1.05 | TCI* | TCI* |
| 3540 | 12 | 43.26% | 20.00% | 45.52% | 10.00% | -2.26% | 10.00% | Benchmark | TCI | 1.14 | 1.27 | TCI* | TCI* |
| 3577 | 12 | 32.46% | 15.00% | 32.87% | 20.00% | -0.40% | -5.00% | Benchmark | Benchmark | 0.90 | 0.86 | Benchmark | Benchmark |
| 3630 | 12 | 36.73% | 20.00% | 40.18% | 20.00% | -3.44% | 0.00% | Benchmark | - | 0.94 | 0.96 | Benchmark* | Benchmark* |
| 3663 | 24 | 47.74% | 3.33% | 47.74% | 3.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3674 | 12 | 40.50% | 25.00% | 40.50% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3711 | 72 | 43.73% | 16.55% | 43.20% | 15.12% | 0.53% | 1.43% | TCI | TCI | 1.03 | 1.05 | TCI | TCI |
| 3751 | 12 | 42.68% | 10.00% | 45.26% | 5.00% | -2.58% | 5.00% | Benchmark | TCI | 1.05 | 1.13 | TCI* | TCI* |
| 3942 | 12 | 42.68% | 10.00% | 40.18% | 20.00% | 2.51% | -10.00% | TCI | Benchmark | 0.88 | 0.78 | Benchmark* | Benchmark* |
| 3944 | 12 | 40.18% | 20.00% | 36.73% | 20.00% | 3.44% | 0.00% | TCI | - | 1.06 | 1.04 | TCI* | TCI* |
| 3949 | 12 | 45.52% | 10.00% | 39.21% | 5.00% | 6.31% | 5.00% | TCI | TCI | 1.26 | 1.33 | TCI | TCI |
| 4210 | 12 | 32.87% | 20.00% | 36.37% | 15.00% | -3.50% | 5.00% | Benchmark | TCI | 1.03 | 1.10 | TCI* | TCI* |
| 4400 | 24 | 42.69% | 11.67% | 43.14% | 21.67% | -0.45% | -10.00% | Benchmark | Benchmark | 0.84 | 0.76 | Benchmark | Benchmark |
| 4512 | 84 | 45.46% | 13.42% | 44.24% | 11.64% | 1.23% | 1.78% | TCI | TCI | 1.05 | 1.07 | TCI | TCI |
| 4513 | 12 | 41.46% | 40.00% | 45.79% | 15.00% | -4.32% | 25.00% | Benchmark | TCI | 1.34 | 1.60 | TCI* | TCI* |
| 4700 | 12 | 42.97% | 15.00% | 47.62% | 0.00% | -4.65% | 15.00% | Benchmark | TCI | 1.22 | 1.53 | TCI* | TCI* |
| 4812 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4832 | 12 | 39.21% | 5.00% | 39.53% | 10.00% | -0.32% | -5.00% | Benchmark | Benchmark | 0.89 | 0.83 | Benchmark | Benchmark |
| 5399 | 12 | 47.86% | 5.00% | 40.18% | 20.00% | 7.68% | -15.00% | TCI | Benchmark | 0.88 | 0.72 | Benchmark* | Benchmark* |
| 5500 | 12 | 45.79% | 15.00% | 48.10% | 10.00% | -2.31% | 5.00% | Benchmark | TCI | 1.05 | 1.11 | TCI* | TCI* |
| 5531 | 12 | 47.62% | 0.00% | 47.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5700 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 50.00% | 0.00% | 47.86% | 5.00% | 2.14% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5812 | 180 | 46.24% | 10.96% | 45.34% | 15.42% | 0.89% | -4.46% | TCI | Benchmark | 0.94 | 0.89 | Benchmark* | Benchmark* |
| 5912 | 36 | 43.00% | 20.50% | 43.87% | 18.00% | -0.87% | 2.50% | Benchmark | TCI | 1.03 | 1.05 | TCI* | TCI* |
| 7011 | 24 | 47.98% | 10.00% | 44.29% | 15.00% | 3.70% | -5.00% | TCI | Benchmark | 0.98 | 0.92 | Benchmark* | Benchmark* |
| 7370 | 24 | 41.47% | 18.33% | 44.43% | 18.33% | -2.96% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 7372 | 36 | 39.88% | 19.50% | 37.71% | 20.50% | 2.16% | -1.00% | TCI | Benchmark | 1.02 | 1.00 | TCI* | TCI* |
| 7510 | 12 | 40.82% | 30.00% | 29.42% | 30.00% | 11.40% | 0.00% | TCI | - | 1.19 | 1.13 | TCI* | TCI* |

# Table 47: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 11 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (11) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR & GDPt-1 | | Twitter - CI & AR & GDPt-1 | | Benchmark - CI | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCI Total Cost | Benchmark Total Cost /TCI Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 44.57% | 21.67% | 44.57% | 21.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2033 | 12 | 42.68% | 10.00% | 45.26% | 5.00% | -2.58% | 5.00% | Benchmark | TCI | 1.05 | 1.13 | TCI* | TCI* |
| 2040 | 24 | 42.85% | 16.67% | 42.85% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2080 | 12 | 47.86% | 5.00% | 45.52% | 10.00% | 2.33% | -5.00% | TCI | Benchmark | 0.95 | 0.88 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.67% | 16.67% | 38.53% | 25.00% | 7.14% | -8.33% | TCI | Benchmark | 0.98 | 0.89 | Benchmark* | Benchmark* |
| 2086 | 36 | 42.60% | 11.50% | 43.49% | 10.00% | -0.89% | 1.50% | Benchmark | TCI | 1.01 | 1.03 | TCI* | TCI* |
| 2090 | 12 | 40.18% | 20.00% | 37.45% | 30.00% | 2.72% | -10.00% | TCI | Benchmark | 0.89 | 0.82 | Benchmark* | Benchmark* |
| 2111 | 12 | 42.39% | 5.00% | 36.01% | 10.00% | 6.38% | -5.00% | TCI | Benchmark | 1.03 | 0.94 | TCI* | Benchmark* |
| 2300 | 24 | 45.40% | 10.00% | 45.40% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2320 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 34.08% | 35.00% | 29.88% | 35.00% | 4.21% | 0.00% | TCI | - | 1.06 | 1.04 | TCI* | TCI* |
| 2842 | 12 | 48.33% | 15.00% | 45.79% | 15.00% | 2.55% | 0.00% | TCI | - | 1.04 | 1.03 | TCI* | TCI* |
| 2844 | 36 | 40.96% | 22.50% | 41.90% | 20.00% | -0.93% | 2.50% | Benchmark | TCI | 1.03 | 1.05 | TCI* | TCI* |
| 2890 | 12 | 36.37% | 15.00% | 32.46% | 15.00% | 3.91% | 0.00% | TCI | - | 1.08 | 1.06 | TCI* | TCI* |
| 2911 | 24 | 41.47% | 18.33% | 43.87% | 6.67% | -2.40% | 11.67% | Benchmark | TCI | 1.18 | 1.37 | TCI* | TCI* |
| 3021 | 12 | 36.73% | 20.00% | 36.73% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3100 | 12 | 47.62% | 0.00% | 47.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3140 | 12 | 36.01% | 10.00% | 32.06% | 10.00% | 3.95% | 0.00% | TCI | - | 1.09 | 1.08 | TCI* | TCI* |
| 3540 | 12 | 42.68% | 10.00% | 45.52% | 10.00% | -2.84% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3577 | 12 | 32.87% | 20.00% | 25.10% | 35.00% | 7.77% | -15.00% | TCI | Benchmark | 0.88 | 0.77 | Benchmark* | Benchmark* |
| 3630 | 12 | 33.27% | 25.00% | 33.27% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3663 | 24 | 46.72% | 10.00% | 44.01% | 10.00% | 2.71% | 0.00% | TCI | - | 1.05 | 1.04 | TCI* | TCI* |
| 3674 | 12 | 40.50% | 25.00% | 40.50% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.66% | 14.64% | 43.54% | 12.50% | -0.88% | 2.14% | Benchmark | TCI | 1.02 | 1.05 | TCI* | TCI* |
| 3751 | 12 | 47.86% | 5.00% | 50.00% | 0.00% | -2.14% | 5.00% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 3942 | 12 | 42.68% | 10.00% | 45.52% | 10.00% | -2.84% | 0.00% | Benchmark | - | 0.95 | 0.96 | Benchmark* | Benchmark* |
| 3944 | 12 | 21.43% | 50.00% | 21.43% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 32.46% | 15.00% | 35.65% | 5.00% | -3.19% | 10.00% | Benchmark | TCI | 1.17 | 1.37 | TCI* | TCI* |
| 4210 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 45.67% | 16.67% | 46.85% | 13.33% | -1.18% | 3.33% | Benchmark | TCI | 1.04 | 1.07 | TCI* | TCI* |
| 4512 | 84 | 45.04% | 12.72% | 43.21% | 17.67% | 1.82% | -4.94% | TCI | Benchmark | 0.95 | 0.90 | Benchmark* | Benchmark* |
| 4513 | 12 | 41.46% | 40.00% | 45.79% | 15.00% | -4.32% | 25.00% | Benchmark | TCI | 1.34 | 1.60 | TCI* | TCI* |
| 4700 | 12 | 45.52% | 10.00% | 47.62% | 0.00% | -2.10% | 10.00% | Benchmark | TCI | 1.17 | 1.38 | TCI* | TCI* |
| 4812 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4832 | 12 | 42.39% | 5.00% | 39.53% | 10.00% | 2.86% | -5.00% | TCI | Benchmark | 0.96 | 0.88 | Benchmark* | Benchmark* |
| 5399 | 12 | 47.62% | 0.00% | 39.85% | 15.00% | 7.77% | -15.00% | TCI | Benchmark | 0.87 | 0.68 | Benchmark* | Benchmark* |
| 5500 | 12 | 48.10% | 10.00% | 42.97% | 15.00% | 5.12% | -5.00% | TCI | Benchmark | 1.00 | 0.93 | TCI* | Benchmark* |
| 5531 | 12 | 36.73% | 20.00% | 39.85% | 15.00% | -3.12% | 5.00% | Benchmark | TCI | 1.03 | 1.10 | TCI* | TCI* |
| 5700 | 12 | 42.11% | 0.00% | 42.11% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 50.00% | 0.00% | 47.86% | 5.00% | 2.14% | -5.00% | TCI | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5812 | 180 | 45.16% | 15.92% | 45.02% | 17.00% | 0.15% | -1.09% | TCI | Benchmark | 0.98 | 0.97 | Benchmark* | Benchmark* |
| 5912 | 36 | 41.06% | 22.00% | 40.95% | 19.50% | 0.11% | 2.50% | TCI | TCI | 1.04 | 1.06 | TCI | TCI |
| 7011 | 24 | 46.98% | 16.67% | 44.44% | 20.00% | 2.54% | -3.33% | TCI | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 7370 | 24 | 41.48% | 20.00% | 45.67% | 16.67% | -4.19% | 3.33% | Benchmark | TCI | 0.99 | 1.03 | Benchmark* | TCI* |
| 7372 | 36 | 35.67% | 29.50% | 37.72% | 22.00% | -2.06% | 7.50% | Benchmark | TCI | 1.09 | 1.16 | TCI* | TCI* |
| 7510 | 12 | 43.55% | 25.00% | 25.10% | 35.00% | 18.46% | -10.00% | TCI | Benchmark | 1.14 | 0.98 | TCI* | Benchmark* |

# Table 48: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 6 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (6) | | | | | | | | | |
| | | Benchmark - Salest-1 | | Twitter - CS | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 46.85% | 13.33% | 44.43% | 18.33% | 2.42% | -5.00% | TCS | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 2033 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2040 | 24 | 45.40% | 10.00% | 45.40% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2080 | 12 | 39.85% | 15.00% | 39.85% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.80% | 20.00% | 46.85% | 13.33% | -1.05% | 6.67% | Benchmark | TCS | 1.09 | 1.17 | TCS* | TCS* |
| 2086 | 36 | 43.68% | 15.00% | 43.68% | 14.00% | 0.00% | 1.00% | TCS | TCS | 1.02 | 1.03 | TCS | TCS |
| 2090 | 12 | 43.26% | 20.00% | 36.37% | 15.00% | 6.89% | 5.00% | TCS | TCS | 1.23 | 1.25 | TCS | TCS |
| 2111 | 12 | 47.86% | 5.00% | 41.14% | 35.00% | 6.72% | -30.00% | TCS | Benchmark | 0.69 | 0.52 | Benchmark* | Benchmark* |
| 2300 | 24 | 45.13% | 3.33% | 46.47% | 3.33% | -1.34% | 0.00% | Benchmark | - | 0.97 | 0.97 | Benchmark* | Benchmark* |
| 2320 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 45.00% | 0.00% | 50.00% | 0.00% | -5.00% | 0.00% | Benchmark | - | 0.90 | 0.90 | Benchmark* | Benchmark* |
| 2844 | 36 | 44.44% | 10.00% | 41.79% | 15.50% | 2.65% | -5.50% | TCS | Benchmark | 0.95 | 0.89 | Benchmark* | Benchmark* |
| 2890 | 12 | 39.53% | 10.00% | 39.53% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2911 | 24 | 45.67% | 16.67% | 45.67% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3021 | 12 | 50.00% | 0.00% | 48.10% | 10.00% | 1.90% | -10.00% | TCS | Benchmark | 0.86 | 0.73 | Benchmark* | Benchmark* |
| 3100 | 12 | 39.21% | 5.00% | 39.21% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3140 | 12 | 45.26% | 5.00% | 39.21% | 5.00% | 6.05% | 0.00% | TCS | - | 1.14 | 1.12 | TCS* | TCS* |
| 3540 | 12 | 43.26% | 20.00% | 43.26% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3577 | 12 | 42.39% | 5.00% | 42.39% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3663 | 24 | 41.63% | 21.67% | 44.42% | 16.67% | -2.79% | 5.00% | Benchmark | TCS | 1.04 | 1.09 | TCS* | TCS* |
| 3674 | 12 | 47.86% | 5.00% | 45.26% | 5.00% | 2.60% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 3711 | 72 | 45.13% | 16.31% | 44.58% | 14.05% | 0.55% | 2.26% | TCS | TCS | 1.05 | 1.07 | TCS | TCS |
| 3751 | 12 | 45.26% | 5.00% | 47.62% | 0.00% | -2.36% | 5.00% | Benchmark | TCS | 1.06 | 1.16 | TCS* | TCS* |
| 3942 | 12 | 45.00% | 0.00% | 36.73% | 20.00% | 8.27% | -20.00% | TCS | Benchmark | 0.79 | 0.59 | Benchmark* | Benchmark* |
| 3944 | 12 | 42.68% | 10.00% | 45.00% | 0.00% | -2.32% | 10.00% | Benchmark | TCS | 1.17 | 1.39 | TCS* | TCS* |
| 3949 | 12 | 45.26% | 5.00% | 45.26% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4210 | 12 | 42.68% | 10.00% | 32.87% | 20.00% | 9.82% | -10.00% | TCS | Benchmark | 1.00 | 0.86 | Benchmark* | Benchmark* |
| 4400 | 24 | 40.03% | 20.00% | 41.31% | 15.00% | -1.29% | 5.00% | Benchmark | TCS | 1.07 | 1.12 | TCS* | TCS* |
| 4512 | 84 | 45.85% | 13.69% | 43.63% | 16.86% | 2.22% | -3.16% | TCS | Benchmark | 0.98 | 0.95 | Benchmark* | Benchmark* |
| 4513 | 12 | 45.79% | 15.00% | 45.79% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 45.79% | 15.00% | 48.33% | 15.00% | -2.55% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 4812 | 12 | 40.18% | 20.00% | 39.85% | 15.00% | 0.32% | 5.00% | TCS | TCS | 1.10 | 1.15 | TCS | TCS |
| 4832 | 12 | 42.11% | 0.00% | 39.21% | 5.00% | 2.89% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5399 | 12 | 34.08% | 35.00% | 42.97% | 15.00% | -8.89% | 20.00% | Benchmark | TCS | 1.19 | 1.43 | TCS* | TCS* |
| 5500 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5531 | 12 | 45.26% | 5.00% | 42.39% | 5.00% | 2.87% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 5700 | 12 | 42.68% | 10.00% | 39.53% | 10.00% | 3.15% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 5731 | 12 | 50.00% | 0.00% | 42.97% | 15.00% | 7.03% | -15.00% | TCS | Benchmark | 0.86 | 0.69 | Benchmark* | Benchmark* |
| 5812 | 180 | 45.09% | 13.83% | 44.41% | 15.14% | 0.68% | -1.30% | TCS | Benchmark | 0.99 | 0.97 | Benchmark* | Benchmark* |
| 5912 | 36 | 42.80% | 16.50% | 42.80% | 15.50% | 0.00% | 1.00% | TCS | TCS | 1.02 | 1.03 | TCS | TCS |
| 7011 | 24 | 45.54% | 13.33% | 44.29% | 16.67% | 1.24% | -3.33% | TCS | Benchmark | 0.97 | 0.93 | Benchmark* | Benchmark* |
| 7370 | 24 | 40.68% | 31.67% | 40.52% | 30.00% | 0.16% | 1.67% | TCS | TCS | 1.03 | 1.03 | TCS | TCS |
| 7372 | 36 | 35.16% | 22.50% | 38.43% | 15.00% | -3.26% | 7.50% | Benchmark | TCS | 1.08 | 1.17 | TCS* | TCS* |
| 7510 | 12 | 48.10% | 10.00% | 48.10% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |

# Table 49: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 8 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (8) | | | | | | | | | |
| | | Benchmark - Salest-1 & GDPt-1 | | Twitter - CS & GDPt-1 | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 45.80% | 18.33% | 44.71% | 25.00% | 1.09% | -6.67% | TCS | Benchmark | 0.92 | 0.87 | Benchmark* | Benchmark* |
| 2033 | 12 | 45.26% | 5.00% | 42.39% | 5.00% | 2.87% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 2040 | 24 | 44.15% | 13.33% | 45.40% | 10.00% | -1.25% | 3.33% | Benchmark | TCS | 1.04 | 1.08 | TCS* | TCS* |
| 2080 | 12 | 39.85% | 15.00% | 39.85% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 46.97% | 15.00% | 49.07% | 6.67% | -2.10% | 8.33% | Benchmark | TCS | 1.11 | 1.23 | TCS* | TCS* |
| 2086 | 36 | 40.74% | 15.50% | 42.50% | 10.00% | -1.76% | 5.50% | Benchmark | TCS | 1.07 | 1.15 | TCS* | TCS* |
| 2090 | 12 | 35.65% | 5.00% | 35.65% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 39.53% | 10.00% | 39.21% | 5.00% | 0.32% | 5.00% | TCS | TCS | 1.12 | 1.21 | TCS | TCS |
| 2300 | 24 | 45.40% | 8.33% | 42.40% | 6.67% | 2.99% | 1.67% | TCS | TCS | 1.09 | 1.11 | TCS | TCS |
| 2320 | 12 | 47.62% | 0.00% | 47.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 50.00% | 0.00% | 50.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2842 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2844 | 36 | 42.10% | 24.00% | 38.19% | 30.50% | 3.91% | -6.50% | TCS | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 2890 | 12 | 39.53% | 10.00% | 39.53% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2911 | 24 | 49.07% | 6.67% | 49.07% | 6.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3021 | 12 | 36.73% | 20.00% | 40.50% | 25.00% | -3.77% | -5.00% | Benchmark | Benchmark | 0.87 | 0.85 | Benchmark | Benchmark |
| 3100 | 12 | 35.65% | 5.00% | 39.21% | 5.00% | -3.56% | 0.00% | Benchmark | - | 0.92 | 0.93 | Benchmark* | Benchmark* |
| 3140 | 12 | 45.26% | 5.00% | 42.39% | 5.00% | 2.87% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 3540 | 12 | 45.52% | 10.00% | 42.68% | 10.00% | 2.84% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 3577 | 12 | 42.68% | 10.00% | 39.53% | 10.00% | 3.15% | 0.00% | TCS | - | 1.06 | 1.05 | TCS* | TCS* |
| 3630 | 12 | 33.27% | 25.00% | 37.09% | 25.00% | -3.82% | 0.00% | Benchmark | - | 0.94 | 0.96 | Benchmark* | Benchmark* |
| 3663 | 24 | 46.60% | 6.67% | 47.62% | 0.00% | -1.02% | 6.67% | Benchmark | TCS | 1.12 | 1.26 | TCS* | TCS* |
| 3674 | 12 | 48.10% | 10.00% | 45.79% | 15.00% | 2.31% | -5.00% | TCS | Benchmark | 0.96 | 0.90 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.90% | 19.83% | 44.30% | 19.17% | -1.40% | 0.67% | Benchmark | TCS | 0.99 | 1.00 | Benchmark* | Benchmark* |
| 3751 | 12 | 42.68% | 10.00% | 45.00% | 0.00% | -2.32% | 10.00% | Benchmark | TCS | 1.17 | 1.39 | TCS* | TCS* |
| 3942 | 12 | 45.00% | 0.00% | 45.52% | 10.00% | -0.52% | -10.00% | Benchmark | Benchmark | 0.81 | 0.69 | Benchmark | Benchmark |
| 3944 | 12 | 50.00% | 0.00% | 47.62% | 0.00% | 2.38% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 3949 | 12 | 45.26% | 5.00% | 50.00% | 0.00% | -4.74% | 5.00% | Benchmark | TCS | 1.01 | 1.11 | TCS* | TCS* |
| 4210 | 12 | 40.18% | 20.00% | 37.45% | 30.00% | 2.72% | -10.00% | TCS | Benchmark | 0.89 | 0.82 | Benchmark* | Benchmark* |
| 4400 | 24 | 44.29% | 16.67% | 44.44% | 20.00% | -0.14% | -3.33% | Benchmark | Benchmark | 0.95 | 0.92 | Benchmark | Benchmark |
| 4512 | 84 | 45.54% | 15.58% | 44.36% | 15.58% | 1.18% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 4513 | 12 | 40.82% | 30.00% | 43.84% | 30.00% | -3.02% | 0.00% | Benchmark | - | 0.96 | 0.97 | Benchmark* | Benchmark* |
| 4700 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4812 | 12 | 40.18% | 20.00% | 39.53% | 10.00% | 0.64% | 10.00% | TCS | TCS | 1.21 | 1.35 | TCS | TCS |
| 4832 | 12 | 45.00% | 0.00% | 42.39% | 5.00% | 2.61% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5399 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5500 | 12 | 48.10% | 10.00% | 48.10% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5531 | 12 | 48.10% | 10.00% | 45.52% | 10.00% | 2.57% | 0.00% | TCS | - | 1.05 | 1.04 | TCS* | TCS* |
| 5700 | 12 | 42.11% | 0.00% | 42.11% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 45.52% | 10.00% | 42.68% | 10.00% | 2.84% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 5812 | 180 | 45.44% | 18.19% | 44.24% | 15.14% | 1.20% | 3.05% | TCS | TCS | 1.07 | 1.10 | TCS | TCS |
| 5912 | 36 | 42.40% | 7.50% | 45.36% | 10.00% | -2.96% | -2.50% | Benchmark | Benchmark | 0.90 | 0.88 | Benchmark | Benchmark |
| 7011 | 24 | 44.44% | 20.00% | 43.14% | 21.67% | 1.30% | -1.67% | TCS | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 7370 | 24 | 38.54% | 26.67% | 41.48% | 20.00% | -2.94% | 6.67% | Benchmark | TCS | 1.06 | 1.13 | TCS* | TCS* |
| 7372 | 36 | 30.37% | 30.00% | 34.24% | 25.00% | -3.88% | 5.00% | Benchmark | TCS | 1.02 | 1.07 | TCS* | TCS* |
| 7510 | 12 | 46.05% | 20.00% | 50.00% | 0.00% | -3.95% | 20.00% | Benchmark | TCS | 1.32 | 1.72 | TCS* | TCS* |

# Table 50: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 10 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (10) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR | | Twitter - CS & AR | | Benchmark - CS | | | | (1:1) | (1:2) | (1:1) | (1:2) |
| 4-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Better Model - FP | Better Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 2000 | 24 | 42.98% | 16.67% | 42.70% | 13.33% | 0.28% | 3.33% | TCS | TCS | 1.06 | 1.10 | TCS | TCS |
| 2033 | 12 | 45.52% | 10.00% | 42.68% | 10.00% | 2.84% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 2040 | 24 | 46.72% | 10.00% | 45.40% | 10.00% | 1.32% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 2080 | 12 | 42.39% | 5.00% | 45.26% | 5.00% | -2.87% | 0.00% | Benchmark | - | 0.94 | 0.95 | Benchmark* | Benchmark* |
| 2082 | 24 | 46.72% | 10.00% | 48.95% | 3.33% | -2.23% | 6.67% | Benchmark | TCS | 1.08 | 1.20 | TCS* | TCS* |
| 2086 | 36 | 43.49% | 10.00% | 44.35% | 7.50% | -0.86% | 2.50% | Benchmark | TCS | 1.03 | 1.07 | TCS* | TCS* |
| 2090 | 12 | 37.81% | 35.00% | 40.82% | 30.00% | -3.01% | 5.00% | Benchmark | TCS | 1.03 | 1.07 | TCS* | TCS* |
| 2111 | 12 | 47.86% | 5.00% | 46.31% | 25.00% | 1.55% | -20.00% | TCS | Benchmark | 0.74 | 0.60 | Benchmark* | Benchmark* |
| 2300 | 24 | 42.85% | 16.67% | 41.32% | 16.67% | 1.52% | 0.00% | TCS | - | 1.03 | 1.02 | TCS* | TCS* |
| 2320 | 12 | 47.62% | 0.00% | 47.62% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 29.88% | 35.00% | 37.45% | 30.00% | -7.58% | 5.00% | Benchmark | TCS | 0.96 | 1.02 | Benchmark* | TCS* |
| 2842 | 12 | 50.00% | 0.00% | 47.62% | 0.00% | 2.38% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 2844 | 36 | 35.67% | 29.50% | 34.38% | 28.50% | 1.28% | 1.00% | TCS | TCS | 1.04 | 1.04 | TCS | TCS |
| 2890 | 12 | 36.01% | 10.00% | 39.53% | 10.00% | -3.52% | 0.00% | Benchmark | - | 0.93 | 0.94 | Benchmark* | Benchmark* |
| 2911 | 24 | 42.55% | 10.00% | 41.01% | 10.00% | 1.54% | 0.00% | TCS | - | 1.03 | 1.03 | TCS* | TCS* |
| 3021 | 12 | 42.97% | 15.00% | 46.05% | 20.00% | -3.07% | -5.00% | Benchmark | Benchmark | 0.88 | 0.85 | Benchmark | Benchmark |
| 3100 | 12 | 36.37% | 15.00% | 42.68% | 10.00% | -6.31% | 5.00% | Benchmark | TCS | 0.98 | 1.06 | Benchmark* | TCS* |
| 3140 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3540 | 12 | 43.26% | 20.00% | 43.26% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3577 | 12 | 32.46% | 15.00% | 32.46% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 36.73% | 20.00% | 40.18% | 20.00% | -3.44% | 0.00% | Benchmark | - | 0.94 | 0.96 | Benchmark* | Benchmark* |
| 3663 | 24 | 47.74% | 3.33% | 47.98% | 10.00% | -0.24% | -6.67% | Benchmark | Benchmark | 0.88 | 0.80 | Benchmark | Benchmark |
| 3674 | 12 | 40.50% | 25.00% | 40.50% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3711 | 72 | 43.73% | 16.55% | 42.80% | 17.21% | 0.93% | -0.67% | TCS | Benchmark | 1.00 | 0.99 | TCS* | Benchmark* |
| 3751 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3942 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3944 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 45.52% | 10.00% | 42.68% | 10.00% | 2.84% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 4210 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 42.69% | 11.67% | 42.55% | 10.00% | 0.14% | 1.67% | TCS | TCS | 1.03 | 1.06 | TCS | TCS |
| 4512 | 84 | 45.46% | 13.42% | 43.29% | 17.75% | 2.17% | -4.33% | TCS | Benchmark | 0.96 | 0.92 | Benchmark* | Benchmark* |
| 4513 | 12 | 41.46% | 40.00% | 41.46% | 40.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 42.97% | 15.00% | 45.52% | 10.00% | -2.55% | 5.00% | Benchmark | TCS | 1.04 | 1.11 | TCS* | TCS* |
| 4812 | 12 | 40.18% | 20.00% | 39.53% | 10.00% | 0.64% | 10.00% | TCS | TCS | 1.21 | 1.35 | TCS | TCS |
| 4832 | 12 | 39.21% | 5.00% | 36.01% | 10.00% | 3.20% | -5.00% | TCS | Benchmark | 0.96 | 0.88 | Benchmark* | Benchmark* |
| 5399 | 12 | 47.86% | 5.00% | 40.18% | 20.00% | 7.68% | -15.00% | TCS | Benchmark | 0.88 | 0.72 | Benchmark* | Benchmark* |
| 5500 | 12 | 45.79% | 15.00% | 45.79% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5531 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5700 | 12 | 42.68% | 10.00% | 42.11% | 0.00% | 0.58% | 10.00% | TCS | TCS | 1.25 | 1.49 | TCS | TCS |
| 5731 | 12 | 50.00% | 0.00% | 42.68% | 10.00% | 7.32% | -10.00% | TCS | Benchmark | 0.95 | 0.80 | Benchmark* | Benchmark* |
| 5812 | 180 | 46.24% | 10.96% | 44.41% | 15.41% | 1.83% | -4.45% | TCS | Benchmark | 0.96 | 0.91 | Benchmark* | Benchmark* |
| 5912 | 36 | 43.00% | 20.50% | 42.30% | 27.00% | 0.70% | -6.50% | TCS | Benchmark | 0.92 | 0.87 | Benchmark* | Benchmark* |
| 7011 | 24 | 47.98% | 10.00% | 45.54% | 13.33% | 2.45% | -3.33% | TCS | Benchmark | 0.98 | 0.94 | Benchmark* | Benchmark* |
| 7370 | 24 | 41.47% | 18.33% | 42.85% | 16.67% | -1.38% | 1.67% | Benchmark | TCS | 1.00 | 1.03 | TCS* | TCS* |
| 7372 | 36 | 39.88% | 19.50% | 41.89% | 18.00% | -2.01% | 1.50% | Benchmark | TCS | 0.99 | 1.01 | Benchmark* | TCS* |
| 7510 | 12 | 40.82% | 30.00% | 45.79% | 15.00% | -4.97% | 15.00% | Benchmark | TCS | 1.17 | 1.33 | TCS* | TCS* |

# Table 51: 4-Digit SIC - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 12 and 4)

Error Detection Ability - Alpha = 0.33

| 4-Digit SIC | Number of Observations | (4) Benchmark - Salest-1 & AR & GDPt-1 False Positive | (4) Benchmark - Salest-1 & AR & GDPt-1 False Negative | (12) Twitter - CS & AR & GDPt-1 False Positive | (12) Twitter - CS & AR & GDPt-1 False Negative | Benchmark - CS Difference - FP | Benchmark - CS Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCS Total Cost | (1:2) Benchmark Total Cost /TCS Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2000 | 24 | 44.57% | 21.67% | 43.14% | 21.67% | 1.43% | 0.00% | TCS | - | 1.02 | 1.02 | TCS* | TCS* |
| 2033 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2040 | 24 | 42.85% | 16.67% | 42.85% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2080 | 12 | 47.86% | 5.00% | 47.86% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2082 | 24 | 45.67% | 16.67% | 46.72% | 10.00% | -1.05% | 6.67% | Benchmark | TCS | 1.10 | 1.18 | TCS* | TCS* |
| 2086 | 36 | 42.60% | 11.50% | 45.27% | 7.50% | -2.67% | 4.00% | Benchmark | TCS | 1.03 | 1.09 | TCS* | TCS* |
| 2090 | 12 | 40.18% | 20.00% | 40.18% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2111 | 12 | 42.39% | 5.00% | 39.21% | 5.00% | 3.18% | 0.00% | TCS | - | 1.07 | 1.06 | TCS* | TCS* |
| 2300 | 24 | 45.40% | 10.00% | 41.63% | 21.67% | 3.77% | -11.67% | TCS | Benchmark | 0.88 | 0.77 | Benchmark* | Benchmark* |
| 2320 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2840 | 12 | 34.08% | 35.00% | 37.09% | 25.00% | -3.01% | 10.00% | Benchmark | TCS | 1.11 | 1.20 | TCS* | TCS* |
| 2842 | 12 | 48.33% | 15.00% | 45.79% | 15.00% | 2.55% | 0.00% | TCS | - | 1.04 | 1.03 | TCS* | TCS* |
| 2844 | 36 | 40.96% | 22.50% | 40.75% | 17.50% | 0.21% | 5.00% | TCS | TCS | 1.09 | 1.13 | TCS | TCS |
| 2890 | 12 | 36.37% | 15.00% | 36.37% | 15.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 2911 | 24 | 41.47% | 18.33% | 41.47% | 18.33% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3021 | 12 | 36.73% | 20.00% | 40.50% | 25.00% | -3.77% | -5.00% | Benchmark | Benchmark | 0.87 | 0.85 | Benchmark | Benchmark |
| 3100 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 3140 | 12 | 36.01% | 10.00% | 36.01% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3540 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3577 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3630 | 12 | 33.27% | 25.00% | 33.27% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3663 | 24 | 46.72% | 10.00% | 42.85% | 16.67% | 3.88% | -6.67% | TCS | Benchmark | 0.95 | 0.88 | Benchmark* | Benchmark* |
| 3674 | 12 | 40.50% | 25.00% | 40.50% | 25.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3711 | 72 | 42.66% | 14.64% | 42.65% | 14.05% | 0.00% | 0.60% | TCS | TCS | 1.01 | 1.02 | TCS | TCS |
| 3751 | 12 | 47.86% | 5.00% | 45.26% | 5.00% | 2.60% | 0.00% | TCS | - | 1.05 | 1.05 | TCS* | TCS* |
| 3942 | 12 | 42.68% | 10.00% | 42.68% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3944 | 12 | 21.43% | 50.00% | 21.43% | 50.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 3949 | 12 | 32.46% | 15.00% | 28.04% | 15.00% | 4.42% | 0.00% | TCS | - | 1.10 | 1.08 | TCS* | TCS* |
| 4210 | 12 | 32.87% | 20.00% | 32.87% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4400 | 24 | 45.67% | 16.67% | 45.67% | 16.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4512 | 84 | 45.04% | 12.72% | 43.13% | 15.92% | 1.91% | -3.19% | TCS | Benchmark | 0.98 | 0.94 | Benchmark* | Benchmark* |
| 4513 | 12 | 41.46% | 40.00% | 41.46% | 40.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4700 | 12 | 45.52% | 10.00% | 45.52% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 4812 | 12 | 40.18% | 20.00% | 39.53% | 10.00% | 0.64% | 10.00% | TCS | TCS | 1.21 | 1.35 | TCS | TCS |
| 4832 | 12 | 42.39% | 5.00% | 42.39% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5399 | 12 | 47.62% | 0.00% | 45.26% | 5.00% | 2.36% | -5.00% | TCS | Benchmark | 0.95 | 0.86 | Benchmark* | Benchmark* |
| 5500 | 12 | 48.10% | 10.00% | 48.10% | 10.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5531 | 12 | 36.73% | 20.00% | 36.73% | 20.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5700 | 12 | 42.11% | 0.00% | 42.11% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 5731 | 12 | 50.00% | 0.00% | 42.68% | 10.00% | 7.32% | -10.00% | TCS | Benchmark | 0.95 | 0.80 | Benchmark* | Benchmark* |
| 5812 | 180 | 45.16% | 15.92% | 43.88% | 16.22% | 1.29% | -0.31% | TCS | Benchmark | 1.02 | 1.01 | TCS* | TCS* |
| 5912 | 36 | 41.06% | 22.00% | 39.11% | 27.00% | 1.95% | -5.00% | TCS | Benchmark | 0.95 | 0.91 | Benchmark* | Benchmark* |
| 7011 | 24 | 46.98% | 16.67% | 44.44% | 20.00% | 2.54% | -3.33% | TCS | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 7370 | 24 | 41.48% | 20.00% | 42.85% | 16.67% | -1.37% | 3.33% | Benchmark | TCS | 1.03 | 1.07 | TCS* | TCS* |
| 7372 | 36 | 35.67% | 29.50% | 35.67% | 29.50% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 7510 | 12 | 43.55% | 25.00% | 45.52% | 10.00% | -1.97% | 15.00% | Benchmark | TCS | 1.23 | 1.43 | TCS* | TCS* |

## Table 52: 4-Digit SIC vs. 2-Digit SIC – Error Detection Performance Summary of Traditional and Continuous Substantive Analytical Models

**Cost Ratio - 46 Industries**

| 4 - Digit SIC | Twitter Consumer Interest | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) |
| Traditional - SAP | 18 | 18 | 15 | 19 | 15 | 16 | 14 | 17 |
| Continuous - SAP | 11 | 21 | 20 | 21 | 9 | 20 | 19 | 19 |

| 4 - Digit SIC | Twitter Consumer Sentiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 21 | 19 | 21 | 19 | 18 | 18 | 20 | 19 |
| Continuous - SAP | 17 | 21 | 18 | 14 | 17 | 21 | 20 | 14 |

**Cost Ratio - 24 Industries**

| 2 - Digit SIC | Twitter Consumer Interest | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) |
| Traditional - SAP | 11 | 12 | 8 | 12 | 11 | 12 | 8 | 12 |
| Continuous - SAP | 4 | 13 | 12 | 14 | 5 | 13 | 11 | 12 |

| 2 - Digit SIC | Twitter Consumer Sentiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 13 | 12 | 12 | 12 | 13 | 12 | 12 | 12 |
| Continuous - SAP | 10 | 13 | 10 | 9 | 10 | 13 | 12 | 9 |

**Cost Ratio - 46 Industries**

| 4 - Digit SIC | Twitter Consumer Interest | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) | (1) vs. (5) | (2) vs. (7) | (3) vs. (9) | (4) vs. (11) |
| Traditional - SAP | 39% | 39% | 33% | 41% | 33% | 35% | 30% | 37% |
| Continuous - SAP | 24% | 46% | 43% | 46% | 20% | 43% | 41% | 41% |

| 4 - Digit SIC | Twitter Consumer Sentiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Model | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) | (1) vs. (6) | (2) vs. (8) | (3) vs. (10) | (4) vs. (12) |
| Traditional - SAP | 46% | 41% | 46% | 41% | 39% | 39% | 43% | 41% |
| Continuous - SAP | 37% | 46% | 39% | 30% | 37% | 46% | 43% | 30% |

**Cost Ratio - 24 Industries**

| 2 - Digit SIC | Twitter Consumer Interest | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Traditional - SAP | 46% | 50% | 33% | 50% | 46% | 50% | 33% | 50% |
| Continuous - SAP | 17% | 54% | 50% | 58% | 21% | 54% | 46% | 50% |
| Relative Difference 4 - Digit SIC vs. 2 - Digit SIC | | | | | | | | |
| Traditional - SAP % change | -7% | -11% | -1% | -9% | -13% | -15% | -3% | -13% |
| Continuous - SAP % change | 7% | -9% | -7% | -13% | -1% | -11% | -5% | -9% |

| 2 - Digit SIC | Twitter Consumer Sentiment | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 to 1 | | | | 1 to 2 | | | |
| Traditional - SAP | 54% | 50% | 50% | 50% | 54% | 50% | 50% | 50% |
| Continuous - SAP | 42% | 54% | 42% | 38% | 42% | 54% | 50% | 38% |
| Relative Difference 4 - Digit SIC vs. 2 - Digit SIC | | | | | | | | |
| Traditional - SAP % change | -9% | -9% | -4% | -9% | -15% | -11% | -7% | -9% |
| Continuous - SAP % change | -5% | -9% | -3% | -7% | -5% | -9% | -7% | -7% |

## Table 53: 2-Digit SIC and Advertising Expense - Prediction Performance of Continuous Substantive Analytical Models with TCI and without TCI (Models 5, 7, 9 and 11 and 1, 2, 3, and 4)

| | (1) | (5) | | | (2) | (7) | | | (3) | (9) | | | (4) | (11) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 + Advt-1 | Saletst-1+TweetCI | | | Saletst-1+GDPt-1+Advt-1 | Saletst-1+TweetCI+GDPt-1 | | | Saletst-1+AR+Advt-1 | Saletst-1+AR+TweetCI | | | Saletst-1+AR+GDPt-1+Advt-1 | Salest1+AR+TweetCI+GDPt-1 | | |
| 2-Digit SIC | MAPE1 | MAPE5 | Difference B/W | p-value | MAPE2 | MAPE7 | Difference B/W | p-value | MAPE3 | MAPE9 | Difference B/W | p-value | MAPE4 | MAPE11 | Difference B/W | p-value |
| 20 | 0.064833 | 0.092126 | -0.0273 W | 0.000 | 0.058191 | 0.0545 | 0.0037 B | 0.002 | 0.056 | 0.078 | -0.021 W | 0.000 | 0.057 | 0.040 | 0.017 B | 0.000 |
| 21 | 0.049036 | 0.056455 | -0.0074 W | 0.001 | 0.04896 | 0.0199 | 0.0290 B | 0.001 | 0.048 | 0.056 | -0.008 W | 0.001 | 0.045 | 0.019 | 0.026 B | 0.001 |
| 23 | 0.114574 | 0.137019 | -0.0224 W | 0.000 | 0.092729 | 0.0576 | 0.0351 B | 0.000 | 0.072 | 0.086 | -0.014 W | 0.077 | 0.071 | 0.045 | 0.027 B | 0.000 |
| 28 | 0.046397 | 0.073338 | -0.0269 W | 0.000 | 0.0414 | 0.0346 | 0.0068 B | 0.000 | 0.040 | 0.046 | -0.006 W | 0.000 | 0.040 | 0.030 | 0.010 B | 0.001 |
| 31 | 0.081963 | 0.16913 | -0.0872 W | 0.001 | 0.072091 | 0.1403 | -0.0682 W | 0.001 | 0.070 | 0.086 | -0.017 W | 0.001 | 0.061 | 0.079 | -0.018 W | 0.001 |
| 35 | 0.060249 | 0.110483 | -0.0502 W | 0.000 | 0.063008 | 0.0211 | 0.0419 B | 0.034 | 0.051 | 0.072 | -0.021 W | 0.000 | 0.046 | 0.022 | 0.024 B | 0.000 |
| 36 | 0.07257 | 0.067517 | 0.0051 B | 0.077 | 0.058609 | 0.0816 | -0.0230 W | 0.000 | 0.058 | 0.055 | 0.003 B | 0.599 | 0.057 | 0.064 | -0.008 W | 0.000 |
| 37 | 0.08462 | 0.108134 | -0.0235 W | 0.000 | 0.060339 | 0.0493 | 0.0110 B | 0.016 | 0.085 | 0.103 | -0.018 W | 0.000 | 0.065 | 0.043 | 0.021 B | 0.000 |
| 39 | 0.11173 | 0.300678 | -0.1889 W | 0.000 | 0.087438 | 0.0385 | 0.0490 B | 0.000 | 0.115 | 0.209 | -0.094 W | 0.000 | 0.088 | 0.039 | 0.049 B | 0.000 |
| 44 | 0.100455 | 0.139587 | -0.0391 W | 0.000 | 0.090369 | 0.0808 | 0.0096 B | 0.000 | 0.096 | 0.144 | -0.048 W | 0.000 | 0.089 | 0.071 | 0.018 B | 0.000 |
| 45 | 0.06296 | 0.064703 | -0.0017 W | 0.598 | 0.054265 | 0.0723 | -0.0180 W | 0.416 | 0.062 | 0.062 | 0.000 B | 0.416 | 0.052 | 0.075 | -0.022 W | 0.000 |
| 47 | 0.077699 | 0.129399 | -0.0517 W | 0.001 | 0.077729 | 0.0182 | 0.0596 B | 0.001 | 0.077 | 0.126 | -0.049 W | 0.001 | 0.077 | 0.020 | 0.057 B | 0.001 |
| 48 | 0.027437 | 0.028751 | -0.0013 W | 0.034 | 0.027471 | 0.0184 | 0.0091 B | 0.000 | 0.028 | 0.029 | -0.001 W | 0.034 | 0.028 | 0.018 | 0.010 B | 0.000 |
| 55 | 0.068057 | 0.097769 | -0.0297 W | 0.000 | 0.069535 | 0.0540 | 0.0155 B | 0.000 | 0.074 | 0.076 | -0.002 W | 0.000 | 0.048 | 0.045 | 0.002 B | 0.034 |
| 57 | 0.104231 | 0.098899 | 0.0053 B | 0.001 | 0.078032 | 0.0404 | 0.0377 B | 0.001 | 0.104 | 0.099 | 0.005 B | 0.001 | 0.072 | 0.040 | 0.032 B | 0.001 |
| 58 | 0.05989 | 0.087865 | -0.0280 W | 0.000 | 0.055533 | 0.0200 | 0.0356 B | 0.568 | 0.058 | 0.080 | -0.022 W | 0.000 | 0.055 | 0.020 | 0.035 B | 0.029 |
| 59 | 0.114291 | 0.251458 | -0.1372 W | 0.077 | 0.087587 | 0.0418 | 0.0458 B | 0.077 | 0.069 | 0.251 | -0.182 W | 0.077 | 0.072 | 0.043 | 0.028 B | 0.000 |
| 73 | 0.063342 | 0.079962 | -0.0166 W | 0.000 | 0.053325 | 0.0336 | 0.0197 B | 0.000 | 0.064 | 0.075 | -0.011 W | 0.000 | 0.054 | 0.041 | 0.014 B | 0.000 |
| 75 | 0.070886 | 0.092242 | -0.0214 W | 0.001 | 0.062475 | 0.0673 | -0.0048 W | 0.001 | 0.072 | 0.075 | -0.004 W | 0.001 | 0.069 | 0.068 | 0.002 B | 0.001 |

## Table 54: 2-Digit SIC and Advertising Expense - Prediction Performance of Continuous Substantive Analytical Models with TCS and without TCS (Models 6, 8, 10 and 12 and 1, 2, 3, and 4)

| | (1) | (6) | | | (2) | (8) | | | (3) | (10) | | | (4) | (12) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Saletst-1 + Advt-1 | Saletst-1+TweetCS | | | Saletst-1+GDPt-1+Advt-1 | Saletst-1+TweetCS+GDPt-1 | | | Saletst-1+AR+Advt-1 | Saletst-1+AR+TweetCS | | | Saletst-1+AR+GDPt-1+Advt-1 | Saletst-1+AR+TweetCS+GDPt-1 | | |
| 2-Digit SIC | MAPE1 | MAPE6 | Difference B/W | p-value | MAPE2 | MAPE8 | Difference B/W | p-value | MAPE3 | MAPE10 | Difference B/W | p-value | MAPE4 | MAPE12 | Difference B/W | p-value |
| 20 | 0.064833 | 0.088308 | -0.0235 W | 0.000 | 0.0582 | 0.0576 | 0.0006 B | 0.001 | 0.0564 | 0.0797 | -0.0233 W | 0.000 | 0.0565 | 0.0421 | 0.0144 B | 0.000 |
| 21 | 0.049036 | 0.059124 | -0.0101 W | 0.001 | 0.0490 | 0.0242 | 0.0247 B | 0.001 | 0.0483 | 0.0584 | -0.0101 W | 0.001 | 0.0455 | 0.0201 | 0.0254 B | 0.001 |
| 23 | 0.114574 | 0.136425 | -0.0219 W | 0.077 | 0.0927 | 0.0577 | 0.0350 B | 0.000 | 0.0718 | 0.0977 | -0.0259 W | 0.000 | 0.0714 | 0.0435 | 0.0278 B | 0.000 |
| 28 | 0.046397 | 0.07062 | -0.0242 W | 0.000 | 0.0414 | 0.0336 | 0.0078 B | 0.000 | 0.0402 | 0.0453 | -0.0051 W | 0.000 | 0.0399 | 0.0291 | 0.0108 B | 0.000 |
| 31 | 0.081963 | 0.14937 | -0.0674 W | 0.001 | 0.0721 | 0.1715 | -0.0994 W | 0.001 | 0.0695 | 0.0947 | -0.0252 W | 0.001 | 0.0614 | 0.0730 | -0.0116 W | 0.001 |
| 35 | 0.060249 | 0.121821 | -0.0616 W | 0.000 | 0.0630 | 0.0216 | 0.0414 B | 0.034 | 0.0510 | 0.0731 | -0.0222 W | 0.000 | 0.0461 | 0.0222 | 0.0239 B | 0.034 |
| 36 | 0.07257 | 0.07607 | -0.0035 W | 0.077 | 0.0586 | 0.0842 | -0.0256 W | 0.000 | 0.0576 | 0.0611 | -0.0035 W | 0.000 | 0.0567 | 0.0627 | -0.0061 W | 0.077 |
| 37 | 0.08462 | 0.152261 | -0.0676 W | 0.000 | 0.0603 | 0.0465 | 0.0139 B | 0.000 | 0.0855 | 0.0927 | -0.0072 W | 0.000 | 0.0649 | 0.0425 | 0.0224 B | 0.176 |
| 39 | 0.11173 | 0.338444 | -0.2267 W | 0.000 | 0.0874 | 0.0471 | 0.0404 B | 0.000 | 0.1151 | 0.2157 | -0.1006 W | 0.000 | 0.0880 | 0.0443 | 0.0436 B | 0.000 |
| 44 | 0.100455 | 0.14858 | -0.0481 W | 0.000 | 0.0904 | 0.0819 | 0.0085 B | 0.000 | 0.0960 | 0.1517 | -0.0558 W | 0.000 | 0.0893 | 0.0672 | 0.0221 B | 0.000 |
| 45 | 0.06296 | 0.064235 | -0.0013 W | 0.114 | 0.0543 | 0.0625 | -0.0082 W | 0.104 | 0.0622 | 0.0627 | -0.0005 W | 0.775 | 0.0523 | 0.0621 | -0.0099 W | 0.007 |
| 47 | 0.077699 | 0.159673 | -0.0820 W | 0.001 | 0.0777 | 0.0144 | 0.0633 B | 0.001 | 0.0767 | 0.1323 | -0.0556 W | 0.001 | 0.0768 | 0.0189 | 0.0578 B | 0.001 |
| 48 | 0.027437 | 0.033711 | -0.0063 W | 0.000 | 0.0275 | 0.0192 | 0.0082 B | 0.034 | 0.0275 | 0.0306 | -0.0031 W | 0.000 | 0.0281 | 0.0195 | 0.0086 B | 0.034 |
| 55 | 0.068057 | 0.101863 | -0.0338 W | 0.000 | 0.0695 | 0.0528 | 0.0168 B | 0.000 | 0.0739 | 0.0844 | -0.0105 W | 0.000 | 0.0475 | 0.0466 | 0.0009 B | 0.000 |
| 57 | 0.104231 | 0.109696 | -0.0055 W | 0.001 | 0.0780 | 0.0439 | 0.0342 B | 0.001 | 0.1041 | 0.1135 | -0.0094 W | 0.001 | 0.0722 | 0.0436 | 0.0286 B | 0.001 |
| 58 | 0.05989 | 0.075983 | -0.0161 W | 0.000 | 0.0555 | 0.0236 | 0.0319 B | 0.203 | 0.0576 | 0.0765 | -0.0189 W | 0.000 | 0.0548 | 0.0237 | 0.0311 B | 0.084 |
| 59 | 0.114291 | 0.255776 | -0.1415 W | 0.599 | 0.0876 | 0.0402 | 0.0474 B | 0.077 | 0.0688 | 0.2543 | -0.1855 W | 0.077 | 0.0719 | 0.0399 | 0.0320 B | 0.000 |
| 73 | 0.063342 | 0.082186 | -0.0188 W | 0.000 | 0.0533 | 0.0333 | 0.0200 B | 0.000 | 0.0638 | 0.0755 | -0.0116 W | 0.000 | 0.0545 | 0.0375 | 0.0169 B | 0.000 |
| 75 | 0.070886 | 0.152845 | -0.0820 W | 0.001 | 0.0625 | 0.0612 | 0.0013 B | 0.001 | 0.0715 | 0.1006 | -0.0291 W | 0.001 | 0.0695 | 0.0605 | 0.0090 B | 0.001 |

# Table 55: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 5 and 1)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (1) Benchmark - Salest-1 False Positive | (1) Benchmark - Salest-1 False Negative | (5) Twitter - CI False Positive | (5) Twitter - CI False Negative | Benchmark - CI Difference - FP | Benchmark - CI Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCI Total Cost | (1:2) Benchmark Total Cost /TCI Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 39.99% | 22.17% | 43.07% | 17.53% | -3.08% | 4.64% | Benchmark | TCI | 1.03 | 1.08 | TCI* | TCI* |
| 21 | 12 | 45.52% | 25.00% | 47.86% | 25.00% | -2.33% | 0.00% | Benchmark | - | 0.97 | 0.98 | Benchmark* | Benchmark* |
| 23 | 36 | 44.43% | 5.00% | 47.03% | 14.00% | -2.59% | -9.00% | Benchmark | Benchmark | 0.81 | 0.73 | Benchmark* | Benchmark* |
| 28 | 72 | 41.23% | 14.40% | 43.54% | 5.71% | -2.31% | 8.69% | Benchmark | TCI | 1.13 | 1.27 | TCI* | TCI* |
| 30 | 12 | 36.77% | 0.00% | 48.33% | 0.00% | -11.57% | 0.00% | Benchmark | - | 0.76 | 0.76 | Benchmark* | Benchmark* |
| 31 | 24 | 39.83% | 0.00% | 45.21% | 0.00% | -5.38% | 0.00% | Benchmark | - | 0.88 | 0.88 | Benchmark* | Benchmark* |
| 35 | 24 | 36.55% | 25.00% | 41.47% | 6.67% | -4.93% | 18.33% | Benchmark | TCI | 1.28 | 1.58 | TCI* | TCI* |
| 36 | 48 | 46.34% | 21.50% | 42.09% | 16.50% | 4.25% | 5.00% | TCI | TCI | 1.16 | 1.19 | TCI | TCI |
| 37 | 84 | 41.61% | 16.61% | 43.35% | 14.05% | -1.74% | 2.56% | Benchmark | TCI | 1.01 | 1.05 | TCI* | TCI* |
| 39 | 36 | 29.18% | 30.00% | 43.87% | 20.50% | -14.70% | 9.50% | Benchmark | TCI | 0.92 | 1.05 | Benchmark* | TCI* |
| 44 | 24 | 44.29% | 10.00% | 40.99% | 20.00% | 3.30% | -10.00% | TCI | Benchmark | 0.89 | 0.79 | Benchmark* | Benchmark* |
| 45 | 96 | 44.30% | 15.91% | 44.79% | 16.28% | -0.49% | -0.38% | Benchmark | Benchmark | 0.99 | 0.98 | Benchmark* | Benchmark* |
| 47 | 12 | 37.45% | 10.00% | 47.62% | 15.00% | -10.17% | -5.00% | Benchmark | Benchmark | 0.76 | 0.74 | Benchmark* | Benchmark* |
| 48 | 24 | 40.67% | 10.00% | 40.80% | 15.00% | -0.13% | -5.00% | Benchmark | Benchmark | 0.91 | 0.86 | Benchmark* | Benchmark* |
| 55 | 24 | 42.55% | 16.67% | 40.20% | 3.33% | 2.35% | 13.33% | TCI | TCI | 1.36 | 1.62 | TCI | TCI |
| 57 | 24 | 45.26% | 5.00% | 42.68% | 15.00% | 2.58% | -10.00% | TCI | Benchmark | 0.87 | 0.76 | Benchmark* | Benchmark* |
| 58 | 180 | 44.16% | 13.78% | 44.80% | 12.81% | -0.64% | 0.97% | Benchmark | TCI | 1.01 | 1.02 | TCI* | TCI* |
| 59 | 36 | 41.88% | 14.50% | 39.22% | 11.50% | 2.67% | 3.00% | TCI | TCI | 1.11 | 1.14 | TCI | TCI |
| 73 | 60 | 34.89% | 23.63% | 40.15% | 18.69% | -5.25% | 4.94% | Benchmark | TCI | 0.99 | 1.06 | Benchmark* | TCI* |
| 75 | 12 | 39.17% | 15.00% | 29.88% | 10.00% | 9.30% | 5.00% | TCI | TCI | 1.36 | 1.39 | TCI | TCI |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 56: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 7 and 2)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (2) Benchmark - Salest-1 & GDPt-1 False Positive | (2) Benchmark - Salest-1 & GDPt-1 False Negative | (7) Twitter - CI & GDPt-1 False Positive | (7) Twitter - CI & GDPt-1 False Negative | Benchmark - CI Difference - FP | Benchmark - CI Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCI Total Cost | (1:2) Benchmark Total Cost /TCI Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 40.02% | 24.82% | 42.66% | 20.67% | -2.63% | 4.15% | Benchmark | TCI | 1.02 | 1.07 | TCI* | TCI* |
| 21 | 12 | 45.00% | 0.00% | 45.00% | 0.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 23 | 36 | 45.45% | 11.50% | 45.27% | 6.50% | 0.18% | 5.00% | TCI | TCI | 1.10 | 1.17 | TCI | TCI |
| 28 | 72 | 39.11% | 17.98% | 41.54% | 23.21% | -2.43% | -5.24% | Benchmark | Benchmark | 0.88 | 0.85 | Benchmark* | Benchmark* |
| 30 | 12 | 45.51% | 10.00% | 33.63% | 30.00% | 11.87% | -20.00% | TCI | Benchmark | 0.87 | 0.70 | Benchmark* | Benchmark* |
| 31 | 24 | 46.79% | 0.00% | 43.52% | 8.33% | 3.27% | -8.33% | TCI | Benchmark | 0.90 | 0.78 | Benchmark* | Benchmark* |
| 35 | 24 | 34.69% | 21.67% | 41.17% | 13.33% | -6.47% | 8.33% | Benchmark | TCI | 1.03 | 1.15 | TCI* | TCI* |
| 36 | 48 | 43.68% | 15.00% | 43.09% | 21.50% | 0.59% | -6.50% | TCI | Benchmark | 0.91 | 0.86 | Benchmark* | Benchmark* |
| 37 | 84 | 38.94% | 27.26% | 42.59% | 12.86% | -3.64% | 14.40% | Benchmark | TCI | 1.19 | 1.37 | TCI* | TCI* |
| 39 | 36 | 24.99% | 39.50% | 43.49% | 9.00% | -18.49% | 30.50% | Benchmark | TCI | 1.23 | 1.69 | TCI* | TCI* |
| 44 | 24 | 41.32% | 16.67% | 46.85% | 13.33% | -5.53% | 3.33% | Benchmark | TCI | 0.96 | 1.02 | Benchmark* | TCI* |
| 45 | 96 | 43.62% | 19.14% | 43.88% | 17.58% | -0.25% | 1.56% | Benchmark | TCI | 1.02 | 1.04 | TCI* | TCI* |
| 47 | 12 | 37.45% | 30.00% | 45.52% | 10.00% | -8.07% | 20.00% | Benchmark | TCI | 1.21 | 1.49 | TCI* | TCI* |
| 48 | 24 | 42.70% | 13.33% | 39.86% | 16.67% | 2.84% | -3.33% | TCI | Benchmark | 0.99 | 0.95 | Benchmark* | Benchmark* |
| 55 | 24 | 39.71% | 16.67% | 45.40% | 8.33% | -5.68% | 8.33% | Benchmark | TCI | 1.05 | 1.18 | TCI* | TCI* |
| 57 | 24 | 39.21% | 5.00% | 42.68% | 10.00% | -3.47% | -5.00% | Benchmark | Benchmark | 0.84 | 0.79 | Benchmark* | Benchmark* |
| 58 | 180 | 44.84% | 17.79% | 45.37% | 16.78% | -0.53% | 1.01% | Benchmark | TCI | 1.01 | 1.02 | TCI* | TCI* |
| 59 | 36 | 39.77% | 18.00% | 43.39% | 7.50% | -3.62% | 10.50% | Benchmark | TCI | 1.14 | 1.30 | TCI* | TCI* |
| 73 | 60 | 32.25% | 37.66% | 38.13% | 26.49% | -5.88% | 11.17% | Benchmark | TCI | 1.08 | 1.18 | TCI* | TCI* |
| 75 | 12 | 36.37% | 15.00% | 25.62% | 40.00% | 10.75% | -25.00% | TCI | Benchmark | 0.78 | 0.63 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 57: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 9 and 3)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (3) Benchmark - Salest-1 & AR: False Positive | (3) Benchmark - Salest-1 & AR: False Negative | (9) Twitter - CI & AR: False Positive | (9) Twitter - CI & AR: False Negative | Benchmark - CI: Difference - FP | Benchmark - CI: Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCI Total Cost | (1:2) Benchmark Total Cost /TCI Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 39.28% | 22.39% | 43.61% | 14.41% | -4.33% | 7.98% | Benchmark | TCI | 1.06 | 1.16 | TCI* | TCI* |
| 21 | 12 | 47.86% | 25.00% | 47.10% | 25.00% | 0.76% | 0.00% | TCI | - | 1.01 | 1.01 | TCI* | TCI* |
| 23 | 36 | 38.99% | 20.50% | 38.65% | 23.00% | 0.34% | -2.50% | TCI | Benchmark | 0.96 | 0.94 | Benchmark* | Benchmark* |
| 28 | 72 | 37.56% | 9.29% | 40.32% | 11.79% | -2.76% | -2.50% | Benchmark | Benchmark | 0.90 | 0.88 | Benchmark* | Benchmark* |
| 30 | 12 | 38.75% | 0.00% | 44.26% | 10.00% | -5.51% | -10.00% | Benchmark | Benchmark | 0.71 | 0.60 | Benchmark* | Benchmark* |
| 31 | 24 | 43.52% | 0.00% | 40.29% | 6.67% | 3.23% | -6.67% | TCI | Benchmark | 0.93 | 0.81 | Benchmark* | Benchmark* |
| 35 | 24 | 27.57% | 33.33% | 38.87% | 20.00% | -11.30% | 13.33% | Benchmark | TCI | 1.03 | 1.19 | TCI* | TCI* |
| 36 | 48 | 41.48% | 29.00% | 41.79% | 10.00% | -0.31% | 19.00% | Benchmark | TCI | 1.36 | 1.61 | TCI* | TCI* |
| 37 | 84 | 40.84% | 19.29% | 42.76% | 15.71% | -1.92% | 3.57% | Benchmark | TCI | 1.03 | 1.07 | TCI* | TCI* |
| 39 | 36 | 36.17% | 32.50% | 39.66% | 21.50% | -3.50% | 11.00% | Benchmark | TCI | 1.12 | 1.22 | TCI* | TCI* |
| 44 | 24 | 42.85% | 10.00% | 44.57% | 15.00% | -1.72% | -5.00% | Benchmark | Benchmark | 0.89 | 0.84 | Benchmark* | Benchmark* |
| 45 | 96 | 43.87% | 15.62% | 44.34% | 18.02% | -0.46% | -2.40% | Benchmark | Benchmark | 0.95 | 0.93 | Benchmark* | Benchmark* |
| 47 | 12 | 39.17% | 10.00% | 47.62% | 15.00% | -8.45% | -5.00% | Benchmark | Benchmark | 0.79 | 0.76 | Benchmark* | Benchmark* |
| 48 | 24 | 40.67% | 10.00% | 39.50% | 18.33% | 1.17% | -8.33% | TCI | Benchmark | 0.88 | 0.80 | Benchmark* | Benchmark* |
| 55 | 24 | 42.85% | 11.67% | 45.67% | 13.33% | -2.82% | -1.67% | Benchmark | Benchmark | 0.92 | 0.91 | Benchmark* | Benchmark* |
| 57 | 24 | 44.71% | 5.00% | 42.68% | 15.00% | 2.03% | -10.00% | TCI | Benchmark | 0.86 | 0.75 | Benchmark* | Benchmark* |
| 58 | 180 | 43.96% | 13.86% | 45.16% | 16.38% | -1.20% | -2.52% | Benchmark | Benchmark | 0.94 | 0.92 | Benchmark* | Benchmark* |
| 59 | 36 | 35.11% | 17.00% | 42.06% | 9.00% | -6.95% | 8.00% | Benchmark | TCI | 1.02 | 1.15 | TCI* | TCI* |
| 73 | 60 | 38.05% | 29.27% | 38.75% | 21.55% | -0.70% | 7.72% | Benchmark | TCI | 1.12 | 1.18 | TCI* | TCI* |
| 75 | 12 | 37.77% | 15.00% | 33.68% | 35.00% | 4.10% | -20.00% | TCI | Benchmark | 0.77 | 0.65 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 58: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCI and without TCI (Models 11 and 4)

Error Detection Ability - Alpha = 0.33

| 2-Digit SIC | Number of Observations | (4) Benchmark - Salest-1 & AR & GDPt-1: False Positive | (4) Benchmark - Salest-1 & AR & GDPt-1: False Negative | (11) Twitter - CI & AR & GDPt-1: False Positive | (11) Twitter - CI & AR & GDPt-1: False Negative | Benchmark - CI: Difference - FP | Benchmark - CI: Difference - FN | Better Model - FP | Better Model - FN | (1:1) Benchmark Total Cost /TCI Total Cost | (1:2) Benchmark Total Cost /TCI Total Cost | (1:1) Better Model - Cost Ratio | (1:2) Better Model - Cost Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 144 | 39.86% | 26.50% | 41.38% | 20.31% | -1.52% | 6.19% | Benchmark | TCI | 1.08 | 1.13 | TCI* | TCI* |
| 21 | 12 | 42.39% | 5.00% | 36.37% | 15.00% | 6.02% | -10.00% | TCI | Benchmark | 0.92 | 0.79 | Benchmark* | Benchmark* |
| 23 | 36 | 41.69% | 15.00% | 42.89% | 17.00% | -1.20% | -2.00% | Benchmark | Benchmark | 0.95 | 0.93 | Benchmark* | Benchmark* |
| 28 | 72 | 37.50% | 20.48% | 41.07% | 24.17% | -3.57% | -3.69% | Benchmark | Benchmark | 0.89 | 0.88 | Benchmark* | Benchmark* |
| 30 | 12 | 50.00% | 0.00% | 33.63% | 30.00% | 16.37% | -30.00% | TCI | Benchmark | 0.79 | 0.53 | Benchmark* | Benchmark* |
| 31 | 24 | 41.74% | 11.67% | 41.74% | 11.67% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 35 | 24 | 26.85% | 30.00% | 37.13% | 31.67% | -10.28% | -1.67% | Benchmark | Benchmark | 0.83 | 0.86 | Benchmark* | Benchmark* |
| 36 | 48 | 43.68% | 15.00% | 38.88% | 23.00% | 4.80% | -8.00% | TCI | Benchmark | 0.95 | 0.87 | Benchmark* | Benchmark* |
| 37 | 84 | 39.01% | 27.66% | 45.33% | 9.52% | -6.32% | 18.13% | Benchmark | TCI | 1.22 | 1.47 | TCI* | TCI* |
| 39 | 36 | 26.65% | 39.00% | 36.90% | 29.50% | -10.25% | 9.50% | Benchmark | TCI | 0.99 | 1.09 | Benchmark* | TCI* |
| 44 | 24 | 39.71% | 16.67% | 46.85% | 13.33% | -7.14% | 3.33% | Benchmark | TCI | 0.94 | 0.99 | Benchmark* | Benchmark* |
| 45 | 96 | 43.19% | 18.34% | 42.96% | 21.33% | 0.22% | -2.99% | TCI | Benchmark | 0.96 | 0.93 | Benchmark* | Benchmark* |
| 47 | 12 | 37.81% | 35.00% | 47.62% | 0.00% | -9.81% | 35.00% | Benchmark | TCI | 1.53 | 2.26 | TCI* | TCI* |
| 48 | 24 | 42.70% | 13.33% | 39.54% | 11.67% | 3.16% | 1.67% | TCI | TCI | 1.09 | 1.10 | TCI | TCI |
| 55 | 24 | 31.44% | 33.33% | 41.31% | 15.00% | -9.87% | 18.33% | Benchmark | TCI | 1.15 | 1.38 | TCI* | TCI* |
| 57 | 24 | 39.21% | 5.00% | 42.68% | 10.00% | -3.47% | -5.00% | Benchmark | Benchmark | 0.84 | 0.79 | Benchmark* | Benchmark* |
| 58 | 180 | 45.47% | 18.84% | 44.08% | 19.31% | 1.39% | -0.47% | TCI | Benchmark | 1.01 | 1.01 | TCI* | TCI* |
| 59 | 36 | 37.84% | 23.50% | 41.99% | 19.50% | -4.15% | 4.00% | Benchmark | TCI | 1.00 | 1.05 | Benchmark* | TCI* |
| 73 | 60 | 35.74% | 32.94% | 39.36% | 25.00% | -3.62% | 7.94% | Benchmark | TCI | 1.07 | 1.14 | TCI* | TCI* |
| 75 | 12 | 40.18% | 20.00% | 29.88% | 35.00% | 10.30% | -15.00% | TCI | Benchmark | 0.93 | 0.80 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 59: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 6 and 1)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | | (6) | | | | | | | | | |
| 2-Digit SIC | Number of Observations | Benchmark - Salest-1 | | Twitter - CS | | Benchmark - CS | | Better Model - FP | Better Model - FN | (1:1) | (1:2) | (1:1) | (1:2) |
| | | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | | | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 39.99% | 22.17% | 43.27% | 15.77% | -3.28% | 6.40% | Benchmark | TCS | 1.05 | 1.13 | TCS* | TCS* |
| 21 | 12 | 45.52% | 25.00% | 43.84% | 5.00% | 1.68% | 20.00% | TCS | TCS | 1.44 | 1.77 | TCS | TCS |
| 23 | 36 | 44.43% | 5.00% | 45.99% | 7.50% | -1.55% | -2.50% | Benchmark | Benchmark | 0.92 | 0.89 | Benchmark* | Benchmark* |
| 28 | 72 | 41.23% | 14.40% | 42.69% | 5.71% | -1.47% | 8.69% | Benchmark | TCS | 1.15 | 1.29 | TCS* | TCS* |
| 30 | 12 | 36.77% | 0.00% | 46.06% | 0.00% | -9.29% | 0.00% | Benchmark | - | 0.80 | 0.80 | Benchmark* | Benchmark* |
| 31 | 24 | 39.83% | 0.00% | 37.11% | 0.00% | 2.72% | 0.00% | TCS | - | 1.07 | 1.07 | TCS* | TCS* |
| 35 | 24 | 36.55% | 25.00% | 41.47% | 3.33% | -4.93% | 21.67% | Benchmark | TCS | 1.37 | 1.80 | TCS* | TCS* |
| 36 | 48 | 46.34% | 21.50% | 45.54% | 17.50% | 0.80% | 4.00% | TCS | TCS | 1.08 | 1.11 | TCS | TCS |
| 37 | 84 | 41.61% | 16.61% | 42.76% | 16.13% | -1.16% | 0.48% | Benchmark | TCS | 0.99 | 1.00 | Benchmark* | Benchmark* |
| 39 | 36 | 29.18% | 30.00% | 44.81% | 17.50% | -15.63% | 12.50% | Benchmark | TCS | 0.95 | 1.12 | Benchmark* | TCS* |
| 44 | 24 | 44.29% | 10.00% | 40.03% | 15.00% | 4.27% | -5.00% | TCS | Benchmark | 0.99 | 0.92 | Benchmark* | Benchmark* |
| 45 | 96 | 44.30% | 15.91% | 44.68% | 18.40% | -0.38% | -2.50% | Benchmark | Benchmark | 0.95 | 0.93 | Benchmark* | Benchmark* |
| 47 | 12 | 37.45% | 10.00% | 46.05% | 5.00% | -8.60% | 5.00% | Benchmark | TCS | 0.93 | 1.03 | Benchmark* | TCS* |
| 48 | 24 | 40.67% | 10.00% | 37.77% | 21.67% | 2.90% | -11.67% | TCS | Benchmark | 0.85 | 0.75 | Benchmark* | Benchmark* |
| 55 | 24 | 42.55% | 16.67% | 42.84% | 10.00% | -0.29% | 6.67% | Benchmark | TCS | 1.12 | 1.21 | TCS* | TCS* |
| 57 | 24 | 45.26% | 5.00% | 39.53% | 10.00% | 5.73% | -5.00% | TCS | Benchmark | 1.01 | 0.93 | TCS* | Benchmark* |
| 58 | 180 | 44.16% | 13.78% | 43.83% | 12.54% | 0.33% | 1.24% | TCS | TCS | 1.03 | 1.04 | TCS | TCS |
| 59 | 36 | 41.88% | 14.50% | 41.36% | 11.50% | 0.52% | 3.00% | TCS | TCS | 1.07 | 1.10 | TCS | TCS |
| 73 | 60 | 34.89% | 23.63% | 40.78% | 23.43% | -5.89% | 0.20% | Benchmark | TCS | 0.91 | 0.94 | Benchmark* | Benchmark* |
| 75 | 12 | 39.17% | 15.00% | 41.14% | 25.00% | -1.97% | -10.00% | Benchmark | Benchmark | 0.82 | 0.76 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 60: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 8 and 2)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (2) | | (8) | | | | | | | | | |
| 2-Digit SIC | Number of Observations | Benchmark - Salest-1 & GDPt-1 | | Twitter - CS & GDPt-1 | | Benchmark - CS | | Better Model - FP | Better Model - FN | (1:1) | (1:2) | (1:1) | (1:2) |
| | | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | | | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 40.02% | 24.82% | 43.00% | 17.55% | -2.98% | 7.26% | Benchmark | TCS | 1.07 | 1.15 | TCS* | TCS* |
| 21 | 12 | 45.00% | 0.00% | 42.11% | 0.00% | 2.89% | 0.00% | TCS | - | 1.07 | 1.07 | TCS* | TCS* |
| 23 | 36 | 45.45% | 11.50% | 42.70% | 15.00% | 2.74% | -3.50% | TCS | Benchmark | 0.99 | 0.94 | Benchmark* | Benchmark* |
| 28 | 72 | 39.11% | 17.98% | 40.43% | 22.38% | -1.32% | -4.40% | Benchmark | Benchmark | 0.91 | 0.88 | Benchmark* | Benchmark* |
| 30 | 12 | 45.51% | 10.00% | 46.06% | 20.00% | -0.56% | -10.00% | Benchmark | Benchmark | 0.84 | 0.76 | Benchmark* | Benchmark* |
| 31 | 24 | 46.79% | 0.00% | 43.92% | 15.00% | 2.87% | -15.00% | TCS | Benchmark | 0.79 | 0.63 | Benchmark* | Benchmark* |
| 35 | 24 | 34.69% | 21.67% | 41.48% | 20.00% | -6.78% | 1.67% | Benchmark | TCS | 0.92 | 0.96 | Benchmark* | Benchmark* |
| 36 | 48 | 43.68% | 15.00% | 46.42% | 13.00% | -2.74% | 2.00% | Benchmark | TCS | 0.99 | 1.02 | Benchmark* | TCS* |
| 37 | 84 | 38.94% | 27.26% | 43.41% | 18.27% | -4.47% | 8.99% | Benchmark | TCS | 1.07 | 1.17 | TCS* | TCS* |
| 39 | 36 | 24.99% | 39.50% | 47.78% | 5.00% | -22.79% | 34.50% | Benchmark | TCS | 1.22 | 1.80 | TCS* | TCS* |
| 44 | 24 | 41.32% | 16.67% | 45.67% | 16.67% | -4.35% | 0.00% | Benchmark | - | 0.93 | 0.94 | Benchmark* | Benchmark* |
| 45 | 96 | 43.62% | 19.14% | 43.95% | 19.23% | -0.33% | -0.09% | Benchmark | Benchmark | 0.99 | 0.99 | Benchmark* | Benchmark* |
| 47 | 12 | 37.45% | 30.00% | 45.52% | 10.00% | -8.07% | 20.00% | Benchmark | TCS | 1.21 | 1.49 | TCS* | TCS* |
| 48 | 24 | 42.70% | 13.33% | 38.01% | 15.00% | 4.69% | -1.67% | TCS | Benchmark | 1.06 | 1.02 | TCS* | TCS* |
| 55 | 24 | 39.71% | 16.67% | 47.62% | 0.00% | -7.90% | 16.67% | Benchmark | TCS | 1.18 | 1.53 | TCS* | TCS* |
| 57 | 24 | 39.21% | 5.00% | 45.52% | 10.00% | -6.31% | -5.00% | Benchmark | Benchmark | 0.80 | 0.75 | Benchmark* | Benchmark* |
| 58 | 180 | 44.84% | 17.79% | 44.50% | 14.18% | 0.34% | 3.61% | TCS | TCS | 1.07 | 1.10 | TCS | TCS |
| 59 | 36 | 39.77% | 18.00% | 44.44% | 10.00% | -4.67% | 8.00% | Benchmark | TCS | 1.06 | 1.18 | TCS* | TCS* |
| 73 | 60 | 32.25% | 37.66% | 38.13% | 26.01% | -5.87% | 11.65% | Benchmark | TCS | 1.09 | 1.19 | TCS* | TCS* |
| 75 | 12 | 36.37% | 15.00% | 50.00% | 0.00% | -13.63% | 15.00% | Benchmark | TCS | 1.03 | 1.33 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 61: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 10 and 3)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | | (10) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR | | Twitter - CS & AR | | Benchmark - CS | | Better | Better | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Model - FP | Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 39.28% | 22.39% | 44.35% | 14.91% | -5.07% | 7.48% | Benchmark | TCS | 1.04 | 1.13 | TCS* | TCS* |
| 21 | 12 | 47.86% | 25.00% | 43.84% | 5.00% | 4.02% | 20.00% | TCS | TCS | 1.49 | 1.82 | TCS | TCS |
| 23 | 36 | 38.99% | 20.50% | 44.15% | 9.00% | -5.16% | 11.50% | Benchmark | TCS | 1.12 | 1.29 | TCS* | TCS* |
| 28 | 72 | 37.56% | 9.29% | 38.20% | 11.43% | -0.64% | -2.14% | Benchmark | Benchmark | 0.94 | 0.92 | Benchmark* | Benchmark* |
| 30 | 12 | 38.75% | 0.00% | 44.95% | 10.00% | -6.20% | -10.00% | Benchmark | Benchmark | 0.71 | 0.60 | Benchmark* | Benchmark* |
| 31 | 24 | 43.52% | 0.00% | 41.74% | 3.33% | 1.78% | -3.33% | TCS | Benchmark | 0.97 | 0.90 | Benchmark* | Benchmark* |
| 35 | 24 | 27.57% | 33.33% | 35.47% | 15.00% | -7.90% | 18.33% | Benchmark | TCS | 1.21 | 1.44 | TCS* | TCS* |
| 36 | 48 | 41.48% | 29.00% | 41.80% | 15.00% | -0.31% | 14.00% | Benchmark | TCS | 1.24 | 1.39 | TCS* | TCS* |
| 37 | 84 | 40.84% | 19.29% | 43.12% | 18.69% | -2.27% | 0.60% | Benchmark | TCS | 0.97 | 0.99 | Benchmark* | Benchmark* |
| 39 | 36 | 36.17% | 32.50% | 41.89% | 25.50% | -5.73% | 7.00% | Benchmark | TCS | 1.02 | 1.09 | TCS* | TCS* |
| 44 | 24 | 42.85% | 10.00% | 42.69% | 15.00% | 0.16% | -5.00% | TCS | Benchmark | 0.92 | 0.86 | Benchmark* | Benchmark* |
| 45 | 96 | 43.87% | 15.62% | 43.40% | 17.45% | 0.47% | -1.84% | TCS | Benchmark | 0.98 | 0.96 | Benchmark* | Benchmark* |
| 47 | 12 | 39.17% | 10.00% | 45.52% | 10.00% | -6.35% | 0.00% | Benchmark | - | 0.89 | 0.90 | Benchmark* | Benchmark* |
| 48 | 24 | 40.67% | 10.00% | 37.77% | 21.67% | 2.90% | -11.67% | TCS | Benchmark | 0.85 | 0.75 | Benchmark* | Benchmark* |
| 55 | 24 | 42.85% | 11.67% | 44.44% | 15.00% | -1.59% | -3.33% | Benchmark | Benchmark | 0.92 | 0.89 | Benchmark* | Benchmark* |
| 57 | 24 | 44.71% | 5.00% | 42.68% | 10.00% | 2.03% | -5.00% | TCS | Benchmark | 0.94 | 0.87 | Benchmark* | Benchmark* |
| 58 | 180 | 43.96% | 13.86% | 45.04% | 14.77% | -1.08% | -0.91% | Benchmark | Benchmark | 0.97 | 0.96 | Benchmark* | Benchmark* |
| 59 | 36 | 35.11% | 17.00% | 40.69% | 9.00% | -5.58% | 8.00% | Benchmark | TCS | 1.05 | 1.18 | TCS* | TCS* |
| 73 | 60 | 38.05% | 29.27% | 38.68% | 24.03% | -0.63% | 5.24% | Benchmark | TCS | 1.07 | 1.11 | TCS* | TCS* |
| 75 | 12 | 37.77% | 15.00% | 43.26% | 40.00% | -5.49% | -25.00% | Benchmark | Benchmark | 0.63 | 0.55 | Benchmark* | Benchmark* |

*Better model determined based on the ratio of costs of FP and FN errors.

# Table 62: 2-Digit SIC and Advertising Expense - Error Detection Performance for Continuous Substantive Analytical Models with TCS and without TCS (Models 12 and 4)

| Error Detection Ability - Alpha = 0.33 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (4) | | (12) | | | | | | | | | |
| | | Benchmark - Salest-1 & AR & GDPt-1 | | Twitter - CS & AR & GDPt-1 | | Benchmark - CS | | Better | Better | (1:1) | (1:2) | (1:1) | (1:2) |
| 2-Digit SIC | Number of Observations | False Positive | False Negative | False Positive | False Negative | Difference - FP | Difference - FN | Model - FP | Model - FN | Benchmark Total Cost /TCS Total Cost | Benchmark Total Cost /TCS Total Cost | Better Model - Cost Ratio | Better Model - Cost Ratio |
| 20 | 144 | 39.86% | 26.50% | 43.97% | 17.07% | -4.11% | 9.43% | Benchmark | TCS | 1.09 | 1.19 | TCS* | TCS* |
| 21 | 12 | 42.39% | 5.00% | 42.39% | 5.00% | 0.00% | 0.00% | - | - | 1.00 | 1.00 | Benchmark* | Benchmark* |
| 23 | 36 | 41.69% | 15.00% | 40.96% | 21.50% | 0.73% | -6.50% | TCS | Benchmark | 0.91 | 0.85 | Benchmark* | Benchmark* |
| 28 | 72 | 37.50% | 20.48% | 39.28% | 21.55% | -1.78% | -1.07% | Benchmark | Benchmark | 0.95 | 0.95 | Benchmark* | Benchmark* |
| 30 | 12 | 50.00% | 0.00% | 46.06% | 20.00% | 3.94% | -20.00% | TCS | Benchmark | 0.76 | 0.58 | Benchmark* | Benchmark* |
| 31 | 24 | 41.74% | 11.67% | 46.79% | 0.00% | -5.05% | 11.67% | Benchmark | TCS | 1.14 | 1.39 | TCS* | TCS* |
| 35 | 24 | 26.85% | 30.00% | 33.10% | 26.67% | -6.25% | 3.33% | Benchmark | TCS | 0.95 | 1.00 | Benchmark* | TCS* |
| 36 | 48 | 43.68% | 15.00% | 43.87% | 19.00% | -0.19% | -4.00% | Benchmark | Benchmark | 0.93 | 0.90 | Benchmark* | Benchmark* |
| 37 | 84 | 39.01% | 27.66% | 43.93% | 16.90% | -4.92% | 10.75% | Benchmark | TCS | 1.10 | 1.21 | TCS* | TCS* |
| 39 | 36 | 26.65% | 39.00% | 31.95% | 32.50% | -5.30% | 6.50% | Benchmark | TCS | 1.02 | 1.08 | TCS* | TCS* |
| 44 | 24 | 39.71% | 16.67% | 45.67% | 16.67% | -5.96% | 0.00% | Benchmark | - | 0.90 | 0.92 | Benchmark* | Benchmark* |
| 45 | 96 | 43.19% | 18.34% | 42.06% | 25.80% | 1.12% | -7.46% | TCS | Benchmark | 0.91 | 0.85 | Benchmark* | Benchmark* |
| 47 | 12 | 37.81% | 35.00% | 45.52% | 10.00% | -7.71% | 25.00% | Benchmark | TCS | 1.31 | 1.65 | TCS* | TCS* |
| 48 | 24 | 42.70% | 13.33% | 38.01% | 15.00% | 4.69% | -1.67% | TCS | Benchmark | 1.06 | 1.02 | TCS* | TCS* |
| 55 | 24 | 31.44% | 33.33% | 41.47% | 18.33% | -10.03% | 15.00% | Benchmark | TCS | 1.08 | 1.26 | TCS* | TCS* |
| 57 | 24 | 39.21% | 5.00% | 45.52% | 10.00% | -6.31% | -5.00% | Benchmark | Benchmark | 0.80 | 0.75 | Benchmark* | Benchmark* |
| 58 | 180 | 45.47% | 18.84% | 43.28% | 19.58% | 2.19% | -0.74% | TCS | Benchmark | 1.02 | 1.01 | TCS* | TCS* |
| 59 | 36 | 37.84% | 23.50% | 39.99% | 22.00% | -2.15% | 1.50% | Benchmark | TCS | 0.99 | 1.01 | Benchmark* | TCS* |
| 73 | 60 | 35.74% | 32.94% | 37.99% | 25.00% | -2.25% | 7.94% | Benchmark | TCS | 1.09 | 1.15 | TCS* | TCS* |
| 75 | 12 | 40.18% | 20.00% | 45.52% | 10.00% | -5.35% | 10.00% | Benchmark | TCS | 1.08 | 1.22 | TCS* | TCS* |

*Better model determined based on the ratio of costs of FP and FN errors.