Exogenous Data in Accounting and Auditing in the Rutgers Series in Accounting Information Systems

Edited by

Miklos A. Vasarhelyi, Hanxin (Alice) Hu
and Ann F. Medinets

Exogenous Data in Accounting and Auditing in the Rutgers Series in Accounting Information Systems

This page intentionally left blank

Exogenous Data in Accounting and Auditing in the Rutgers Series in Accounting Information Systems

EDITED BY

MIKLOS A. VASARHELYI
Rutgers Business School, USA

HANXIN (ALICE) HU
Kean University, USA

AND

ANN F. MEDINETS
Rutgers Business School, USA

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL.

First edition 2025

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83608-735-9 (Print)
ISBN: 978-1-83608-734-2 (Online)
ISBN: 978-1-83608-736-6 (Epub)

Contents

About the Editors

Miklos A. Vasarhelyi is the KPMG Distinguished Professor of Accounting Information Systems at Rutgers University. He is a leading Researcher in continuous audit application with over 200 journal articles and 20 books published in the field. He is also Editor of the *Journal of Emerging Technologies in Accounting* and the *Artificial Intelligence in Accounting and Auditing* series. As Head of the Rutgers Accounting Research Center (RARC), the Continuous Auditing & Reporting (CAR) Lab, and the Rutgers AICPA Data Analytics Research (RADAR) initiative, he has worked on projects for Siemens, KPMG, Procter & Gamble, AICPA, CA Technologies, and Brazil's Itau-Unibanco.

Hanxin (Alice) Hu is an Assistant Professor in the Department of Accounting and Finance at Kean University. She holds a PhD in Accounting Information Systems from Rutgers Business School. Her research interests focus on financial reporting, audit risk assessment, and the application of artificial intelligence in accounting and auditing. She has published in peer-review journals, including *Accounting Horizons*, *Journal of Information Systems*, and *Journal of Emerging Technologies in Accounting*. She enjoys teaching financial and managerial accounting, audit analytics, and financial statement analysis.

Ann F. Medinets is an Associate Professor of Professional Practice at Rutgers University. Her paper "Say-on-Pay: Is Anybody Listening?" with Stephani Mason and Dan Palmon (*Multinational Finance Journal*, 2016) was featured on Harvard's Forum on Corporate Governance and Financial Regulation. Another paper "Disclosure Regulations Work: The Case of Regulation G" with Dan Palmon and Yu-An Chen (*Review of Quantitative Finance and Accounting*, 2022) received a Bright Ideas Award in Accounting. Other publications include "Giving Shareholders Real Rights, Reducing Excessive Executive Pay, and Reducing One-Size-Fits-All Regulation" with Dan Palmon and Gary Kleinman (Rutger*s Business Review*, 2022), "Developing Continuous Assurance at Siemens" with Jason Gross and Gerard Brennan (Audit Analytics and Continuous Audit: Looking toward the Future, 2015), and "A Theoretical Model of Cognitive Factors that Affect Auditors' Performance and Perceived Independence" with Dan Palmon, Gary Kleinman, and Asokan Anandarajan (*International Journal of Behavioural Accounting and Finance*, 2010).

This page intentionally left blank

About the Contributors

Talha Afzal is a PhD Student in the Department of Accounting and Information Systems at Rutgers University. His research interests include AI with a focus on neural networks and generative AI, as well as text mining. His co-authored article, "Fair Value Estimates for Illiquid Cryptocurrency," was published in the *International Journal of Accounting Information Systems*.

Irina Jie Bao is a PhD Candidate in the Department of Accounting and Information Systems at Rutgers University. She earned her bachelor's degree from Fudan University in China. Her research interests encompass a broad spectrum of areas, including environmental, social, and governance (ESG) criteria, artificial intelligence (AI), large language models (LLMs), auditing, the economics of regulation, accounting standards, and the gamification of accounting education. She is particularly focused on exploring the potential applications of LLMs in the ESG context, such as extracting and analyzing data from ESG reports and enhancing ESG assurance.

Melissa A. Dardani is a licensed Certified Public Accountant (CPA), Certified Fraud Examiner (CFE), and is pursuing a PhD in Accounting Information Systems from Rutgers University Graduate Business School in Newark, New Jersey. At Rutgers, she focuses on integrating artificial intelligence (AI) into business reporting, measurement, and assurance. Her research primarily investigates the applications of AI for continuous auditing and assurance, emphasizing a holistic, risk-focused approach to better preserve and protect business value. Bringing a decade of practical experience to Rutgers, her expertise includes forensic accounting, specializing in assisting businesses in distress due to fraud, mismanagement, and waste. She has conducted forensic audits and investigations and has designed and implemented systems of control for the finance function.

Leonardo Flach is a Full Professor in Accounting at the Federal University of Santa Catarina (UFSC), Brazil. His research interests include corporate sustainability and cooperative management.

Haoyun (Harry) Gao is a third-year PhD Student in Accounting & Information Systems at Rutgers University, with research interests at the intersection of auditing and emerging technologies. His work includes a published article titled

"Extracting Financial Data from Unstructured Sources: Leveraging Large Language Models" in the *Journal of Information Systems*.

Hanchi Gu is a Lecturer at the Shanghai University of Finance and Economics. He earned his PhD from Rutgers University in the Accounting and Information Systems program. His research focuses on artificial intelligence applications in accounting and auditing (e.g., large language models, outlier detection, and supervised-learning models), continuous monitoring, and other topics on accounting information systems.

Benita Gullkvist is currently employed as Senior Researcher at the University of Vaasa, Finland, after being an Associate Professor of Accounting at Hanken School of Economics, Finland, for over 10 years. She received her PhD degree from Abo Akademi University in 2005, where she was appointed as Adjunct Professor in April 2012. Her research interests include behavioral, social, and organizational issues related to accounting and accounting information systems. She is a Member of the editorial board of the *International Journal of Accounting Information Systems*, and she also serves as a Reviewer for journal and conference submissions. She has published in renowned journals such as the *European Accounting Review*, *Management Accounting Research*, and *Critical Perspectives on Accounting*. In addition, she has made contributions to several books and actively presented her research at international accounting conferences.

Qing Huang is an Assistant Professor at Marshall University. She earned her PhD in Accounting and Information Systems from Rutgers University and is a certified public Accountant (inactive) in China. Her research focus is on continuous auditing, audit analytics, outlier detection, graph neural networks, machine learning, and fraud detection.

Lanxin Jiang is a PhD Candidate at Rutgers University, specializing in archival auditing and accounting information systems. Her research centers on harnessing the power of machine learning to advance accounting and auditing practices, as well as leveraging emerging technology to enhance ESG reporting and assurance. Her ongoing projects include the application of machine learning to accounting research, audit data outlier detection, text mining of ESG reports, and blockchain- and lifecycle assessment enabled greenhouse gas (GHG) reporting. One of her dissertation chapters utilizes explainable AI to analyze audit partner characteristics extracted from LinkedIn.

Xin Jin is a second-year PhD Candidate at the Escola Brasileira de Administração Pública e de Empresas (Brazilian School of Public and Business Administration, EBAPE), which is part of the Fundação Getulio Vargas (Getulio Vargas Foundation, FGV) in Brazil. Her research areas include accounting, auditing, ESG reporting, and exogenous data. Specifically, her interests focus on corporate social performance, the correlation between firms' corporate governance and their environmental performance, and the use of exogenous data to predict firms' performance.

Steven Katz is an Accounting Information Systems PhD Candidate at Rutgers Business School. He has an abiding interest in natural language processing, corporate responsibility, and the future of ethical technological innovation. He has presented papers exploring the role of technology in governmental accounting and the automated extraction of information from corporate sustainability reports.

Huaxia Li is an Assistant Professor of Accounting in the Department of Accounting and Finance at the University of Michigan-Dearborn. He earned his Ph.D. in Management (Accounting Information Systems) from Rutgers, The State University of New Jersey. Prior to his current role, he served as a Researcher at the Rutgers Accounting Research Center and the Continuous Auditing and Reporting Lab. Huaxia's research interests include accounting information systems, accounting data accessibility, the application of emerging technologies in accounting and auditing, governmental financial accounting and disclosure, the municipal bond market and fiscal distress, and corporate social responsibility. His work has been published in the *Journal of Information Systems, Sustainability Accounting, Management, and Policy Journal*, and the *Journal of Emerging Technologies in Accounting*. He received his M.S. in Information Technology and Analytics from Rutgers, The State University of New Jersey, and his B.B.A in Accounting and Finance from the University of Macau.

Shaoyu (Shawn) Liu is a PhD Candidate in the Department of Accounting and Information Systems at Rutgers Business School. His research focuses on leveraging emerging technologies, such as large language models (LLMs) and robotic process automation (RPA), to explore strategic corporate disclosures and enhance financial reporting transparency. His work delves into the use of ESG disclosures as a strategic tool prior to earnings announcements, uncovering the impact of misclassifications in Form 8-K filings on market outcomes. His research has been presented at major conferences, including the AAA Annual Meeting. He is also interested in examining misinformation on social media, the detection of malicious content, and its influence on consumer behavior and stock returns. His work aims to enhance transparency in digital business environments and provide innovative solutions for distinguishing between genuine and manipulated content.

Fabricia Silva da Rosa has a PhD in Production Engineering. She is a Professor at the Department of Accounting Sciences and the Accounting Program at the Federal University of Santa Catarina (UFSC), Brazil, and she leads the Center for Studies in Innovation, Accounting and the Environment (NEIMAC).

Maristela Roza is a PhD Candidate in the Graduate Program in Accounting (PPGC) at the Federal University of Santa Catarina (UFSC) in Brazil where she is a Member of the Research Group on Work and Teaching in Accounting (NETEC). Her research interests include environmental sustainability, artificial intelligence, and electric vehicles (EVs), as well as public policy and regulations.

Fangbing Xiong is a PhD candidate at Rutgers University who specializes in information system design and LLM improvement based on her expertise in emerging technologies, such as text mining, data visualization, process mining, process automation, and retrieval-augmented generation. Her research includes the development of a digital role-playing game (D-RPG) that combines behavioral experiments and gamification to enhance accounting education. The first paper from this project has been submitted to *Journal of Emerging Technologies in Accounting* and received the Jack and Maye Stewart Student Project Award. Additionally, she is developing an LLM-powered AI agent framework for accounting and auditing using a design science approach. She explores advanced emerging technologies and their applications in business, aiming to benefit accounting education and practice through innovation.

Maria Guangyue Zhang is a PhD Candidate at Rutgers Business School. Her main research focus is on accounting information systems. Her work includes continuous monitoring for risk management and consumer privacy in the healthcare setting. She also studies financial accounting and disclosure using exogenous data sources, such as financial performance evaluation in multi-segment firms and cryptocurrency fair value measurement.

Introduction

Miklos A. Vasarhelyi[a], *Hanxin (Alice) Hu*[b], *Ann F. Medinets*[a] *and Steven Katz*[a]

[a]*Rutgers Business School, USA*
[b]*Kean University, USA*

Traditionally, accounting information has relied on data sourced directly from firms of interest. A host of elaborate and sensitive measures have been devised to parse endogenous data with the understanding that they were, to some extent, flawed. Reported earnings may be managed. Managerial disclosures and analyses may carefully omit relevant information. Earnings calls may be unnecessarily optimistic. All of these types of information share the same endogenous source and, therefore, are burdened with the potential for bias. However, the digital age has seen an enormous increase in the volume and variety of relevant data available from sources external to a given reporting firm. This exogenous data offers independent, timely insights into endogenous business outcomes, allowing for unbiased confirmation or contradiction of corporate claims. Exogenous data is also often voluminous, frequently updated, and varied in format, allowing for sensitive analyses conducted on demand using data that would be cumbersome to collect manually (e.g., Yelp reviews, compared to conducting a randomized survey).

Any source of information that is both external to an organization of interest and correlated with that company's reported outcomes can provide useful exogenous data. For example, social media posts, Google Trends, and weather data can all proxy for business-relevant factors (e.g., customer's impressions of products, broad social trends, and localized weather events that impact operations). Accountants can use these sources to supplement and substantiate their analyses with data that does not share the same inherent conflicts of interest carried by endogenous data. For example, if a firm reports making changes to its manufacturing process that minimize its factories' impact on the local ecology, satellite images of the environment surrounding those factories can be used to assure their claims (Gu et al., 2023). Similarly, a retailer's claims about sales revenue may be tested by analyses that use weather data to estimate likely customer behavior.

Exogenous Data in Accounting and Auditing in the Rutgers Series in Accounting Information Systems, 1–4

doi:10.1108/978-1-83608-734-220251014

If reported revenue does not match customers' expected behavioral patterns in the face of inclement weather, it may be a sign that revenue is being misstated (Yoon et al., 2024).

This book explores several innovative sources and uses of exogenous data, organized into three parts. The first part provides an overview of applications for exogenous data in business research. The second part focuses on extracting insights from analyzing search volume index (SVI) data, which represents how often specific phrases are searched for on search engines like Google. The third part explores the use of social media and public internet activity to infer users' sentiments, behaviors, and intentions.

Part One: Exogenous Data in Business Research

In Chapter 1, Melissa A. Dardani argues for the utility of exogenous data in modern reporting environments. She suggests that accounting analyses can be improved by combining vast exogenous data sources with data-hungry machine-learning algorithms that improve as they incorporate more input and that these analyses could theoretically be conducted on a continuous basis to create a nearly real-time assurance tool.

Chapter 2 by Irina Jie Bao and Benita Gullkvist discusses how Large Language Models like ChatGPT can be used to extract and analyze various types and sources of exogenous data. They demonstrate this process by using ChatGPT to obtain and use both endogenous and exogenous environmental data for ExxonMobil, laying out a step-by-step explanation of their procedure.

In Chapter 3, Hanchi Gu and Qing Huang examine published policies and actual practices related to corporate websites' distribution of traceable packets of internet data colloquially known as "cookies." Their comparison between official and observed practices suggests that there may be privacy concerns on corporate websites, adding a hidden cost to the use of this endogenous, non-financial information.

Part Two: Exogenous Data from Google Trends

In Chapter 4, Fangbing Xiong explores how firms and auditors can leverage search engine query frequency, commonly measured using the SVI in general and referred to as Google Trends specifically. Her paper reviews SVI literature and demonstrates how to use Google Trends to gather information about consumer habits.

In a related area, Chapter 5 by Maria Guangyue Zhang uses SVI to demonstrate that investors have shifted away from firms' annual financial reports toward newer, third-party intermediaries that offer more timely and digestible information. In addition, she shows that investors in the cryptocurrency market react materially to relevant news headlines, further suggesting that exogenous sources of information have tangible effects on financial markets.

In Chapter 6, Xin Jin and Haoyun (Harry) Gao predict Brazilian airlines' quarterly revenue using SVI statistics, both concurrently and for the subsequent

quarter. They also use SVI to predict flight-related carbon emissions both in the same month and two months in advance, rationalizing that customers are likely searching on Google either for flights on short notice or for trips far in advance.

Chapter 7 by Maristela Roza, Leonardo Flach, and Fabricia Silva da Rosa examines the impact of EU policy changes on consumers in the German electronic vehicle market through the lens of Google Trends. She also uses Trends to proxy for consumers' stress in response to the policy changes, gaining insight into the psychology of a large population of potential customers.

Part Three: Exogenous Data from Social Media

Shaoyu (Shawn) Liu catalogs the different types of misleading reviews on social media in Chapter 8, including compensated endorsements and targeted slander campaigns in the context of past research on social media and the integrity of review systems. Detection systems struggle to keep up with innovative manipulators, suggesting that social media platforms and users alike should be prepared to grapple with widespread misinformation when conducting business online.

Chapter 9 by Huaxia Li describes exogenous data sources that may be useful in governmental accounting, with a particular emphasis on governmental environmental, social, and governance outcomes. This includes foot traffic data, geopolitical and economic data, and measures of public sentiment toward governmental policy. A case study uses text mining to extract negative sentiment in response to New York City street cleanliness, providing novel information that may have implications for future legislation and city planning.

In Chapter 10, Steven Katz discusses flaws in existing predictors for movies' box office revenue and proposes two potential exogenous leading indicators: illegal movie downloads and activity on piracy-focused social media websites. In this framework, legitimate consumers' intent to view is proxied by tracking downloads and uploads on illegal torrent streams, while piracy-focused social media posts can be mined for their sentiment and volume of discussion to represent interest and viewer response at large.

Talha Afzal proposes a framework for mining customers' textual internet reviews in Chapter 11 by decomposing them into topics using machine-learning algorithms. While numerical reviews are straightforward to process, incorporating the context offered by customers' stated opinions has the potential to yield much more informative analytics.

In Chapter 12, Lanxin Jiang reviews past studies that analyze different aspects of accounting employees' LinkedIn profile pages, including profile photos, number of connections, and individual characteristics. LinkedIn's unique status as a combined social media platform and professional hub creates an unusual convergence of data that can inform studies ranging from the financial to the deeply psychological.

We are all adrift in a sea of data. An incredible amount is being generated, stored, and lost every day. Exogenous data analyses present opportunities to supplement traditional analyses by pulling something relevant out of the great shifting sea of data. Realizing that potential will require ingenuity, passion, and

scientific rigor. This volume is an attempt to demonstrate and inspire all three of these traits in this tentative analytical realm. The world is full of interesting, useful information if we are prepared to look for it.

References

Gu, Y., Dai, J., & Vasarhelyi, M. A. (2023). Audit 4.0-based ESG assurance: An example of using satellite images on GHG emissions. *International Journal of Accounting Information Systems*, *50*(C), 100625.

Yoon, K., Kogan, A., Vasarhelyi, M. A., & Pearce, T. (2024). External nonfinancial measures in substantive analytical procedures: Contributions of weather information. *Journal of Information Systems*, *38*(2), 143–162.

Part One

Exogenous Data in Business Research

This page intentionally left blank

Chapter 1

Accounting's Resurrection: An Exploration of Future Business Reporting with Exogenous Data

Melissa A. Dardani

Rutgers Business School, USA

Abstract

As the relevance and utility of traditional accounting data decline, emerging types of data are proving valuable for predictive analytics. The accounting profession must contend with this data's increased capability to enhance business reporting in the face of antiquated reporting methods and standards. Simultaneously, a disconnect between financial reports and stakeholder needs arises from inherent information asymmetry and conflicting reporting priorities among stakeholders. Management tends to focus on factors affecting share price, while investors may be concerned with the horizon and sustainability of an investment's value. When coupled with the subjective nature of accounting principles and the limitations of auditing practices, concerns about the reliability and usefulness of modern financial reports are exacerbated. Despite the potential of predictive analytics to enhance business reporting, the profession is lagging with respect to the integration of these new, valuable data sources. Not embracing exogenous data and artificial intelligence for their utility seems inconsistent with the conceptual framework for financial reports, which defines the characteristics of useful financial information. Bridging this gap requires accounting practices, standards, and frameworks to evolve. This chapter is designed to encourage the exploration of expanded data sources and analytical methodologies in accounting and auditing by outlining conceptual

Exogenous Data in Accounting and Auditing in the Rutgers Series in Accounting Information Systems, 7–23

doi:10.1108/978-1-83608-734-220251001

considerations for this proposed enhancement to financial reporting and assurance.

Keywords: Audit; accounting; financial reporting; artificial intelligence; data analysis; predictive analysis; exogenous data; big data

Introduction

As the value of accounting data is diminishing in terms of its usefulness, timeliness, and understandability, other types of data are emerging as valuable sources of predictive information. The modern problem with financial reports should be attributed, at least in part, to the inherent dichotomy among various groups of stakeholders and their conflicted reporting priorities. For example, organizational management is primarily concerned with attributes that impact share price, such as analyst ratings and institutional interest, while prospective investors tend to focus on a company's fair market value and ability to generate returns over time. Paired with the asymmetrical nature of business information, this problem should be of concern to the accounting profession, especially when considering the subjective nature of accounting principles and the limitations of auditing practices. Consequently, today's business filings have questionable usefulness for stakeholders.

The increase in data and the ease of access to analytical tools present an opportunity to improve the business reporting cycle with predictive analytics, a shift that is already driving business performance and market price discovery. Analytics are used internally by market participants, such as companies and analysts, for growth, profitability, strategy, and planning. Analysts also use a wide array of market-related financial and non-financial data to explore investment performance.

Yet accounting measurements, perhaps rightfully, tend to favor fair representation and consistency in reporting and have, therefore, been slow to incorporate these new sources of available information. It is argued that current financial reporting standards and their historical focus are hindering the reports' usefulness by excluding significant, even material components of value (Schmidt et al., 2020). The field of computer science is expanding the usefulness of data through the creation of predictive models (Kureljusic & Karger, 2024). These models require data that capture the predictive qualities surrounding the prediction target. For financial markets, these sources come in a wide variety, but for reporting purposes, these abundant data sources must be filtered to include only those impacting actual value, rather than reactionary attributes that cause fickle swings in market price.

In light of the widely accessible nature of exogenous data today, simply continuing to ignore these external predictive variables leads to an inexcusable information gap between market performance and financial reporting. For example, as of the date of this writing, artificial intelligence computing leader NVIDIA is