MIDWEST DECISION SCIENCES INSTITUTE

2025 Midwest Decision Sciences Institute Conference

Hosted by Penn State Behrend, Erie, PA

April 11-12, 2025

Responsible Leadership and Decision Making for Resilient Businesses Through Analytics

Conference Chair

Sourish Sarkar, Penn State Behrend

Student Case Competition Organizers

Carol Putman, Penn State Behrend

Rajkamal Kesharwani, Mercyhurst University

PennState Behrend

Black School of Business

Conference Organization

Conference Chair
Sourish Sarkar, Penn State Behrend

Undergraduate Student Case Competition Organizers
Carol Putman, Penn State Behrend
Rajkamal Kesharwani, Mercyhurst University

Conference Organizing Committee
Xianghui (Richard) Peng, Penn State Behrend
Ginny Schwartz, Penn State Behrend
Carol Peterson, Penn State Behrend
Carol Putman, Penn State Behrend
Rajkamal Kesharwani, Mercyhurst University
Osman Aydas, Oakland University
Jeong Hoon Choi, Youngstown State University
Alina Marculetiu, Youngstown State University

Award Committees and Review Panels

Stan Hardy Best Conference Paper Award — Review Panel
Yi Tan, University of Alabama in Huntsville
Yixiu (Ashley) Yu, Ball State University
Sumanta Singha, Texas Tech University
Shubham Singh, Purdue University Fort Wayne
Cherry Singhal, University of New Mexico

Student Best Paper Award — Review Panel
Junhyun Bae, Oakland University
Yaman Roumani, Oakland University
Joseph Nwankpa, Miami University
Rahul Venkat, Purdue University
Sharan Shirodkar, Prediction Guard

Undergraduate Student Case Competition — Organizers
Carol Putman, Penn State Behrend
Rajkamal Kesharwani, Mercyhurst University

MWDSI Board of Directors

Serving at the time of the 2025 conference

Osman Aydas, Oakland University	*President*
Jeong Hoon Choi, Youngstown State University	*President-Elect*
Qiannong (Chan) Gu, Ball State University	*Past President*
Sourish Sarkar, Penn State Behrend	*Secretary*
Jiangxia (Renee) Liu, Valparaiso University	*Treasurer*
Yi (Maggie) Guo, University of Michigan-Dearborn	*Vice President for Communications*
Jason Davidson, Butler University	*Vice President for Technology*
Alina Marculetiu, Youngstown State University	*Vice President for Member Services*
Ata Karbasi, Valparaiso University	*Vice President for Industry Liaison*
Matthew Lanham, Purdue University	*Vice President for Student Liaison*

Conference Awards

Stan Hardy Best Conference Paper Award

Winner 74
Does News Coverage of ESG Violations Affect Brand Sales? An Empirical Analysis
Yao Chen, Bucknell University; Rakesh Mallipeddi, The Ohio State University; Serkan Akturk, Clemson University

First Runner-Up 43
Trade credit contracts under weather risk
Piyal Sarkar, Pennsylvania State University, Harrisburg; Mohamed Wahab Mohamed Ismail, Toronto Metropolitan University; Liping Fang, Toronto Metropolitan University

Second Runner-Up 171
Sustaining The Surge: Twitter Topics and Persistent Momentum of Meme Stocks
Shaoqing Zhang, Le Moyne College; Xinyuan Tao, New Jersey Institute of Technology

Student Best Paper Award

Undergraduate Division — First Place 163
Who is More Likely to Enjoy Football as Their Favorite Sport? The Impact of Gender & Previous Sports Experience
Eileen Zhan, Calvin University; Yong Xu, Ferris State University

Graduate Division — First Place, and MWDSI Social Media Award 268
Mapping the Path to Translations: Empowering SIL to Reach New Frontiers in Language Translations
Vandana Brahmasa, Amy Fischbach, Parvathy Nair, Gitanjali Nambiar, Aryan Ritesh, Huilin Xu, Rohan Ajay, Matthew A. Lanham, Purdue University

Graduate Division — Second Place 259
Detecting and Predicting Phantom Inventory Using Random Forest and LSTM Models
Hung-Chen Hsu, Kratika Jain, Da Fang Lin, Yiran Liu, Kaushalya B. Naidu, Namitha Ramakrishnan, Rohan Ajay, Matthew A. Lanham, Purdue University

Graduate Division — Third Place 233
Simulating B2B Sales Process Using Agentic AI
Chhaya Tundwal, Kshitij Chauhan, Tirth Patel, Vishnu Anand, Yash Kothari, Yu-Yang Ho, Rohan Ajay, Matthew A. Lanham, Purdue University

Undergraduate Student Case Competition

First Place
Bowling Green State University
Stephen Sostakowski, Shannon Kelley, Sabrine Encina, Hai Le

Second Place
University of Michigan-Dearborn
Mariana Molinari, Maddie Johnson, Colin Sethi, John Hakanson

Third Place
John Carroll University
Maria Julien, Bene Caputo, Alexis Kirby, Kelly Faber

Sponsors

Corporate Sponsor

Logistics Plus

Student Paper Competition Sponsor

Purdue University, Mitchell E. Daniels, Jr. School of Business

Student Case Competition Sponsor

Erie Brewing Company

Host Sponsor

Penn State Behrend, Black School of Business

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Abstracts

Diversity, equity, equality, and inclusion from the perspective of sustainable development goals: A bibliometric analysis

Mariana Nicolae, Russell Merz, Sufian Qrunfleh

Eastern Michigan University, Ypsilanti, USA

Abstract

The United Nations (UN) and the Association to Advance Collegiate Schools of Business (AACSB) recognize the importance of research in creating societal impact. This research aims to examine the development of diversity, equity, equality, and inclusion (DEEI) literature through the lens of the UN Sustainable Development Goals (SDGs), which offer a universal framework for addressing critical global economic, environmental, and social issues. A bibliometric analysis of business school scholarship retrieved from the Web of Science database is conducted using the R-based Bibliometrix tool.

Reliability of supply chain networks under uncertain failures

Ashish Chandra

Illinois State University, Normal, USA

Abstract

In today's rapidly evolving and uncertain supply chain landscape, designing networks that consistently perform under fluctuating demand and unexpected disruptions remains a critical challenge. The vast number of potential failure scenarios makes comprehensive evaluation infeasible using traditional methods. To address this, we present an optimization-based framework that quantifies the worst-case performance of supply chain networks while incorporating flexible routing strategies. By explicitly modeling disruptions and demand uncertainty, our approach overcomes computational intractability and establishes valid upper bounds on worst-case link utilization and unmet demand—key metrics for ensuring operational resilience. Leveraging robust optimization and convexification techniques, we optimize network design to withstand disruptions and adapt to demand variations. Through extensive evaluations across diverse supply chain topologies, our results highlight the effectiveness of our method in enhancing network reliability. This framework not only provides a systematic approach for performance certification but also serves as a valuable decision-support tool for supply chain architects seeking to build more resilient networks in an increasingly volatile global environment.

Exposing the Productivity Challenge in the MENA region: evidence from a 50-years scientometric analysis

Nasreddine Saadouli

Gulf University for Science and Technology, Mishref, Kuwait

Abstract

The Middle East and North Africa (MENA) region often makes the World news; however, it is rarely business news. A significant number of the countries in this region are plagued by unstable political systems and inefficient economies. The effects of macro and micro factors on productivity have been studied scarcely in the MENA region. Nonetheless, micro factors generally have a positive effect on productivity; whereas macro factors have been found to exert a negative impact on total factor productivity. This paper examines the research related to productivity in the MENA region and through a scientometric analysis depicts the broad intellectual and conceptual trends. Consequently, a thematic map is developed to categorize the key findings and discern the challenges facing productivity in the MENA region and the directions for future research.

The Impact of Funding and Manager's Compensation on Innovation: SMEs in Ireland

Carlos Liard

Central Connecticut State University, New Britain, USA

Abstract

The literature is not clear about whether internal or external financing of the day-to-day operations of firms is better for promoting innovation among Small and Medium-sized Enterprises (SMEs), as the effectiveness of each source of funding may depend on various factors such as the firm's stage of development, industry, and business model. On the one hand, internal financing, such as reinvesting profits or using existing resources, can provide greater flexibility and control over the innovation process. On the other hand, external funding, such as venture capital, can bring in expertise, networks, and resources that may not be available internally. At the same time, some evidence suggests that performance compensation for managers also affects innovation. While some studies have found a positive relationship between performance-based compensation and innovation, others have found no significant or negative relationship. This paper implements a Logit model to analyze the impact of both sources of financing, as well as manager's compensation, on innovation. Following the World Bank's Enterprise Survey, innovation is defined as new methods of manufacturing products, logistics, or distribution methods for inputs.

Employee Motivation in Iran's E&P Sector: A Quantitative Exploration

Hamed Mehrzadegan[1], Vahid Ghomi[2]

[1]University of Bolton, London, United Kingdom. [2]Penn State, Mont Alto, USA

Abstract

Employee motivation is a critical determinant of organizational performance, particularly in high-risk and resource-intensive sectors like Exploration and Production (E&P). This study investigates the interplay of organizational, cultural, and economic factors influencing employee motivation within the Iranian E&P sector. Using a quantitative research approach, data were collected from employees of three leading E&P companies through structured questionnaires. Hayes' Process Model was applied to explore mediation and moderation effects within the dataset.

The findings reveal that organizational factors, including leadership quality and safety culture, significantly mediate the impact of broader contextual variables on employee motivation. However, cultural and economic factors demonstrate nuanced qualitative influences, shaping motivational dynamics indirectly. The results contribute to motivation theory by integrating sector-specific and cultural-economic perspectives and suggest practical strategies for policymakers and industry leaders. Key recommendations include leadership development programs, enhanced safety policies, and tailored motivation strategies responsive to Iran's socio-economic context.

These insights offer valuable contributions to both academic literature and practical frameworks for optimizing workforce motivation in challenging industrial environments.

Teaching Operations through Real World Case Studies of On-campus Businesses

Rajkamal Kesharwani[1], Sourish Sarkar[2]

[1]Mercyhurst University, Erie, USA. [2]Penn State University, Erie, USA

Abstract

It is widely believed that students absorb and retain concepts better through hands on experience. In this presentation, we study the projects of undergraduate students in a sophomore level operations class, where the students form groups to study various businesses on campus to understand how the textbook concepts relate to the real world. A total of seven businesses were studied such as cafeterias, stadiums, art theaters and gymnasium facilities. The students were asked specific questions and the goal was to enable the students to compare these businesses and list out their similarities and differences from an operations standpoint. It was found that this approach to operations resulted in not only a higher experience for the students, but the students who were the customers of these businesses were also able to act like consultants and give improvement suggestions to the businesses.

Understanding Business College Students' Perceptions of the Classroom Climate and Relationship-rich Education: An Experiential Study

Recep Pekdemir

University of Wisconsin La Crosse, La Crosse, USA

Abstract

The paper aims to document the experience of classroom climate and relationship-rich education practices in undergraduate-level accounting courses. The experience was realized during the Fall of 2023, the Spring of 2024, and the Fall of 2024 in an upper Midwest middle-class university. The study has two purposes: The first states the position of the classroom climate and relationship-rich education environments in the accounting courses. The second seeks consistency between the results of the Learning Environment Surveys conducted by the university and those of the survey the instructor conducted. The results show the demographics of those surveys, and a comparative analysis is performed to understand consistency.

College educators are liable to influence student growth and behavior, and classrooms or learning environments will be essential, as well as social climate and social and physical elements. Both college administrators and educators should ensure a welcoming classroom and learning environment in which students are engaged, connected, motivated, satisfied, and successful. Thus, college administrations and educators survey to understand students' thoughts and perceptions for a better learning environment.

The study's findings are interesting and encouraging, so the instructor will keep going with such studies to understand the classroom climate and relationship-rich education environments during the classroom sessions. The study might be interesting to some other college educators as well.

Agile Project Management Theory Development

Ying Cao[1], Xinyu Wei[2], Xianghui Peng[1]

[1]Penn State Erie, The Behrend College, Erie, USA. [2]California State University, Chico, Chico, USA

Abstract

This paper scrutinizes the prevailing methods of agile project management (APM) in the industry, aiming to transition APM from a practice-oriented to a theory-oriented approach. A research model is developed through theoretical justifications and tested using survey data from project management professionals via structural equation modeling. This study establishes APM as a mature organizational initiative fostering excellence and transformation. By building upon existing validated theories, it identifies critical constructs, validates measurements, and tests theoretical frameworks, thereby advancing APM theory development. Furthermore, a holistic investigation of APM with theoretical foundations facilitates the diffusion of its principles among industry practitioners, enhancing knowledge dissemination for apprentices. Ultimately, this study provides a comprehensive understanding of APM's theoretical underpinnings, offering practical guidance for its implementation in diverse organizational contexts.

Leading Transformative Change: Harnessing AI for Strategic Innovation and Organizational Growth

Osman Aydas

Oakland University, Rochester, MI, USA

Abstract

This presentation outlines a course that empowers executives to lead AI-driven transformations, equipping them with the knowledge and strategies needed to harness the transformative potential of Artificial Intelligence (AI) in modern organizations. Through an engaging blend of foundational concepts, real-world applications, and strategic frameworks, participants will explore AI's role in reshaping industries, enhancing decision-making, and driving innovation.

Key topics include a comprehensive understanding of AI, its evolution, and its impact on human decision-making. Practical applications are highlighted through case studies, offering participants actionable insights into data analytics and opportunity identification. Additionally, the course delves into strategies for organizational transformation in the AI era, ethical considerations, and leadership imperatives necessary for fostering AI innovation. Participants will leave this course equipped to drive AI adoption, capitalize on emerging opportunities, and position their organizations as leaders in the age of AI.

Diagnostic Services: Online versus Face-to-Face

Zafer Ozdemir

Miami University, Oxford, USA

Abstract

Diagnostic services are increasingly being offered online. A unique characteristic of such services is that, a consumer may seek to validate an initial diagnosis by purchasing another one. Using a game theoretic model, we derive the value of a diagnosis for a buyer and then characterize the conditions under which a provider would utilize the face-to-face versus the online channel, as well as those that render the sale of "validation" services preferable to the sale of "initial" services. Using monopoly and duopoly settings and a dynamic information search process, we show that (i) the cost to acquire the face-to-face service of a high-quality provider and (ii) a negative impact on the provider's diagnostic accuracy reduces the value of an initial service more than that of a validation service. Accordingly, this research provides a framework to understand when a high-quality provider should focus on the sale of validation as opposed to initial services, and whether to deliver them online or face-to-face.

Inventory Shortages and Customer Service Levels in Retail Outlers

Amelia Carr

Bowling Green State University, Bowling Green, USA

Abstract

Throughout the midwestern United States, inventory stock outs in retail outlets are visible. Consumers experience delays in being able to purchase certain goods and services as they shop these retail outlets. The supply chain shortages for certain items impact manufacturers production of goods when companies are not able to acquire raw materials from suppliers. There were inventory shortages worldwide beginning in 2020 during the pandemic due to plant closures and border closings. Years since the pandemic was declared over, the are inventory shortages that persist are driven by other factors. In this study, we explore what some of the shortages are, where they occur in the supply chain, and what is the cause of the inventory shortages. The research question concerns whether it is a supply chain decision to have less inventory on hand? Or, is it an issue in the supply chain that is causing the inventory shortages? The study includes visits to retail outlets to observe what the inventory shortages are and interviews with the retailers to understand the reason for the shortages. A survey of customers provide insights on how the consumers are responding to the inventory shortages. Using statistical analysis, we assess what level of inventory shortage leads to lower cost for the retailer and while maintaining an acceptable customer service levels. Based on this study, theoretical and managerial implications are discussed and directions for future research.

Utilizing Market Competition for Managerial Decisions

Sima Fortsch[1], Jeong Hoon Choi[2], Elena Khapalova[3], Chia-Hung Chuang[4]

[1]Northern Kentucky University, HilHighland Heights, KY, 41076, USA. [2]Youngstown State University, YoungstownYoungstown, OH, 44555,, USA. [3]State University of New York at Canton, Canton, NY 13617, USA. [4]University of North Dakota, North Dakota, USA

Abstract

Purpose – Amid frequent job transitions and an abundance of information, managers increasingly rely on data-driven decisions. Leveraging competitors' sales data is crucial, but managing multiple competitors' information can be both challenging and costly. This research aims to demonstrate that effective data-driven decisions can be made with less information by exploring how sales data from one competitor can be used as a substitute for another.

Design/methodology/approach – This study focuses on the North American automotive industry, utilizing monthly sales data from five major players: GM, Ford, Honda, Toyota, and Nissan. Four scenarios are examined: (1) access to complete sales information from all competitors, (2) address the missing data by substituting variables using the historical marginal rate of substitution between two variables, (3) employ a practical and time-efficient (direct) substitution method to fill data gaps, and (4) disregard the missing information entirely.

Findings – Our results demonstrate that managers do not require complete information to make effective, data-driven decisions. A straightforward approach to addressing data gaps proves sufficient, allowing for efficiency without compromising the quality of information-led strategies.

Keywords elasticity; autoregressive models; time series regression; seemingly unrelated regression.

Sales competition among automakers: A comparative study of five firms

Sima Fortsch[1], Jeong Hoon Choi[2], Mohsen Ghodrat[3]

[1]Northern Kentucky University, Highland Heights, KY, 41076, USA. [2]Youngstown State University, Youngstown, OH, USA. [3]University of Canada West, Vancouver, BC, Canada

Abstract

This study gains insight into the competitive dynamic in the North American auto industry. We explore the intensity of competition by analyzing time series sales data over a 10-year period from 2009 to 2019. We used two regression methodologies: vector autoregressive regression and seemingly unrelated regression models. Contrary to the expected fierce competition, the results show moderate competition between the automakers' sales. While there are some interrelations between the sales, the rivalry does not affect all members equally. These findings challenge the prevailing notion of intense intra-group competition within the North American auto industry.

Keywords: autoregressive models; time series regression.

Helping Students Contextualizing OM Concepts through Instructional Means

Esmail Mohebbi

University of West Florida, Pensacola, USA

Abstract

This talk explores the effectiveness of using structured field interviews and simulations as instructional tools to facilitate students' ability to contextualize topics and related concepts, thereby cultivating a more profound learning environment in an undergraduate operations management course. Performance- and perception-based learning assessment measures are utilized to present a comparative analysis of the utility of alternative means of instruction.

Agent AI: Revolutionizing Industries and Reshaping the Future of Work

Nikhil Katiki

Herndon, ARLINGTON, USA

Abstract

The advent of Artificial Intelligence is revolutionizing the global landscape, offering innovative solutions that enhance automation while minimizing human intervention. Among the emerging technologies, "Agent AI" stands out for its potential to impact organizations both vertically and horizontally. This paper delves into the prospective role of AI Agents as powerful tools for organizations in the near future, exploring their capacity to drive efficiency across various industries. Attendees will gain valuable insights into the influence of Agent AI on sectors such as Customer Service and Sales, as well as its potential to transform individual workflows through automation. While the prospects are promising, this talk also addresses the critical aspects of AI hallucination and potential pitfalls in the Agent AI revolution. Furthermore, it highlights ongoing research areas, providing a balanced perspective on this transformative technology and its implications for the future of work and industry.

Optimizing Reverse Logistics for Circular Economy: Strategies for Efficient Material Recovery and Resource Circularity.

Oluwabukola Daramola[1], Olumide Ibikunle[2]

[1]Tennessee State Uni, Nashville, USA. [2]Vanderbilt University, Nashville, USA

Abstract

Optimizing reverse logistics is essential for advancing the circular economy by enabling efficient material recovery and resource circularity. This study explores strategic frameworks and innovative practices to enhance reverse logistics systems, integrating technologies such as IoT, blockchain, and AI to improve traceability, streamline operations, and support data-driven decision-making.

Collaboration among manufacturers, distributors, consumers, and policymakers is crucial in developing seamless reverse logistics networks. Regulatory frameworks like extended producer responsibility (EPR) and eco-design mandates further drive resource-efficient systems. Consumer participation is also emphasized, with awareness campaigns and incentives playing a key role in responsible disposal behaviors.

Case studies from industries such as electronics, textiles, and automotive highlight successful reverse logistics implementations, demonstrating reductions in waste and resource inefficiencies. Performance metrics, including material recovery rates, cost savings, and carbon footprint reduction, are used to assess economic, environmental, and social benefits.

Despite these advantages, challenges such as logistical complexity, high initial investments, and fragmented supply chains persist. Solutions include leveraging digital platforms for real-time data sharing, adopting closed-loop supply chain models, and fostering public-private partnerships to share risks and resources.

By optimizing reverse logistics and reinforcing circular economy principles, organizations can achieve sustainable growth, reduce dependency on finite resources, and contribute to global environmental goals. This research offers actionable insights for policymakers, businesses, and academia to collaboratively drive a more sustainable future.

The Legitimacy Journey of Generative AI in Sport: Media Perceptions, Adoption, and Industry Impact.

Eric Brownlee[1], Steven Creek[2], Josh Maurer[1], Benyawarath Nithithanatchinnapat[1], Jinhee Yoo[1]

[1]Gannon University, Erie, USA. [2]Appalachian State University, Boone, USA

Abstract

The emergence of Generative AI has begun reshaping industries, including sport business and affiliated sectors. As with any transformative technology, legitimacy, adaptation, and trust are critical hurdles before it becomes widely accepted and customer-facing. This study examines the first two years of Generative AI following the introduction of ChatGPT in November 2022, focusing exclusively on its intersection with the sporting domain. Given the media's central role in shaping public perception, informing on the availability of new advances, and influencing stakeholder attitudes, we investigate the viewpoints, adoption, and uses of this new technology on its path towards legitimacy in the sports sector. Through a content analysis of news articles sourced from Lexis-Nexis, we employ sentiment and thematic analysis to assess media portrayals of this technology. Our findings reveal the varying degrees of support for Generative AI within these industries over time, highlighting both the positive and negative impacts linked to different sporting domains and forms of legitimacy. This study contributes to the understanding of how an emerging technology's legitimacy and uses evolve in complex, high-visibility industries and offers insights into the role of media in shaping technological adoption.

Leveraging Corporate Social Performance for Servitization: A Dynamic Capabilities Perspective

Sean Yim[1], Catherine Ha[2], Ben Lee[1]

[1]Penn State Behrend, Erie, USA. [2]Pacific Lutheran University, Tacoma, USA

Abstract

Servitization, a strategic decision where firms shift from product-centric to service-centric business models, has become increasingly critical for sustainable competitive advantage in today's dynamic business environment. Literature indicates that the service transition strategy (hereafter servitization) represents a fundamental organizational transformation that requires substantial firm resources. Drawing on Dynamic Capabilities theory, we argue that successful servitization depends on firms' abilities to sense service opportunities, seize these opportunities, and reconfigure resources accordingly. We propose that Corporate Social Performance (CSP) develops these essential dynamic capabilities that facilitate servitization. Using a panel dataset of over 21,000 firm-year observations, our empirical study reveals that CSP positively influences servitization (measured by service business ratio). However, this relationship is weakened in high-tech industries where technological capabilities may be more crucial. Conversely, firms with strong innovation values can better leverage CSP-based dynamic capabilities to achieve superior servitization outcomes.

Designing Decision Points in AI-Guided Learning: The PALS Framework

Kyle Chalupczynski

Penn State, Erie, USA

Abstract

This paper presents the development of a novel decision-based framework for AI-guided learning assignments. Current AI integration approaches in education often lack structured interaction points, potentially leading to passive engagement. The PALS (Personalized AI Learning Simulations) framework addresses this by embedding strategic decision points throughout the learning process. Each assignment incorporates specific decision tollgates where students must make and justify choices based on course concepts before proceeding. The framework's design emphasizes systematic progression: each decision point builds on previous choices, creating a coherent learning path while enabling assessment of student understanding. Currently being implemented in an introductory MIS course, this approach offers a theoretical model for integrating decision-making skills into AI-enhanced education. The paper discusses the framework's design principles, implementation considerations, and potential implications for AI integration in higher education.

Developing Faculty Tools for AI Integration: A Systematic Design Approach

Kyle Chalupczynski

Penn State, Erie, USA

Abstract

This paper presents the development of a toolkit designed to support faculty AI integration efforts in higher education. The system includes four complementary tools: a narrative arc creator, assignment builder, implementation guide, and custom instruction configurator. Each tool aims to automate technical aspects while preserving instructor control over educational content. The toolkit's design emphasizes flexibility - supporting various implementation levels from single assignments to full courses. Currently in beta testing, this approach explores systematic methods for scaling AI integration across disciplines while maintaining pedagogical quality. The research contributes to the emerging discussion on practical approaches to AI integration in higher education, offering a structured framework for faculty implementation. The paper examines design considerations, initial testing feedback, and potential pathways for broader adoption.

Revolutionizing production processes: The impact of generative AI on the Korean webtoon industry

Jeong Hoon Choi

Youngstown State University, Youngstown, USA

Abstract

The rapid advancement of generative artificial intelligence (AI) is transforming production processes across industries, including the digital content sector. This study explores the impact of generative AI on the webtoon industry, with a particular focus on South Korea, a global leader in webtoon production. Traditionally, the webtoon creation process has been characterized by a division of labor across different countries, where Korean creators developed the story, initial sketches were outsourced to lower-cost regions, and final touches were completed in Korea. However, the integration of generative AI tools is disrupting this workflow by automating sketching, coloring, and even storyboarding, significantly reducing dependence on human labor and geographic constraints. By examining these changes, this study aims to provide a theoretical framework for understanding how AI-driven production reconfigures the webtoon industry. The findings offer insights into the broader implications of AI adoption, shedding light on the evolving dynamics of labor, supply chains, and creative autonomy in the digital era.

A Generalized PERT for Subjective Distributions

Byung Cheol Kim

Erie, Erie, USA

Abstract

The Program Evaluation and Review Technique (PERT) is a widely used tool for project risk modeling and simulation. However, it has been criticized for its limited adaptability to general subjective distributions, primarily due to its assumptions of constant variance and limited tail behaviors. In this study, we explore an approach to generalizing the classic PERT model, originally developed by the U.S. Navy for the Polaris project in the late 1950s. By revisiting PERT's foundational principles, we reconstruct its core structure into a more robust algorithm. Specifically, we redefine the variance constraint of the classic PERT as a decision parameter, allowing for greater flexibility in modeling uncertainty. Through comprehensive regression analyses, we derive a simplified formula that facilitates the practical implementation of the generalized PERT across a wide range of applications. This refined approach enhances PERT's adaptability while preserving its inherent simplicity. Finally, we present preliminary experimental results that demonstrate the effectiveness of this extended framework.

Marketing Operations: A Strategic Framework for Organizational Success and Decision-Making

Jeff Darville

LaRoche University, Pittsburgh, USA

Abstract

This research examines the evolving role of marketing operations in modern organizations by analyzing public marketing data and existing literature. Drawing from multiple industry case studies and empirical research, this investigation explores how marketing operations contribute to organizational effectiveness, promotion campaign execution, and strategic planning. The study integrates technology, performance measurement, and data analytics in marketing decision-making processes. The analysis reveals that organizations with well-structured marketing operations exhibit superior campaign performance and more effective resource allocation. Through the examination of both B2B and B2C contexts, key success factors in marketing operations are identified, including integrated technology systems, standardized measurement frameworks, and data-driven decision-making protocols. This research builds upon a framework for marketing operations efficiency, extending it to incorporate modern digital transformation considerations. Organizations implementing comprehensive marketing operations show improved cross-functional collaboration, enhanced integrated marketing communications consistency, and better alignment between marketing activities and business objectives. The paper also explores how marketing operations bridge strategic planning and tactical execution, enabling a more agile response to market changes. Findings suggest successful marketing operations require a balanced approach combining technological infrastructure, organizational processes, and human expertise. This research contributes to the growing body of knowledge regarding the strategic importance of marketing operations. It provides practical insights for organizations aiming to optimize their marketing function through improved operational efficiency and data-driven decision-making.

Beyond Traditional Indicators: Exploring Macroeconomic and Institutional Determinants of Logistics Performance Using Lasso Regression

Bitanya Adane, Bulent Erenay, Dekuwmini Mornah

Northern Kentucky University, Highland Heights, USA

Abstract

The Logistics Performance Index (LPI) serves as a critical benchmark for assessing the efficiency of global trade logistics. While existing LPI components—such as customs efficiency, infrastructure quality, and shipment timeliness—offer valuable insights, they may not fully capture the broader economic and institutional factors influencing logistics performance. Previous research suggests that additional macroeconomic, governance, and innovation-driven factors significantly impact a country's logistics capabilities. This study employs Lasso regression modeling to identify key determinants of LPI by incorporating the World Governance Index (WGI), Global Innovation Index, Human Development Index (HDI), and regional logistics performance alongside the traditional LPI indicators. By leveraging Lasso regression, we enhance variable selection and mitigate multicollinearity among predictors, ensuring a robust and interpretable model. This research expands the understanding of LPI determinants, demonstrating the importance of institutional and innovation-driven factors in shaping logistics outcomes.

Underpaid and Overperforming: Examining Gender Disparities in Executive Leadership and Compensation

Jiangxia Liu

Valparaiso University, Valparaiso, USA

Abstract

Women occupy a disproportionately low percentage of executive leadership positions and, on average, receive lower compensation than their male counterparts, despite empirical evidence suggesting that they frequently outperform men in similar roles. While some of the observed gender pay disparity can be attributed to measurable factors such as hierarchical position and tenure, a significant portion remains unaccounted for. This study seeks to examine the underlying causes of this paradox, exploring structural, organizational, and socio-cultural factors that contribute to persistent gender disparities in executive compensation and leadership representation.

Critical aspects of data in AI

Hongjiang Xu

Butler University, Indianapolis, USA

Abstract

AI are built based on large scale of data. There are some critical aspects of data that need attention in AI development and usage. With data-centric and data-driven AI applications, they use large amount of data to train and test the AI models. The AI models use large training data to identify patterns and trends to help them learn and make predictions to produce outcomes and assist with decision making. The critical data related aspects in AI include data quality, data collection, data access, data processing, data storage, data management, data governance and data privacy and security. Each of those aspects also includes many different sub-dimensions. For example, for data quality in the model building process, it involves many data quality controls and assurance measurements, such as, data cleaning, credibility check, completeness check, data validation, data bias check, data preparation, algorithmic fairness check, continuous monitoring, and human expert oversight. Organizations need to pay attention to those data aspects in order to have successful development and full usage of credible AI applications.

A Unified Framework for Decision-Making in Retail Evaluation: Integrating AI-Driven Sentiment Analysis and Expert Insights with Fuzzy MCDM

Adem Pinar

The School of Business, Shenandoah University, Winchester, USA

Abstract

This research presents a hybrid method that leverages AI-driven sentiment analyses alongside a fuzzy decision-making process to assess retailer performance from the customer's perspective. A large collection of customer reviews was processed using AI techniques to identify key themes that capture both positive experiences and areas for improvement. Sentiment analyses were conducted to extract emotional tones and opinions from the textual data. Expert judgment was subsequently applied to refine these themes and assign appropriate weights, ensuring that the significance of each aspect was clearly represented. The fuzzy approach, based on q-rung orthopair fuzzy sets, converted subjective opinions into straightforward numerical scores, thereby enabling a balanced comparison of retailers.

Critical aspects such as customer service, delivery efficiency, and return policies were examined while uncertainties inherent in traditional evaluation techniques—relying solely on numerical data or isolated expert opinions—were reduced. By combining insights from AI-driven sentiment analyses with expert evaluations, the proposed framework offers a practical tool for decision-makers seeking to enhance retail services and maintain competitiveness in a rapidly changing market. The results underscore the value of integrating human expertise with automated AI usage, leading to more informed and effective strategic decisions. A case study involving major U.S. retailers illustrates the framework's effectiveness, highlighting distinct performance differences and pinpointing specific areas for improvement, while laying the groundwork for future research in dynamic retail environments.

The AI Space Race: The Economic impact of DeepSeek on the AI Oligopoly

Jason Davidson

Butler University, Indianapolis, USA

Abstract

January 2025 the LLM DeepSeek R1 was released in direct competition to the industry leader OpenAI's platform, ChatGPT. DeepSeek gained rapid notoriety not only due to its mathematical reasoning architecture but its reduced cost and training time. While DeepSeek does not outperform ChatGPT in all tasks, it has proven superior in technical functions such as code generation. Additionally, while ChatGPT-4 cost $100 million dollars and utilized state of the art GPUs, allegedly, its counterpart used fewer and inferior computational technology due to export restrictions to China. Furthermore, the creators of DeepSeek claim it cost $6 million dollars to create and train which is 94% cheaper than GPT-4. While DeepSeek's performance is undeniable, development techniques, monetary investment, training equipment, and potential trade secret violations have been brought into question. Some believe DeepSeek was created using unauthorized technology from ChatGPT, others believe its developers are not honestly disclosing the development cost or access to advanced GPUs. No matter the truth, the performance of DeepSeek and the fact that its creators have made its design and code open source deeply impacts the current oligopoly in the AI LLM industry. What will be the impact on the industry if DeepSeek did create a state-of-the-art LLM for less capital, with inferior hardware, in less time? Alternatively, what are the consequences if the claims of design costs, training time, and development technology are false? This paper will explore these questions to determine the potential impacts of these recent developments on marketshare, competition, and potential legal recourse.

Costs and Benefits of Professional Skepticism

Ashutosh Deshmukh

Penn State Behrend, Erie, USA

Abstract

This paper investigates the effects of error rates on costs associated with professional skepticism. The concept of professional skepticism is a critical element of audit quality and is embedded in US and international auditing standards (see Hurtt (2010); Robinson (2018); Khan and Oczkowski (2019)). The extant research explores ethical disposition, personality traits, and situational factors.

This paper focuses on the costs and benefits associated with professional skepticism. The base error rates (unintentional and intentional) vary in each client's case. If an error is found, the increased audit cost is justified; otherwise, the auditor may be penalized by the superiors, or the audit may become uneconomical. The power of audit technology will determine the type I and type II error rates. Type I error can be defined as the auditor determining that an error exists when there is no error, and type II error can be defined as the auditor determining that an error does not exist when there is an error. Each error will have associated costs, such as increased audit costs or missed errors.

The correct application of professional skepticism needs to balance the type I and type II errors, as these errors are reciprocally related. This paper uses an analytical approach to analyze base rates of errors, costs associated with type I and type II errors, and the auditor’s judgment to either extend audit procedures or accept the account balance. The paper discusses the practice implications of the results.

Parallels and Challenges Between Supply Chain Trends and Generations

Carol Putman

Penn State Behrend, Erie, USA

Abstract

Supply chain management (SCM) is a dynamic and evolving field, both in business practice and academic study. As SCM trends have progressed, so have the generations managing these systems. Key challenges currently shaping the field include sustainability, resilience, data analytics, and artificial intelligence. Addressing these challenges in the college classroom is crucial for developing students' skills and preparing them for their careers.

The presence of multiple generations in both the workplace and the classroom adds complexity to the development of business decision-making skills. This paper presents parallel timelines of SCM trends and workforce/student generations, with a focus on the current environment to provide a foundation for exploring future developments. Enhancing effective communication and teamwork skills is essential for positive outcomes.

Teaching Service Design

Carol Putman

Penn State Behrend, Erie, USA

Abstract

Product development is crucial for the competitive success of manufacturing companies, requiring collaboration across various departments such as engineering, marketing, and operations. Similarly, service design is essential but has only recently gained prominence. Unlike product development, service improvements often remain siloed, addressing individual service experiences rather than the gaps between customer interactions.

While project management and other technical courses enhance skills relevant to both product development and service design, service-specific topics often receive less attention. Recent advancements in information technology, particularly agile project management and user experience (UX) design, have improved service design. However, a comprehensive customer experience perspective for service systems is still lacking, and suitable college textbooks are rare.

This paper presents a framework, resources, and technologies beneficial for teaching service design. It identifies common themes with traditional project management and discusses topics unique to service design. Integrating these skills into the business curriculum is helpful for preparing students to make improvements in service-oriented careers.

Corporate Social Strategy of Korean Business Groups

Donghyup Woo[1], Hyungsoo Woo[2,3]

[1]University of Pittsburgh at Greensburg, Greensburg, USA. [2]The Polish Academy of Sciences, Warsaw, Poland. [3]KAIST, Daejeon, Korea, Republic of

Abstract

Large corporations in Korea have focused on those parts of CSR activities that are more relevant to business and attractive to management beyond fulfilling normative expectations of society. Drawing on instrumental stakeholder theory and institutional economics, this study seeks to explain how Korean business groups strategically respond to stakeholder demands and corporate social contributions. The current paper examines the impact of sub-dimensions of corporate social performance (i.e., corporate governance, corporate integrity, community, customer, environment, and employee relations) on corporate financial performance.

An investigation and review of consumer perceptions of the threats associated with AI supported product development and AI containing consumer products

Alan Brandyberry

Kent State University, Ambassador Crawford College of Business and Entrepreneurship, Information Systems and Business Analytics Dept., Kent, OH, USA

Abstract

There has been a great deal of discussion about what threats, dangers, and risks are associated with Artificial Intelligence (AI) and its associated technologies; however, there has been little resolution. Certainly, consumers and others have expressed concerns about the privacy, security, and trust issues associated with the data that AI may employ to inform non-AI-containing product development as well as the innovation of new AI-containing products that are targeted towards consumers. At the extreme, existential threats have been proposed. There does seem to be parallels to past consumer trust issues with other technologies such as the issues in the late 1990's concerning consumer acceptance of electronic commerce (EC) during its introduction. This research first explores and categorizes the historical sources for perceived AI threats, including from literature and media as well as from consumer experiences with other technological innovations. Second, general technology threats associated with AI and those threats that appear to be unique to AI technologies will be discussed and categorized. Finally, forward projections concerning the future of consumer perceptions about AI and AI threats will be presented with suggestions for how industry may facilitate the development of consumer perceptions that enhance the ability of industry to successfully innovate AI products and reduce consumer resistance due to perceptions of AI threats. Additionally, the ethical considerations associated with activities designed to decrease perceived threats associated with AI will be addressed. This research will support planned future empirical research into the topic.

Driving Change Success: Readiness, Cynicism, and Emotions as Predictors of Organizational Citizenship Behaviors During Implementation

Andrés Raineri B.[1], Bernardo F. Quiroga[2]

[1]Pontificia Universidad Católica de Chile, Santiago, Chile. [2]West Virginia University, Morgantown, USA

Abstract

This study investigates how employees' Change Readiness, Change Cynicism, and Affects influence Organizational Citizenship Behaviors towards organization (OCBO) and individuals (OCBI) during change implementation. Data from 479 employee-manager dyads were analyzed using covariance-based factor analysis. Results revealed that Change Readiness and Change Cynicism directly influence both OCBO and OCBI, but through different pathways: OCBO was mediated by perceived benefits and activating pleasure, while OCBI was directly impacted by Self-Efficacy. The findings suggest that employee attitudes towards change are complexly interrelated, emphasizing the need for balanced change management strategies that address both positive and negative attitudinal components.

Developing Data-Literate Communicators: A Descriptive Analytics Course Model

Matthew Swinarski, Sara Luttfring, Dani McCauley

Penn State Behrend, Erie, USA

Abstract

This study describes the design of DA 201W, an undergraduate Descriptive Analytics course that integrates technical analysis with effective communication of insights. The course structure draws from the Financial Times Visual Vocabulary Guide, which was developed to help journalists, analysts, and data communicators select appropriate visuals for presenting different types of data analyses. The guide categorizes visualizations based on key analytical concepts, including Distribution, Deviation, Part-to-Whole, Ranking, Magnitude, Correlation, Spatial, Flow, and Change Over Time. Lessons in the course are structured around these categories and are further divided into three primary sections: Statistics, Tables, and Graphs and Charts. These sections introduce students to essential analytical methods, guiding them in understanding their purpose, calculation, and interpretation. Students apply these concepts through Excel-based exercises using a core dataset, supplemented by additional data sources for practice and assignments.

A key feature of the course is its emphasis on writing and communication skills. Through a series of developmental assignments, students learn to translate technical analysis into clear, meaningful narratives. The course culminates in a final project where students develop an 800–1000 word paper that integrates their analytical work into a compelling narrative.

By balancing technical proficiency with communication, the course equips students with the skills needed to analyze data effectively and communicate insights clearly—an essential capability in today's data-driven workforce.

International Trade and The Environment: Sticks, Carrots and Terms of Trade.

Val Vlad

Penn State, Behrend College., Erie, USA

Abstract

A North and South two-country model of international trade is utilized to examine the effects of trade and capital mobility on the environment. Both countries produce a range of goods that are traded internationally. In this model, the South handles pollution abatement through efforts by both the public sector and private industries. Capital flows from the North to the South, while pollution migrates in the opposite direction, from the South to the North. Each country employs a set of policy instruments to manage pollution levels and their transboundary movement. The North implements a tariff (stick) and provides foreign aid (carrot) to incentivize pollution abatement in the South. Conversely, the South uses public funds for abatement (carrots) and imposes an emission tax (stick). The Nash equilibrium levels of these policy instruments are analyzed under two scenarios: (1) with exogenous prices, and (2) with endogenous prices. The study also explores how these policies impact the relationship between terms of trade (TOT) and welfare in both economies. One key finding is that, with trade, the policies are likely to become more stringent.

JEL Classifications: F35, Q28, H41

The Impact of Team Formation on Overfitting in Crowdsourced Machine Learning Solutions

Huiyan Chen[1], Brian Lee[2], Jan Stallaert[3]

[1]Penn State Behrend, Erie, USA. [2]Penn State University, University Park, USA. [3]University of Connecticut, Storrs, USA

Abstract

While machine learning solutions are commonly crowdsourced through contest platforms, their effectiveness remains understudied. Overfitting is one of the fundamental challenges in machine learning, particularly in settings where models can be iteratively refined based on immediate feedback. While technical solutions to overfitting have been extensively studied, overfitting remains prevalent in practice. This work takes a behavioral approach to investigate when overfitting is most likely to occur in the context of crowdsourced machine learning contests, where it is common for data scientists to form teams and collaboratively develop solutions. We aim to understand whether and how their collaboration influences overfitting.

Analyzing a large dataset from Kaggle, one of the largest online machine learning communities, we find that the overfitting of machine learning solutions is significantly influenced by team formation in heterogenous ways. Specifically, solutions from inexperienced teams overfit more than those from individuals, while experienced teams overfit less. These results suggest that team formation is beneficial only when members have sufficient experience; otherwise, it may lead to negative outcomes. Moreover, contest designs play a significant role in the impact of team formation. We show that when the contest enabling knowledge sharing and involving less competitors are more likely to help teams avoid producing overfitted solutions. This suggests that crowdsourcing platforms can mitigate overfitting by fostering a more collaborative and less pressured environment.

Our findings extend the current understanding of the socio-technical aspects of model development in machine learning and provide practical approaches for improving model prediction through crowdsourcing.

User Assessment of Security Threats: Initial Findings

Yi Guo, Young K. Ro

University of Michigan -- Dearborn, Dearborn, USA

Abstract

Computer security breaches are increasingly common, making it crucial to understand how end users assess and respond to such threats. Users typically evaluate threats based on severity, vulnerability, and avoidability, which guide their decisions. Preventative actions include installing antivirus software, updating systems, and activating firewalls. Threat assessments are influenced by factors like potential financial loss and psychological distance. This study investigates how the seriousness and perceived proximity of security threats shape users' assessments and behaviors. With this research, we know how different factors of threats are viewed and better understand the user decision process, which may help in training to better increase its effectiveness. Data were collected via a controlled experiment using a set of security threat scenarios. By manipulating threat severity and psychological distance we examined how these factors, along with individual characteristics, influence perceptions of severity, vulnerability and avoidability, and determine the preventative measures users are likely to take. Over 80 participants were involved in the study. Measures showed good reliability. Findings will be reported during the conference.

Optimization Model Constraint Considerations Under Uncertainty

Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

We provide and examine various optimization model constraint considerations that could be considered best practices to handle predictive model uncertainty. The motivation of this research is to help practitioners have a minimum set of data-supported constraints to examine when integrating prediction with optimization avoid to unintended consequences. Most academic courses will have standards for developing and evaluating predictive models, as well as optimization models, but there is a lack of best practices when integrating the two domains.

The Role of Information Usage Between Information Sharing and Performance.

Abdurrezzak Sener[1], Mutlu Yuksel Avcilar[2], Mehmet Barut[3], M. Bayram Yildirim[3]

[1]Penn State University, Pittsburgh, USA. [2]Osmaniye Korkut Ata Üniversitesi, Osmaniye, Turkey.
[3]Wichita State University, Wichita, USA

Abstract

The purpose of this study is to provide a holistic approach of extending dyadic relationships between information sharing and operational performance as well as between operational and organizational performances into an extended relationship between ultimate suppliers and ultimate customers from suppliers' point of view. We introduce mediating impacts of information usage between information sharing and operational performance, and perceived logistics performance between operational and organizational performances. The empirical data obtained from 209 retailers via surveys reveal that the integrated relationships are valid and significant. The mediating impacts of both information usage and perceived performance of logistics play a significant role in performance relationships and are paramount to the success of organizations.

Impact of cocreation under supply chain encroachment setting

Abhishek Roy[1], Piyal Sarkar[2], Shubham Gupta[3]

[1]Temple University, Philadelphia, USA. [2]Penn State University, Harrisburg, USA. [3]New Jersey Institute of Technology, New Jersey, USA

Abstract

The increasing complexity of supply chains has led firms to redesign supply chain strategies based on the collaborative value created through cocreation with other supply chain members. Cocreation is an economic strategy that binds supply chain members to jointly improve the output. Previous literature has investigated the impact of cocreation under different settings, but to the best of our knowledge, the impact of cocreation under supply chain encroachment setting has not been investigated so far. Through this research, we try to address this major gap in the literature. Supplier encroachment mitigates double marginalization and secure Pareto improvements for both parties. Our research investigates on how cocreation can add value under supplier encroachment environment for a two party supply chain. The study provides several managerial insights by integrating the cocreation and encroachment strategies.

Trade credit contracts under weather risk

Piyal Sarkar[1], Mohamed Wahab Mohamed Ismail[2], Liping Fang[2]

[1]Pennsylvania State University, Harrisburg, USA. [2]Toronto Metropolitan University, Toronto, Canada

Abstract

Trade credit is one of the widely used financing schemes today. Trade credit acts as an incentive to encourage buyers to increase ordering quantities as it allows delayed payments. The coordination mechanism under trade credit has become a hot research topic. However, no study to date has explored the role of trade credit financing in the case of supply chains under weather risk, even though the profitability of many industries is impacted due to weather. Results show that the trade credit contract with a weather rebate and financial hedging performs the best among the three contracts. This is a new idea in the area of trade credit financing. Our study contributes to the understanding of trade credit contracts and their implementations to withstand weather and credit risks in supply chains. It explores how weather rebate and financial hedging strategies can be implemented to enhance the performance of trade credit contracts under weather risk. Industries can use the contracts to coordinate supply chains under weather risk. Our analysis further shows that the designed trade credit contracts can be implemented under any market risks, which can be correlated or independent to the weather risk. This further generalizes the applications of the contracts under different risk scenarios.

The full paper is not published in the conference proceedings at the authors' request.

The Underground Ballroom Challenge: Minecraft-based Case Exercises for On-line Project Management Classes

Natalie Simpson

University at Buffalo, Buffalo, USA

Abstract

Asynchronous online classes do not as readily support team-based learning when compared to their in-person counterparts, due to the loss of a physical classroom as a shared problem-solving space. To overcome this challenge in a core MBA course in project management, the best-selling video game Minecraft provides a virtual community space within which teams of students apply project management processes to a suite of assignments. These assignments, collectively known as the "Underground Ballroom Challenge," require application of class-related techniques ranging from risk assessment to work breakdown structure to adaptation and change management. In this presentation we will share the component assignments, the particulars of harnessing Minecraft for instructional purposes, and the practices developed to muster large numbers of asynchronous online students into effective teams inside of six-week semesters.

The Impacts of the Social Capital and Mobile Technology Usage on Knowledge Workers' Work-Family Role Integration

Xiaodong Deng[1], Mengxiao Li[2]

[1]Oakland University, Rochester, USA. [2]Jiangxi University of Finance and Economics, Nanchang, China

Abstract

The use of mobile technology (MT) enables knowledge workers to work at flexible time and location even after their work hours. This practice could potentially help improve knowledge workers' productivity and maintain a good balance between their work and family. However, literature has suggested that this practice also creates knowledge workers' workloads previously not existed or infeasible. These new workloads may demand additional efforts or even new types of skills or resources (Duxbury and Smart, 2011; Doherty, 2016) and further blur their work-family boundary. This study presents a research model to explore how knowledge workers could use mobile technology and their social capitals to integrate their roles of work and family. Social capital is the goodwill embedded in an individual's social network and the potential resources available to the individual (Adler and Kwon, 2002). In the context of mobile technology usage, the ways that a knowledge worker uses the mobile technology will play a significant role in facilitating his/her efforts of coordinating the integration of work and family activities. Hypotheses are developed to elaborate the relationships among the MT usage patterns, knowledge workers' social capitals, and their levels of work-family role integration. The theoretical and practical implications are discussed.

Circular Economy Strategies in U.S. Airlines: Analysis of Full-Service and Low-Cost Carriers

Ata Karbasi

Valparaiso University, Valparaiso, USA

Abstract

As the airline industry faces increasing environmental pressures, circular economy principles offer a viable pathway toward sustainability. This study examines how U.S. airlines integrate circular economy strategies into their operations, with a particular focus on the differences between full-service and low-cost carriers. By analyzing sustainability reports from U.S. airlines this study identifies trends in sustainability commitments, evaluates the effectiveness of these strategies in lowering carbon footprints, and assesses their financial and operational implications. Findings indicate notable variations in circular economy adoption across different airline business models, with some carriers leading in sustainable aviation fuel investment and waste reduction while others prioritize cost-driven efficiency measures. This research offers actionable insights for industry leaders, policymakers, and sustainability advocates seeking to drive meaningful change in aviation's environmental impact.

Integrating Leadership Learning into Business School Curricula

Jinjing Zhu, Todd Palmer

St. Bonaventure University, St. Bonaventure, USA

Abstract

As the demand for effective leadership in the business world continues to grow, business schools must adapt their curricula to equip their students with the necessary leadership skills. This research explores the integration of leadership learning into business school curricula, examining the benefits, challenges, and strategies for embedding leadership development within academic programs. Using the practices from our senior core course in strategic management to empower student leaders and develop their leadership skills as a case study, this research highlights the importance of moving beyond traditional management-focused courses to include experiential learning opportunities that foster leadership competences. Student leaders were given a major task (managing a company in trouble) and smaller tasks/tools (eight workbooks created by Google Docs) to lead the strategic management process step by step. Each week, student leaders had to assign team members to work on certain parts of the workbooks independently, lead discussions to finish the parts requiring team cooperation and consensus, and ensure that the work was completed on time. Feedback was provided to student leaders on a timely basis. Our practice received positive feedback from students.

A Data-driven Learning Approach to Diffusion Prediction in Social Networks

Wenjun Wang

University of Arkansas at Little Rock, Little Rock, USA

Abstract

Influence diffusion in a social network refers to the expected number of social entities who will be activated to adopt a product or an opinion, given a set of initial adopters who trigger the cascade of further adoptions through word-of-mouth-based information propagation in the network. In our previous study, we developed a Multiple-path Asynchronous Threshold (MAT) model for influence diffusion, in which we quantify influence and track its propagation and aggregation. Our MAT model captures not only direct influence from neighboring influencers but also indirect influence passed along by messengers. Moreover, our MAT model takes into consideration influence attenuation along diffusion paths, temporal influence decay, network activeness, and individual diffusion dynamics. In this paper, we propose an effective data-driven learning approach to influence-diffusion prediction based on the MAT model and datasets that contain both network connectivity and action traces of a large number of episodes. Specifically, we partition the episodes into a training set and a test set. We use the training set and maximum likelihood estimation (MLE) to learn the model parameters, including the temporal influence decay rate and network activeness. Then we use the learned parameter values in the MAT model to predict the influence diffusion for the episodes in the test set, and evaluate the prediction accuracy based on the actual influence spread. Our work is an important step toward a more realistic diffusion model, and provides insights and implications for diffusion research and marketing practice.

Using AI to Reimagine Case-Based Learning: A Collaborative, Student-Driven Approach

Scott Stroupe

Penn State Univ., Erie, USA

Abstract

Case-based learning is central to business education; however, traditional methods can lead to lopsided participation and uneven analysis. This presentation introduces a student-driven approach that replaces static cases with real-time case development, drawing on current business events and student interests.

During a two-week cycle, students develop a case through structured independent research, peer collaboration, and in-class discussions that apply key course concepts. By using a variety of AI tools, particularly Notebook LM, we streamline data collection and synthesis, enabling students to concentrate their time and energy on critical analysis. The objective is to increase student engagement and create greater analytic rigor among all students by giving them increased agency and ownership over the cases they examine. In practice, this approach also improves their familiarity with essential strategic analysis tools and processes.

The presentation will provide practical insights into adopting this approach, including obstacles and pitfalls encountered before and during implementation.

Required Skills of Middle-Level Managers in China

Tongwei Song[1], Lai Liu[2], Kai Koong[3]

[1]China Mobile Information Technology, Xi'an, China. [2]Univerity of Texas Rio Grande Valley, Edinburg, USA. [3]Tuskegee University, Tuskegee, USA

Abstract

According to the literature, middle-level managers represented the critical link between top management and lower-level supervisor and corporate staff. Middle-level managers are responsible for the rendering of decisions that affect the short-term performance of the firm. Typically, these managers are involved with the making of tactical decisions. They must be able to coordinate projects assigned to supervisors and can follow directives and implement strategies laid out by their executives. Based on the data extracted from two of the most popular job portals in China, this study examined the traits and skills expected of middle-level managers. Specifically, this study analyzed the job descriptions and their requirements and attempted to develop a model consisting of the various skill sets and their common terminologies that employers and employees can use as a common language for job opportunities in the middle management labor market. This study found that the requirements relating to experience and academic preparations were different. A total of seventeen attributes were identified. First, the most important set of 7 critical skills are Team building, Commitment to Excellence, Collaboration, Vision, Logical Thinking, Problem Solving, and Learning and Innovation. Second, the next set of important skills includes Ability to Handle Pressure, Team Leadership, Responsibility, Personal Character, Dedication and Hard Work, Analysis and Decision Making, and Teamwork. Finally, the last set of three skills consists of Familiarity with Work Processes, Related Professional Background, and Communication Skills. These characteristics can easily be fitted into a three-tier pyramid model.

Resilient supply chains considering recent technological innovations

Raghavan Srinivasan[1], Vinay Gonela[2]

[1]Ball State University, Muncie, USA. [2]Texas A&M University – Central Texas, Killeen, USA

Abstract

Supply chain disruptions, though often unexpected, can vary in severity. While minor disruptions may be relieved through strategies that restore normalcy, more extreme events can lead to significant penalties, including bankruptcies. These disruptions can also affect the public by causing shortages of goods or driving up prices, with food and medicine being particularly critical. Therefore, it is vital for organizations to be prepared for such events by building resilience within their supply chains. Although implementing resilient strategies can incur additional operational costs, these costs are often inevitable. This study examines supply chain mitigation strategies, focusing on the potential of emerging technologies such as Blockchain (BC), Internet of Things (IoT), Analytics, Artificial Intelligence (AI), and Machine Learning (ML) in the context of the next industrial revolution. We review literature from the past decade, summarize key insights related to various supply chain processes, and conclude with a discussion on future research paths aimed at developing sustainable and resilient supply chains.

Personalization in Omnichannel Supply Chain Networks: Impact on Consumer Understanding

Shubham Singh[1], Zaid Masri[1], Sushil Sharma[2]

[1]Purdue University Fort Wayne, Fort Wayne, USA. [2]Texas A&M Texarkana, Texarkana, USA

Abstract

This study explores the role of personalization within omnichannel supply chain networks and its impact on consumer understanding of product availability, delivery expectations, and overall satisfaction. While existing research often emphasizes marketing aspects of personalization, its implications for supply chain operations and consumer perceptions of operational effectiveness are underexplored. By examining consumer experiences across multiple channels, this research aims to understand how personalized interactions influence consumer perceptions related to inventory reliability, product accessibility, and service dependability.

Utilizing survey-based research, the study investigates how consumers perceive personalized experiences in relation to supply chain responsiveness and inventory management. The research addresses the dual implications of personalization, recognizing it as a strategic approach that enhances consumer satisfaction but also increases operational expectations and complexity. This study underscores the necessity of aligning personalization initiatives with operational capabilities to ensure positive consumer perceptions and effective supply chain management.

This presentation contributes to the supply chain management literature by highlighting the critical intersection between consumer perception and operational strategy in personalized omnichannel environments. The insights provided can help supply chain managers better understand consumer expectations and optimize personalization strategies to achieve operational efficiency and enhanced consumer satisfaction.

Supply Chain Disruptions and Economic Growth

Sudip Ghosh

Penn State Berks, Reading, USA

Abstract

There has been a plethora of literature showing the effects of macro-economic shocks on business cycle analysis.

We develop a new argument for business cycle model prediction. We use changes in inventory-to-shipment ratio as a new determinant to predict business cycle ratio using data from 1975 to 2019 for the U.S. economy. Our model shows that two-thirds of the cyclical variability of inventory investment is strongly pro-cyclical, and production is more volatile than sales. The co-movement between inventory investment and final sales is often interpreted as evidence that inventories amplify aggregate fluctuations. Our model contradicts this view. Despite the positive correlation between sales and inventory investment, we find that inventory accumulation has minimal consequence for the cyclical variability of GDP.

There has been a plethora of literature showing the effects of macro-economic shocks on the business cycle analysis. The fluctuations in macro activities could be due to large or small shocks (e.g., Friedman and Romer 1963; Romer and Romer 1989). Credit market conditions and underlying frictions should be examined carefully in determining the strength and scope of business cycle (Bernanke, & Gertler 1995). Bernanke, Gertler, and Gilchrist (1999) showed asymmetric information and agency costs in lending relationship, led to the "financial accelerator" model where credit markets create macroeconomic shocks. Our paper another dimension to the seminal debate of macro-economic shocks and prediction of economic slowdown.

Challenges in Creating a Sustainable Embedded Mentorship Experience in a Business Curriculum.

Scott Stroupe

Penn State Univ., Erie, USA

Abstract

Last year, an embedded mentorship program was introduced that paired students with external professionals as part of a core business class assignment. The goal was to leverage an underutilized resource—a large network of willing mentors—to make mentorship a visible and integral part of the student experience. By increasing awareness, reducing barriers to access, and offering incentives, we saw a dramatic rise in participation. One year in, the results remain positive, yet challenges have emerged regarding the sustainability and scalability of this model. Managing the program alongside teaching is time-intensive, and maintaining a strong mentor network requires continuous effort. Legitimate questions remain about the long-term practicality of this approach.

This presentation builds on our initial findings by sharing qualitative insights from students and mentors and highlighting key takeaways for improving engagement and effectiveness. We will explore the limitations of scalability and provide best practices for embedding mentorship into the curriculum, rather than treating it as an optional extracurricular activity. Strategies such as peer mentoring and student-alumni involvement will be examined as potential solutions for sustaining the program alongside regular coursework. This session contributes to the broader discussion on mentorship in higher education, offering practical strategies for long-term integration into business courses.

4

Increasing Role for Business Decision Making in a Time of Macroeconomic Challenges.

Ionut Vlad[1], Val Vlad[2]

[1]Indiana University of Pennsylvania, Indiana, USA. [2]Penn State, The Behrend College., Erie, USA

Abstract

The US faces a set of macroeconomic challenges that have become more a new normal than an incidental, short-run, downturn. Many western economies are facing more dire circumstances. (German economy has been in recession in the last two years.) The forecasts are discouraging for all economies. The global economy is settling at a low growth rate. In this project, we analyze the data on three main indicators of economic performance, from 1970 until 2024. The data suggests that the U.S economy faces three major challenges: slow economic growth, inflation, and high levels of indebtedness. This critical triad makes policy making an almost impossible decision-making process inoperative and inefficient, leaving the economy vulnerable, at high risk for recession and even stagflation.

In a time when macroeconomic policies have little to offer, a significant balancing, positive, impact can come from cutting edge decision making. In a future where supply chains might face major headwinds generated by underperforming economies and other exogenous factors, analytics-based decision making may be a major direction to increase the ingenuity of supply chain processes, to increase their efficiency and offer cost cutting strategies to maintain the trade fluency (that might be adversely impacted by slow growth, tariffs and geopolitical divergences.)

Digital Transformations and Its implications on Knowledge Ecosystems

Yohannes Haile

Penn State Behrend, Erie, USA

Abstract

Digital Transformation [DT] has significant implications on innovation, entrepreneurship, and solutions development disrupting conventional business and operating models of enterprises. Digital business and operating models are transforming the pathways value is created, captured, and delivered encompassing modifying/sensing new opportunities, developing unique value propositions, value flow, and monetizing schemes. The relationships between innovation, entrepreneurship, and solutions development [outcome markers] and digital transformation [the antecedent] are mediated by learning-to-learn, creativity, and problem solving, which are the major constituents of the knowledge ecosystem. As such DT has significant implications on knowledge ecosystems necessitating recalibrations on learning engagements, goal setting, feedback, and reward mechanisms as learning is this new paradigm requires more experiential learning, frequent pivots, balancing between short & long-term goals, differentiated rewards, systems thinking/leadership, and flourishing under uncertainty.

Designing Cost-Effective Shipping Routes with Time Windows

Xi Jiang[1], Xihui Chen[2], Xiuling Hei[3], Sanjay Kumar[4]

[1]School of Business, Shanghai Normal University Tianhua College, Shanghai, China. [2]Hangzhou International Urbanology Research Center, Hangzhou, China. [3]Ningxia University, Yinchuan, China. [4]Central Michigan University, Mt. Pleasant, USA

Abstract

Current research on liner shipping network design often overlooks strict port time window constraints, making schedules impractical. This paper introduces a MILP-based method for designing routes that accommodate multiple weekly port time windows. Our model optimizes port call sequences, schedules, and sailing speeds to reduce overall costs. Numerical experiments confirm its solvability and reveal significant cost savings versus initial sequences. Findings indicate that bunker prices, port time windows, and shipping demand affect the optimal sequence.

Can Lessons Learned from the Data Science/Business Analytics Program Divide Translate to the Teaching of Artificial Intelligence?

Richard Grenci

John Carroll University, University Heights, OH, USA

Abstract

Undergraduate business analytics minor programs show distinct differences from their data science counterparts. Such programs differ not only in focus and content, but also in terms of accessibility to business majors who desire a deeper dive into data analysis. Lessons learned from this divide can offer insights into designing undergraduate artificial intelligence (AI) minor programs specifically targeted at business majors. As AI minor programs are being rolled out, differences in approaches are quickly emerging. AI programs are being offered out of departments as distinct as philosophy and computer science; and just like with business analytics programs, some AI minors are being created with the business major in mind. The differences between programs will be discussed and will be used as a lens for analyzing the strategies and the pros and cons of the varied approaches.

Using Students as Guest Speakers in an MBA Supply Chain Management Course: A Study on Engagement and Impact

Bulent Erenay, Dekuwmini Mornah

Northern Kentucky University, Highland Heights, USA

Abstract

Guest speaker sessions are widely used in business education to provide students with industry insights and practical applications of theoretical concepts. This study examines the impact of including students with supply chain management (SCM) experience as guest speakers in their own MBA course. A survey was conducted to assess the value of these sessions, their relevance to course content, and their professional benefits.

Results show that 39% of respondents participated as guest speakers, while 86% attended the sessions live or in a hybrid format. Networking was a key benefit, as students valued the opportunity to connect with peers and learn about different career paths. Many respondents said the sessions helped them see the connections between SCM concepts and real-world business practices. The LinkedIn networking component also provided a way for students to expand their professional connections.

These findings suggest that student guest speakers create an interactive learning environment and strengthen the link between coursework and industry practice. By drawing on the experiences of MBA students, this approach adds depth to class discussions and encourages professional connections. Future research could examine how this type of engagement influences students' careers over time.

Procurement Option Contracts to Mitigate Spot Price Volatility and Breach Risks

Ozgun Demirag[1], Jie Wang[2], Yong-Hong Kuo[2]

[1]Pennsylvania State University, Erie, USA. [2]The University of Hong Kong, Hong Kong, China

Abstract

We examine procurement option contracts in the presence of potential breaches caused by price fluctuations. Through a three-stage game-theoretic framework, we explore how a supplier and manufacturer navigate spot market price volatility and contract breaches. Our findings indicate that contract effectiveness diminishes as demand and spot price volatility increase but improves with higher supplier search costs in the spot market. Among different breach mechanisms, renegotiation proves to be the most effective in mitigating contract inefficiencies. Suppliers capture a greater share of supply chain profits when demand and price fluctuations are moderate and when search costs are elevated. Our study highlights the critical role of well-structured contracts, particularly those incorporating renegotiation, in sustaining supply chain profitability under breach risk.

Empowering Older Adults with Personal Cyber Protection Through Microlearning

Kathleen Noce

Penn State Erie, Erie, USA

Abstract

As older adults increasingly engage with digital technology, they become more vulnerable to cyber threats, including scams, identity theft, and online fraud. However, traditional training methods can be overwhelming or ineffective for this demographic (Ahmad et al., 20220). This study aims to develop a microlearning curriculum designed to equip older adults with essential cybersecurity knowledge and practical skills in a digestible, accessible format. By leveraging short, focused learning modules, the curriculum will enhance digital safety awareness and build confidence in navigating online environments securely. The research will explore best practices in microlearning, assess the specific cybersecurity challenges faced by older adults, and evaluate the effectiveness of this approach in improving knowledge retention and behavioral change which have been found to be affected by numerous factors (Kebede et al., 2022). The findings will contribute to the development of targeted educational strategies that empower older adults to protect themselves in an increasingly vulnerable digital world. By incorporating principles of cognitive load theory and experiential learning, this curriculum ensures that senior learners can confidently recognize online threats, safeguard personal information, and navigate the digital world securely. The findings from this research will contribute to best practices in cybersecurity education for older adults and support lifelong digital literacy.

Enhancing Student Engagement in Business Statistics: Evaluating the Impact of an Escape Room on ANOVA Learning

Elizabeth McQuillen

Wayne State University, Detroit, USA

Abstract

Maintaining student engagement in statistics courses is a persistent challenge in business education. Traditional assessments often prioritize procedural accuracy over active learning, which can limit student enthusiasm and deep understanding of statistical concepts. This study examines the impact of a team-based Escape Room activity as an alternative assessment method for ANOVA, comparing it to a traditional individual assignment. Unlike prior research comparing separate cohorts, this study employs a within-subjects design, assessing the same students' performance and perceptions across two learning formats.

Undergraduate business students first completed a traditional ANOVA assignment, working either together or independently to compute and interpret an ANOVA table. The following week, they participated in Escape the ANOVA Room, a structured, collaborative game designed to reinforce statistical concepts through teamwork and sequential problem-solving. Performance data from both assignments were analyzed using paired t-tests, and a post-activity survey captured student attitudes toward engagement, collaboration, and AI-assisted game design.

Preliminary results suggest that the Escape Room format increases engagement, with students reporting greater motivation and confidence in applying ANOVA concepts. This study contributes to the growing body of research on active learning strategies in business statistics, providing insights into how interactive assessments can enhance participation and conceptual retention. Implications for innovative instructional design and the integration of AI in classroom activities will be discussed.

Vision, Mission, and Values...is in the wrong order.

Christopher Harben

Penn State Behrend, Erie, USA

Abstract

In my word with companies on strategic planning, I have come to recognize that ineffective strategy often begins with an an issue with the concepts of Mission, Vision, and Values. These items are crtitical to the foundation of a strategy. Yet, they are in the wrong order. The things an organization does, the choices it makes, the paths it follows, the people it hires, all SHOULD flow from the Values the organization hold. In this presentation, I will argue that the foundational element of strategic planning is the values that the organization holds. There can be no Mission or Vision without first understanding the Values of the organization. Yet, the literature (as well as corporate practice) is to relegate Values to the thrid position.

By following my suggestion, strategic goals and tasks can be more easily aligned with the organization throughout all levels. The Values helps to answer the question "why are we doing this?" And the answer is much more than "because this is how we make money."

A Data Breach's Impact on Marketing Capabilities

Benita Bowden, Amit Agarwal, Ben Lee

PSU Behrend, Erie, USA

Abstract

Managers and academics have long considered a firm's marketing capabilities – reflecting how well a firm utilizes its marketing resources to achieve targeted outcome – to be a sustainable competitive advantage contributing to firm performance. Therefore, it is not surprising that Chief Marketing Officer (CMO) surveys regularly report that enhancing marketing capabilities is one of the top priorities of firms. Nonetheless, the question of "how to enhance marketing capabilities" remains surprisingly unanswered in the literature, leaving firms struggling to effectively enhance these critical capabilities. In this paper, relying on the marketing-finance and strategic leadership literatures, we examine two research questions: (1) Does a data breach incident harm a firm's marketing capabilities? (2) If so, what factors alleviate or exacerbate such negative influence? We plan to collect a large representative panel dataset by combining multiple secondary databases. We plan to analyze the dataset using a panel data econometric approach. Our project would extend the extant interdisciplinary data breach literature and marketing capability literature. Furthermore, our project would help managers accomplish their top strategic priority: managing marketing capabilities. Lastly, our project will address the societal implications of cybersecurity risks by highlighting the negative consequences these risks pose and motivating firms to mitigate them.

Understanding Self-Congruence and Motivation in Travel Experience Sharing: A Generational Perspective

Parichat Kheawcha-um[1], Lanlung Chiang[2], Hsunchi Chu[1]

[1]College of Management, Yuan Ze University, Taoyuan, Taiwan. [2]Department of Business Administration, National Taiwan University of Science and Technology, Taipei, Taiwan

Abstract

In the digital era, social media serves as a powerful platform for tourists to share their travel experiences, influencing potential visitors and destination branding. While past research highlights the impact of user-generated content (UGC), not all tourists actively share their experiences online. This study examines the role of self-congruence and expectation confirmation in shaping tourists' intention to share travel experiences on social media. Using a survey of 413 Generation Y and Z tourists, regression analysis reveals significant generational differences. For Generation Y, both intrinsic and extrinsic motivation moderate the relationship between satisfaction and sharing intention, while for Generation Z, neither motivation type has a significant effect. These findings suggest that marketing strategies should be adapted to generational preferences, leveraging external incentives for Generation Y and alternative engagement strategies for Generation Z. Theoretical and practical implications are discussed.

Predicting Ransomware Attacks

Yaman Roumani

Oakland University, Rochester, USA

Abstract

Ransomware attacks pose a major cybersecurity threat, causing financial and operational damage. Despite their impact, predictive measures remain underdeveloped. This study addresses the gap by using time series analysis to forecast ransomware incidents. Analyzing 1,880 reported cases, the data is decomposed into trend, seasonal, and residual components to build a predictive model. Results show effective forecasting performance. The findings highlight the value of time series forecasting in cybersecurity, offering organizations actionable intelligence to enhance preparedness and situational awareness against ransomware threats.

Generative artistry: Integrating Fine Art Principles into Small Business Marketing

Monali Jathar[1], Anup Nair[2], Vallari Chandna[2]

[1]Freelance Consultant, Green Bay, USA. [2]University of Wisconsin- Green Bay, Green Bay, USA

Abstract

Our conceptual paper explores the potential of fine art principles and Generative AI (Gen-AI) to act synergistically in enhancing small business branding and marketing. In today's visually saturated digital environment, small businesses must differentiate themselves to capture and retain customer attention. We propose that the strategic integration of fine art concepts—composition, color theory, texture, and narrative—within AI-generated visual content can create a powerful emotional resonance, fostering deeper customer connections.

Given the affordability and easy access to Gen-AI tools, these can greatly empower small businesses This access is enabling the creation of tailored and customized visual content for various platforms, including social media. By leveraging Gen-AI, businesses can generate diverse and personalized visuals that align with their brand identity and resonate with specific target demographics. This allows for experimentation with various artistic styles, ensuring that the visuals lead to the desired emotional responses and tell compelling brand stories.

We conceptualize that integration of fine art principles while utilizing easy to use Gen-AI tools, can impact the desired outcomes of customer satisfaction and brand loyalty. By creating visually engaging content, particularly for social media platforms, small businesses can enhance their brand and create memorable customer experiences. These translate into tangible outcomes, including increased customer satisfaction, positive e-word-of-mouth, stronger brand loyalty, and higher levels of social media engagement. This conceptual framework demonstrates how small businesses can leverage fundamental artistic priciples alongside the latest innovations to build authentic customer relationships and cultivate lasting brand value in the digital age.

Big Five Personality Traits and Podcast Visuals

Monali Jathar[1], Anup Nair[2], Vallari Chandna[2]

[1]Freelance Consultant, Green Bay, USA. [2]University of Wisconsin- Green Bay, Green Bay, USA

Abstract

In this paper we explore the potential connection between the Big Five personality traits and the visual aesthetics of podcast cover art. We propose that specific design elements attract listeners with corresponding personality profiles. In the competitive podcast landscape, cover art serves as a crucial first impression, acting as a visual gateway to a podcast's content and tone.

We propose that consumers are driven by their inherent personality traits i.e. the Big 5 traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism when it comes to selection of podcasts. They are drawn to distinct visual styles that resonate with their psychological makeup.

We conceptualize how various design elements, such as color palettes, typography, imagery, and composition, align with these personality dimensions. For example, individuals high in Openness will likely gravitate towards abstract, unconventional designs, while those high in Conscientiousness might prefer structured, minimalist aesthetics. We propose that these visual cues act as non-verbal signals, effectively communicating the podcast's essence and attracting listeners who share similar personality traits. This paper provides a theoretical framework for understanding how personality influences visual preferences in podcast marketing, suggesting that strategic design choices can enhance audience targeting and foster stronger connections between creators and their listeners. By recognizing the personality traits as they relate to podcast covers i.e. visual attraction elements as they relate to personality factors, podcast creators can optimize their cover art to effectively reach their desired audience.

Building a Relational Model of Online Graduate Education

Ana Rosado Feger, Amy Taylor-Bianco

Ohio University, Athens, USA

Abstract

Business colleges are facing enrollment challenges as the traditional student population shrinks. At the same time, the dynamic employment environment is driving more working adults into re-credentialing or pursuing a different career path. Many of these students are seeking online degrees that fit into their schedules. However, students often struggle to complete online degrees, and attrition rates are high. We use our experience with an online graduate degree to develop a model of relational graduate online education, from curricular design to faculty selection and student recruitment. We use student feedback to demonstrate the effectiveness of the program.

Why Does Your Employment History Really Matter? An Application of the Data Science Literature on Provenance to Employee Selection

Mark Brown, Gregory Jetton

Bradley University, Peoria, USA

Abstract

The human resource management literature has long been concerned with understanding the factors that determine effective employee selection, One area that is particularly interesting today, in the era of artificial intelligence, is the usefulness of employees' past employment histories in the selection process. Specifically, are applicants who have worked for some organizations better applicants than those who have not worked for these organizations? One widely explored area of research, which may be useful in understanding this phenomenon, is the literature on provenance. Provenance speaks to both the history of something and the relevance of that history. The notion of provenance is present in the art history, geology, and data science literature. In this manuscript, we will in particular focus on the data science literature and its usefulness in employee selection. Herschel et al. (2017) provides an overview of how data scientists view provenance, its types and applications. They conclude that the main question to be focused on in data provenance are: "What for?", What form?", and "What from?" Just as these questions apply to the field of data science, they may also apply to other field as well. Moreover, while analogies may hold for better or worse across academic disciplines, these questions are applicable to the selection process and this is the focus of this manuscript.

Synthetic Data, Real Impact: Using AI to Generate Useful Datasets for Industry and Academia

Cody Baldwin

University of Wisconsin-Madison, Madison, USA

Abstract

Analytics is driven by data. It's not possible without it. However, within industry and academia, we often find that data is scarce, inaccessible, sensitive, or imbalanced – creating a roadblock to our analytics progress. This paper explores how artificial intelligence (AI) can be used to create realistic synthetic datasets that can help address these data limitations.

Predicting Hospital Length-of-Stay for Adolescents with Multimorbidity Patterns Adjusting for Hospital-Acquired Conditions Risk Profile, Individual Demographic and Neighborhood Structural Disparity Factors

Ajit Appari

Northeastern University, Boston, USA

Abstract

Multimorbidity is a major concern worldwide because of its undesirable effects on the individuals' health and health care needs, and the challenges it poses to healthcare providers in effectively caring for patients with multiple, interlinked diseases and conditions especially during hospitalization. Notwithstanding the rich literature on multimorbidity among adults, understanding of multimorbidity in adolescent children and its effect on health care is still in the nascent stage (van den Akker et al., 2022). More importantly, machine learning modeling is becoming a necessity for delivering effective and equitable health care services while ensuring better health care outcomes for patients.

This study using all-payers administrative data on 100,240 adolescent hospitalizations (patient age: 10-14 years= 48,735; 15-17 years=53,665) during 2015Q4-2016Q4 across 447 Texas hospitals evaluates Random Forest (RF) and Support Vector Machine models for predicting length of stay (LoS) at admission [short (LoS<3days), moderate (LoS =3-5days), prolonged (LoS >5days)] for patients with multimorbidity pattern conditional on their source and type of admission, major diagnostic category, hospital-acquired conditions (HACs) risk profile (RF classifier for HACs count: none, 1-2; 3-5; 6-10; >10), multimorbidity patterns burden, individual disparity factors (age, race/ethnicity, insurance), and neighborhood structural disparity factors (zip-code tabulation area level socioeconomic disadvantage, affluence, and ethnic/immigrant concentration) controlling for provider and admission factors. Patients with multiple chronic conditions were segmented for 15 multimorbidity patterns using Bernoulli Mixture clustering.

Predicting Metastatic Cancer Diagnosis: A Data-Driven Approach to Health Equity

Ilyas Ustun

DePaul University, Chicago, USA

Abstract

This research explores predictive modeling for metastatic cancer diagnosis using a real-world dataset containing patient demographics, diagnosis, treatment options, insurance details, and socio-economic factors. The dataset is enriched with zip code-level toxicology data from NASA/Columbia University. The challenge involves predicting whether patients receive a metastatic cancer diagnosis within 90 days of screening. By leveraging machine learning, this study aims to identify disparities in healthcare access and assess environmental influences on timely diagnosis and treatment. Insights from this research can support equitable healthcare interventions and improve outcomes for aggressive cancer cases.

Does News Coverage of ESG Violations Affect Brand Sales? An Empirical Analysis

Yao Chen[1], Rakesh Mallipeddi[2], Serkan Akturk[3]

[1]Bucknell University, Lewisburg, USA. [2]the Ohio State University, Columbus, USA. [3]Clemson University, Clemson, USA

Abstract

This study examines the effects of news coverage related to environmental, social, and governance (ESG) violations on consumer purchasing behavior at a large U.S. retailer. Whereas prior studies suggest that firms' ESG engagement affects their financial performance, there is limited guidance on how news of ESG violations affects consumer purchasing behavior. We assemble a unique dataset combining transaction data from a major retailer with 311 stores in the U.S. with news coverage data of ESG violations. Our findings show that news coverage of ESG violations significantly and negatively impacts brand sales at the retail level. However, the relationship between ESG news and sales is more nuanced, depending on the type of violation, i.e., environmental, social, and governance. While our findings indicate that news on social-related incidents negatively impacts consumer purchase behavior, we uncover a counter-intuitive effect: news coverage of negative environmental and governance issues is associated with an increase in purchases. Our study also highlights the important role of brand- and market-specific factors, such as brand value, consumer demographics, and political leanings based on store locations, in moderating the impact of negative ESG news on consumer purchases. We validate our main findings via a series of robustness analyses that account for potential endogeneity issues, alternative model specifications, and estimation strategies, including difference-in-differences with propensity score weighting.

The full paper is not published in the conference proceedings at the authors' request.

Meeting in the Middle: Data-Driven Strategic Planning for Organizational Success

Deborah DeLong, Keratiloe Sishoma Mogotsi

Chatham University, Pittsburgh, USA

Abstract

Higher education is changing rapidly as competition for student enrollment and retention intensifies. Many colleges and universities face budget challenges that require difficult choices to be made about resource allocation. Strategic planning and decision support models play an increasingly important role in helping universities to navigate these complex industry conditions, but significant challenges can impede progress within the institution. Resistance to top-down strategic planning directives can arise from limited stakeholder buy-in, poor alignment between department goals and institutional priorities, unclear definitions of success, and tension between long-term goals and daily operations. This paper presents a customizable decision support methodology to guide bottom-up strategic planning efforts that align effectively across units and with administration priorities. Leadership within disparate academic units generate insights, targeted interventions, and outcome priorities with a series of introspective and analytical team exercises. The work facilitates resource allocation decisions that support the specific needs within units while ensuring alignment with enterprise goals. The proposed approach aims to overcome the inherent obstacles of top-down institution-wide strategic planning with a compelling and complimentary bottom-up unit-specific process.

Building AI Solutions in Snowflake

Cody Baldwin

University of Wisconsin-Madison, Madison, USA

Abstract

Snowflake is arguably the most popular cloud data warehouse in the marketplace. It is efficient, scalable, secure, and easy to use. This workshop will showcase the LLM functions and other AI features in Snowflake that allow organizations to extract more value from their data – without that data ever leaving the data warehouse. Thus ensuring security and privacy.

Refilling the effect of supply chain disruptions

Ceyhun Ozgur

Valparaiso university, Valparaiso, USA

Abstract

Thod manuscripts defidripmonsttsted how supply chain dogru doisdrutoond rggrvt the cte performance of s firm The changes om the levels of inventory can effect the performance of the firm positively or negativity The changes in the levels of inventory can have a n upward or downward change in the amount change in supply chain disruptions.

This manuscript is going to ask as a form of questionnaire to managers or the owner of the firm how the changes in the levels of inventory effect the supply chain disruptions.

Who is More Likely to Enjoy Football as Their Favorite Sport? The Impact of Gender & Previous Sports Experience

Eileen Zhan[1], Yong Xu[2]

[1]Forest Hills Central High, Grand Rapids, USA. [2]Ferris State University, Big Rapids, USA

Abstract

This study analyzes data from over a thousand U.S. adults to examine how gender and past sports experience influence football preference. While men's football preference correlates with participation in football, soccer, hockey, and golf, no such pattern exists for women. Logistic regression shows that prior football experience is a key predictor of preference, with men favoring the sport more due to cultural and participation norms. Promoting inclusivity, family-oriented initiatives, and sustainable practices can help increase female engagement and support football's long-term popularity.

A multinational study of the role of food banks in food supply chains

Ryan Atkins[1], Kimberly Deranek[2], Guénola Nonet[3], Felicitas Schneider[4]

[1]Duquesne University, Pittsburgh, USA. [2]Nova Southeastern University, Fort Lauderdale, USA. [3]Jönköping University, Jönköping, Sweden. [4]Thünen Institute of Market Analysis, Braunschweig, Germany

Abstract

Food banks serve a vital role in distributing food to individuals facing food insecurity. Previous research has focused primarily on the operational challenges of managing food banks, while the motivational aspects of food bank operations have received little attention. Through a series of 24 interviews with food bank leaders around the world, this study has more clearly delineated the role of food banks in managing the motivations of key stakeholders, while also providing insight into innovative food bank practices.

Encouraging Pre-Class Preparation Through Targeted Reading Quizzes

Linda Hajec

Penn State - Behrend, Erie, USA

Abstract

Many of us have experienced a decline in student effort to prepare for class discussions. As a result, instructors face challenges in delivering deeper, more engaging lessons because class time is required to cover more basic, foundational concepts. To address this issue, I implemented a brief five-question multiple-choice reading quiz at the beginning of each new chapter. These quizzes are based on end-of-chapter questions from the textbook, requiring students to engage with the material without relying on rote memorization. By providing a clear focus for their reading which is intended to help them obtain a baseline understanding, this strategy encourages students to come to class with foundational knowledge, allowing for richer discussions and more efficient use of class time. This presentation will explore the effectiveness of this approach in fostering student engagement and improving classroom dynamics, including practical preparation and application of the idea in two different course settings.

Bringing the Work-Place to Life: An Interactive First-Day Activity for Accounting Information Systems

Linda Hajec

Penn State - Behrend, Erie, USA

Abstract

Many students entering an Accounting Information Systems (AIS) course have limited exposure to manufacturing environments, making concepts like the production cycle, inventory flows, and internal controls more abstract. To bridge this gap, I modified an activity from a peer presentation at a conference to create an engaging first-day activity that immerses students in a simulated manufacturing experience through an interactive play.

The activity centers on an intern's first day at a manufacturing company, where three students act out a scripted tour of the facility. A slide deck with images of key manufacturing areas—such as loading docks, quality assurance labs, and production lines—accompanies the performance, providing students with a visual context for discussion. Throughout the play, the class participates in guided group discussions prompted by embedded questions which require them to recall and apply knowledge from prior courses, encouraging them to stay engaged.

This dynamic approach achieves three key objectives: (1) engaging students on the first day and fostering active participation, (2) introducing foundational real-world topics in a memorable and accessible way, and (3) establishing a practical framework that students will build upon throughout the course. By transforming abstract concepts into a visual experience, this activity enhances student comprehension and engagement, ultimately improving their ability to apply AIS principles as we move into a course that is vocabulary intensive for the first several chapters.

Sustaining The Surge: Twitter Topics and Persistent Momentum of Meme Stocks

Shaoqing Zhang[1], Xinyuan Tao[2]

[1]Le Moyne College, Syracuse, USA. [2]New Jersey Institute of Technology, Newark, USA

Abstract

We explore the impact of social media on meme stock price movements, particularly in the post-GameStop short squeeze period. By analyzing 2.6 million tweets, we examine how investor discussions on platforms like Twitter influence stock momentum. Using a customized BERT model for sentiment classification, we identify significant correlations between social media discourse and price fluctuations. Our findings suggest that meme stock momentum persists beyond major events, offering insights into market dynamics and predictive applications for financial analysts. This research bridges gaps in understanding retail investor behavior, highlighting the role of online discourse in shaping financial market trends.

Generative AI for Grading in Business Courses: Opportunities and Challenges

Sourish Sarkar[1], Sumanta Singha[2], Rajkamal Kesharwani[3]

[1]Penn State Behrend, Erie, USA. [2]Texas Tech University, Lubbock, USA. [3]Mercyhurst University, Erie, USA

Abstract

Generative AI tools can offer opportunities for efficient grading in business-related courses. By providing detailed feedback, these tools can reduce instructors' workload and deliver timely insights to students. However, questions remain about how accurately AI can assess complex cognitive tasks and address bias, fairness, and ethics. This preliminary study discusses the possibility of using AI for evaluating students' work and the challenges in aligning AI-based grading with human scoring on routine tasks. Pairing AI-generated assessments with human oversight may optimize grading workflows, enhance student engagement, and foster equitable outcomes. Using AI for evaluating student assessment may serve as both a valuable instructional tool and a catalyst for deeper learning.

Navigating Sustainability Controversies: Implications for SSCM and Firm Performance in Turbulent Environments

Amir Naderpour[1], Gregory Frazier[2]

[1]Eastern Kentucky University, Richmond, USA. [2]University of Texas at Arlington, Arlington, USA

Abstract

Stakeholders today expect companies to take responsibility for sustainability throughout their supply chains and respond quickly to any issues. This study examines how sustainable supply chain management (SSCM) controversies impact sustainability practices and firm performance. Using stakeholder theory, we explore direct and indirect effects, while considering industry turbulence influence in this relationship. Our findings show that SSCM controversies do not directly harm firm performance but encourage better governance, social, and environmental practices. However, these benefits weaken in unstable industries. This research underscores the need for adaptive sustainability strategies that enable companies to maintain robust performance in dynamic business environments.

Age, Education, and Ethics: Reassessing CEO Influence on Global Supply Chain Violations

Amir Naderpour, Ki-Jung Kim

Eastern Kentucky University, Richmond, USA

Abstract

Global supply chains face increasing scrutiny for social compliance, highlighting the importance of executive leadership. Drawing on Upper Echelons Theory, we investigate how CEO age and MBA education predict social compliance violations across supplier factories. Using Fair Labor Association audit data, our findings show that older CEOs, likely to rely on legacy practices, are associated with higher violation counts. However, an MBA moderates this effect by introducing contemporary management insights that reduce noncompliance, even for older executives. Our study advances theoretical understanding of leadership's impact on ethical operations and offers practical implications for fostering responsible management in global contexts ultimately.

Best Practices in Data and Security Governance in AI

Jason Davidson, Hongjiang Xu

Butler University, Indianapolis, USA

Abstract

An AI adoption survey conducted by McKinsey global spanning the past six years found that business interest in generative AI has increased by 72 percent in the past year. With 65% of the companies surveyed reporting their employees utilize AI on a daily basis. With this growth in adoption rates and the vast options for deployment in the workplace, IT management must balance issues with data governance and security with AI driven systems with the demands of the evolving workforce. Data used in AI are different than in the traditional systems. It works in the big data environment, and it is often open, linked, varied, dynamic, and highly volatile. This work seeks to explore the current best practices in AI data governance in business with a mixed method approach or quantitative surveys and qualitative interviews. The goal being the creation of an AI data and security governance framework that could help drive decision making in deployment and management of AI systems. We will identify and explore the key factors that must be considered in corporate using AI and answer the question what are the best practices for data and security governance in corporate AI deployment and management? In doing so, there are many areas of considerations, such as: legal, ethical, technical, managerial, organizational, financial and social.

Predicting 90-Day Post Liver Transplant Mortality Using Machine Learning

Jeff Liggett, Ilyas Ustun

DePaul University, Chicago, USA

Abstract

The current liver allocation system for transplant candidates relies on the MELD-Na score, a composite value derived from four biological parameters: serum creatinine, serum sodium, bilirubin, and the international normalized ratio of prothrombin time. Originally developed to predict mortality in patients undergoing transjugular portosystemic intrahepatic shunt (TIPS) procedures, the MELD-Na score has also been applied to liver transplant candidates since its implementation in 2002. However, we contend that the MELD-Na score is outdated and underutilizes the wealth of available data. To address this, we propose leveraging machine learning techniques to enhance the accuracy of 90-day mortality predictions. Specifically, we plan to employ a logistic regression model that surpasses the predictive capabilities of the MELD-Na score. By doing so, we aim to facilitate better donor-recipient matches, ultimately improving post-transplant results. Using the wealth of the available data, we plan to build a model that is effective and differentiates from other state of the art models that have been proposed. By continuing to iterate and validate our predictive model, we strive to establish a comprehensive and reliable tool for risk stratification in liver transplantation, ultimately improving patient outcomes and informing clinical decision-making in this critical medical domain.

The full paper is not published in the conference proceedings at the authors' request.

Unlocking the Commercialization of SAF through Integration of Industry 5.0: A Technological Perspective

Sajad Ebrahimi

Grand Valley State University, Grand Rapids, USA

Abstract

Sustainable aviation fuel (SAF) has demonstrated promising potential in reducing carbon emissions within the aviation industry. Various initiatives have been proposed to expedite the adoption of SAF within the industry. However, the technological advancements required for its improvement necessitate in-depth insights to guide policy makers and investors. Industry 5.0 has emerged as a promising platform that can provide the most advanced technologies. This study aims to review the existing literature and identify similar studies and case studies that can assist in advancing SAF supply chains and production processes utilizing technologies offered by Industry 5.0.

The full paper is not published in the conference proceedings at the authors' request.

Carbon Emission Modeling of Consumer Trip Chaining

Jay Brown[1], Maxum Bushuev[2], Alfred Guiffrida[3]

[1]Loyola University Maryland, Baltimore, USA. [2]Morgan State University, Baltimore, USA. [3]Kent State University, Kent, USA

Abstract

The procurement of goods today is highly customer-centric with the consumers enjoying multiple strategies for obtaining products in hand. Two key strategies include direct delivery to the home by third party logistics providers and self-pickup at brick-and-mortar stores/designated pick-up locations. Trip chaining is often used by consumers engaged in self-pickup. A trip chain involves the consumer visiting multiple distinct locations to acquire a set of products and services. Trip chaining is popular since it often leads to time savings from fewer total vehicle traveled miles, better navigation of congested travel corridors and efficiencies in combining work and shopping schedules. Despite growing and intensifying environmental concerns, the trip chaining literature has made only causal mention of the carbon emissions associated with trip chaining. In this paper we present a methodology for examining the trip chain carbon footprint associated with internal combustion and electric vehicles.

A supply chain view to technology transfer

Benjamin Agyei-Owusu

Arizona State University, Phoenix, USA

Abstract

There has been growing interest in technology transfer and supply chain management. However, it has been noted that majority of technology transfer initiatives actually fail to meet expected outcomes. A failure to effectively consider supply chain issues may have accounted for this. This study explores the intersection of supply chain management and technology transfer, and conceptualizes a new supply chain view to technology transfer. The study then highlights the research, management and policy implications of the supply chain view to technology transfer.

The full paper is not published in the conference proceedings at the authors' request.

Optimizing Agricultural Sustainability and Ecological Protection in the Yangtze River Economic Belt: A Coordinated Approach

Lili Gan[1], Xiangling Hu[2], Jaideep Motwani[2]

[1]East China Jiao-tong University, Nanchang, China. [2]Grand Valley State University, Grand Rapids, USA

Abstract

This study explores the relationship between agricultural ecological protection and food sustainability within the Yangtze River Economic Belt (YREB), a critical region for China's environmental and food security. Utilizing an integrated evaluation framework, the study applies analytical tools such as entropy weight, comprehensive index models, and coupling coordination to assess these interactions. Findings indicate that while ecological protection faced slight declines from 2002 to 2022, food sustainability demonstrated an overall upward trend, though with fluctuations. The study highlights an increasing coordination between these factors, underscoring the essential role of ecological considerations in sustainable food production. Recommendations emphasize improved land management, enhanced policy support, and integrated agricultural practices to strengthen the balance between ecological sustainability and food security.

Using AI in Strategic Formulation to create Competitive Advantage

William Johnson

Penn State, Erie, USA

Abstract

Utilizing an RBV (Resource-based View) logic, this paper explores the possibility of whether the use of Artificial Intelligence (AI) can help generate strategic competitive advantage for a firm. The RBV perspective used is the VRIO model developed by Jay Barney, which can help one determine the potential competitive advantage a specific resource might give an organization. The analysis suggests that in and of itself AI cannot provide a specific company with a specific competitive advantage. However, by taking a complex systems approach that involves the innovation processes of an organization, AI can be used to develop an intangible competency. When combined with other resources such as specific industry knowledge and the innovation processes of the firm, this can have the potential for creating truly sustainable competitive advantage. It is concluded that, given the trajectory of AI development, no company can afford to ignore the impact and potential use of AI on its strategic business models but its effective use in creating a strategy for competitive advantage must still be discerned.

Fear and Optimism in AI

Parag Uma Kosalge

Grand Valley State University, Grand Rapids, USA

Abstract

With the rapid growth of artificial intelligence (AI), concerns about its potential to cause widespread harm have also intensified. Prominent voices in the tech industry have publicly expressed these fears, fueling anxiety around AI's future impact. However, there is limited empirical research examining whether such fears truly exist among everyday users, or how these fears coexist with perceptions of AI's benefits. This paper proposes a model that considers both fear and optimism toward AI. We suggest that fear, rather than deterring engagement, may actually encourage individuals to explore and adopt AI technologies, driven by curiosity or a desire to stay informed and prepared. Our study offers a structural model capturing insights into how users may perceive AI, balancing apprehension with hope, and outlines key expectations users may have regarding AI applications. It aims to contribute to a deeper understanding of public sentiment toward AI and inform strategies for technology adoption and communication.

Understanding user perception about Webtop services

Parag Uma Kosalge

Grand Valley State University, grand rapids, USA

Abstract

This paper presents a model for understanding user perceptions of Webtop services, which offer a full desktop experience within a web browser. Early providers such as Cloudo and ZeroPC struggled with adoption and eventually exited the market. However, the pandemic dramatically shifted this landscape, with businesses embracing Webtop services to ensure employees could access consistent software and systems from home, regardless of their operating systems. This shift raises an important question: what factors shape user perceptions of Webtop services?

In this study, we identify six key factors within a structural model that influence user perception and adoption. Given that the core functionality of Webtop services remained consistent before and after the pandemic, we propose that the model also remains stable — with the primary difference being a heightened perceived importance of the technology in the post-pandemic world. While this paper focuses on one dimension of Webtop services, it offers valuable insights into user expectations and perceptions. Our findings not only contribute to academic research in this evolving area but also provide practical guidance for technology developers and service providers looking to better meet user needs.

Using Games in Education

Parag Uma Kosalge

Grand Valley State University, Grand Rapids, USA

Abstract

This research explores the potential and effectiveness of game-based course delivery in higher education. During the pandemic, educators were required to record lectures for online delivery. These recorded lectures now offer an opportunity to be repurposed, giving educators more flexibility to use in-person class time for active, engaging, and collaborative assignments — including game-based learning activities. Existing literature on Games-Based Learning (GBL) emphasizes the role of games in enhancing motivation, engagement, and concept retention. However, a fundamental barrier to implementing such courses is determining whether they can be both, feasible as well as effective, in practice.

This study examines a course designed with game-based delivery to address key concerns: Do students find this approach valuable and interesting? Does it influence attendance, improve concept retention, and enhance overall learning effectiveness? Moreover, do students express a preference for such courses compared to traditional formats? To explore these questions, the study develops a seven-factor model that captures critical elements influencing student perceptions and outcomes.

As the first study to directly examine game-based course delivery in this context, the research not only defines what constitutes a game-based course but also provides actionable guidance for educators who wish to integrate gaming elements into their pedagogy. It can inform future course design, helping institutions and instructors leverage recorded content and in-person time more effectively while fostering higher student engagement, participation, and learning outcomes.

Physician Workload and Medical Malpractice Risk: Insights from an Empirical Investigation

Denise Bergdolt, Prakash Mirchandani

University of Pittsburgh, Pittsburgh, USA

Abstract

Medical malpractice - professional negligence resulting in patient harm - is a critical issue in the United States healthcare system leading to payments amounting billions of dollars annually. Utilizing hospital's, department's and physician-specific data, previous research has shown that varying stress factors, mental workload, and burnout negatively correlate with job performance in medical settings. We construct an extensive, national retro-perspective longitudinal database to determine how physician workload intensity impacts medical malpractice case frequency in the United States. Applying a combination of principal component analysis and generalized propensity score stratification, we find a concave relationship between ln(workload intensity) and case intensity. We run multiple robustness checks corroborating the results of our primary investigation. These include a generalized propensity score analysis with a subset of our data that excludes extreme values of workload and case intensities, a generalized propensity score method using workload intensity instead of ln(workload intensity), and dividing our data into high and low workload intensity and conduct propensity score matching to estimate the effect on case intensity. Our findings have significant impacts on policy recommendations and mitigation strategies to minimize the prevalence of medical malpractice.

The full paper is not published in the conference proceedings at the authors' request.

Strategic Timing and Operations for Investment-Ready Manufacturing Startups

Xiangling Hu[1], Xinxin Hu[2], Ping Su[3]

[1]Grand Valley State University, Grand Rapids, USA. [2]University of Houston – Downtown, Houston, USA. [3]Hofstra University, New York, USA

Abstract

Manufacturing startups face complex challenges in securing funding and ensuring long-term survival. This study develops a dynamic, quantitative framework to guide decision-making around operations and investment timing, tailored to the startup's financial position and market trajectory. Integrating insights from entrepreneurial finance and operations management, the model identifies optimal strategies under varying market conditions—growing, shrinking, or peaking markets. The findings highlight how fixed costs, capital constraints, and market forecasts influence when and how startups should seek external investment. This work offers actionable guidance to founders and investors by linking operational choices with funding opportunities and firm valuation.

Simulating B2B Sales Process Using Agentic AI

Chhaya Tundwal, Kshitij Chauhan, Tirth Patel, Vishnu Anand, Yash Kothari, Yu-Yang Ho, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

This study examines how companies can effectively balance AI automation and human oversight in B2B sales pipelines to maximize efficiency. We deploy Hermes-Llama-70B AI agents, leveraging Prediction Guard's API and LangChain, to automate the sales pipeline. Our analysis compares AI-driven pipelines against traditional human workflows using synthetically generated user personas and interactions. Results reveal key trade-offs across varying autonomy levels, offering practical insights. The project aims to enhances Prediction Guard's internal processes and provides a clear model for businesses seeking responsible AI integration alongside essential human roles.

Enhancing Judicial Efficiency and Access to Justice Using AI

Shubham Gaddi, Saquib Hussain, Nandini Devalla, Hugo Bizcek, Ayush Gupta, Darshan Upadhyay, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

This study explores how AI can enhance productivity in Indiana's legal system. The motivation is that Integrating AI into court workflows requires balancing efficiency gains with accountability and protecting sensitive data from AI feedback loops. We use survey data from over 100 Indiana judges, and apply Python-based NLP and Azure Language AI to extract sentiments and key concerns, offering guidance for responsible AI integration in judicial processes.

Leveraging Analytics & Modeling to Derive Insights for Point-based Loyalty Campaigns

Yifan Gu, Muskan Aggarwal, Sai Nithin Godi, Dhairya Dedhia, Harshal Amin, Nikitha Balaji, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

This study examines how loyalty program design influences customer engagement and spending for a U.S. supercenter chain. We analyze the effectiveness of different offer types and delivery methods, controlling for pre-campaign behavior, to identify strategies driving incremental impact. Predictive models assess the ability of customer behavior to forecast campaign-period sales. Results indicate that per-visit incentives slightly outperform point multipliers, though differences are modest. A simple linear model performs comparably to more complex techniques, offering actionable insights for optimizing loyalty strategies. Future improvements could leverage granular segmentation and enriched data for enhanced program effectiveness.

Detecting and Predicting Phantom Inventory Using Random Forest and LSTM Models

Hung-Chen Hsu, Kratika Jain, Da Fang Lin, Yiran Liu, Kaushalya Naidu, Namitha Ramakrishnan, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

Phantom inventory refers to discrepancies where inventory exists in records but not in reality, which can arise due to a variety of factors, such as misplacement, system glitches, theft, shrinkage, and inaccurate data entry. This paper will focus on exploring real business inventory data and modeling on detection and prediction. We aim to develop powerful machine learning models, random forest, time series, and clustering, by analyzing transaction data, stock level movement patterns, and sale anomalies to uncover hidden discrepancies and eventually manage to predict stockouts. Detecting and predicting phantom inventory case is crucial to optimize business performance and enhance profitability.

Mapping the Path to Translations: Empowering SIL to Reach New Frontiers in Language Translations

Vandana Brahmasa, Amy Fischbach, Parvathy Nair, Gitanjali Nambiar, Aryan Ritesh, Huilin Xu, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

We developed a mapping tool for SIL International, a global leader in language translation, to identify unmet translation needs and guide market expansion efforts. The motivation for this research is that despite global connectivity, many low-resource languages lack Bible translations; SIL seeks to address this gap using AI and data science. We design and build a mapping tool with Python and HTML which identifies underserved areas, integrates LLMs to help identify barriers, and offers insights applicable to both SIL and broader global outreach initiatives.

Risk Benchmarking for LLMs

Abhishek Bagepalli, Avanti Chandratre, Chan-Yen Hsiung, Jayesh Rajendra Chaudhari, Ramya Chowdhary Polineni, Tsung-Yu Lu, Rohan Ajay, Matthew Lanham

Purdue University, West Lafayette, USA

Abstract

We evaluate the robustness and security risks of Large Language Models (LLMs) using adversarial prompt benchmarking. As generative AI gains traction in enterprise settings, ensuring model reliability is vital. We introduce a risk benchmarking framework that measures prompt deviation rate (PDR) under adversarial modifications at letter and sentence levels. Using Hermes-3-Llama-3.1-70B for generating adversarial samples and the 8B variant for evaluation, we test sensitivity to subtle perturbations. Results show minimal impact from letter-level changes but greater effects from structural modifications. Despite this, statistical tests reveal consistent PDR across strategies, offering insights to strengthen LLM resilience.

Papers

DECISION SCIENCES INSTITUTE

Sales competition among automakers: A comparative study of five firms

Sima M. Fortsch
Haile College of Business
Northern Kentucky University
1 Louie B Nunn Drive, Highland Heights, KY 41099
Email: fortschs1@nku.edu
Phone: 859-572-6582

Jeong Hoon Choi, Ph.D.
Williamson College of Business Administration
Youngstown State University
1 Tressel Way, Youngstown, OH 44555
Email: jhchoi@ysu.edu
Phone: 330-941-2401

Mohsen Ghodrat, Ph.D.
University of Canada West
Vancouver, BC, Canada
Email: mohsen.ghodrat@ucanwest.ca
Phone: +18774316887

ABSTRACT

This study gains insight into the competitive dynamic in the North American auto industry. We explore the intensity of competition by analyzing time series sales data over a 10-year period from 2009 to 2019. We used two regression methodologies: vector autoregressive regression and seemingly unrelated regression models. Contrary to the expected fierce competition, the results show moderate competition between the automakers' sales. While there are some interrelations between the sales, the rivalry does not affect all members equally. These findings challenge the prevailing notion of intense intra-group competition within the North American auto industry.

KEYWORDS: Autoregressive models, Time series regression

INTRODUCTION

Competition in the automotive industry is a key driver of innovation, market expansion, and consumer choice. Though it is often viewed as a source of pressure, it drives innovation, enhances efficiency, and promotes better products for customers by pushing firms to improve their quality of products and services. Understanding and leveraging competition can provide strategic advantages for businesses. For example, Fortsch et al. (2021) highlight the potential benefits of leveraging competitors' sales data to enhance sales predictions. Their study demonstrates that incorporating competitors' sales information into demand forecasting can lead to more accurate predictions, helping to mitigate challenges such as overstocking or understocking expensive inventories, particularly automotive.

Literature on rivalry dynamics within industries has traditionally focused on studying strategic groups. The concept of strategic groups, which was introduced by Hunt (1972) and then developed by Porter (1979; 1980), emphasizes that businesses in the same strategic group often use comparable tactics and, as a result, deal with comparable difficulties brought on by external economic variables (Choi et al., 2024). This framework is vital because it emphasizes that firms within the same group frequently compete for similar resources and market opportunities, such as raw materials, production capacity, and customer bases. As such, these firms must closely monitor their competitors' moves, particularly sales performance, to make informed decisions regarding strategic adjustments in areas like inventory management and supply chain optimization. The conventional wisdom suggests that the fiercest competition occurs within strategic groups, where firms are believed to resemble one another on key strategic dimensions (Cool & Dierickx, 1993; McNamara et al., 2003; McGee & Thomas, 1986; 1992; Nair & Filer, 2003). It is believed that this fierce competition encourages innovation and drives companies to improve their efficacy and efficiency to remain competitive (Dranove et al., 1998).

Over the last century, the auto industry has undergone several transformations, moving from craft production to mass production and then lean production (a.k.a. Toyota Production System). Competition in the auto industry has been shaped by technological developments and changes in customer demand, resulting in the current environment where big auto companies pursue cost leadership as well as customization and diversification. The intensity of competition has led to a shift from mass production to more customer-oriented build-to-order (BTO) systems (Holweg, 2008), which has led to inventory and cost reduction and has improved responsiveness to customer needs.

This study addresses straightforward inquiries about firms' performance in a competitive environment, such as: Do firms in the North American auto industry compete in sales? Can a company's sales data identify its major competitors rather than relying solely on market share and product diversity? To answer these questions, we collected seasonally adjusted monthly data on domestic automobile sales from the U.S. Automotive News database, covering the period from January 2009 to December 2019. Here, 'domestic' refers to North American sales, specifically within the United States and Canada. We focused on domestic sales due to the unavailability of continuous worldwide sales data. The study period is carefully chosen to represent the true understanding of competition without the variability introduced by the worldwide COVID-19 pandemic. The dataset encompasses five major automakers in the North American auto industry: General Motors (GM), Ford, Honda, Toyota, and Nissan. We excluded companies for which continuous data was inaccessible during our study window, or the data had many missing observations. For instance, databases from some automakers like BMW and Volkswagen, had significant gaps in coverage, while others contained too many missing entries for the period to be considered. To ensure the reliability of our findings, we decided to exclude these companies from our study. Time series data for GM, Ford, Honda, Toyota, and Nissan were persistently reliable for the duration of the study.

The rest of this research is organized as follows. Section 2 offers a comprehensive review of the relevant literature. Section 3 outlines the research methodologies, datasets, and econometric tests employed. Section 4 presents the results, and in Section 5, we discuss the managerial implications. Section 6 concludes the research and discusses the limitations.

LITERATURE REVIEW

Competition drives firms to become more efficient by forcing them to reduce operation costs and improve product quality to maintain profitability. Auh and Menguc (2005) define competitive intensity as “a situation where competition is fierce due to the number of competitors in the market and the lack of potential opportunities for further growth.” The results of a firm’s behavior will become stochastic rather than deterministic as competition intensifies because a firm's behavior is greatly impacted by the actions and contingencies made by rival firms. A competitive context typically refers to situations where organizations are likely to find themselves in zero-sum relationships, either directly or indirectly (Barnett, 1997). For example, competition tends to be more intense among "structurally equivalent" organizations (Burt, 1992) or those within the same "niche" (Hannan & Freeman, 1989). As a result, competition varies depending on the environment, becoming stronger when there is a higher likelihood of zero-sum dynamics between organizations. Galvin et al. (2020) explored the relationship between network rivalry, competition, and innovation. The authors investigated how rivalry within networks—such as business ecosystems or industries—affects competitive behavior and drives innovation. Their study examines the dynamics of competition within these networks and how different competitive strategies influence the pace and direction of innovation, with implications for both firms and industries.

Competition in the automobile industry has long been an intriguing topic for numerous business scholars. One stream of the prior research investigated pricing strategy and competition within the auto industry. Busse et al. (2010) investigated pricing promotions in the U.S. automotive industry, while Zeng et al. (2016) analyzed Toyota's implementation of "no-haggle" pricing and its impact on sales, particularly in markets where competitors do not offer fixed pricing policies. Karabakal et al. (2000) focused on optimizing supply chain performance and distribution systems for Volkswagen of America. Hu et al. (2014) explored the relationship between advertising, consumer information searches, and subsequent sales using consumer prepurchase search data and U.S. automobile industry sales data. Borah and Tellis (2016) investigated automobile recalls and their adverse effects on sales and other metrics, not only for the recalling company but also for its competitors. Nagler (2014) examined negative externalities and their competitive implications in the motor vehicle market.

On the other spectrum, prior researchers investigated the competition in the auto industry in context of other factors. Meckling and Nahm (2019) explored the political dynamics and implications of technology bans in the auto industry, particularly in the context of green goals and industrial policy competition. The authors investigate how governments use technology bans to promote environmental objectives, such as reducing greenhouse gas emissions, while fostering industrial competitiveness. Their study examines the intersection of environmental policies and economic strategies, highlighting the challenges and opportunities that arise from such bans, especially in the global competition for leadership in green technologies within the auto industry. Grieco et al. (2021) examined how market power and its changes over time in the U.S. automobile industry. The authors quantified the evolution of market power, investigated its determinants, and provided the implications for consumers and the industry as a whole. Their study used pricing, production, and market share data to analyze shifts in competition, pricing power, and industry dynamics over several decades. Morris et al. (2020) analyzed the cybersecurity challenges the automotive industry faces. The authors focused on the tensions and issues within the knowledge environment related to cybersecurity threats. They sought to understand how these threats impact the industry, explore the gaps in knowledge and preparedness, and propose strategies for addressing these challenges to enhance the security of automotive systems and protect against potential cyberattacks. Pichler et al. (2021) focused on modernization and transformation within the auto sector. The authors analyzed the policy measures and strategies implemented by the EU, assess their effectiveness in fostering innovation, sustainability, and competitiveness, and explore the broader implications for the industry's evolution in response to these policies. Liu (2021) analyzed the relationship between competition and Tesla Motors' valuation. The study investigated how competitive dynamics in the automotive industry influenced Tesla's market value and financial performance. Through a case study, the author seeks to provide insights into the factors driving Tesla's valuation, including competitive strategies, market positioning, and the impact of industry trends on the company's financial metrics.

The literature indicates that another stream of the previous studies focused on automobile competition, sales, inventory, and sales forecasting. For example, Abdel-Khalik and El-Sheshai (1984) conducted a comparative study on the predictive capabilities of ARIMA models for forecasting sales revenue across various industries. The authors studied the concept of stock price behavior concerning information uncertainty. Their research focused on how different levels of information uncertainty impact the relationship between accounting information and stock prices. They explored the varying effects of uncertainty on investor behavior and how it influences the market's response to financial disclosures. Their study aimed to understand how information asymmetry and uncertainty affect stock valuation, contributing to the broader understanding of financial markets and investor decision-making. On the other hand, Syntetos et al. (2009) provided a comprehensive review of inventory planning and forecasting, integrating diverse strands of literature such as dynamic system methods, control theory, forecast theory, and judgment methods.

Aviv (2003) also introduced models designed to capture enriched information beyond past demand data for forecasting and inventory management. Specifically, the author considered demand within a vector autoregressive (VAR) time series framework. Similarly, Shoesmith and Pinder (2001) explored forecasting techniques employing vector autoregression (VAR) and Bayesian vector autoregression (BVAR) methods. Their findings suggest that BVAR proves beneficial in scenarios with limited historical data, with both VAR and BVAR outperforming exponential smoothing and seasonal decomposition methods. Furthermore, Fisman (2006) investigated the impact of lack of foreign competition on sales forecasts and inventory decisions, noting a tendency towards optimistic sales forecasts and significant implications for inventory management. Barlas and Gunduz (2011) delved into the bullwhip effect and inventory fluctuations, emphasizing the importance of basing order placement decisions on known demand patterns rather than forecasts. Neale and Willems (2009) analyzed inventory management amidst nonstationary demand, focusing on a comprehensive inventory model that accounts for inventory locations and levels across a multi-stage supply chain.

Some of the exploratory studies within the automotive industry examined the sales interconnectivity. For instance, Wochner et al. (2016) focused on balancing supply and demand while introducing new products in the automotive sector. Jayarajan et al. (2018) empirically investigated the demand for new and used cars in the U.S. market and its impact on product durability. Jan and Hsiao (2004) analyzed the automotive industry in Taiwan, highlighting its development and general characteristics, particularly the support provided by government policies in developing countries. Liu and Chen (2013) explored supplier risk-sharing contracts in the Taiwanese automotive industry, emphasizing risk shifting and absorption. Esteban-Bravo and Lado (2011) examined horizontal alliances in the Spanish auto industry, focusing on their impact on brand value. Kirca et al. (2019) utilized U.S. automotive industry data from 2007 to 2013 to study brand performance based on interactions between product and brand portfolio characteristics. Hahn et al. (2000) investigated how the Hyundai Motor Company coordinated production planning and scheduling through a centralized production-and-sales-control department. Goolsbee and Krueger (2015) reviewed the rescuing and restructuring efforts concerning GM and Chrysler. Despite these valuable contributions, none of the above studies have specifically addressed the sales dynamics among members of the same strategic group to discover the intensity of the members' rivalry.

Various other studies encompass a range of methodologies, including case studies, theoretical and simulation research, and empirical data-based analyses. The breadth of topics covered in scholarly literature is often contingent on data availability. For instance, Cachón and Olivares (2010) investigated the factors influencing differences in finished goods inventory in the U.S. automotive industry, identifying production flexibility and dealership numbers as key determinants. Their study included companies such as Toyota, Chrysler, Ford, and GM. Copeland et al. (2011) developed a model for inventory and pricing decisions within the automotive sector, highlighting the significant impact of inventory policies, particularly the "build to stock" approach, on prices and sales. Their focus was on the inventory-to-sales ratio and ordering costs. Chiang et al. (2016) also examined demand forecasting and its implications for the bullwhip effect using data from the U.S. auto industry. Olivares and Cachón (2009) conducted an empirical study focusing on inventory levels at a GM dealership within an isolated U.S. market. Their investigation centered on how local competition influences demand, sales, and inventory dynamics. Ramey and Vine (2006) examined the substantial decrease in production and sales variance within the automobile industry and the consequential impact on inventory investment, particularly its covariance with sales. They noted a shift towards inventories better insulating production from sales, resulting in a decline in the inventory-to-sales ratio, coinciding with changes in GDP growth around 1984. Alessandria et al. (2010) highlighted the drop in international trade during the 2008-2009 economic crisis and the significant inventory adjustments within the auto industry. They observed a negative correlation between the inventory-to-sales ratio and industrial production, emphasizing how fluctuations in this ratio affect imports and exports. The authors concluded that inventory management is crucial when considering international trade dynamics.

It is important to emphasize that all the above studies share a different objective than our research. While they provide valuable insights into various aspects of the automotive industry, including inventory management, pricing strategies, international trade dynamics, and sales, none specifically address the central focus of our investigation, which is to examine the intensity of competitiveness in the auto industry by analyzing their sales. This research examines the competitive dynamics among five major automakers—GM, Ford, Honda, Toyota, and Nissan.

SALES AND ITS CONNECTION TO COMPETITION

Why study the company's sales data when focusing on rivalry in the North American auto industry? The purpose of this study is to investigate whether or not all firms in the auto industry compete with similar intensity by examining the companies' sales over time. However, studying one company's sales performance does not provide vital information about its rivalry and the intensity of competition. The interconnections among the companies' sales give us a more sophisticated approach than studying one company's sales prediction, which merely relies on the sales information from the same company. Consider a scenario where GM misestimates its sales because its competitor over- or under-performs in the market. This could have a significant impact on GM's available on-hand inventory, not caused by the company's own sales predictions but by how its competitor performed. In such cases, it is believed to use models that include the competitors' sales information. In econometrics, the most common methodologies used for interconnected variables are Vector Autoregressive Regression (VAR) or Seemingly Unrelated Regression (SUR) models. Both methods are linear representations of demand predictions. One could also adopt a nonlinear or log-log model, but to make our results straightforward, we adopted the linear representations. Both VAR and SUR are prediction models that can be used to forecast a company's sales while allowing us to include the competitors' sales data in the regression system of equations. Aside from demand predictions, the VAR and SUR models can also be used to study the interconnectivity of various variables because they represent a causal relationship. In this study, we target the casual relationship instead of focusing on demand predictions. Then, the following question arises. Does the company have access to the sales information of its competitors to include them in their causal models? The answer is yes; in many cases, they do. For example, in the automotive news database, the sales of many automaker's competitors are accessible. Also, we can assume that companies in the same industry have access to reports regarding the sales of their rivals by utilizing the firm's competitive intelligence.

VAR and SUR models: Insights into competitive dynamics

VAR and SUR models offer distinct advantages when analyzing the intensity of rivalry among the competitors. The advantages of these models are as follows. First, they capture interdependencies among the members of the same strategic group. In the context of rivalry among firms in the same strategic group, VAR and SUR can model how the performance (or strategies) of one firm affects others over time using the sales performance. This helps us understand the competitive dynamics in more detail. Second, flexibility in modeling. VAR does not require the identification of a clear dependent or independent variable, making it ideal for situations where all firms' actions or performances are mutually influential as they relate to sales. This flexibility is crucial in rivalry, where firms often react to each other's moves (or sales) quickly. Third, impulse response analysis is useful. VAR allows for impulse response analysis, which can be used to trace the effects of a shock from one firm on the other firms within the group over time. This helps in understanding the ripple effects of sales performance across a group of competitive firms. Fourth, efficiency in estimation. SUR is advantageous when dealing with multiple equations that are potentially related through error terms (miscalculations of beta coefficients in causality relationships). In the auto industry, different firms' performance might be influenced by common external factors (e.g., market conditions). SUR accounts for these cross-equation correlations, leading to more efficient and potentially more accurate coefficient parameters than running separate regressions for each firm. Fifth, handling cross-equation correlation. By explicitly modeling the correlation between the error terms of different equations, SUR can capture the interconnectedness of firms' sales performances and outcomes. This is particularly relevant in a strategic group where firms might be influenced by similar unobservable factors, making SUR a robust tool for examining rivalry. Sixth, simultaneous consideration of multiple firm's reactions to sales. SUR allows the simultaneous estimation of multiple firms' equations, providing a holistic view of how strategies within the group are interrelated. This is beneficial when analyzing strategic interactions and the collective behavior of firms in the same group.

In sum, VAR is ideal for understanding dynamic interdependencies and the impact of time-lagged variables effects among competitive firms, while SUR provides more efficient estimation in the presence of cross-sectional correlations, making it useful for examining simultaneous strategic interactions among firms. Together, VAR and SUR models can offer an important understanding of the competitive dynamics within the auto industry. For this study, while we rely on sales coefficients, the goal is to get an insight into how the competitors' sales are connected instead of the actual sales predictions. Thus, this study investigates the causality or interaction among companies' sales within the auto industry.

Leveraging VAR and SUR modeling

VAR originated in econometrics in the 1970s, particularly with the work of Clive Granger and Paul Newbold. Granger, who won the Nobel Prize in Economics in 2003, played a significant role in developing VAR models. The basic idea behind VAR analysis is to model each variable in a system as a linear function of its own lagged values and the lagged values of all other variables in the system. In other words, a VAR model captures the past behavior of each variable and how it influences the current values of all variables in the system. The VAR equations resemble ordinary least squares (OLS) equations but with the distinction that all individual equations are consolidated into a unified system to estimate the beta coefficients. This holistic approach enables us to capture the relationship among prediction errors across the entire system. The higher the intensity of competition, the higher the relationship among the companies' sales would be. SUR, similar to VAR, is a linear prediction model that incorporates competitors' sales data to forecast sales. The difference between VAR and SUR is their independent variables. While VAR models have the lagged dependent variables on the right side of the equation, SUR allows us to include both lagged dependent variables and lagged (or current) independent variables on the right side of the equation. Like VAR, SUR models are dynamic and stochastic. The emphasis of the SUR model lies in the forecast errors, which some scholars believe to have a lower forecast error than VAR models (Zellner, 1962; 1963). The standard procedures employed to estimate SUR beta coefficients include Feasible Generalized Least Square (FGLS), Maximum Likelihood (ML), 3SLS (3 Stage Least Square) estimations, and Iterative Generalized Least Squares (IGLS) methods. Regardless of the technique utilized to estimate SUR's beta coefficients, the process adheres to a two-step iterative approach aimed at minimizing regression residuals and enhancing regression coefficients.

The structural form of VAR regression is as follows. , where is the estimates of sales units for company *k* at time t. We used companies initials for *k* (i.e., F, G, H, N, T for Ford, GM, Honda, Nissan, and Toyota respectively). The term u_k refers to the errors of the sales estimates for company *k*. The reduced form of the VAR equations is displayed in equations 1.1 through 1.5. For simplicity, we included two lagged sales, t-1 and t-2, using the company's previous month's sales and the month before that. As a reminder, the preliminary showed that any previous sales data beyond t-2 were not significant.

(1.1)
(1.2)
(1.3)
(1.4)
(1.5)

SUR equations systems differ from VAR in a way that the right-hand side calls for including exogeneous variables; hence, we can better identify the key casual players impacting the company's sales. In this system of regression equations, a company's sales are the dependent variable on the left side, while the competitor's sales are represented as independent variables on the right. A general form of SUR regression is: . Due to the complexity of the SUR equation, we only use the time t and t-1 (skipping t-2). The reduced form of the SUR model is as follows.

(2.1)

(2.2)

(2.3)

(2.4)

(2.5)

In the above equations, both VAR and SUR, the OLS matrices are used to compute the regression coefficients, where the procedures are related to variables and their beta coefficients.

= + = (S)B + U

Econometric tests to validate results

This study investigates whether the automakers' competition is evident in their sales. If we could find such intense rivalry, we could verify the firms in the North American auto industry compete vigorously in sales. Furthermore, using the sales data, we examine the duration of a competitive move and the speed of the response from the competitors. The United States is a home base for manufacturing several foreign automakers. According to 'Select USA Automotive Spotlight,' domestic and international automakers have invested billions in the United States in the recent decade. For example, in 2015, the U.S. produced 2.6 million vehicles sold in 200 countries with a total revenue of over $65 billion. Simultaneously, the United States automakers controlled nearly 3.2 million local jobs related directly and indirectly to auto manufacturing. The importance of this industry, hence, makes it attractive for research. Furthermore, the automakers' sales data presents companies in the same strategic group, implying high competitiveness. Consequently, we utilize the seasonally adjusted monthly data obtained from 'U.S. Automotive News' from January 2009 to December 2019 for GM, Ford, Honda, Toyota, and Nissan. The duration of the data is selected to be pre-coronavirus period so that the impact and noise related to the worldwide shutdown be eliminated because our goal is mainly to target the intensity of rivalry in a strategic group as it relates to sales. We refer the readers to the reasons for not including the other automobile companies in the Introduction Section, where we explain this in more detail.

Econometric tests to validate results

Since our obtained dataset is a time series, various econometric tests must be performed to ensure the accuracy of results. For all analyses, the STATA software was utilized. First, we examined the normality of the data distribution using the Jarque-Bera test (known as the JB test). The JB test is a valuable tool for assessing the normality of data by examining its skewness and kurtosis and comparing these to what would be expected in a normal distribution. This test uses the maximum likelihood function to determine whether a given probability resembles the normal distribution (Hamilton, 1994). The null hypothesis signifies that the data is normally distributed, while the alternative hypothesis rejects the normality of the data. Similarly, to ensure accuracy for the computed outcome, it is reasonable to conduct the Augmented Dickey-Fuller (ADF) and Phillip-Perron (PP) tests (Dickey & Fuller, 1979; Phillips & Perron, 1988). These tests focus on finding evidence of autocorrelation and unit roots, which use maximum likelihood and Chi-square to determine the evidence of the problem. In the presence of serial correlation, computed variables may be correlated with errors of the past variables, making the outcome inaccurate (Hamilton, 2014).

Table 1a presents that the unit root is present before taking the first difference of the obtained data. This is shown as the values of the test statistics are smaller than the critical values. Hence, it is inadvisable to trust the regression output because a 'random walk' exists. Consequently, we took the first difference in the data to overcome the unit root problem. Table 1b indicates that after taking the first difference of the data, the unit problem disappears (the values of test statistics are larger than the critical values). Furthermore, we confirmed ADF with the Philip-Perron (PP) test. Both tests resulted in the same outcome.

The next test is to validate the accuracy of our results by conducting the stability tests on the VAR and SUR outcome after the regression is performed. Figure 1 and 2 present that all eigenvalues from VAR and SUR fall within the unit circle, which indicates that our regression output is reliable.

Table 1a. Results of the ADF (Augmented Dickey-Fuller) test before taking the first differenced of the sales data.

Company Sales	Test Statistics	Critical Value 5%
GMS	1.077	7.919
FS	0.652	7.920
TS	0.487	7.900
HS	0.629	7.199
NS	0.243	7.812

Table 1b. Results of the ADF (Augmented Dickey-Fuller) test after taking the first differenced of the sales data.

Company Differenced Sales	Test Statistics	Critical Value 5%
GMSd	60.644	7.918
FSd	24.742	7.919
TSd	48.203	7.188
HSd	46.821	7.998
NSd	28.499	7.918

Figure 1. VAR stability graph: The eigenvalues are inside the unit circle, indicating that the results can be trusted.

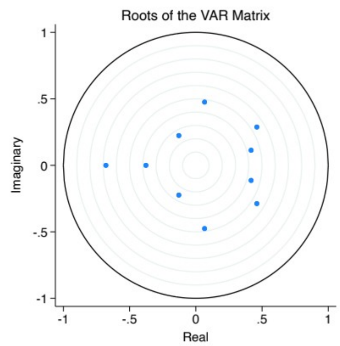
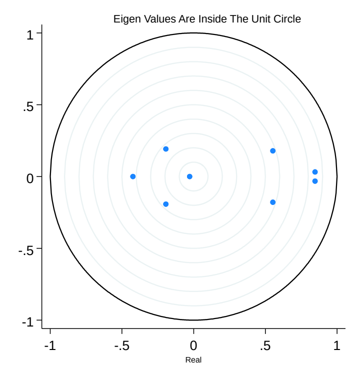


Figure 2. SUR stability graph: The eigenvalues are inside the unit circle, indicating that the results can be trusted.



RESULTS AND DISCUSSIONS

The main objective of this study is to validate the existence of an intense rivalry in the North American auto industry. Analyzing companies' sales could provide insights into their competitive behavior, even without complete information on their strategic decisions. While it is true that strategic decisions made by companies often occur behind closed doors and are not fully disclosed to the public, the outcomes of these decisions frequently manifest in observable metrics like sales figures. Companies' sales are the direct result of how the market responds to a company's product offerings, pricing strategies, promotional efforts, and distribution tactics—all of which are strategic decisions. If a company experiences a significant increase or decrease in sales, it can often be traced back to the success or failure of these strategic initiatives.

Similarly, by comparing the sales performance of companies within the same industry, one can infer competitive shifts. If one company's sales are increasing while another is stagnant or declining, it could suggest that the first company is gaining a competitive advantage, perhaps through innovation, better customer service, or more strategic pricing. Furthermore, sales patterns over time can reveal shifts in consumer preferences that companies are responding to through their strategies, even when they're correlated with known or unknown external events or actions. Therefore, while sales data alone cannot provide a complete picture of a company's strategic decisions, it serves as a crucial indicator of how these strategies are playing out in the competitive landscape. Consequently, we targeted the monthly sales data for five automakers

(GM, Ford, Honda, Toyota, and Nissan) as these companies offered available continuous time series data for the North American region (USA and Canada). The obtained data spanned from 2009 to 2019. The results of our analysis are categorized into two sections: VAR and SUR results.

VAR results

Utilizing the VAR model, a linear equation system that uses the maximum likelihood estimates to minimize the errors of the beta coefficients, we explored the interrelationships among the sales of the five competitive automakers (Equations 1.1 through 1.5). We anticipated that these companies fiercely compete in sales because they are in the same or similar strategic group.

The results summarized in Table 2 indicate that the Root Mean Square Errors (RMSEs) for all companies have improved by combining the sales of all companies into a VAR system of equation instead of a single equation for each company (AR, autoregressive models). RMSE for GM when applying simple regression was 45,024, but it decreased to 41,817 when we employed VAR regression. This is an improvement of 7 percent. Ford's RMSE improved from 32,076 to 28,340, Honda's from 19,670 to 18,685, Toyota's from 29,872 to 27,915, and Nissan's from 17,819 to 15,895 units. With low regression errors, we can infer that the VAR regression produces successful outcomes compared to simple regressions. The improvement in the regression error implies that these competitor's monthly sales are interrelated, and these companies might compete for sales. However, the small improvement in RMSEs (from 5% to 12% across the companies) infers that intense competition is not evident, at least not by analyzing the company's sales over time utilizing VAR models.

On the other hand, the improvements in R squared values were more significant, ranging from 12% to 40% improvement (Table 2). For example, we noticed that GM's R squared value increased from 0.63 to 0.70, a 12% improvement. The R squared for Ford increased from 0.51 to 0.64, for Honda from 0.29 to 0.49, for Toyota from 0.37 to 0.49, and for Nissan from 0.30 to 0.49. Across the board, the R squared increased from 12 to 40 percent.

Hence, the results in Table 2 indicate that including the competitor's sales could improve the coefficient of determination with the minimum improvement for GM and the maximum for Honda. However, R squared is merely an indication of the proportion of variance explained by the model; hence, the regression error, RMSE, might be a better indication of whether these companies' sales impact each other. Yet, the small improvement in RMSE indicates that the sales of one company might not impact the sales of another despite being in the same industry. The results conflict against the conventional wisdom of fierce competition in the auto industry, but rather a weak or moderate rivalry among the firms.

Table 2. AR (individual firms) versus VAR (combined firms) regressions outcome.

Automaker	GM	Ford	Honda	Toyota	Nissan
RMSE (AR)	45,024	32,076	19,670	29,872	17,819
RMSE (VAR)	41,817	28,340	18,685	27,915	15,895
Improvement	7%	12%	5%	7%	11%
R^2 (Single Firm)	0.6333	0.5119	0.2930	0.3734	0.3058
R^2 (VAR)	0.7077	0.6451	0.4904	0.4953	0.4907
Improvement	12%	20%	40%	24%	37%

After examining the regression errors and the R squared values, we now analyze the beta coefficients. Table 3 presents the variations in beta coefficients and their p-values among automakers. To compare all companies, we must rely on the VAR question's beta coefficients. Table 3 highlights that while VAR analysis is essential for capturing rivalry among members of a strategic group, it is evident that not all companies equally compete in sales. For instance, the first thing we noticed is that many variables were insignificant at 95% level (p-value is greater than 0.05). Second, we noticed some irregularities such as while Nissan's sales, for example, did not impact any other company's sales (Table 3, far right column), it was significantly impacted by all sales (Table 3, bottom row). On the other hand, Honda's sales were not impacted by other competitors (third row), but it affects the competitor's sales (third column). One might conclude that Honda is a major player, impacting all sales, while Nissan is the opposite. However, this statement does not match further analysis when we discuss the impulse response function in the next section. The rest of the variables were hit and miss as some were significant and others were not. Therefore, we summarize that we do not find any evidence of intense rivalry among the strategic group, at least not by examining their sales over time, but moderate rivalry is evident.

Table 3. VAR regression outcome: the beta coefficients and p-values.

Sales	GM	Ford	Honda	Toyota	Nissan
GM	**0.5769** (0.000)*	0.6662 (0.004)*	-0.8522 (0.049)*	0.4409 (0.190)	-0.2811 (0.403)
FORD	0.1929 (0.026)*	**0.4972** (0.003)*	-0.5828 (0.047)*	-0.9277 (0.678)	0.1206 (0.586)
HONDA	-0.5422 (0.334)	-0.1951 (0.058)	**0.3300** (0.007)*	0.3896 (0.792)	0.01746 (0.907)
TOYOTA	0.3145 (0.713)	0.1155 (0.045)*	-0.6315 (0.028)*	**0.5832** (0.008)*	-0.3211 (0.883)
NISSAN	-0.9312 (0.051)*	0.2004 ((0.022)*	-0.4322 (0.008)*	0.2758 (0.031)*	**0.6117** (0.000)*

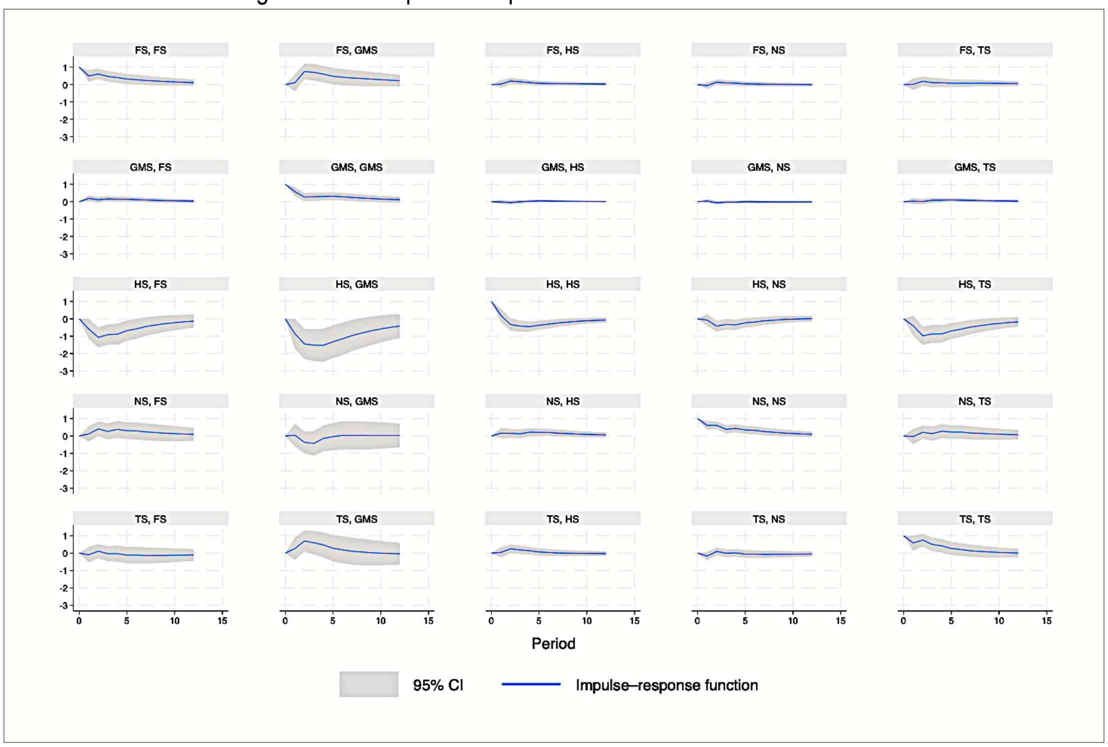
The model's beta coefficients are **bolded** for the company's own previous sales.

* Represent the variable is significant at the 5% level.

Next, we discuss the impulse response function (IRF). IRF traces the effect of a one-time shock (any unexpected changes: i.e. price change, innovation, etc.) from one variable on other variables (current and future). The main goal of an IRF is to quantify a shock and its duration over time. It illustrates the time path of the impact—how long the effect lasts and whether it strengthens, diminishes, or reverses over time. Therefore, IRF is a powerful tool for analyzing how shocks or changes in one variable affect other variables over time and how long this impact can last. This can be particularly useful for examining rivalry among competitors within the same strategic group. If we believe that the members of a strategic group compete fiercely, we must anticipate that the IRF would present a significant impact that last for a while.

The impulse response function is surprising if we believe that fierce coemption exists within the auto industry, but it is not surprising given our VAR regression that showed many variables were insignificant. Figure 3 displays the IFR graphically. The first row is Ford, the second is GM and the rest are Honda, Nissan, and Toyota respectively. The way to read the above graph is the following. The first row, first column indicates the impact of unexpected changes in sales at the Ford company on its own sales prediction. The gray area presents the disturbance in prediction (increase in error of sales forecast). We start with the first row from left to right. The first row, second, third, fourth, and fifth columns present the impact on GM, Honda, Nissan, and Toyota respectively. For instance, we notice that if an unexpected shock occurs in Ford's sales (first row), the sales' prediction for GM would be impacted, but the other automakers' sales predictions would not because only GM (the second column) shows the gray area increases.

Figure 3. The Impulse Response Function for five automakers.



In fact, monitoring the columns in Figure 3, we discover that changes in any automaker's sales predictions would impact the sales predictions at GM (second column). Hence, we can conclude that while GM's sales predictions are more sensitive to what happens to its competitors' sales, GM's sales don't impact the competitors' sales predictions. This is because the gray area (errors in predictions) is increased in the second column, where it shows the influence of the competitors' sales on GM's sales, but remains unchanged in the second row, where it shows the influence of GM's sales on others. It's also important to point out that IRF shows prediction/estimation errors. Therefore, we can't say whether the sales' quantities are ever impacted but rather the sales' predictions become inaccurate. We also cannot know the direction of the prediction errors.

So, what's the conclusion after studying VAR regression and IRF output? We believe that Figure 3 and the insignificance of many beta coefficients in VAR regression raise some doubts regarding the existence of a fierce rivalry among the members of the same strategic group. However, there must be some competition present because VAR methodology outperformed the simple regression where the competitors' sales were not included in the regression model (Table 2). However, we conclude that both VAR and IRF (which is tied to

VAR) did not depict a fierce competition in the North American auto industry, at least not by studying monthly sales.

SUR results

The SUR equations were listed from equations 2.1 through 2.5. The difference between the VAR regression and SUR was the inclusion of exogenous variables in the SUR. To understand the dynamic of sales' interdependencies and their impact on each other, we discussed in detail that SUR might be more useful than VAR as it can examine the simultaneous interactions among firms' sales. Again, we start by analyzing the RMSE and R squares.

Table 4 displays the outcome of SUR. We noticed that the RMSE for Ford, GM, Honda, Toyota, and Nissan were 15,210, 29,163, 8,338, 11,167, and 11,181, respectively. While all R squares are high, the highest R square belonged to Toyota at 0.9, and the lowest to Nissan at 0.72. A comparison of VAR and SUR is presented in Table 5. We noticed that using the SUR methodology significantly improved the regression errors. Table 5 confirms that utilizing SUR can reduce the regression errors (RMSE) for GM, Ford, Honda, Toyota, and Nissan by 30, 46, 55, 60, and 30 percent, respectively. The same was true for R squares. SUR improved the R squares across all companies. The highest change was noticed for Toyota, a 60 percent increase, and the lowest was tied for GM and Nissan, a 30 percent increase for each.

Another noticeable change was the improvement of the variable significance in the SUR. We discovered that most variables that were not significant in VAR regression became significant in the SUR, implying 'simultaneous' rivalry. Therefore, SUR could capture a more intense rivalry among the competitors in the same strategic group than the VAR regression could detect. The problem is that we don't have the option of IRF with SUR methodology. As we noticed with the VAR regression, for SUR, the variables were not as significant which meant that they might not impact other variables (as was confirmed by the IRF, with the exception of Honda). Hence, we still don't know whether an intense competition exists, at least not by examining the sales.

On the other hand, we must conclude that moderate rivalry must exist among the firms belonging in the same strategic group, but with at a lesser intensity. This is because the comparison of Tables 2, 3, and 5, indicated that the VAR regression outperformed the simple regression, and the SUR outperformed the VAR. Since VAR and SUR allow for inclusion of the competitor's sales, the main reason for having a superior output must be due to competitors' sales impacting the companies' sales predictions (Tables 7 through 10 present the detail of the SUR results all automakers in the Appendix). To further examine the existence of rivalry among the members, we can rely on the correlation matrix and Granger causality.

Table 4. The SUR results.

Equation	RMSE	R^2	Chi^2	$P(Chi^2)$
GM	29,163	0.8437	8ac9	0.000
Ford	15,210	0.8885	1,323	0.000
Honda	8,338	0.8710	1,081	0.000
Toyota	11,167	0.9110	1,676	0.000
Nissan	11,181	0.7223	406	0.000

Table 5. VAR and SUR comparison.					
Automaker	GM	Ford	Honda	Toyota	Nissan
RMSE (VAR)	41,817	28,340	18,685	27,915	15,895
RMSE (SUR)	29,163	15,210	8,338	11,167	11,181
Improvement	30%	46%	55%	60%	30%
R^2 (VAR)	0.7077	0.6451	0.4904	0.4953	0.4907
R^2 (SUR)	0.8437	0.8885	0.8710	0.9110	0.7223
Improvement	19%	27%	43%	84%	47%

Correlation matrix and Granger causality results

Table 6 displays the results of the correlation matrix for the automakers' monthly sales. A correlation coefficient greater than 0.7 is considered strong, between 0.4 and 0.7 is moderate, between 0.2 and 0.4 is weak, and below 0.2 is considered no correlation. Table 6 presents the strongest correlation between sales of Honda and Toyota at 0.79. Ford's sales and the sales of Honda and Toyota are correlated moderately at 0.55 and 0.58, respectively. Some weak correlations exist between the sales of Ford and GM at 0.41 and between Toyota and Nissan at 0.48. All other correlations are insignificant.

Table 6. Correlation matrix for sales.					
	GM	Ford	Honda	Toyota	Nissan
GM	1.000				
Ford	0.418	1.000			
Honda	0.327	0.556	1.000		
Toyota	0.246	0.585	0.790	1.000	
Nissan	0.037	0.355	0.352	0.480	1.000
Note: Correlations above 0.7 is strong, between 0.5 and 0.7 is moderate, and below 0.5 is weak.					

However, correlation refers to a statistical relationship between two variables, where changes in one variable are associated with changes in another (positively or negatively). If two variables are correlated, it means they move together (in the same direction or the opposite). Correlation does not imply that one variable causes the change in the other. A common mistake is to assume that because two variables are correlated, one must cause the other. This is not necessarily accurate. Both variables might be correlated with a third variable. For example, some economic factors or governmental policies. On the other hand, causality implies a cause-and-effect relationship between two variables, where a change in one variable directly causes a change in another. Therefore, while some moderate correlation exists, to uncover the intensity of the rivalry among the competitors, analyzing causal relationships (VAR, SUR, or Granger Causality) are more important.

There are two types of causal relationships. First, is the true causality, a direct cause-and-effect relationship between two variables, where changes in one variable can directly lead to changes in another variable. Establishing such causality requires understanding the underlying mechanisms, conducting controlled experiments, or other explanations. This information is not readily available to the present scholars. The second is the Granger causality, a statistical concept used to determine whether one time series can predict another. While Granger is not a true causality, this temporal causal relationship is based on the predictability. For time series, with several potential predictors, to determine which variables provide significant predictive information about a target variable, studying the Granger causality is significant. If the Granger test is positive, we can say that the rivalry among the members of the same strategic group is significant because their sales impact each other's; otherwise, we cannot conclude that a member's sales have a significant impact on other members' sales.

Table 7 displays the results of the Granger causality among the sales of the five automakers, where the p-values show the significance. According to the Granger test, Honda's sales can impact GM's (p-value, 0.06) and be impacted by GM's sales (0.04). While Ford doesn't impact GM's sales, GM's sales can impact Ford's (0.05). The rest of the variables, however, display no significant impact on other variables at 90 and 95 percent confidence interval. Hence, we can conclude that we did not discover a strong causal relationship among the sales of these companies within the same strategic groups. additionally, we close by mentioning that studying the monthly sales did not portray an intense rivalry among the members of the same strategic group.

Table 7. The Granger Causality test results for sales. The beta coefficients are omitted due to clutter.

Company	Impacted by	P-value	Company	Impacted by	P-value
GM	Ford	0.223	Honda	GM	0.042**
GM	Honda	0.063*	Honda	Ford	0.591
GM	Toyota	0.341	Honda	Toyota	0.192
GM	Nissan	0.335	Honda	Nissan	0.310
Ford	GM	0.055*	Toyota	GM	0.338
Ford	Honda	0.263	Toyota	Ford	0.774
Ford	Toyota	0.419	Toyota	Honda	0.130
Ford	Nissan	0.107	Toyota	Nissan	0.216
			Nissan	GM	0.100
** Significant at 95%			Nissan	Ford	0.130
* Significant at 90%			Nissan	Honda	0.888
			Nissan	Toyota	0.784

MANAGERIAL IMPLICATIONS

This study discovered that while sales can indicate rivalry among the members within the North American auto industry, studying eleven years of sales data for GM, Ford, Honda, Toyota, and Nissan did not produce an intense rivalry as anticipated. However, the comparison of Tables 2, 3, and 5 indicated that some rivalry was present (at the weaker intensity). Nevertheless, in compliance with Fortsch et al. (2021), our analysis showed that it is beneficial for the companies to utilize the SUR regression because it outperformed the VAR and simple regressions. Moreover, we found that utilizing VAR and SUR significantly enchased all the regressions' R squared values by allowing the models to include other companions' sales information. The highest R squared value improvement was for Honda at 40 percent after using VAR, and 84 percent improvement for Toyota after using SUR. Similarly, all the regression errors (RMSEs) significantly decreased after using VAR and SUR.

The improvement in R squared values and the regression errors indicate that some competition exists among the strategic group because VAR and SUR models allow us to include the information about the competitor's sales into the model. However, many regression beta coefficients remained insignificant and IRF indicated that a shock in the sales of one company doesn't necessary impact any other company. The exception to this rule was Honda, whose sale's unpredictability impacted the other companies' sales prediction.

In summary, we believe that we could not conclude an intense competition among the firms in the North American auto industry, but rather a moderate rivalry. While the auto industry is known to be highly competitive, we were unable to demonstrate a high degree of competition. The lack of evidence of a strong competition could be due to our chosen methods (VAR, SUR, and Granger) or due to lack of existence of a fierce competition.

Finally, our findings suggest that studying time series data for sales may not be sufficient to study the competition withing the same strategic group. Other factors could indicate intense competition. For example, pricing, frequent price changes, and promotional activities can signal aggressive competition withing the same strategic group. We did not have access to the above information. Similarly, as the strategic group diagram indicates, market share and product diversity could be an indicator of fierce competition. Other factors such as product innovation, differentiation, capital investment, marketing, advertising, customer loyalty, mergers and acquisitions, and the customer switching costs could paint a more comprehensive picture of the competitive dynamics within the strategic group. We state that the strategic group diagram does not include many important indicators of competitiveness.

CONCLUSION AND RESEARCH LIMITATION

In this study, we focused on domestic sales data to understand the competition among North American Automakers. Contrary to conventional wisdom, our findings suggested that a strong rivalry among the five automakers was not evident by studying the companies' sales data. However, we believe that the unavailability of data may have limited our understanding as we could not include all automakers, or the sales data did not capture the rivalry that could have been captured by studying pricing, innovation, or other key indicators. Similarly, we obtained our data from the automotive website, and our conclusions are contingent on the accuracy of the data.

However, we emphasize that the rivalry is present, even though it might not be as strong as we had anticipated, because VAR model outperformed the simple regression as R squares and RMSEs improved significantly. Furthermore, the results of SUR outperformed the results of VAR, which indicates that the righthand side of the regression equation must include the competitors' sales' variables. Putting these findings together, we must conclude that some (moderate) rivalry is indicated.

Finally, we have considered a simple approach to linear relationships. Specifically, we employed the vector autoregressive regression (VAR) and seemingly unrelated regression (SUR). Due to the complex nature of the auto industry, our results may have been different if a nonlinear modeling approach had been considered. This means that the competitors' sales may have not been linearly connected. We suggest that further research should explore alternative approaches to uncovering the competition within the North American Automakers. We specifically encourage scholars to expand the scope of the analysis to include non-traditional models, pricing, advertising initiatives, and other key indicators to offer a better understanding of the competition.

APPENDIX

Table 8. SUR results for GM.			
GM	Coefficient	Error	P(value)
FS(t)	1.582	0.125	0.000
HS(t)	0.192	0.296	0.517
TS(t)	-0.063	0.223	0.775
NS(t)	-0.855	0.219	0.000
GMS(t-1)	0.385	0.075	0.000
FS(t-1)	-0.696	0.148	0.000
HS(t-1)	-0.124	0.293	0.671
TS(t-1)	0.532	0.229	0.020
NS(t-1)	-0.093	0.228	0.683
Constant	26,680	15,100	0.077

Table 9. SUR results for Ford.			
Ford	Coefficient	Error	P(value)
GMS(t)	0.423	0.033	0.000
HS(t)	-0.030	0.149	0.840
TS(t)	0.189	0.107	0.076
NS(t)	0.584	0.106	0.000
GMS(t-1)	-0.120	0.043	0.005
FS(t-1)	0.506	0.066	0.000
HS(t-1)	-0.059	0.150	0.604
TS(t-1)	-0.365	0.111	0.001
NS(t-1)	-0.046	0.116	0.691
Constant	14,497	7,697	0.060

Table 10. SUR results for Honda.			
Honda	Coefficient	Error	P(value)
GMS(t)	0.016	0.025	0.517
FS(t)	-0.009	0.048	0.840
TS(t)	0.655	0.038	0.000
NS(t)	-0.081	0.063	0.203
GMS(t-1)	-0.047	0.025	0.057
FS(t-1)	0.049	0.461	0.279
HS(t-1)	0.477	0.077	0.000
TS(t-1)	-0.362	0.061	0.000
NS(t-1)	0.155	0/066	0.019
Constant	1,515	4,511	0.737

Table 11. SUR results for Toyota.			
Toyota	Coefficient	Error	P(value)
GMS(t)	-0.009	0.033	0.775
FS(t)	0.107	0.060	0.076
HS(t)	1.134	0.066	0.000
NS(t)	0.404	0.076	0.000
GMS(t-1)	0.039	0.033	0.230
FS(t-1)	-0.112	0.058	0.056
HS(t-1)	-0.576	0.105	0.000
TS(t-1)	0.638	0.070	0.000
NS(t-1)	-0.405	0.083	0.000
Constant	6,701	5,923	0.258

Table 12. SUR results of Nissan.			
Nissan	Coefficient	Error	P(value)
GMS(t)	-0.132	0.033	0.000
FS(t)	0.338	0.061	0.000
HS(t)	-0.143	0.112	0.203
TS(t)	0.413	0.077	0.000
GMS(t-1)	0.220	0.033	0.516
FS(t-1)	-0.185	0.059	0.002
HS(t-1)	0.227	0.114	0.048
TS(t-1)	-0.294	0.085	0.001
NS(t-1)	0.577	0.072	0.000
Constant	4,802	5,938	0.419

REFERENCES

Abdel-Khalik, A.R. & El-Sheshai, K.M. (1984). Sales revenues: Time-series properties and predictions. *Journal of Forecasting*, 2(4), 351-362.

Alessandria, G., Kaboski, J.P., & Midrigan, V. (2010). The Great Trade Collapse of 2008-09: An Inventory Adjustment? IMF Economic Review, 58(2), 254-294.

Auh, S. and Menguc, B. (2005). Balancing exploration and exploitation: The moderating role of competitive intensity. Journal of Business Research, 58 (12), 1652-1661. https://doi.org/10.1016/j.jbusres.2004.11.007.

Aviv, Y. (2003). A Time-Series Framework for Supply-Chain Inventory Management. Operations Research, 51(2), 210-227.

Barlas, Y. & Gunduz, B. (2011). Demand forecasting and sharing strategies to reduce fluctuations and the bullwhip effect in supply chains. *Journal of the Operational Research Society*, 62(3), 458-473.

Barnett, W.P. (1997). The Dynamics of Competitive Intensity. *Administrative Science Quarterly*, 42 (1) (Mar 1997), 128-160.

Borah, A. & Tellis, G. J. (2016). Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands? *Journal of Marketing Research*, 53(2), 143–160. doi: 10.1509/jmr.13.0009.

Burt, R.S. (1992). Structural Holes: The Social Structure of Competition. Cambridge, MA: Harvard University Press.

Busse, M.R., Simester, D.I., & Zettlmeyer, F. (2010). The best price you'll ever get: The 2005 employee discount pricing promotions in the U.S. automotive industry. *Marketing Science*, 29(2), 268-209.

Cachón, G.P. & Olivares, M. (2010). Drivers of Finished-Goods Inventory in the U.S. Automobile Industry. *Management Science*, 56(1), 202-216.

Chiang, C-Y., Lin, W.T., & Suresh, N.C. (2016). An empirically-simulated investigation of the impact of demand forecasting on the bullwhip effect: Evidence from U.S. auto industry. *International Journal of Production Economics*, 177, 53–65.

Choi, J.H., Choi, S., Suresh, N. (2024). Inventory and financial performance in the pharmaceutical industry: what transpired over the last two decades? International Journal of Operations & Production Management. 44 (10), 1863-1890.

Cool, K., & I. Dierickx (1993). Rivalry, strategic groups and firm profitability. Strategic Management Journal, 14(1), pp. 47-61.

Copeland, A., Dunn, W., & Hall, G. (2011). Inventories and the automobile market. *The RAND Journal of Economics*, 42(1), 121-149.

Dickey, D.A. & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74, 427–431.

Dranove, D., Peteraf, M.A., & Shanley, M.T. (1998). Do strategic groups exist? an economic framework for analysis. *Strategic Management Journal*, 19, 1029-1044.

Esteban-Bravo, M. & Lado, N. (2011). Brand value in horizontal alliances: the case of twin-cars. *Journal of the Operational Research Society*, 62, 1533–1542.

Fisman, R. (2006). The Effect of Foreign Competition on Forecasting Bias. *The Review of Economics and Statistics*, 88(1), 61-68.

Galvin, P., Burton, N., Singh, P., Sarpong, D., Bach, N., & Teo, S. (2020). Network rivalry, competition and innovation. *Technological Forecasting and Social Change*, 161: 120253. https://doi.org/10.1016/j.techfore.2020.120253

Goolsbee, A.D. & Krueger, A.B. (2015). A retrospective looks at rescuing and restructuring General Motors and Chrysler. *The Journal of Economic Perspective*, 29(2), 3-23.

Grieco, P. L.E., Murry, C., & Yurukoglu, A. (2021). The evolution of market power in the US auto industry. *The Quarterly Journal of Economics*, 139 (2), 1201-1253. https://doi.org/10.1093/qje/qjad047

Hahn, C.K., Duplaga, E.D., & Hartley, J.L. (2000). Supply-Chain Synchronization: Lessons from Hyundai Motor Company. *Interfaces*, 30(4), 32–45.

Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press, Princeton, NJ.

Hannan, M.T., and Freeman, J. (1989) Organizational Ecology. Cambridge, MA: Harvard University Press.

Holweg, M. (2008). The Evolution of Competition in the Automotive Industry. In: Parry, G., Graves, A. (eds) Build To Order. Springer, London. https://doi.org/10.1007/978-1-84800-225-8_2

Hu, W.M., Xiao, J., & Zhou, X. (2014). Collusion or Competition? Interfirm Relationships in the Chinese Auto Industry. The Journal of Industrial Economics 62.1: 1-40.

Hunt, M. S. (1972). Competition in the major home appliance industry 1960-1970, doctoral dissertation, Harvard University.

Jan, T-S., & Hsiao, C-T. (2004). A four-role model of the automotive industry development in developing countries: a case in Taiwan. *Journal of the Operational Research Society*, 55, 1145–1155.

Jayarajan, D., Siddarth, S., & Silva-Risso, J. (2018). Cannibalization vs. competition: An empirical study of the impact of product durability on automobile demand. *International Journal of Research in Marketing*, 35(4), 641-660.

Karabakal, N., Günal, A., & Ritchie, W. (2000). Supply-Chain Analysis at Volkswagen of America. *Interfaces*, 30(4), 46-55.

Kirca, A.H., Randhawa, P., Talay, M.B., & Akdeniz, M.B. (2019). The interactive effects of product and brand portfolio strategies on brand performance: Longitudinal evidence from the U.S. automotive industry. *International Journal of Research in Marketing*, 37(2), 421-439. https://doi.org/10.1016/j.ijresmar.2019.09.003.

Liu, C-L. & Chen, S-L. (2013). Risk sharing in the supplier relations for the Taiwanese automotive industry. *Journal of the Operational Research Society*, 64, 365–371.

Liu, S. (2021). Competition and valuation: a case study of Tesla Motors. *IOP Conference Series: Earth and Environmental Science*, 692(2), 022103. IOP Publishing.

McGee, J. & Thomas, H. (1986). Strategic Groups: theory, research, & taxonomy. *Strategic Management Journal*, 7(2), 141-160.

McGee, J., & Thomas, H. (1992). Strategic groups and intra-industry competition. *International Review of Strategic Management*, 3, 77–99.

McNamara, G. M., Deephouse, D. L., & Luce, R. A. (2003). Competitive positioning within & across a strategic group structure: the performance of core, secondary & solitary firms. *Strategic Management Journal*, 24(2), 161-181.

Meckling, J. & Nahm, J. (2019). The politics of technology bans: Industrial policy competition and green goals for the auto industry. *Energy policy*, 126, 470-479.

Morris, D., Madzudzo, G., & Garcia-Perez, A. (2020). Cybersecurity threats in the auto industry: Tensions in the knowledge environment. *Technological Forecasting and Social Change*, 157, 120102.

Nagler, M.G. (2014). The Strategic Significance of Negative Externalities. *Managerial and Decision Economics*, 35(4), 247–57. doi:10.1002/mde.2611.

Nair, A., & Filer L. 2003. Cointegration of firm strategies within groups: a long run analysis of firm behavior in the Japanese steel industry. *Strategic Management Journal*, 24(2), 145–159.

Neale, J.J. & Willems S.P. (2009). Managing Inventory in Supply Chains with Nonstationary Demand. *Interfaces*, 39(5), 388-399.

Olivares, M. & Cachón, G.P. (2009). Competing Retailers and Inventory: An Empirical Investigation of General Motors 7 Dealerships in Isolated U.S. Markets. *Management Science*, 55(9), 1586-1604.

Phillips, P.C.B. & Perron, P. (1988), Testing for a unit root in time series regression. *Biometrika*, 75, 335-346.

Pichler, M. Krenmayr, N, Schneider, E., & Brand, U. (2021). EU industrial policy: Between modernization and transformation of the automotive industry. *Environmental Innovation and Societal Transitions*, 38, 140-152.

Porter, M.E. (1979). The structure within industries and companies' performance. *Review of Economics and Statistics*, 61, 214-219.

Porter, M.E. (1980). Competitive Strategy. The Free Press: New York.

Ramey, V.A. & Vine, D.J. (2006). Declining Volatility in the U.S. Automobile Industry. *The American Economic Review*, 96(5), 1876-1889.

Shoesmith, G.L. & Pinder, J.P. (2001). Potential Inventory Cost Reductions Using Advanced Time Series Forecasting Techniques. *Journal of the Operational Research Society*, 52(11), 1267-1275.

Syntetos, A.A., Boylan, J.E. & Disney, S.M. (2009). Forecasting for Inventory Planning: A 50-Year Review. *Journal of the Operational Research Society*, 60, s149-s160.

USA Automotive Spotlight (news): (https://www.selectusa.gov/automotive-industry-united-states). Automotive News data center: http://www.autonews.com/section/datacenter

Wochner, A., Grunow, M., Staeblein, T., & Stolletz, R. (2016). Planning for ramp-ups and new product introductions in the automotive industry: Extending sales and operations planning. *International Journal of Production Economics*, 182, 372–383.

Zellner, A. (1962). An efficient method of estimating Seemingly Unrelated Regression equations and tests of aggregation bias. *Journal of the American Statistical Association*, 57, 500–5009.

Zellner, A. (1963). Estimations for seemingly Unrelated Regression Equations: some finite sample results. *Journal of the American Statistical Association*, 58, 977–992.

Zeng, X., Dasgupta, S., & Weinberg, C.B. (2016). The competitive implications of a “no-haggle” pricing strategy when others negotiate: Findings from a natural experiment. *International Journal of Research in Marketing*, 33(4), 907-923.

DECISION SCIENCES INSTITUTE

Understanding Self-Congruence and Motivation in Travel Experience Sharing: A Generational Perspective

Parichat Kheawcha-um
College of Management, Yuan Ze University, Taoyuan, Taiwan
Email: parichat.kh@gmail.com

Lanlung Chiang
Department of Business Administration, National Taiwan University of Science and Technology, Taipei, Taiwan
Email: luke114ba@mail.ntust.edu.tw

Hsunchi Chu
College of Management, Yuan Ze University, Taoyuan, Taiwan
Email: chsunchi@saturn.yzu.edu.tw

ABSTRACT

In the digital era, social media serves as a powerful platform for tourists to share their travel experiences, influencing potential visitors and destination branding. While past research highlights the impact of user-generated content (UGC), not all tourists actively share their experiences online. This study examines the role of self-congruence and expectation confirmation in shaping tourists' intention to share travel experiences on social media. Using a survey of 413 Generation Y and Z tourists, regression analysis reveals significant generational differences. For Generation Y, both intrinsic and extrinsic motivation moderate the relationship between satisfaction and sharing intention, while for Generation Z, neither motivation type has a significant effect. These findings suggest that marketing strategies should be adapted to generational preferences, leveraging external incentives for Generation Y and alternative engagement strategies for Generation Z. Theoretical and practical implications are discussed.

KEYWORDS: Generational differences, Travel experience sharing, Self-congruity theory, Extrinsic motivation, Intrinsic motivation

1. INTRODUCTION

1.1 Research Background

In the digital era, technology has become increasingly integrated into various aspects of people's lives. It is undeniable that technological advancements have made life more convenient and comfortable. With the advancement of technology, people can connect more easily via social media. Social media platforms function as virtual spaces that facilitate human interaction, content sharing, and exchange thoughts online. (Islam, 2021). The increasing significance of social media in the tourism industry is evident as its applications continue to expand across both the business and consumer sides. (Nusair, 2020; Afren, 2024). On the business side, social media content provides valuable sources for destination managers and businesses sectors, aiding them in managing, creating, and planning their marketing strategies (Arica & Corbaci, 2020; Chu et al., 2020). Businesses strategically utilize social media to produce and manage professional content (firm-generated content) through their marketing teams, then deliver it widely to their target audience (Colicev et al., 2019). Moreover, when tourists share their experiences, photos, and reviews, it not only enhances the sense of community but also builds trust and credibility for brands. Engaging with this content enables tourism providers to connect more personally with their audience, respond to feedback, and showcase real experiences, ultimately strengthening brand loyalty and appeal (Lam et al., 2020). Thus, tourism service providers should continuously update and integrate social media strategies to improve engagement with potential visitors (customers), boosting market share and enhancing competitiveness (Keelson, 2024). From the consumer side, tourists engage with social media in two primary ways. Initially, they use it to find inspiration and reliable information for trip planning, shaping their interests and making choices based on content from travel businesses and fellow travelers. As mentioned by Dedeoglu et al. (2020), User-generated content serves as a crucial determinant in shaping these decisions. Second, tourists use social media for sharing purposes both in real-time and post-trip. The shared content can take various forms, including travel information, photos and videos documenting their experiences .(Arica & Corbaci, 2020; Sotiriadis, 2017).

Prior studies in the tourism and hospitality field have explored the significance of user-generated content (Hua et al., 2024). Several studies confirmed that tourists sharing their experiences on social media can improve tourist satisfaction and loyalty (Abbasi et al., 2024), enhance electronic word-of-mouth (eWOM) (Armutcu et al., 2023), and boost sales of products and services (Mariani & Borghi, 2021). Consequently, businesses should motivate satisfied guests to post their experiences in order to attract potential travelers (Arasli et al., 2024)."
Interestingly, while both tourism businesses and tourists create content on social media, studies indicate that content created by users is generally regarded as more credible than that produced by travel companies (Septiar & Omar, 2022). Yamagishi et al. (2024) found that tourist-generated content is often viewed as more reliable than information provided by travel companies and mainstream media. Similarly, Lien and Cao (2014) showed that content shared by close friends is inclined to be trusted by viewers. Although the UCG can be beneficial, tourism businesses and destination managers often find it challenging to motivate satisfied

tourists to contribute their own contents. Previous studies highlight that while many tourists use social media as an information source for travel, only a few actively share their experiences post-trip (Munar & Jacobsen, 2014; Chopra et al., 2024;). Hene, tourists are often less willing to post their experiences than businesses expect (Oliveira et al., 2020; Muharrir et al., 2023). This study aims to assist tourism businesses and destination managers in understanding the factors that influence travelers' willingness to share journeys experience on digital platforms. Consequently, it enables them to leverage tourists' voices to enhance promotion efforts, reach a wider audience, and attract potential travelers.

1.2 Research Objectives

This study aims to achieve the following research objectives:

1. To investigate tourists' sharing intention on social media platforms through the lens of Expectation Confirmation Theory (ECT)
2. To examine the role of self-congruence in influencing tourist satisfaction and travel experience sharing intention.
3. To explore the moderating effects of intrinsic and extrinsic motivation on the link between tourist satisfaction and experience sharing intention.
4. To examine how intrinsic and extrinsic motivation moderate the relationship between self-congruence and the intention to share experiences.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1 Expectation Confirmation Theory (ECT)

Expectation Confirmation Theory (ECT), introduced by Oliver (1980), explains how consumer satisfaction arises from the comparison between expectations and actual experiences. Initially developed for consumer behaviour and social psychology, ECT has been widely applied in tourism to understand how travellers' expectations influence satisfaction and decision-making. When experiences align with or exceed expectations, satisfaction increases; otherwise, dissatisfaction occurs (Oh et al., 2022). This framework is critical for tourism marketing, as aligning service performance with expectations can enhance satisfaction, loyalty, and positive word-of-mouth (Hoang et al., 2023). Hence, two variables were derived from this theory.

2.1.1 Expectation Confirmation (CF)

Expectation confirmation occurs when tourists compare their actual travel experiences with pre-trip expectations. This judgment influences satisfaction, which in turn impacts post-trip behaviours like revisiting and recommending destinations (Oliver, 1980). Shukla et al. (2023) identified three levels of confirmation: positive disconfirmation (expectations exceeded), negative disconfirmation (expectations unmet), and simple confirmation (expectations met). Prior studies highlight confirmation as a key driver of satisfaction and behavioural intentions, such as sharing experiences online (Nwatu et al., 2021). However, some scholars argue that direct post-purchase evaluations might be more reliable than expectation-based assessments

(Prakash, 1984). In this study, we propose that confirmation positively influences tourist satisfaction.

Hypothesis 1: Confirmation positively influences tourist satisfaction.

2.1.2 Tourist Satisfaction (SAT)

Satisfaction is a critical post-consumption factor in marketing and tourism research (Smith, 1994). It arises from the cognitive comparison between expected and actual experiences (Oliver, 1980). While some models assess satisfaction through expectation-performance comparisons, others focus on service quality (Cronin & Taylor, 1992) or fairness perceptions (Oliver & Swan, 1989). Prior studies suggest that satisfied tourists are inclined to post experiences online (Mohamad et al., 2022). However, some researchers question the strength of this relationship, with mixed findings regarding satisfaction's impact on electronic word-of-mouth (Serra-Cantallops, 2020). In this study, we examine the effects of tourist satisfaction and intention to share

Hypothesis 2: Tourist satisfaction positively influences the intention to share travel experiences on social media.

2.2 Self-Congruity Theory (SC)

Self-congruity theory explains how individuals prefer brands or destinations that align with their self-image (Sirgy, 1985). In tourism, self-congruence affects satisfaction, loyalty, and word-of-mouth intentions (Kim & Thapa, 2018; Cifci, 2022). This theory has been applied in diverse tourism contexts, including wine tourism (Pratt & Sparks, 2014), theme parks (Fu et al., 2020), and cultural destinations (Meeprom & Fakfare, 2021). Research indicates that travelers who see a strong alignment between their self-image and a destination's image tends to experience greater satisfaction (Cifci, 2022) and greater intention to share experiences (Yang & Lee, 2016). However, the results are not conclusive. Some studies argue that self-congruence influences revisitation more than recommendation behaviours (Malhotra, 1981). Hence, two hypotheses were developed based on this theory.

Hypothesis 3a: Self-congruence positively influences tourist satisfaction.
Hypothesis 3b: Self-congruence positively influences the intention to share travel experiences on social media.

2.3 Self-Determination Theory (SDT): Intrinsic and Extrinsic Motivation

Self-determination theory (Deci & Ryan, 1985) suggests that individuals are driven by intrinsic motivation, such as personal interest and enjoyment, or extrinsic motivation, which stems from external rewards and pressures. This framework has been widely applied to tourism research, particularly in understanding tourists' travel behaviours and social media engagement (Manhas & Ashraf, 2024). In this study, two moderating variables were developed from this theory.

2.3.1 Intrinsic Motivation (INM)

In the tourism context, several studies have examined intrinsic motivation in travel experience sharing. Arica et al. (2022) found that altruism, personal fulfilment, and self-actualization significantly encourage tourists' experience sharing on social networks. Intrinsic motivation drives individuals to share content for enjoyment, personal fulfilment, or altruism (Septiari & Omar, 2022). Prior studies confirm that intrinsic motivations, such as knowledge-sharing and self-expression, strongly influence travel experience sharing (Cetinsöz et al., 2024). Similarly, Oliveira et al. (2020) and Munar and Jacobsen (2014) highlighted some intrinsic motivations, such as altruism, as antecedents of sharing travel experiences on social media. However, research on the moderating role of intrinsic motivation in shaping satisfaction-driven sharing behaviours remains scarce.Interestingly, while García-Haro et al. (2024) proposed that altruism (an intrinsic motivation) is recognized as one of the key motivations for information sharing, their study revealed that altruism did not moderate the link between the perceived usefulness of social media and the intention to share travel content. Conversely, self-interest, which refers to personal benefits (i.e., tangible rewards and intangible incentives), was found to strengthen this relationship. In this study, we develop two hypotheses based on intrinsic behaviour.

> Hypothesis 4a: Intrinsic motivation positively moderates the relationship between tourist satisfaction and intention to share travel experiences on social media.
> Hypothesis 4b: Intrinsic motivation positively moderates the relationship between self-congruence and intention to share travel experiences on social media.

2.3.2 Extrinsic Motivation (EXM)

Extrinsic motivation stems from external rewards, such as monetary incentives, recognition, or social status (Ryan & Deci, 2000). Similar to economic exchange theory, individuals anticipate economic rewards for sharing their knowledge. (Bock and Kim,2002). Prior studies show that extrinsic motivators can enhance content-sharing behaviours, particularly in online communities (Daxböck, 2021). However, monetary incentives, in particular, have been debated, while some studies show that they boost participation, others argue they undermine intrinsic motivation and content authenticity (Bakshi et al., 2019).Interestingly, Nguyen et al. (2024) highlighted that Generation Z is not strongly motivated by financial compensation and benefits, as they often receive financial support from their parents. Similarly, Arriva (2025, February 11) revealed that extrinsic motivations, such as traditional loyalty programs, are not attractive to Generation Z if they do not resonate with their self-identity. In contrast, Abbasi et al. (2024) found that referral rewards and social recognition are key extrinsic motivators in digital tourism marketing. Despite these inconclusive findings and the limited study of extrinsic motivation as a moderator, this study proposes the following two hypotheses:

Hypothesis 5a: Extrinsic motivation positively moderates the relationship between tourist satisfaction and intention to share travel experiences on social media.
Hypothesis 5b: Extrinsic motivation positively moderates the relationship between self-congruence and intention to share travel experiences on social media.

2.4 Intention to Share Travel Experiences on Social Media (ITS)

Modern travellers actively share their experiences on social media and travel platforms, posting trip details, personal stories, photos, videos, and reviews. They also provide feedback, ratings, and recommendations to help others make informed travel decisions (Kang & Schuett, 2013; Hou et al., 2024; Tsai et al., 2020). Previous studies have identified several key factors influencing travel-related sharing behaviour, including altruism (Zucco et al., 2020), self-representation (Koo et al., 2015), hedonic enjoyment (Bakshi et al., 2019), monetary incentives (Li et al., 2024), and content reliability (Arica et al., 2022).

While both Generation Y and Z are internet-savvy, their motivations for sharing travel experiences differ significantly. Research highlights key distinctions in platform preferences (Șchiopu et al., 2023), motivational patterns (Agoes & Safari, 2024), and decision-making behaviors (Uysal, 2022), suggesting that each generation engages with social media in unique ways. Millennials are primarily driven by public acknowledgment, frequent positive feedback, gratification, and recognition, reflecting their reliance on extrinsic motivation (Pinzaru et al., 2016). In contrast, Generation Z appears less influenced by financial rewards or traditional perks, as many receive financial support from their parents (Nguen et al., 2024). Additionally, Arriva (2025) found that Gen Z disengages from traditional loyalty programs when rewards fail to align with their personal values. Rather than seeking external validation, Gen Z is more self-focused, using social media as a tool for self-expression and identity formation (Alqvist & Klaus, 2018; Yoanita et al., 2022; Andrew & Herdiansyah, 2024). Their sharing behavior is shaped by factors such as the need for uniqueness, self-actualization, travel experiences, and reflected self-appraisal (Anggarawati & Saputra, 2023). Moreover, social media serves as a crucial communication platform for this generation, reinforcing its role beyond simple content sharing (Siagian & Yuliana, 2023).

Despite growing research on Gen Z's social media engagement, literature gaps persist due to the lack of a solid theoretical foundation (Garcia-Haro et al., 2024). To address this, the present study integrates Expectation Confirmation Theory, Self-Congruence Theory, and Self-Determination Theory to explore the factors shaping tourists' sharing intentions, offering a deeper understanding of how motivations differ across generations.

3. CONCEPTUAL FRAMEWORK AND METHODOLOGY

This study integrates Expectation Confirmation Theory (ECT) by Oliver (1980), Self-Congruence Theory (SCT) by Sirgy & Su (2000), and Self-Determination Theory (SDT) (Deci & Ryan, 1985) to explore tourists' experience-sharing intention. ECT links expectation fulfillment to satisfaction, which, in turn, influences tourists' experiences and sharing intentions. However, satisfaction alone does not fully explain sharing behaviour. SCT extends ECT by adding a psychological dimension, showing that when tourists perceive a strong alignment between their self-identity and a destination, their satisfaction is reinforced, strengthening their intention to share. Yet, since only a small fraction of satisfied tourists contribute content (Oliveira et al., 2020), incorporating motivational factors from SDT enhances the model's practical application. Understanding tourists' motivational preferences helps marketers develop targeted strategies that encourage engagement based on whether intrinsic or extrinsic motivation drives their behaviour.

Figure 3-1 Conceptual Framework

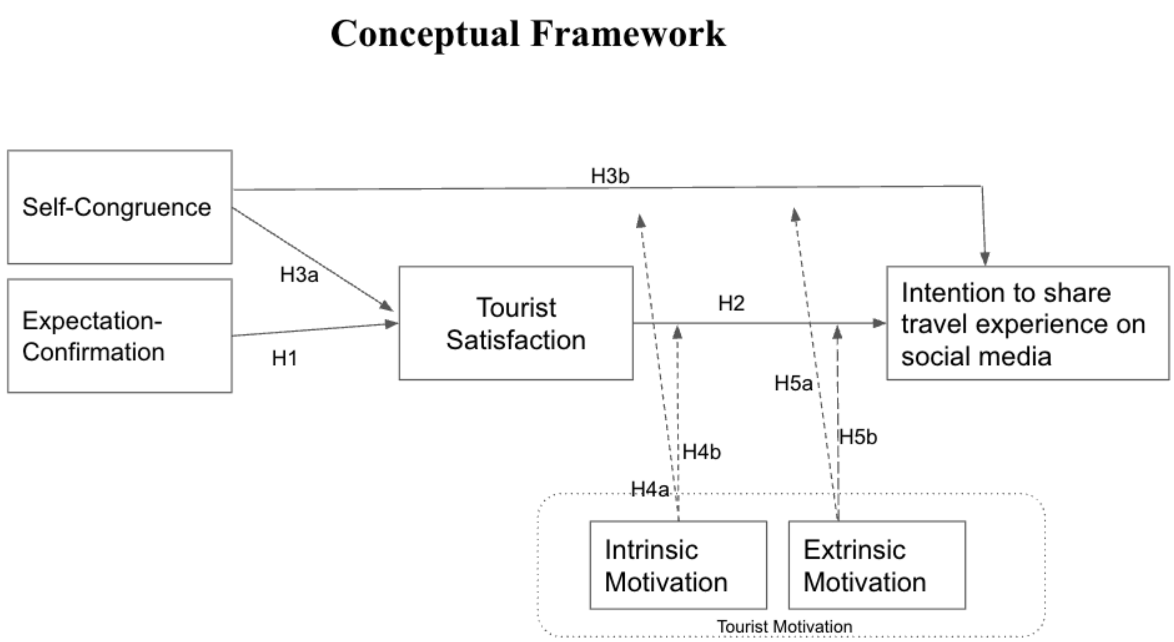


Figure 3-1 Conceptual model

3.1 Measurements

As presented in Figure 3-1, the conceptual framework of this study consists of six constructs, which were derived from and integrated with three established theories from past literature. Expectation Confirmation and Tourist Satisfaction were based on the Expectation-Confirmation

Theory (Oliver, 1980), which explains how expectations influence post-consumption satisfaction. Self-congruence represents four dimensions of alignment between self-image and a destination, following the Self-Congruity Theory (Sirgy & Su, 2000). Extrinsic and Intrinsic Motivation served as moderating variables and were developed from the Self-Determination Theory (Deci& Ryan, 1985). Lastly, the Intention to Share Travel Experiences was adapted from prior research on social sharing behaviour (Kumar et al., 2021). Four items for each construct were derived from previous research to ensure measurement ensures validity and conceptual alignment. Respondents rated their agreement levels on a five-point Likert scale, with responses ranging from 1 (strongly disagree) to 5 (strongly agree), respectively. Since the target respondents were Thai natives, the questionnaire was translated into Thai by a bilingual language expert fluent in both Thai and English. To ensure content validity, a pilot study was conducted with a Thai sample to assess readability and reliability. Minor refinements were made to improve clarity before the final online survey distribution. The questionnaire items used for measurement are shown in Table 3-1

Table 3-1 Construct and Items

Construct	Items	Source
Expectation confirmation (CF)	1. My trip experience was better than what I had expected. 2. The service I received during this holiday is better than I had expected. 3. The service delivered during this holiday is consistent with what I had expected 4. Overall, most of my trip expectations were confirmed	Sedera et al.(2017)
Tourist Satisfaction (SAT)	1. I feel satisfied with the service I received during this trip	Sedera et al.(2017)
	2. I am pleased with the experience I had during this trip	
	3. I am very content about choosing this trip	
	4. Overall, I am very satisfied with this trip	
Self Congruence (SC)	1. This [destination x] is consistent with how I see myself. [Actual SC]	Sirgy & Su, C. (2000)
	2. This [destination x] is consistent with how I like to see myself. [Ideal SC]	
	3. This [destination x] is consistent with how I believe others see me. [Social SC]	
	4. This [destination x] is consistent with how I would like others to see me. [Ideal Social SC]	

Intrinsic Motivation (INM)	I like to share my impressions through social media.	Wang et al. (2014)
	I want to help others.	
	Sharing my travel knowledge and information through social media is pleasant.	
	I want to prevent people from facing bad experiences	
Extrinsic Motivation (EXM)	I want to be more recognize for my experience	Wang et al. (2014), Park & Lee (2021)
	I want to earn money with my social media views.	
	I want to receive monetary reward in returns for my sharing	
	I want to receive additional rewards in returns for me sharing experience	
Intention to share travel experience on social media (ITS)	I will upload my opinion about my trip on SNSs	Kumar et al.(2021)
	I will share my reviews about my trip on SNSs	
	On SNSs, I will inform friends about my trip	
	I will post information about my trip on SNSs	

As the questionnaire is self-reported, Common Method Bias (CMB) was assessed using the Marker Variable Approach with an income-related variable (INCOME), which was theoretically unrelated to the independent variable (IV), dependent variable (DV), and moderators. Bivariate Correlation Analysis confirmed INCOME as a valid marker variable, showing low and non-significant correlations with key constructs, ranging from -0.072 to 0.078, $p > 0.05$. (Lindell & Whitney, 2001). Then, Partial Correlation Analysis further tested whether controlling for INCOME affected relationships between the main constructs. Results showed minimal changes in Pearson's coefficients (±0.001), and significance levels remained stable, indicating that common method variance did not substantially influence the findings. Thus, CMB does not pose a significant threat to this study's validity.

3.2 Data Collection and Analysis Procedure

Data were gathered using an online survey shared across multiple digital platforms, including social media and travel-related communities, enabling broad participation. To enhance data reliability and prevent duplicate responses, the survey system was configured to restrict multiple submissions by limiting one response per IP address. Initially, 542 responses were obtained. Then, three screening criteria were applied to ensure the target sample: (1) respondents must belong to either Generation Y or Generation Z based on their birth year, (2) they must actively use at least one personal social media platform, and (3) they must have travelled to a different city within the past six months. After applying these criteria, 413 valid responses were retained for analysis.

4. DATA ANALYSIS AND RESULTS

For data analysis, Jamovi was used to assess model fit indices, as well as to evaluate the reliability and validity of the constructs. SPSS was employed for descriptive analysis of respondent demographics. Direct effects were tested using regression analysis, while the moderating effects of intrinsic and extrinsic motivation were analyzed through interaction effect testing in regression analysis. All data analyses were performed using a p-value threshold of 0.05 to determine the strength and direction of relationships among variables.

4.1 Descriptive analysis of respondent Characteristics

A total of 542 responses were collected, with 413 valid after excluding 24 respondents outside Generation Y and Z and 105 who had taken their last trip more than six months ago. The sample comprised 57.9% males (239 respondents) and 42.1% females (169 respondents). Generation Z (ages 13–28) comprised 57.1%, while Generation Y (ages 29–44) accounted for 42.9%. Most respondents had an income below 15,000 Thai baht (44.1%), followed by 15,001–30,000 baht (36.1%). The top occupations were students (33.9%) and office workers (25.4%), followed by business owners, freelancers, and civil servants, respectively. The descriptive analysis is as follows.

Table 4-1 Respondent Characteristics

Item	Category	Frequency	Percentage rates (%)
Gender	Male	239	57.9
	Female	169	40.9
	other	5	1.2
	Total	413	100
Age	GenZ	236	57.1
	GenY	177	42.9
	Total	413	100
Income	0-15,000	182	44.1
	15,001-30,000	149	36.1
	30,001-45,000	49	11.9
	45,001-60,000	19	4.6
	>60,000	14	3.4
	Total	413	100
Occupation	student	140	33.9
	office_worker	105	25.4
	business_owner	49	11.9

	freelance	48	11.6
	civil servant	46	11.1
	other	25	6.1
	Total	413	100
Education	< high_school	5	1.2
	high_school	65	15.7
	Bachelor_equivalent	311	75.3
	Master	27	6.5
	PhD	5	1.2
	Total	413	100

4.2 Measurement Model Assessment

Jamovi program was used to analyze the measurement model to ensure that all constructs were measured reliably and validly, confirming their appropriateness for further analysis.

4.2.1 Model Fit Indices

The results from the model fit indices show that RMSEA of 0.07 falls within the ≤ 0.08 threshold, indicating an acceptable fit (MacCallum et al., 1996). Similarly, the χ^2/df ratio of 3.02 meets the ≤ 5 criteria, confirming a reasonable fit (Marsh & Hocevar, 1985). The GFI of 0.962 surpasses the ≥ 0.90 benchmark, supporting strong model fitness (Hu & Bentler, 1998), while the AGFI of 0.95 meets the ≥ 0.90 standard for acceptability (Tabachnick 2007). Additionally, the CFI of 0.95 exceeds the ≥ 0.95 threshold, reinforcing strong data validity (West et al., 2012). These results indicate that the model is well-fitted and suitable for further analysis.

Table 4-2 Results of model fit

Index	Values
Chi-square/degrees of freedom	3.02
RMSEA	0.07
GFI	0.96
AGFI	0.95
CFI	0.95

4.2.2 Reliability and Validity Assessment of Constructs

After confirming model fit, reliability and validity assessments were conducted. Cronbach's Alpha (α) values exceed the 0.70 threshold, ranging between 0.904 and 0.946, indicating strong reliability in the measurement model. Composite Reliability (CR) values span 0.91 to 0.95. Average Variance Extracted (AVE) values fall between 0.71 and 0.82, confirming that all constructs effectively represent their associated indicators.Hence, these results establish that the constructs are both reliable and valid for hypothesis testing.

Table 4-3. Summary of the Cronbach's α, Composite Reliability and Average Variance Extracted

Variables	Items	Cronbach's α	CR	AVE
Expectation Confirmation	CF1, CF2, CF3,CF4	0.93	0.93	0.77
Tourist Satisfaction	SAT1,SAT2,SAT3,SAT4	0.95	0.95	0.82
Self Congruence	SC1, SC2, SC3, SC4	0.94	0.94	0.78
Intrinsic Motivation	IM1, IM2,IM3,IM4,	0.90	0.91	0.71
Extrinsic Motivation	EX1, EX2, EX3,EX4	0.91	0.92	0.75
Intention to share	INT1, INT2, INT3, INT4	0.93	0.93	0.78

4.3 Hypothesis Testing

4.3.1 Results of Direct Effect Testing (H1, H2, H3a, H3b)

The analysis of direct effects reveals significant associations between the key constructs. H1: Confirmation positively influences tourist satisfaction was supported, as the results revealed a significant impact of confirmation on tourist satisfaction ($\beta = 0.782$, $t = 25.477$, $p < 0.001$), accounting for 61.2% of the variance. Furthermore, H2: Tourist satisfaction positively influences the sharing intention was also supported ($\beta = 0.506$, $t = 11.884$, $p < 0.001$), accounting for 25.6% of the variance. Additionally, H3a: Self-congruence positively influences tourist satisfaction was confirmed. The relationship between self-congruence and tourist satisfaction was substantial ($\beta = 0.680$, $t = 18.823$, $p < 0.001$), with 46.3% of the variance explained. Similarly, H3b: Self-congruence positively influences the sharing intention was confirmed. The findings indicate that self-congruence had a significant effect on tourist sharing intention ($\beta = 0.571$, $t = 14.096$, $p < 0.001$), explaining 32.6% of the variance.

Table 4-4 Hypotheses Testing. (Direct effects)

Hypothesis	Beta (β)	t-value	p-value	R^2	Results
H1: Confirmation → Tourist Satisfaction	0.782	25.477	<0.001	0.612	Supported
H2: Tourist Satisfaction → Intention to Share	0.506	11.884	<0.001	0.256	Supported

H3a: Self-Congruence → Tourist Satisfaction	0.68	18.823	<0.001	0.463	Supported
H3b: Self-Congruence → Intention to Share	0.571	14.096	<0.001	0.326	Supported

4.3.2. Moderating Effect Results on Generation Y Tourists

For H4a, the results indicate that Intrinsic Motivation (INM) significantly moderates the link between Tourist Satisfaction (SAT) and experience-sharing intention (ITS) ($\beta = 0.154$, $t = 2.629$, $p = 0.009$), showing that the effect of satisfaction on sharing intention changes based on the level of intrinsic motivation. However, for H4b, the moderating effect of Intrinsic Motivation in the association between Self-Congruence (SC) and sharing intention was not supported ($\beta = -0.013$, $t = -0.227$, $p = 0.821$), indicating that intrinsic motivation does not influence the impact of self-congruence on tourists' sharing intention.
For Extrinsic Motivation (EXM), the findings highlight a moderating influence on H5a and H5b. For H5a, Extrinsic Motivation significantly shapes the link between Tourist Satisfaction and Sharing Intention ($\beta = 0.120$, $t = 2.059$, $p = 0.041$), suggesting that external rewards or recognition enhances the effect of satisfaction on sharing behaviour. For H5b, Extrinsic Motivation substantially moderates the connection between Self-Congruence and Intention to Share ($\beta = 0.122$, $t = 2.136$, $p = 0.034$), implying that individuals driven by extrinsic motivation tend to share experience if they feel aligned with the destination or travel experience.

Table 4-5 Results of Moderation Analysis For Generation Y tourists

Hypothesis	Path (Moderation Effect)	B (Beta Coefficient)	t-value	p-value	Result
H4a	Z_SAT_INM_interaction → ITS	0.154	2.629	0.009	Supported
H4b	Z_SC_INM_interaction → ITS	-0.013	-0.227	0.821	Not Supported
H5a	Z_SAT_EXM_interaction → ITS	0.12	2.059	0.041	Supported
H5b	Z_SC_EXM_interaction → ITS	0.122	2.136	0.034	Supported

4.3.3 Moderating Effect Results on Generation Z Tourists

The results indicate that neither Intrinsic Motivation (INM) nor Extrinsic Motivation (EXM) significantly moderates the relationships between Tourist Satisfaction (SAT), Self-Congruence (SC), and Intention to Share Travel Experiences (ITS) among Generation Z respondents.

For H4a, the moderating effect of Intrinsic Motivation on the link between Tourist Satisfaction and sharing intention was not statistically significant (β = 0.001, t = 0.027, p = 0.978), indicating that satisfaction's influence on sharing intention remains unchanged regardless of intrinsic motivation levels. Similarly, for H4b, Intrinsic Motivation did not moderate the link between Self-Congruence and Intention to Share (β = 0.019, t = 0.694, p = 0.488). Hence, it did not alter the influence of self-congruence on Generation Z tourists' sharing intention.

For Extrinsic Motivation, the findings similarly show no significant moderation effects among Generation Z tourists. Results from H5a show that Extrinsic Motivation did not alter the effect of Tourist Satisfaction on Sharing Intention (β = -0.008, t = -0.248, p = 0.804), implying that external rewards or incentives do not enhance the effect of satisfaction on sharing behaviour. Likewise, Extrinsic Motivation did not significantly moderate the link between Self-Congruence and Sharing Intention (β = 0.006, t = 0.192, p = 0.848). Hence, Hypothesis 5b was not supported, suggesting that external motivators do not strengthen the influence of self-congruence on sharing intention.

Table 4-6 Results of Moderation Analysis For Generation Z tourists

Hypothesis	Moderation Effect	B (Beta Coefficient)	t-value	p-value	Results
H4a	Z_SAT_INM_interaction → ITS	0.001	0.027	0.978	Not Supported
H4b	Z_SC_INM_interaction → ITS	0.018	0.694	0.488	Not Supported
H5a	Z_SAT_EXM_interaction → ITS	-0.007	-0.248	0.804	Not Supported
H5b	Z_SC_EXM_interaction → ITS	0.006	0.192	0.848	Not Supported

5. DISCUSSION AND CONCLUSION

5.1 Discussion

As H1 confirms, the study shows that Expectation Confirmation positively influences Tourist Satisfaction among both Generation Y and Z tourists. This suggests that when tourists' pre-trip expectations align with their actual travel experiences, they experience higher satisfaction. This result is consistent with previous studies by Oh et al. (2022), Leou and Wang (2023), and Hoang et al. (2021), who highlighted that expectation confirmation significantly affects tourist satisfaction. Hence, the findings underscore the importance of managing tourist expectations effectively. To ensure high tourist satisfaction, destination marketers should focus on transparent and engaging pre-trip communication, including realistic yet appealing representations of destinations through social media, travel blogs, and official websites.

For H2, the results indicate that satisfied travelers tend to post their experience online. This is similar to a previous study by Widiana & Novani (2022) who indicated that satisfied tourists tend to have a stronger intention to share user-generated content (UGC) (). Additionally, this is supported by studies such as Kang, & Schuett (2014) found that tourists with positive travel experiences often intend to share journeys online. To encourage more tourists to share experiences online , marketers should prioritize engaging with satisfied tourists, as their stories generate a more positive impact compared to those shared by unsatisfied travelers. However, prior research highlights that not many tourists actually share their experiences online (Munar & Jacobsen, 2014), suggesting that additional stimuli are needed to strengthen this relationship.

As H3a and H3b are supported, the study finds that self-congruence significantly influences both tourist satisfaction and the sharing intention. This means when tourists perceive a destination aligns with their self-image, they experience greater satisfaction and feel more motivated to share experiences online. This finding is supported by several previous studies in various tourism contexts. Regarding self-congruence and satisfaction, Kim and Thapa (2018) confirmed that self-congruence influenced satisfaction in cultural tourism, while Cifci (2022) confirmed that self-congruence plays a crucial role in enhancing satisfaction at faith-based tourism sites, where travelers feel a strong personal connection with the experience. In terms of sharing intention, Yang and Lee (2016) demonstrated that self-congruence significantly influences tourists' intention to recommend destinations, while Luna-Cortés et al. (2019) reveal that travelers who perceive a strong personal connection with their travel destinations have greater chances to enhance sharing behavior.
As self-congruence is proven to positively influence tourist satisfaction and sharing intention , marketers should focus on enhancing destination branding to align with travelers’ self-concept. Instead of promoting what tourists can see at the destination, change it to why people like you will fall in love with our destination. Suggested marketing campaigns include creating authentic narratives that reflect diverse traveler personalities, emphasizing unique aspects of the destination that resonate with specific audience segments, and utilizing storytelling techniques that allow tourists to see themselves in the experience. Strengthening brand associations through personalized travel recommendations and relatable social media content can further solidify the connection between tourists and destinations, increasing their willingness to share experiences.

Remarkably, the results significantly demonstrate that Generation Y and Z tourists require different stimuli to enhance the relationships between satisfaction and sharing intention, as well as self-congruence and intention to share. For Generation Y, H4a confirms that intrinsic motivation strengthens the link between tourist satisfaction and the intention to share travel experiences on social media. This suggests that when Generation Y tourists are intrinsically motivated, their satisfaction with a destination has an even stronger impact on their likelihood of sharing experiences online. While the moderation effects of intrinsic motivation are limited in this context, previous studies highlight its importance, which could support this finding. Septiari & Omar (2022) and Oliveira et al. (2020) confirm that personal fulfillment and enjoyment drive travel-sharing behavior. Additionally, Wang et al. (2014) found that intrinsically motivated travelers tend to voluntarily share their experiences online. Therefore, marketers should enhance tourists' intrinsic motivation by showing how their recommendations influence others on social media. Campaigns like "Share Your Hidden Gem" or "Must Try...!" can inspire altruism and make travel sharing more impactful. Notably, while H4b is not supported, this suggests that although tourists may feel personally connected to a destination, their intrinsic motivation alone does not necessarily increase their likelihood of sharing their experiences online. As a result, marketers should focus on extrinsic motivation instead of only focus on enhancing tourist intrinsic motivation. Hence the moderating effect of extrinsic motivation were proved as significant in H5a and H5b.

For Generation Y tourists, the results from H5a and H5b confirm that extrinsic motivation enhances the influence of both tourist satisfaction and self-congruence on the sharing intention. This suggests that external rewards, recognition, and incentives further encourage satisfied and self-congruent tourists in sharing experiences online. Although studies on the moderating effect of extrinsic motivation on tourist behavioral intention are scarce, findings from direct-effect studies support the significant role of extrinsic motivation. For instance, Le (2022) and Wang et al. (2014) found that extrinsic motivators significantly increase tourists' willingness to share information online. Likewise, Chen & Law (2016) confirmed that social recognition and referral reward systems encourage content creation on digital platforms.To maximize social media engagement, marketers should strategically implement extrinsic incentives that motivate tourists to inspire them. Reward programs, such as discounts, referral bonuses, or exclusive perks, can encourage sharing behavior. Enhancing social recognition through "elite traveler" programs or featuring top contributors on official platforms can further appeal to status-driven tourists. Additionally, gamification strategies, such as leaderboards, loyalty points, and travel-sharing challenges, can create a fun and competitive environment that sustains engagement.

The lack of support for H4a, H4b, H5a, and H5b suggests that neither intrinsic nor extrinsic motivation significantly moderates the relationships between tourist satisfaction, self-congruence, and sharing intention among Gen Z tourists. This highlights a generational shift, where Gen Z is less influenced by internal fulfillment or external rewards in travel sharing.The insignificance of extrinsic motivation aligns with Arriva (2025), who noted Gen Z’s disengagement from traditional loyalty programs when rewards lack personal relevance. Similarly, Nguyen et al. (2024) found that financial support from parents reduces their reliance on compensation and benefits. Prior studies emphasize generational differences in platform preferences (Șchiopu et al., 2023), motivational patterns (Agoes & Safari, 2024), and decision-making behaviors (Uysal, 2022), reinforcing the need for tailored engagement strategies across generations. Unlike Generation Y, who are driven by self-fulfilment and identity creation Alqvist & Klaus (2018) making their sharing intention be moderated by recognition and incentives, Generation Z's sharing behaviour is primarily self-driven, guided by their digital identity and self-expression (Audrew & Herdiansyah, 2024). To engage this cohort, brands should focus on empowering user-generated content (UGC) in a way that enhances self-expression. Strategies such as co-creation campaigns (e.g., letting users contribute to destination storytelling), interactive filters, and exclusive content personalization features will align with their desire to craft a unique digital persona. Additionally, marketers should engage this generation through their favourite platforms such as Tiktok and Instagram (Tubalawony, 2023; Levak, T., & Barić Šelmić, S. 2018). and content patterns they like (e.g., Instagram Stories, TikTok challenges) to maximize their engagement. Lastly, traditional monetary incentives, check-in discounts, and loyalty points should be replaced with personalized, experience-driven rewards (Arrivia 2025, February 11). Hence, Gen Z is highly selective in what they share, as their social media presence is an extension of their identity rather than just a means of communication (Siagian & Yuliana,2023). Consequently the alternative motivational factors should be designed to resonate with generation Z preferences rather than focus on financial reward or enhancing their altruistic motivation.

5.2 Conclusion

This study examines the factors influencing travel experience sharing intention on social media by integrating Expectation Confirmation Theory, Self-Congruence Theory, and Self-Determination Theory. The findings confirm that expectation confirmation and self-congruence significantly enhance tourist satisfaction and sharing intention, emphasizing the importance of positive travel experiences and personal identity alignment with destinations. However, generational differences emerged, particularly in the moderating role of intrinsic and extrinsic motivation. For Generation Y, extrinsic motivation strengthens the link between satisfaction and sharing intention, as well as self-congruence and sharing intention, while intrinsic motivation moderates only the link between satisfaction and sharing. In contrast, for Generation Z, neither intrinsic nor extrinsic motivation significantly moderates any of the proposed relationships, highlighting the need for alternative engagement strategies. These insights suggest that while intrinsic and extrinsic motivation remain effective for Generation Y, marketers must explore new approaches to encourage travel content sharing among Generation Z.

Theoretically, this study makes a theoretical contribution by filling three key research gaps in the literature: (1) the lack of a comprehensive model integrating multiple perspectives to explore tourists' behavior in sharing experiences online (2) the limited exploration of intrinsic and extrinsic motivations as moderators between satisfaction and sharing intention, and (3) the absence of research on how self-congruence interacts with these motivations to influence sharing behaviour. Moreover, this study sheds new light on generational differences, revealing that while intrinsic and extrinsic motivation effectively moderate sharing intention among Generation Y, they do not play the same role for Generation Z. This challenges existing motivational frameworks and calls for new theoretical models to explain the sharing behaviour of young tourists.

Practically, while this study highlights the generational differences in motivational influences, tourism businesses and destination marketer could utilize the results to create their marketing accordingly. Hence, the suggested marketing plans are as follows: While expectation confirmation and self-congruence significantly influence travel experiences sharing online among both Generation Y and Z, marketers should design campaigns reinforcing these psychological drivers. To enhance expectation confirmation, campaigns should deliver realistic yet engaging travel portrayals through user-generated content, immersive previews, and real-time testimonials to bridge the gap between expectations and actual experiences. Interactive trip-planning tools and AI-driven personalized recommendations can help ensure that travellers' pre-trip expectations align with their actual journeys. In addition, since self-congruence plays a significant role by influencing sharing intention, marketing campaigns should be structured to reflect travellers' identities and values. Rewards that align with self-image can enhance engagement, like an eco-friendly tote bag for sustainable travellers or a welcome mocktail in an elegant glass for those who value prestige.

Next, to effectively motivate Generation Y sharing intention, marketers should leverage extrinsic incentives such as recognition, monetary rewards, and exclusive perks.

The UCG competition campaign that judges by anonymous vote and grants by money reward is recommended since they simultaneously enhance the need for recognition and monetary rewards for Generation Y tourists. Moreover, A "Travel Influencer for a Day" campaign can feature top sharers on brand pages, offer free upgrades, or provide VIP experiences, reinforcing status and achievement. To enhance intrinsic motivation, marketers can launch a "Travel for Good" initiative, where users share experiences about sustainable travel or community impact, with top posts earning charity donations in their name or exclusive eco-tour experiences. A "Memories That Matter" campaign can invite travellers to share meaningful travel moments, rewarding heartfelt stories with personalized keepsakes or exclusive experience upgrades. By combining financial rewards, social validation, and purpose-driven engagement, marketers can effectively inspire Gen Y. Since neither intrinsic nor extrinsic motivation drives Gen Z's sharing behavior, marketers should consider replacing traditional incentives with identity-based engagement when targeting this cohort. Rather than offering discounts or loyalty points, brands might design rewards that align with Gen Z's self-expression. Experiential incentives, such as concert tickets for their favorite bands or VIP access to exclusive events that resonate with their self-identity, could enhance their sharing intention by reinforcing their self-representation. Further engagement with Gen Z may require marketers to implement interactive and community-driven campaigns, as this generation appears to prefer short-lived, dynamic content compared to others. For example, branded TikTok challenges that encourage users to showcase their unique travel experiences could drive engagement and virality, especially when supported by influencers and trendsetters. Similarly, Instagram Reels campaigns featuring interactive filters, AR-based experiences, or remixable content might create shareable moments that seamlessly integrate with their digital identity.

Lastly, while this study lacks an exploration of alternative motivational factors, as it only examined the moderating effects of extrinsic and intrinsic motivation, future research should explore additional psychological and digital influences that may drive Generation Z's travel-sharing behaviour. Factors such as social comparison, digital algorithms, and fear of missing out (FOMO) could offer deeper insights into the mechanisms shaping Generation Z's engagement with travel content. Furthermore, conducting in-depth interviews with Gen Z travellers could provide a more nuanced understanding of their preferences, allowing for the identification of effective content strategies and marketing approaches tailored to their digital behaviours. By integrating these perspectives, future studies can offer deeper insights and ultimately inspire Generation Z to create more content.

REFERENCES

Abbasi, A. Z., Rather, R. A., Hooi Ting, D., Nisar, S., Hussain, K., Khwaja, M. G., & Shamim, A. (2024). Exploring tourism-generated social media communication, brand equity, satisfaction, and loyalty: A PLS-SEM-based multi-sequential approach. *Journal of Vacation Marketing, 30*(1), 93–109.

Afren, S. (2024). The role of digital marketing in promoting the tourism business: A study of the use of social media in promoting travel. *World Journal of Advanced Research and Reviews, 21*(1), 272–287.

Agoes, A., & Safari, A. (2024, July). Motivational patterns for engaging in heritage bus city tours among youth tourists (Generation Y and Z): An unsupervised machine learning analysis. *IOP Conference Series: Earth and Environmental Science, 1366*(1), 012044. IOP Publishing.

Alqvist, J., & Klaus, E. (2018). Watch out for Generation Z: A qualitative research about generational differences and similarities towards influence marketing within lifestyle brands. *Lund University Student Papers*. http://lup.lub.lu.se/student-papers/record/8962272

Anggarawati, S., & Saputra, F. E. (2023). If I don't share my unforgettable journey, I'll lose it! Young travellers' propensity to share e-WOM on social media. *Frontiers in Business and Economics, 2*(1), 1–9. https://doi.org/10.56225/finbe.v2i1.189

Arasli, H., Saydam, M. B., Gunay, T., & Jafari, K. (2021). Key attributes of Muslim-friendly hotels' service quality: Voices from Booking.com. *Journal of Islamic Marketing, 14*(1), 106–127.

Arica, R., & Çorbaci, A. (2020). The mediating role of the tourists' citizenship behavior between value co-creation and satisfaction. *Advances in Hospitality and Tourism Research (AHTR), 8*(1), 125–150.

Arica, R., Cobanoglu, C., Cakir, O., Corbaci, A., Hsu, M. J., & Della Corte, V. (2022). Travel experience sharing on social media: Effects of the importance attached to content sharing and what factors inhibit and facilitate it. *International Journal of Contemporary Hospitality Management, 34*(4), 1566–1586.

Armutcu, B., Tan, A., Amponsah, M., Parida, S., & Ramkissoon, H. (2023). Tourist behavior: The role of digital marketing and social media. *Acta Psychologica, 240*, 104025.

Arrivia. (2025, February 11). Adapting loyalty programs for Gen Z travelers. *Arrivia*. https://www.arrivia.com/insights/adapting-loyalty-programs-gen-z-travelers/

Audrew, E., & Herdiansyah, H. (2024). The impact of self-esteem and media literacy on personal branding strategies on Instagram among Generation Z. *Ettisal: Journal of Communication, 9*(2). https://doi.org/10.21111/ejoc.v9i2.11758

Bakshi, S., Dogra, N., & Gupta, A. (2019). What motivates posting online travel reviews? Integrating gratifications with technological acceptance factors. *Tourism and Hospitality Management, 25*(2), 335–354.

Bentler, P. M. (1995). *EQS structural equations program manual* (Vol. 6). Multivariate Software.

Bock, G. W., & Kim, Y. G. (2002). Breaking the myths of rewards: An exploratory study of attitudes about knowledge sharing. *Information Resources Management Journal (IRMJ, 15*(2), 14–21.

Cetinsöz, B. C., Cakır, O., Durdu, K. M., & Arıca, R. (2024). Examining the effects of importance attached to content sharing and knowledge-sharing facilitators on tourists' actual travel experience-sharing behavior. *Tourism: An International Interdisciplinary Journal, 72*(4), 566–580.

Chen, Y. F., & Law, R. (2016). A review of research on electronic word-of-mouth in hospitality and tourism management. *International Journal of Hospitality & Tourism Administration, 17*(4), 347–372.

Chopra, I. P., Lim, W. M., & Jain, T. (2024). Electronic word-of-mouth on social networking sites: What inspires travelers to engage in opinion seeking, opinion passing, and opinion giving? *Tourism Recreation Research, 49*(4), 726–739.

Chu, S. C., Deng, T., & Cheng, H. (2020). The role of social media advertising in hospitality, tourism, and travel: A literature review and research agenda. *International Journal of Contemporary Hospitality Management, 32*(11), 3419–3438.

Cifci, I. (2022). Testing self-congruity theory in Bektashi faith destinations: The roles of memorable tourism experience and destination attachment. *Journal of Vacation Marketing, 28*(1), 3–19.

Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user-generated content and the stages of the marketing funnel. *International Journal of Research in Marketing, 36*(1), 100–116.

Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56*(3), 55–68.

Daxböck, J., Dulbecco, M. L., Kursite, S., Nilsen, T. K., Rus, A. D., Yu, J., & Egger, R. (2021). The implicit and explicit motivations of tourist behavior in sharing travel photographs on Instagram: A path and cluster analysis. In *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference, January 19–22, 2021* (pp. 244–255). Springer International Publishing.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer.

Dedeoğlu, B. B., Taheri, B., Okumus, F., & Gannon, M. (2020). Understanding the importance that consumers attach to social media sharing (ISMS): Scale development and validation. *Tourism Management, 76*, 103954.

Dela Cerna, J. T., Sonido, J. M. C., Rubio, D. K. M., & Sisles, L. F. V. (2024). Analyzing Gen Z's travel influence in the digital age: The correlational effect of social media. *International Journal of Multidisciplinary Research and Analysis, 7*(10). https://doi.org/10.47191/ijmra/v7-i10-01

Elkatmiş, M. (2024). Examination of social media usage habits of Generation Z. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2024.1370823

Fu, X., Kang, J., Hahm, J. J., & Wiitala, J. (2020). Investigating the consequences of theme park experience through the lenses of self-congruity and flow. *International Journal of Contemporary Hospitality Management, 32*(3), 1181–1199.

García-Haro, M. Á., Ruiz-Palomino, P., Martínez-Cañas, R., & Martínez-Ruiz, M. P. (2024). Assessing the drivers to share content on social media in tourism: The empirical evidence on a world heritage city. *European Journal of Management and Business Economics.*

Hoang, D. S., Tučková, Z., Huy, D. T. N., & Pham, P. T. (2021). Factors to enhance tourist's ecotourism loyalty: An empirical study in Vietnam. *Webology.*

Hoang, S. D., Dey, S. K., Ratilla, M., & Tučková, Z. (2023). Investigating the antecedents of ecotourism and destination loyalty among Vietnamese tourists: An expectation–confirmation perspective. *Global Business Review*. https://doi.org/10.1177/09721509231174740

Hou, Z., Liu, B., Liu, S., & Liu, F. (2024). How does the knowledge association of online travel reviews affect tourism purchase behavior? An empirical study based on a knowledge network perspective. *Journal of the Knowledge Economy*, 1–28.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hua, Y., Ding, L., Dong, H., & Lin, Z. (2024). Influence of user-generated content (UGC) in social media on the intangible cultural heritage preservation of Gen Z tourists in the digital economy era. *International Journal of Tourism Research, 26*(5), e2743.

Islam, M. R., Abdul Kader Jilani, M. M., Miah, S. J., Akter, S., & Ulhaq, A. (2021). Discovering tourist preference for electing destinations: A pattern mining-based approach. *Asia Pacific Journal of Tourism Research, 26*(10), 1081–1096.

Kang, M., & Schuett, M. A. (2013). Determinants of sharing travel experiences on social media. *Journal of Travel & Tourism Marketing, 30*(1–2), 93–107.

Keelson, S. A., Bruce, E., Egala, S. B., Amoah, J., & Bashiru Jibril, A. (2024). Driving forces of social media and its impact on tourists' destination decisions: A uses and gratification theory. *Cogent Social Sciences, 10*(1), 2318878.

Kim, M., & Thapa, B. (2018). The influence of self-congruity, perceived value, and satisfaction on destination loyalty: A case study of the Korean DMZ. *Journal of Heritage Tourism, 13*(3), 224–236.

Kumar, T. B. J., Goh, S. K., & Balaji, M. S. (2021). Sharing travel-related experiences on social media: Integrating social capital and face orientation. *Journal of Vacation Marketing, 27*(2), 168–186.

Lam, J. M., Ismail, H., & Lee, S. (2020). From desktop to destination: User-generated content platforms, co-created online experiences, destination image, and satisfaction. *Journal of Destination Marketing & Management, 18*, 100490.

Le, X. C. (2022). Charting sustained usage toward mobile social media applications: The criticality of expected benefits and emotional motivations. *Asia Pacific Journal of Marketing and Logistics, 34*(3), 576–593.

Leou, E. C., & Wang, H. (2023). A holistic perspective to predict yoga tourists' revisit intention: An integration of the TPB and ECM model. *Frontiers in Psychology, 13*, 1090579.

Levak, T., & Barić Šelmić, S. (2018). Escaping the "virtual promenade": New trends in the use of social networks by members of Generation Z. *Media, Culture and Public Relations, 9*(1–2), 37–55.

Li, F., Ma, J., & Tong, Y. (2024). Livestreaming in tourism: What drives tourism live streamers to share their travel experiences? *Tourism Review, 79*(3), 780–795.

Lien, C. H., & Cao, Y. (2014). Examining WeChat users' motivations, trust, attitudes, and positive word-of-mouth: Evidence from China. *Computers in Human Behavior, 41*, 104–111.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*(1), 114–121. https://doi.org/10.1037/0021-9010.86.1.114

Luna-Cortés, G., López-Bonilla, J. M., & López-Bonilla, L. M. (2019). Self-congruity, social value, and the use of virtual social networks by Generation Y travelers. *Journal of Travel Research, 58*(3), 398–410.

MacCallum, R. C., & Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research, 32*(2), 193–210.

Malhotra, N. K. (1981). A scale to measure self-concepts, person concepts, and product concepts. *Journal of Marketing Research, 18*(4), 456–464.

Manhas, P. S., & Ashraf, F. (2024). Measuring destination reputation through travel motivation and tourists' attitude: An empirical approach. *Journal of Hospitality Application & Research, 19*(1).

Mariani, M., & Borghi, M. (2021). Customers' evaluation of mechanical artificial intelligence in hospitality services: A study using online reviews analytics. *International Journal of Contemporary Hospitality Management, 33*(11), 3956–3976. https://doi.org/10.1108/ijchm-06-2020-0622

Meeprom, S., & Fakfare, P. (2021). Unpacking the role of self-congruence, attendee engagement, and emotional attachment in cultural events. *International Journal of Event and Festival Management, 12*(4), 399–417.

Mohamad, N., Tan, V., & Tan, P. P. (2022). Travel experience on social media: The impact towards tourist destination choice. *Social and Management Research Journal (SMRJ, 19*(2), 21–52.

Muharrir, M., Ikramuddin, I., Nurainun, N., Heriyana, H., & Iis, E. Y. (2023, December). Sharing tour travel experiences on social media at tourist destinations. In *Proceedings of the International Conference on Finance Economics and Business (ICOFEB, 1*, 0023–0023).

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management, 43*, 46–54.

Nguyen, C. N., Tho, A., & Nguyen, T. B. (2024). Understanding Generation Z's job engagement and performance in a generationally diverse workplace. *Ho Chi Minh City Open University Journal of Science-Economics and Business Administration, 14*(2), 93–108.

Nusair, K. (2020). Developing a comprehensive life cycle framework for social media research in hospitality and tourism: A bibliometric method 2002–2018. *International Journal of Contemporary Hospitality Management, 32*(3), 1041–1066.

Nwatu, B. C., Nwaizugbo, I. C., & Rahim, A. G. (2021). Antecedents of tourists' behavioral intentions: Perspectives of the expectations confirmation model—A study of select tourism sites in South-East Nigeria.

Oh, S., Ji, H., Kim, J., Park, E., & del Pobil, A. P. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. *Information Technology & Tourism, 24*(1), 109–126.

Oliveira, T., Araujo, B., & Tam, C. (2020). Why do people share their travel experiences on social media? *Tourism Management, 78*, 104041.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research, 14*(March), 495–507.

Oliver, R. L., & Swan, J. E. (1989). Consumer perceptions of interpersonal equity and satisfaction in transactions: A field survey approach. *Journal of Marketing, 53*(2), 21–35.

Park, D. H., & Lee, S. (2021). UGC sharing motives and their effects on UGC sharing intention from quantitative and qualitative perspectives: Focusing on content creators in South Korea. *Sustainability, 13*(17), 9644.

Pinzaru, F., Vătămănescu, E. M., Mitan, A., Săvulescu, R., Vițelar, A., Noaghea, C., & Bălan, M. (2016). Millennials at work: Investigating the specificity of Generation Y versus other generations. *Management Dynamics in the Knowledge Economy, 4*(2), 173–192.

Prakash, V. (1984). Validity and reliability of the confirmation of expectations paradigm as a determinant of consumer satisfaction. *Journal of the Academy of Marketing Science, 12*, 63–76.

Pratt, M. A., & Sparks, B. (2014). Predicting wine tourism intention: Destination image and self-congruity. *Journal of Travel & Tourism Marketing, 31*(4), 443–460.

Putri, F. S. E., Khumayah, S., & Nurfalah, F. (2023). Formation of self-identity of Generation Z through Instagram social media amongst students. *Imagionary, 2*(1), 1–6. https://doi.org/10.51353/jim.v2i1.814

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

S'Hail, M., & Benabdelouahed, R. (2024). Exploring user-generated content in tourism: Insights from Moroccan travelers. *International Journal of Applied Management and Economics, 2*(07), 076–103. https://doi.org/10.5281/zenodo.11045564

Șchiopu, A. F., Nica, A. M., Pădurean, A. M., & Ţală, M. L. (2023). Generation Z vs. Generation Y: Different from or similar? A comparison of centennials and millennials regarding the use of social media for travel purposes. *Cactus Tourism Journal, 5*(1), 20–35.

Sedera, D., Lokuge, S., Atapattu, M., & Gretzel, U. (2017). Likes—The key to my happiness: The moderating effect of social influence on travel experience. *Information & Management, 54*(6), 825–836.

Septiari, E. D., & Omar, A. (2022). Examining the impacts of travellers' intrinsic motivation in social media travel sharing behavior in Indonesia. *Journal of Tourism, Hospitality and Culinary Arts, 14*(2), 168–182.

Serra-Cantallops, A., Ramón Cardona, J., & Salvi, F. (2020). Antecedents of positive eWOM in hotels: Exploring the relative role of satisfaction, quality, and positive emotional experiences. *International Journal of Contemporary Hospitality Management, 32*(11), 3457–3477.

Shukla, A., Mishra, A., & Dwivedi, Y. (2023). Expectation confirmation theory: A review. In S. Papagiannidis (Ed.), *TheoryHub Book*. Available at https://open.ncl.ac.uk / ISBN: 9781739604400

Siagian, N. R., & Yuliana, N. (2023). The role of social media in Generation Z communication. *Interdisciplinary Journal of Advanced Research and Innovation, 1*(3), 109–118.

Sirgy, M. J. (1985). Using self-congruity and ideal congruity to predict purchase motivation. *Journal of Business Research, 13*(3), 195–206.

Sirgy, M. J., & Su, C. (2000). Destination image, self-congruity, and travel behavior: Toward an integrative model. *Journal of Travel Research, 38*(4), 340–352.

Sotiriadis, M. D. (2017). Sharing tourism experiences in social media: A literature review and a set of suggested business strategies. *International Journal of Contemporary Hospitality Management, 29*(1), 179–225.

Stahl, C. C., & Literat, I. (2023). #GenZ on TikTok: The collective online self-portrait of the social media generation. *Journal of Youth Studies, 26*(7), 925–946.

Tabachnick, B. G. (2007). *Using multivariate statistics.* Alyn and Bacon.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson Education, Allyn & Bacon, Inc.

Tsai, C.-F., Chen, K., Hu, Y.-H., & Chen, W.-K. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management, 80*, 104122. https://doi.org/10.1016/j.tourman.2020.104122

Tubalawony, J. (2023). The role of social media marketing with TikTok in Generation Z: A study on the choice of fitness centers by Generation Z in Ambon, Indonesia. *Open Access Indonesia Journal of Social Sciences, 6*(7), 1242–1250.

Uysal, D. (2022). Gen-Z's consumption behaviors in the post-pandemic tourism sector. *Journal of Tourism Leisure and Hospitality, 4*(1), 67–79.

Wang, X., Li, X., Li, Q., & Peng, L. (2014). Intention of sharing travel experiences on social media: Motivations and the moderating effects of face orientation. In *Proceedings of the 25th Australasian Conference on Information Systems, 8th–10th December*. Auckland, New Zealand: ACIS.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. *Handbook of Structural Equation Modeling, 1*(1), 209–231.

Widiana, R., & Novani, S. (2022). Factors influencing tourists' travel experience sharing on social media. *Indonesian Journal of Business and Entrepreneurship (IJBE), 8*(1), 104–104.

Yamagishi, K., Canayong, D., Domingo, M., Maneja, K. N., Montolo, A., & Siton, A. (2024). User-generated content on Gen Z tourist visit intention: A stimulus-organism-response approach. *Journal of Hospitality and Tourism Insights, 7*(4), 1949–1973.

Yang, Q., & Lee, Y. C. (2016). The effects of self-congruity and functional congruity on e-WOM: The moderating role of self-construal in tourism. *The Journal of Information Systems, 25*(1), 03.

Yoanita, D., Chertian, V. G., & Ayudia, P. D. (2022). Understanding Gen Z's online self-presentation on multiple Instagram accounts. *Jurnal Studi Komunikasi, 6*(2), 603–616. https://doi.org/10.25139/jsk.v6i2.4922

Zucco, F. D., Falaster, C., Theiss, T., & Kraus, C. B. (2020). Behavioral profiles in sharing tourism experiences in virtual networks. *e-Review of Tourism Research, 17*(6), 934–961.

DECISION SCIENCES INSTITUTE

Predicting Ransomware Attacks

Yaman Roumani
Oakland University
Email: yamanroumani@oakland.edu

ABSTRACT

Ransomware attacks pose a major cybersecurity threat, causing financial and operational damage. Despite their impact, predictive measures remain underdeveloped. This study addresses the gap by using time series analysis to forecast ransomware incidents. Analyzing 1,880 reported cases, the data is decomposed into trend, seasonal, and residual components to build a predictive model. Results show effective forecasting performance. The findings highlight the value of time series forecasting in cybersecurity, offering organizations actionable intelligence to enhance preparedness and situational awareness against ransomware threats.

KEYWORDS: ransomware, forecasting, time series, STL decomposition, ARIMA, exponential smoothing

INTRODUCTION

Ransomware attacks continue to severely disrupt organizations, resulting in significant financial losses and data breaches. A 2024 crypto crime report highlights ransomware as a highly profitable tool for cybercriminals, with extorted payments reaching a record $1.1 billion in 2023 (Chainalysis, 2024). Despite increased awareness, the number of victims surged by 143% from 2022 to 2023 (Akami, 2023). Additionally, cybercriminals are continuously refining their tactics to avoid detection and exert greater pressure on victims (FBI, 2023).

With ransomware attacks showing no signs of slowing down, cybersecurity investment has become a strategic necessity for organizations. A survey of 569 cybersecurity leaders revealed that 91% anticipate increased budgets for technologies like centralized monitoring, AI, machine learning, and next-generation firewalls to enhance threat detection and response (Fortinet, 2023). However, cybercriminals evade detection using techniques such as packers, crypters, polymorphism, and metamorphism (Geng et al., 2024). Thus, beyond detection, additional security measures are needed. This study focuses on identifying ransomware trends, analyzing attack patterns, and forecasting future threats, enabling organizations to allocate resources proactively and strengthen their defenses.

While previous research has largely focused on detecting and preventing ransomware attacks, studies on predicting future threats remain limited (Razaulla et al., 2023; Al-Rimy et al., 2018). First, past research has not utilized historical time-series data to analyze trends, seasonality, and noise in ransomware incidents—key factors in understanding attack patterns. Second, existing prediction methods have been constrained to specific ransomware cases, limited strains, and narrow technological scopes (Quinkert et al., 2018; Mathane & Lakshmi, 2021; Gazzan & Sheldon, 2023), resulting in small sample sizes and non-generalizable findings. Therefore, broader investigations across diverse ransomware strains and technologies are needed.

This paper aims to bridge the gap in existing research by developing a ransomware prediction model using time series analysis. By analyzing past attack components—trend, seasonality, and noise—it forecasts future threats. The model is built using publicly reported ransomware incidents from the Critical Infrastructure Ransomware Attacks (CIRA) dataset (2013–2024). This study advances prior research by emphasizing recent data, accounting for key attack patterns, and considering diverse ransomware families for broader applicability. Notably, it is the first to apply time series analysis based on targeted industries, offering valuable insights for security teams and decision-makers.

LITERATURE REVIEW

Previous studies have proposed various strategies to counter ransomware attacks, which can be categorized into detection, prevention, and prediction. Ransomware detection focuses on identifying suspicious behaviors, such as file activities and registry key changes, and creating models to recognize these patterns in attacks (Begovic et al., 2023). However, detecting ransomware is challenging due to its evolving nature and the wide range of variants with distinct signatures (Beaman et al., 2021; Razaulla et al., 2023). Prevention involves proactive and reactive measures to stop attacks in real-time, but these manual procedures are error-prone, especially in the fast-changing ransomware landscape (McIntosh et al., 2021). Lastly, ransomware prediction aims to forecast attacks by monitoring early indicators and assessing the likelihood of future incidents (Lanza et al., 2024).
Within the ransomware prediction literature, previous studies have proposed various models to predict different outcomes, such as ransomware deployment methods, families, and attacks. For instance, Hull et al. (2019) developed a predictive model for ransomware deployment methods, analyzing 18 ransomware families with a classifier to predict early warning signs at different stages of deployment. Jeon et al. (2023) utilized deep learning to predict ransomware behavior in healthcare IoT by analyzing operation codes and API call sequences. In the realm of ransomware attacks, studies have employed various predictive models, including time series and machine learning. For example, Quinkert et al. (2018) proposed a model using malicious web domains related to ransomware attacks as external signals for time series forecasting to predict future activity, though their focus was restricted to a single ransomware strain. Albulayhi and Al-Haija (2022) used a time-series regression method with a neural network algorithm to predict ransomware and malware attacks. However, their model focused on yearly attacks and did not account for key time-series components such as trend, seasonality, and noise. Rhode et al. (2018) developed a prediction model using recurrent neural networks to determine whether an executable file is malicious or benign within the first five seconds of execution, achieving 94% accuracy with 3,000 ransomware samples. Mathane and Lakshmi (2021) introduced a context-aware technique using context ontology to build attack profiles for ransomware and applied AI and machine learning to predict attacks on IoT systems. Li (2024) compared various machine learning models for predicting malware attack trends in cyber supply chains, finding neural networks to be the most effective. Gogineni et al. (2022) analyzed ransomware event series at the program binary level and used deep neural networks to forecast future threats.

METHODOLOGY

Time series analysis refers to a set of statistical techniques used to extract insights from time-ordered data. In information security, time series analysis has been applied to predict various threats, including software vulnerabilities (Roumani et al., 2015), stages of cyber-intrusions (Rege et al., 2018), and distributed denial-of-service attacks (Falowo & Abdo, 2024). While this analysis helps in understanding historical trends, identifying patterns, forecasting future values, and making informed decisions (Box et al., 2015), it does not aim to explain or quantify the causal factors behind the data behavior. Unlike other statistical models that treat observations as independent, time series models account for temporal dependencies, recognizing that current values are often influenced by past values (Chatfield, 2013). In this study, time series analysis helps identify key components of ransomware attacks, such as trends, seasonality, and noise, by detecting upward or downward movements, periodic fluctuations, cyclical behaviors, and random variations (Yaffee & McGee, 2000). We apply the following time series models to our dataset: seasonal and trend decomposition using LOESS, autoregressive integrated moving average, and exponential smoothing.

Seasonal and Trend Decomposition Using LOESS (STL)

STL is a non-parametric smoothing method that decomposes time series data into three components: trend (T), seasonal (S), and residuals (R) (Cleveland et al., 1990). By employing locally estimated scatterplot smoothing (LOESS), a regression technique, STL produces smooth estimates of the three components, enabling it to extract seasonality, trends, and calculate residuals as unexplained variance or random noise in the ransomware data (Cleveland & Devlin, 1988). STL iteratively applies LOESS smoothing to refine and improve the estimates of the trend, seasonal, and residual components of the original time series at each time point.

STL can be integrated with other methods, such as autoregressive integrated moving average (ARIMA) and exponential smoothing, to combine the strengths of each technique for improved forecasting (Hyndman & Athanasopoulos, 2018). ARIMA is a widely used time series forecasting model, particularly in fields like economics, finance, and healthcare. An ARIMA model utilizes lagged values and forecast errors to model the data, represented as (p, d, q), where p is the number of autoregressive terms (AR), d is the number of differences to achieve stationarity, and q is the number of moving average (MA) terms. For example, an ARIMA(1,1,1) model implies that the number of ransomware attacks at time t depends on the number of attacks at time t-1, the data has been differenced twice for stationarity, and the forecast at time t is based on past prediction errors at time t-1.

Exponential smoothing, on the other hand, is a forecasting method that assumes future values are influenced by recent observations. It uses exponentially weighted averages of past data, giving greater weight to recent observations to predict future values (De Livera et al., 2011). In this study, exponential smoothing would emphasize more recent ransomware attacks over older data points. An exponential smoothing model (ETS) requires three parameters: error (E), trend (T), and seasonality (S), each of which can be additive (A), multiplicative (M), or none (N). The ETS model determines whether trend and seasonality are independent of the series' level or whether they scale with the level of the series, causing variations that grow or shrink proportionally to the overall magnitude (Hyndman et al., 2008). For example, an ETS(A,N,N) model suggests that forecasting ransomware attacks involves adding an error term to the previous observation without considering trend or seasonality.

By combining STL with ARIMA or exponential smoothing, STL handles non-constant seasonal patterns, while ARIMA or exponential smoothing captures any remaining correlations in the non-seasonal part of the data by independently fitting a model and forecasting the residual component of the decomposed time series (Hyndman & Athanasopoulos, 2018).

Data Collection

We gathered data on ransomware attacks from the CIRA dataset, maintained by the Cybersecurity in Application, Research, and Education Lab at Temple University. The dataset contains over 1,800 records of publicly disclosed ransomware incidents (Rege, 2024). These incidents, collected from media and security reports, span from November 2013 to August 2024. For each incident, CIRA provides details such as the date, targeted critical infrastructure sector, ransomware strain, duration, and ransom amount. In this study, we aggregate the total number of incidents by month. Figure 1 presents the monthly aggregation of ransomware attacks revealing noticeable increase in attacks during and after the COVID-19 era, attributed to expanded attack surfaces and evolving attack tactics (Lang et al., 2023).

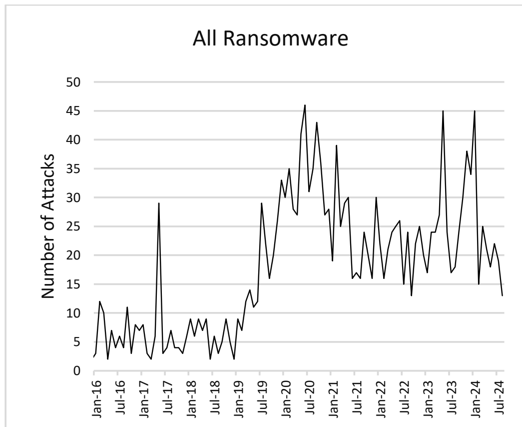


Figure 1. Ransomware attacks

RESULTS

STL Decomposition

We utilized the STLF package in the R statistical software to perform STL decomposition and forecasting. The number of ransomware attacks was used as the dependent variable, while monthly periodicity was the independent variable. Each dataset was divided into a training set and a testing set. The training set was utilized to develop and train the fitted time series model, while the testing set was employed to evaluate the model's performance. Moreover, STL decomposition was applied to the training set to extract the trend, seasonal, and noise components. Based on Figure 2, the dataset shows increasing trend in ransomware attacks starting in 2019.

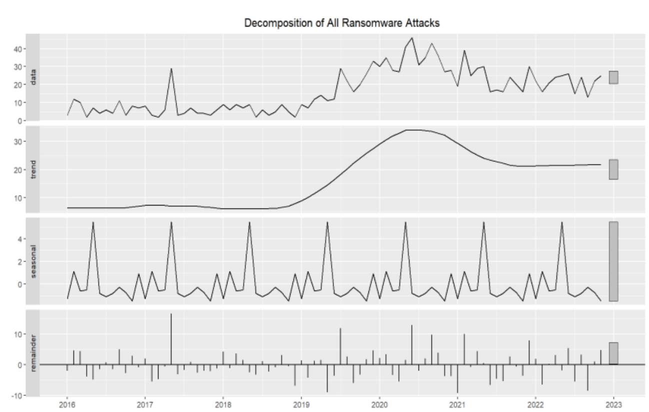


Figure 2. STL decomposition of ansomware attacks.

To determine the contribution of each time series component to the overall variability in the data, we calculated the variance and computed the ratio of each component's variance to the total variance. Based on the results, the trend component was the major contributor to the variability in the dataset (73.86%). The seasonality component, on the other hand, was a minor contributor to the variability. Finally, the remainder (noise) component accounted for 17.26%.

Fitting Time Series Model

After performing STL decomposition, a time series model was developed for the dataset. To determine the optimal model, we compared the normalized Bayesian information criterion (BIC) values. Furthermore, to ensure the time series model satisfies the necessary assumptions, we assessed stationarity using the autocorrelation function (ACF) and the augmented Dickey-Fuller test (Dickey, 2015). Residual diagnostics were also performed to assess model adequacy. Specifically, we applied the Ljung-Box test to evaluate the null hypothesis that the residuals are randomly distributed, which would indicate a good model fit (Ljung & Box, 1978). Based on the fitted model, our analysis identified STL-ARIMA(0,1,1) as the optimal choice. This model implies that predicting future ransomware attacks is based on the difference between consecutive observations while also smoothing out random noise using the most recent past error.

Forecasting

Using the fitted model, we forecasted the number of monthly ransomware attacks. To assess the predictive performance of the forecasting model, we used three metrics: mean absolute error (MAE), root mean square error (RMSE), and symmetric mean absolute percentage error (SMAPE). MAE calculates the average absolute difference between the observed and predicted values, treating all errors equally since it does not square the differences. RMSE also measures the average difference between observed and predicted values but gives more weight to larger errors due to the squaring of the differences, making it more sensitive to outliers. SMAPE averages the percentage deviation between the observed and predicted values, treating overestimations and underestimations equally, and ensures the percentage error remains between 0% and 200%. The training set yielded MAE values of 4.63 and 6.73 for the training and testing sets respectively. RMSE values, on the other hand, were 6.18 for the training set and 8.74 for the testing set. Finally, the SMAPE value was 26.26%. Figure 3 shows the fitted time series model with forecasts.

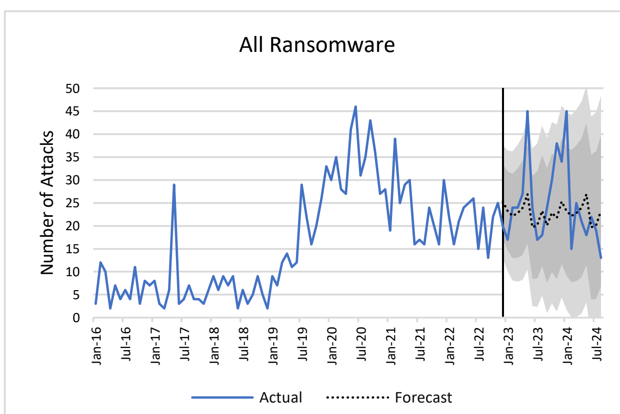


Figure 3. Fitted time series model and forecasts

DISCUSSION

The analysis suggests that STL decomposition and time series modeling are effective for understanding the underlying patterns and trends of ransomware attacks, as well as predicting future threat volumes. Our model, which was based on the count of ransomware attacks, demonstrated its capacity to capture the overall intensity of cyberattacks, making it a promising method for forecasting ransomware threats. The decomposed time series components reveal that the trend for all ransomware attacks has remained flat, with monthly attack numbers fluctuating between 20 and 25. This could be due to the countermeasures implemented by organizations and increasing public awareness, which may be slowing the growth of ransomware threats and discouraging attackers. However, this could also indicate that attackers might soon pivot to other methods or targets, perhaps using more sophisticated, harder-to-detect techniques (Oz et al., 2022). Notably, no seasonal component was found, suggesting ransomware attacks lack a regular monthly pattern. However, it's possible that any seasonality component was too small compared to noise or non-linear components, so future research might investigate if daily, weekly, or quarterly patterns reveal any seasonality.

Implications for Research and Practice

This study offers several contributions to both theory and practice. From a theoretical perspective, it expands the cybersecurity literature by addressing a notable gap, focusing on proactive methods rather than the predominantly reactive measures like mitigation and response seen in previous research (Al-Rimy et al., 2018). Specifically, our study advances the literature by identifying key time series components of ransomware attacks and proposing a novel model for predicting future threats. We show that ransomware threats can be anticipated based on their trends, enhancing our understanding of their evolution and providing a systematic approach to forecasting potential attacks. Additionally, this study responds to the call by Razaulla et al. (2023) for improved ransomware defense solutions. Our predictive model contributes to the theoretical development of predictive strategies and broadens the cybersecurity and risk management literature. Notably, this is the first study to apply STL decomposition and time series analysis to publicly disclosed ransomware incidents.

As ransomware attacks continue to rise, our findings suggest four important implications. First, we recommend that organizations monitor ransomware attack trends to implement proactive risk management. For instance, firms can predict months with higher attack likelihoods, enabling them to prepare and mitigate the impact of potential incidents. Second, our model can assist organizations in prioritizing investments by forecasting future threats, allowing for better resource allocation and reducing unnecessary expenditures. Third, since our model was built using publicly available data, practitioners can replicate them and monitor changes in ransomware patterns. Finally, our research should be of significant value to policymakers, providing insights that can inform regulations and guidelines to protect from ransomware trends.

Limitations and Future Research

This study acknowledges several limitations, which also present opportunities for future research. First, the data was sourced solely from the CIRA database, which includes only publicly reported ransomware incidents. As these reports are voluntary, the findings may not be fully generalizable. Future research could benefit from changes to the Cybersecurity and Infrastructure Security Agency's (CISA) reporting requirements, which will mandate that certain sectors and organizations report ransomware incidents (CISA, 2022).
Second, our analysis relied on a single independent variable—the monthly number of ransomware attacks—without considering other potential factors that could influence the predictive accuracy of the model. This limitation applies to both the ARIMA and exponential smoothing methods. Future studies might consider using the autoregressive integrated moving

average with exogenous inputs (ARIMAX), which extends ARIMA by integrating multiple independent variables.

Building on this study, further research could explore whether external factors, such as geopolitical events, economic conditions, or industry-specific threats, could improve the accuracy of the forecasting model. Additionally, future work could focus on developing predictive model for specific types of ransomwares, such as ransomware-as-a-Service (RaaS), different ransomware families, or attack techniques like double extortion. Finally, upcoming studies might investigate the use of hybrid models that combine traditional time series methods with artificial intelligence techniques, such as deep learning, to enhance the accuracy of ransomware attack predictions.

REFERENCES

Akcora, C. G., Li, Y., Gel, Y. R., & Kantarcioglu, M. (2019). Bitcoinheist: Topological data analysis for ransomware detection on the bitcoin blockchain. arXiv preprint arXiv:1906.07852.

Akami. (2023). Ransomware on the Move: Evolving Exploitation Techniques and the Active Pursuit of Zero-Days. Retrieved from: https://www.akamai.com/lp/soti/ransomware-on-the-move

Alamri, A. H. (2024). Dragos Industrial Ransomware Analysis: Q2 2024. Retrieved from: https://www.dragos.com/blog/dragos-industrial-ransomware-analysis-q2-2024/

Al-Rimy, B. A. S., Maarof, M. A., & Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. Computers & Security, 74, 144-166.

Albulayhi, K., & Al-Haija, Q. A. (2022, December). Early-stage malware and ransomware forecasting in the short-term future using regression-based neural network technique. In 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 735-742). IEEE.

Asghar, M. R., Hu, Q., & Zeadally, S. (2019). Cybersecurity in industrial control systems: Issues, technologies, and challenges. Computer Networks, 165, 106946.

Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., & Khan, M. K. (2021). Ransomware: Recent advances, analysis, challenges and future research directions. Computers & security, 111, 102490.

Begovic, K., Al-Ali, A., & Malluhi, Q. (2023). Cryptographic ransomware encryption detection: Survey. Computers & Security, 132, 103349.

Berardi, D., Giallorenzo, S., Melis, A., Melloni, S., Onori, L., & Prandini, M. (2023). Data flooding against ransomware: Concepts and implementations. Computers & Security, 131, 103295.

Chainalysis. (2024). Ransomware Payments Exceed $1 Billion in 2023, Hitting Record High After 2022 Decline. Retrieved from: https://www.chainalysis.com/blog/ransomware-2024/

Chatfield, C. (2013). The analysis of time series: theory and practice. Springer.

CISA (2022). Cyber Incident Reporting for Critical Infrastructure Act of 2022 (CIRCIA). Retrieved from: https://www.cisa.gov/topics/cyber-threats-and-advisories/information-sharing/cyber-incident-reporting-critical-infrastructure-act-2022-circia

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. J. off. Stat, 6(1), 3-73.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American statistical association, 83(403), 596-610.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American statistical association, 106(496), 1513-1527.

Dickey, D. A. (2015). Stationarity issues in time series models. SAS Users Group International, 30.

Dissanayake, N., Jayatilaka, A., Zahedi, M., & Babar, M. A. (2022). Software security patch management-A systematic literature review of challenges, approaches, tools and practices. Information and Software Technology, 144, 106771.

Falowo, O. I., & Abdo, J. B. (2024). 2019–2023 in Review: Projecting DDoS Threats With ARIMA and ETS Forecasting Techniques. IEEE Access, 12, 26759-26772.

FBI. (2023). Two or More Ransomware Variants Impacting the SameVictims and Data Destruction Trends. Retrieved from: https://www.ic3.gov/Media/News/2023/230928.pdf

Fortinet. (2023). The 2023 Global Ransomware Report. Retrieved from: https://www.fortinet.com/content/dam/fortinet/assets/reports/report-2023-ransomware-global-research.pdf

Gazzan, M., & Sheldon, F. T. (2023). Opportunities for early detection and prediction of ransomware attacks against industrial control systems. Future Internet, 15(4), 144.

Geng, J., Wang, J., Fang, Z., Zhou, Y., Wu, D., & Ge, W. (2024). A Survey of strategy-driven evasion methods for PE malware: transformation, concealment, and attack. Computers & Security, 137, 103595.

Gogineni, K., Derasari, P., & Venkataramani, G. (2022, September). Foreseer: Efficiently forecasting malware event series with long short-term memory. In 2022 IEEE International Symposium on Secure and Private Execution Environment Design (SEED) (pp. 97-108). IEEE.

Hillman, D., Harel, Y., & Toch, E. (2023). Evaluating organizational phishing awareness training on an enterprise scale. Computers & Security, 132, 103364.

Hull, G., John, H., & Arief, B. (2019). Ransomware deployment methods and analysis: views from a predictive model and human responses. Crime Science, 8(1), 1-22.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. Australia: Otexts, 2018.

Jeon, J., Baek, S., Jeong, B., & Jeong, Y. S. (2023). Early prediction of ransomware API calls behaviour based on GRU-TCN in healthcare IoT. Connection Science, 35(1), 2233716.

Lang, M., Connolly, L., Taylor, P., & Corner, P. J. (2023). The evolving menace of ransomware: A comparative analysis of pre-pandemic and mid-pandemic attacks. Digital Threats: Research and Practice, 4(4), 1-22.

Lanza, C., Lahmadi, A., & François, J. (2024). Ransomware Analysis: Knowledge Extraction and Classification for Advanced Cyber Threat Intelligence. CRC Press.

Lee, K. & Popoff, E. (2024). 2024 State EdTech Trends Report. Retrieved from: https://www.setda.org/priorities/state-trends/

Li, S. (2024). Comparative Analysis of Predicting Malware Attack Trends in Cyber Supply Chain Using Multiple Classification Models. IEEE Access.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303.

Mathane, V., & Lakshmi, P. V. (2021). Predictive analysis of ransomware attacks using context-aware AI in IoT systems. International Journal of Advanced Computer Science and Applications, 12(4).

McIntosh, T., Kayes, A. S. M., Chen, Y. P. P., Ng, A., & Watters, P. (2021). Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions. ACM Computing Surveys (CSUR), 54(9), 1-36.

NordLocker (2023). Ransomware statistics: Who is targeted the most? Retrieved from: https://nordlocker.com/ransomware-attack-statistics/

Oz, H., Aris, A., Levi, A., & Uluagac, A. S. (2022). A survey on ransomware: Evolution, taxonomy, and defense solutions. ACM Computing Surveys (CSUR), 54(11s), 1-37.

Quinkert, F., Holz, T., Hossain, K. S. M., Ferrara, E., & Lerman, K. (2018). Raptor: Ransomware attack predictor. arXiv preprint arXiv:1803.01598.

Razaulla, S., Fachkha, C., Markarian, C., Gawanmeh, A., Mansoor, W., Fung, B. C., & Assi, C. (2023). The age of ransomware: A survey on the evolution, taxonomy, and research directions. IEEE Access, 11, 40698-40723.

Rege, A. (2024). Critical Infrastructure Ransomware Attacks (CIRA) Dataset. Version 12.9. Temple University. Retrieved from: https://sites.temple.edu/care/cira/. ORCID: 0000-0002-6396-1066.

Rege, A., Obradovic, Z., Asadi, N., Parker, E., Pandit, R., Masceri, N., & Singer, B. (2018). Predicting adversarial cyber-intrusion stages using autoregressive neural networks. IEEE Intelligent Systems, 33(2), 29-39.

Rhode, M., Burnap, P., & Jones, K. (2018). Early-stage malware prediction using recurrent neural networks. computers & security, 77, 578-594.

Roumani, Y., Nwankpa, J. K., & Roumani, Y. F. (2015). Time series modeling of vulnerabilities. Computers & Security, 51, 32-40.

Sophos. (2024). The State of Ransomware in Healthcare 2024. Retrieved from: https://www.sophos.com/en-us/whitepaper/state-of-ransomware-in-healthcare

Sugden, M., Fixler, A. (2024). Healthcare Cybersecurity Needs a Check Up. Retrieved from: https://cybersolarium.org/wp-content/uploads/2024/06/csc2.0-healthcare-cybersecurity-needs-a-check-up-report.pdf

Yaffee, R. A., & McGee, M. (2000). An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®. Elsevier.

Yeoh, W., Huang, H., Lee, W. S., Al Jafari, F., & Mansson, R. (2022). Simulated phishing attack and embedded training campaign. Journal of Computer Information Systems, 62(4), 802-821.

Xu, S. (2021, May). The application of machine learning in Bitcoin ransomware family prediction. In Proceedings of the 2021 5th International Conference on Information System and Data Mining (pp. 21-27)

DECISION SCIENCES INSTITUTE

Who is More Likely to Enjoy Football as Their Favorite Sport? The Impact of Gender & Previous Sports Experience

Eileen Zhan
Calvin University & Forest Hills Central High School
Email: eileenzhan@gmail.com
Yong Xu
Ferris State University
Email: YongXu@ferris.edu

ABSTRACT

This study analyzes data from over a thousand U.S. adults to examine how gender and past sports experience influence football preference. While men's football preference correlates with participation in football, soccer, hockey, and golf, no such pattern exists for women. Logistic regression shows that prior football experience is a key predictor of preference, with men favoring the sport more due to cultural and participation norms. Promoting inclusivity, family-oriented initiatives, and sustainable practices can help increase female engagement and support football's long-term popularity.

KEYWORDS: Football, Gender, Sport Management, Favorite Sport

INTRODUCTION

Football, known as the all-American favorite pastime, holds a prominent place as the most popular sport in the United States. Many people love to play it; others enjoy watching it. In this paper, we want to know how gender and previous sport experiences impact the preference of football as their favorite sport.

To understand which demographic, males or females, are more inclined to enjoy football the most, we applied a series of plots and Chi-Square tests to a data sample of over 1,000 adults (IPOS. 2021) to explore the answers to these questions.

We then ran a Logit-Regression test to explore further how prior football and other sports experience, and gender affected preferences for football.

LITERATURE REVIEW

One stream of papers looked at how gender influences sports preferences. Robinson and Trail (2005) studied how gender, motives, and attachment points affect preferences in collegiate sports [13]. They found that while gender played a small role in explaining motives and attachment, it still had a meaningful influence on what sports people preferred. Ridinger and Funk (2006) expanded the study by showing that although men and women share some reasons for being sports fans, they also have unique motivators, especially at NCAA basketball games [12]. This highlights the different ways men and women connect with sports. Warm, Waddill, and Dunham (2004) looked at gender roles and found that masculinity is a strong predictor of being a sports fan, while femininity is not. Sveinson and Hoeber (2015) pointed out that most research on sports fans focuses on men, even though women have different ways of being fans that need more attention. Deaner, Balish, and Lombardo (2016) took an evolutionary approach to explain that men may naturally be more interested in sports because aligned with sports culture, men are more competitive and willing to take risks. McDonald et al. (2024) explored why some people aren't interested in sports and suggested ways to make sports more appealing to different groups. This ties in with other research that shows how gender and cultural background affect what kinds of physical activities people enjoy, and how these preferences can change as they get older.

This stream of paper shows that gender affects sports preferences in many ways, from social roles to deeper evolutionary factors. However, these studies did not analyze both men’s and women’s experiences to understand how people connect with sports.

Next, we examine cultural Influences on football preference. Football's prominence in American culture has been well-documented. Correia and Esteves (2007) showed that people’s social and cultural reasons play a big role in why they attend football games. Deaner et al. (2016) explained that men’s interest in sports, like football, may partly come from natural traits like competitiveness. Jakubowska, Antonowicz, and Kossakowski (2020) explored how women are starting to break into what has traditionally been a male-dominated fan culture in football. These studies on cultural diversity in sports psychology explains how culture and gender work together to shape what sports people enjoy and how they behave as fans. These insights help us understand our results from American culture.

In addition to gender impact, another dimension in our study is previous sports experience and its impact. Previous sports experience significantly influences preferences and fandom. Lim et al. (2012) found that prior football experience increases perceived credibility and favorable attitudes toward the sport, aligning with findings in this study that prior experience predicts football preference. Past experiences with sports can strongly influence how people engage with physical activities later in life. Elsborg et al. (2021) studied inactive adults in Denmark and found that those who had played sports or exercised in the past were better equipped with the skills, confidence, and motivation needed to stay active. This suggests that being involved in sports earlier in life can help people stay connected to physical activity as they get older. Similarly, people with past football experience may be more likely to choose football as their favorite sport, as previous exposure can shape long-term preferences.

This study builds on existing research by examining the combined effects of gender and prior sports experience on football preferences, which has not been extensively analyzed in earlier studies. While previous research has explored gender differences and cultural influences on sports preferences (e.g., Robinson & Trail, 2005; Deaner et al., 2016), our study uniquely considers how experiences in other sports, such as volleyball and basketball, interact with gender to shape preferences for football. By applying robust statistical methods, including Chi-Square tests and logistic regression, this research provides new insights into the cultural and experiential factors influencing football fandom. These findings contribute to the broader understanding of sports preferences and offer practical recommendations for promoting inclusivity in sports.

RESULTS

Data Collection

The data utilized in this study is from a IPOS (September 2021) poll. This poll sampled 1,019 adults aged 18 or older in the U.S. The margin of sampling error is plus or minus 3.3 percentage points at the 95% confidence level for results based on the entire sample of adults.

The poll queried respondents on whether they had played various sports or not (football, soccer, hockey, baseball, basketball, golf, volleyball, rugby, water polo, and other sport). Rugby and water polo are less than 1% favorite rate, and we classified them into other sports.

Gender and Sport Experience Cross Impact

We plot the results to the question of “what’s your favorite sports”, excluding four people who put ‘skip’. The results are summarized as below:

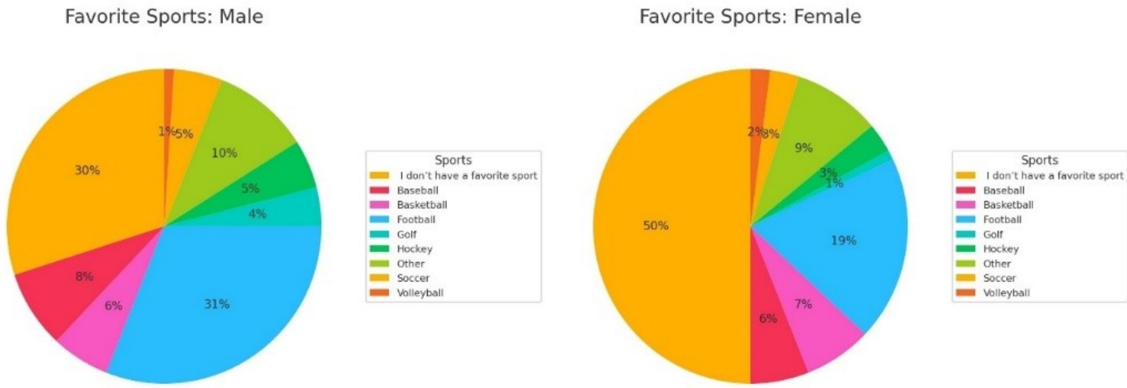


Figure 1: Favorite Sports Among Male and Female

From the plots we can see that football is the most favored sports across both genders. 31% of male responders and 19% of female responders rate football as their favorite sport.

Next, we compare the proportion of males and females participating in various sports, considering both total participants and those whose favorite sport is football, with overall being general proportion regardless of sport preference and "– Football" being the proportion of those who both play the sport and have football as their favorite sport by gender.

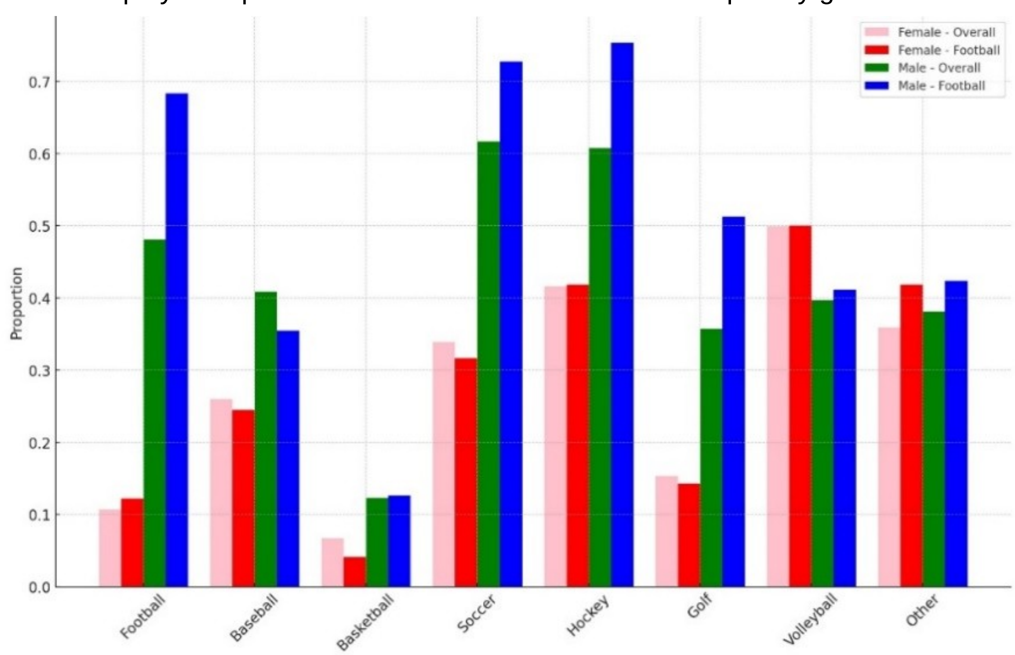


Figure 2: Comparison of Sport Participation by Gender And Football Preference

We compare each sport's participation rate for females and males whose favorite sport is football, the x2 test results are summarized in Table 1:

Table 1: Sport Participation Difference Between Males and Females Who Favor Football

Sport	H0	H1	Chi-Square	P-Value	Conclusion
Football	H0: There is no significant difference in participation rates between males and females for Football.	H1: There is a significant difference in participation rates between males and females for Football.	74.231	6.95E-18	Reject null hypothesis
Baseball	H0: There is no significant difference in participation rates between males and females for Baseball.	H1: There is a significant difference in participation rates between males and females for Baseball.	2.887	0.0892	Fail to reject null hypothesis
Basketball	H0: There is no significant difference in participation rates between males and females for Basketball.	H1: There is a significant difference in participation rates between males and females for Basketball.	4.276	0.0387	Reject null hypothesis
Soccer	H0: There is no significant difference in participation rates between males and females for Soccer.	H1: There is a significant difference in participation rates between males and females for Soccer.	40.136	2.37E-10	Reject null hypothesis
Hockey	H0: There is no significant difference in participation rates between males and females for Hockey.	H1: There is a significant difference in participation rates between males and females for Hockey.	27.516	1.56E-07	Reject null hypothesis
Golf	H0: There is no significant difference in participation rates between males and females for Golf.	H1: There is a significant difference in participation rates between males and females for Golf.	33.875	5.88E-09	Reject null hypothesis
Volleyball	H0: There is no significant difference in participation rates between males and females for Volleyball.	H1: There is a significant difference in participation rates between males and females for Volleyball.	1.581	0.209	Fail to reject null hypothesis
Other	H0: There is no significant difference in participation rates between males and females for other sports.	H1: There is a significant difference in participation rates between males and females for other sports.	~0	1.0000000000	Fail to reject null hypothesis

[The proportion of females who played other sports was 41/59 or 0.694, and the proportion of males who played other sports was 67/91 or 0.736]

Based on these tests and plots, there is no significant difference in participation rates for baseball, volleyball, and other sports among females and males whose favorite sport is football. However, for football, basketball, soccer, hockey, and golf, there is strong evidence for significant differences in participation rates between genders. From Figure 3, it is evident that males' participation rates in football (0.68), basketball (0.13), soccer (0.73), hockey (0.76), and golf (0.51) are considerably higher than those of females (0.12, 0.04, 0.32, 0.42, and 0.14, respectively). There are with 512 males and 507 females, 158 males said football was their favorite spot, while 98 females said football was their favorite spot.

The differences in participation rates can likely be attributed to historical and cultural norms that have traditionally associated sports like football, hockey, and basketball with men. Golf, although less physical, is more associate with male-central business social networks. Limited access and fewer opportunities for females in these sports may also play a role. Media representation and the visibility of male athletes in these sports likely inspire more male participation. In addition, perceptions about physicality and safety may deter females from participating in these traditionally male-dominated sports.

Next, we want to compare the differences in participation rate between general responders and the responders who select football as their favorite sport. Within the male group and within female group, we did the following participation rate tests:

H0: No difference in participation rates between general responders and favor football responders for sport i

H1: Significant difference in participation rates between general responders and favor football responders for sport i
i=1, 2, 3, 4, 5, 6, 7, 8, which are corresponding to Football, Baseball, Basketball, Soccer, Hockey, Golf, Volleyball, and other. Below are the results:

Table 2: Sport Participation Difference Between General Responders and Males Who Favor Football

Sport	Chi-Square	P-Value	Conclusion
Football	19.175	1.19E-05	Significant
Baseball	1.244	0.265	Not Significant
Basketball	0	1	Not Significant
Soccer	5.970	0.014	Significant
Hockey	10.531	0.001	Significant
Golf	11.545	0.001	Significant
Volleyball	0.058	0.809	Not Significant
Other	0.773	0.379	Not Significant

Table 3: Sport Participation Difference Between General Responders and Women who Favor Football

Sport	Chi-Square	P-Value	Conclusion
Football	0.082	0.775	Not Significant
Baseball	0.038	0.846	Not Significant
Basketball	0.567	0.452	Not Significant
Soccer	0.104	0.747	Not Significant
Hockey	0	1	Not Significant
Golf	0.015	0.902	Not Significant
Volleyball	0	1	Not Significant
Other	1.003	0.317	Not Significant

The results show that there is no significant difference in baseball, basketball, volleyball, and other sports participation rates between general participants and those who favor football within the male group. However, those who have played football, soccer, and golf on outdoor grass fields are more likely to rate football as their favorite sport.
On the other hand, for the female group, there is no significant difference in sports participation rates for each of the eight sports between general participants and participants who favor football. This could suggest that female participants have diverse interests and do not overly favor football participation based on their favorite sport.

Logit-Regression Analysis
Based on our previous analysis, we developed the logistic regression model relates the expected log of the odd that favorite sport is football, by introducing football experience, soccer experience, hockey experience, and golf experience as the independent valuables. We find out that significant and higher Pseudo R-squared and lower Log-Likelihood. Therefore a simplified variable of "football experience or not" can already captures key effects more effectively.

$$ln\frac{p}{1-p} = a + b_1x_1 + b_2x_2 \tag{1}$$

Where is the probability of football being the favorite sport, is the intercept, represents gender (0=female, 1=male), indicates if the person has experience with football (0=no, 1=yes).
Results:
=-1.5541, significant (p < 0.001).
: 0.2538, not significant (p=0.124).

: 0.9395, significant ($p < 0.001$).
Pseudo R-squared: 0.04345; LLR p-value: .
From the results, we see that football experience matters the most. People with experience in the sport are much more likely to favor it as their favorite sport.
Gender's Role is marginal. While the coefficient for males suggests a slightly higher likelihood of favoring football, it is not statistically significant. However, in our studies in section 3.2, the percentages of male and female favoriting football are significantly different. However, the proportion of females who have football experience (11%) is much smaller than the proportion of male who have football experience (42%), therefore, heavily deduct the impact of gender different.
The model has a pseudo R-squared of 0.04345, indicating that while football experience is a key factor, other variables not included in the model also influence football preference. While statistically significant, this indicates other factors not included in the model might also influence football preference.
Therefore, efforts to promote football related business should target increasing engagement or participation, as experience strongly correlates with preference.
Strategies might need to address Gen Z's lower interests and participation for football, possibly by aligning the sport with Gen Z's unique values or cultural interests.
Gender differences in football preference appear small. It indicates that promotional strategies need not heavily differentiate between men and women.

DISCUSSION

This study finds that gender and prior sports experience significantly influence football preference, aligning with research on how cultural norms and sports exposure shape interests. Men are more likely than women to favor football, a trend linked to historical and cultural factors. Sports fandom has traditionally been male dominated, often associated with masculinity and competitiveness (Warm, Waddill, & Dunham, 2004). Robinson and Trail (2005) further highlight that while gender is not the main factor, it still plays a role in shaping spectatorship and participation. Likewise, our logistic regression analysis shows that prior football experience is a stronger predictor of preference than gender, likely due to higher male participation rates.
Sports experience also strongly impacts football preference. Men who have played football, soccer, hockey, or golf are more likely to favor football, supporting Lim et al. (2012), who found that prior exposure enhances credibility and positive attitudes toward the sport. Elsborg et al. (2021) further demonstrate that past physical activity strengthens long-term engagement. This connection between experience and preference is stronger in men than women, possibly due to different motivations. Ridinger and Funk (2006) suggest that women often engage with sports for emotional and social reasons rather than competition. Additionally, Deloitte Insights (2023) found that 36% of female sports fans became fans due to influence from male family members, highlighting the role of family dynamics in shaping women's interest in football.
Gender inclusivity is key to football’s long-term popularity. Expanding female participation through increased funding, outreach programs, and better media representation could help grow the sport’s female fanbase. Family-focused initiatives, such as discounted tickets and co-ed youth leagues, could further encourage female engagement. Promoting female athletes and diverse role models may also foster greater inclusivity.
Beyond social factors, football must address its environmental and economic impact. Large stadiums, frequent travel, and event operations contribute to its carbon footprint. Eco-friendly initiatives, such as energy-efficient stadiums, carbon reduction programs, and sustainable event management, could reduce this impact.
Economically, football must adapt to changing consumer habits. As esports and digital engagement grow, new revenue streams beyond ticket sales and broadcasting rights—such as

subscription-based digital content, AI-driven fan experiences, virtual merchandise, and blockchain-based engagement—could help sustain the sport. Additionally, improving gender diversity in leadership positions can support both economic and social sustainability (Lesch et al., 2022).

These findings offer insights for sports marketing, youth engagement, and inclusivity initiatives. Encouraging participation, promoting family-friendly activities, and fostering greater awareness among women can help ensure football's long-term growth and sustainability.

CONCLUSION

This study confirms that gender and prior sports experience significantly influence football preference. Men are more likely to favor football, largely due to higher participation rates and cultural factors. However, prior football experience is a stronger predictor of preference than gender, particularly among men, while women's interest is often shaped by family influence. Women's football preference appears to be influenced more by emotional, social, and family dynamics rather than personal participation. Since men have historically dominated football, increasing female engagement through targeted initiatives, family-oriented programs, and inclusive media representation could help shift this trend.

To ensure football's long-term popularity, efforts should focus on expanding participation, promoting gender inclusivity, and fostering sustainability. Encouraging football as a family-friendly sport and adapting marketing strategies to appeal to a broader audience will be key to maintaining and growing the sport's fanbase.

REFERENCES

Correia, A., & Esteves, S. (2007). An exploratory study of spectators' motivation in football. *International Journal of Sport Management and Marketing*, 2(5-6), 572–590.

Deaner, R. O., Balish, S. M., & Lombardo, M. P. (2016). Sex differences in sports interest and motivation: An evolutionary perspective. *Evolutionary Behavioral Sciences*, 10(2), 73–85. https://doi.org/10.1037/ebs0000069.

Deloitte Insights. (2023). Engaging women sports fans: How teams, leagues, and brands can expand their audience. Deloitte. Retrieved from https://www2.deloitte.com/us/en/insights/industry/technology/female-sports-fans-engagement.html.

Elsborg, P., Heinze, C., Melby, P. S., Nielsen, G., Bentsen, P., & Ryom, K. (2021). Associations between previous sport and exercise experience and physical literacy elements among physically inactive Danes. *BMC Public Health*, 21, 1-10.

Farrell, A., Fink, J. S., & Fields, S. (2011). Women's sport spectatorship: An exploration of men's influence. *Journal of Sport Management*, 25(3), 190–201. https://doi.org/10.1123/jsm.25.3.190.

Gender and Cultural Diversity in Sport, Exercise, and Performance Psychology. (n.d.). Retrieved from https://oxfordre.com/psychology.

Jakubowska, H., Antonowicz, D., & Kossakowski, R. (2020). Female fans, gender relations and football fandom: Challenging the brotherhood culture. Routledge.

IPOS. (2021, September). psos Understanding Society: Wave 21 September 2021 [Roper #31118741]. Retrieved from https://ropercenter.cornell.edu/ipoll/study/31118741.

Lesch, L., Kerwin, S., Thormann, T. F., & Wicker, P. (2022). Critical Masses and Gender Diversity in Voluntary Sport Leadership: The Role of Economic and Social State-Level Factors. *Sustainability*, 14(10), 6208. https://doi.org/10.3390/su14106208.

Lim, C. H., Chung, J., Frederick, E., & Pedersen, P. M. (2012). Investigating the influence of past football experience and perceived football knowledge on attitude, credibility, expertise and trustworthiness, and future search intention. *Journal of Contemporary Athletics*, 6(1), 1–11.
McDonald, H., Pallant, J., Funk, D. C., & Kunkel, T. (2024). Who doesn't like sport? A taxonomy of non-fans. *Sport Management Review*, 27(1), 67–89. https://doi.org/10.1016/j.smr.2023.06.004.
Ridinger, L., & Funk, D. C. (2006). Looking at gender differences through the lens of sport spectators. *Sports Marketing Quarterly*, 15(3), 155–166.
Robinson, M. J., & Trail, G. T. (2005). Relationships among spectator gender, motives, points of attachment, and sport preference. *Journal of Sport Management*, 19(1), 58–80. https://doi.org/10.1123/jsm.19.1.58.
Warm, D. L., Waddill, P. J., & Dunham, M. D. (2004). Using sex and gender role orientation to predict level of sport fandom. *Journal of Sport Behavior*, 27(4), 367–378.
Sveinson, K., & Hoeber, L. (2015). Overlooking the obvious: An exploration of what it means to be a sport fan from a female perspective. *Leisure Studies*, 34(4), 405–419. https://doi.org/10.1080/02614367.2014.938769.
Gender Differences in Physical Activity Motivators and Context Preferences in Adults of Various Ages. (n.d.). Retrieved from https://bmcpublichealth.biomedcentral.com.
Associations Between Previous Sport and Exercise Experience and Physical Literacy in Inactive Adults. (n.d.). Retrieved from https://bmcpublichealth.biomedcentral.com.
Thinkhouse HQ. (2021). How Gen Z is reshaping the sports landscape. Retrieved from https://www.thinkhousehq.com/the-youth-lab/how-gen-z-is-reshaping-the-sports-landscape.

DECISION SCIENCES INSTITUTE

Sustaining The Surge: Twitter Topics and Persistent Momentum of Meme Stocks

Shaoqing Zhang
Le Moyne College
Email: zhangsh@lemoyne.edu

Xinyuan Tao
New Jersey Institute of Technology
Email: xtao@njit.edu

ABSTRACT

We explore the impact of social media on meme stock price movements, particularly in the post-GameStop short squeeze period. By analyzing 2.6 million tweets, we examine how investor discussions on platforms like Twitter influence stock momentum. Using a customized BERT model for sentiment classification, we identify significant correlations between social media discourse and price fluctuations. Our findings suggest that meme stock momentum persists beyond major events, offering insights into market dynamics and predictive applications for financial analysts. This research bridges gaps in understanding retail investor behavior, highlighting the role of online discourse in shaping financial market trends.

KEYWORDS: Social media, Retail investor, Market momentum, Twitter, Meme stock

INTRODUCTION

As an unprecedented event that highlighted the evolving power dynamics between retail investors and Wall Street professionals, the short squeeze of GameStop Corp ($GME) was largely fueled by a massive mobilization of individual investors who coordinated their actions through social media platforms.[The GameStop short squeeze in January 2021 was a remarkable financial event where retail investors, largely coordinated through the Reddit forum r/WallStreetBets, drove up the stock prices of GameStop, a struggling video game retailer. This was primarily aimed at countering hedge funds that had heavily shorted GameStop's stock, betting on its price to fall. As retail investors began purchasing the stock en masse, the price soared, causing significant losses for short sellers who were forced to buy back shares at higher prices to cover their positions. This surge caused the stock price to increase by over 1,500% at its peak.] This remarkable market movement not only resulted in substantial financial consequences for many hedge funds and other institutional investors but also sparked a broader discussion about market dynamics (Zheng et al., 2021), the role of retail investors (Van Kerckhoven & O'Dubhghaill, 2023; Allen et al., 2021), and the impact of social media on financial markets (Khan et al., 2022). Suddenly, stocks similar to GameStop have gained significant attention and experienced dramatic price fluctuations, largely driven by discussions and hype on social media platforms rather than traditional financial fundamentals.

Meme stocks, such as GameStop and AMC, are often highly shorted by short sellers, although this is not their defining characteristic. The label "meme stock" primarily comes from how these stocks are popularized and hyped on social media platforms, particularly within online communities like Reddit and Twitter. These stocks become the subject of widespread attention and discussion, often due to their dramatic and rapid price movements, which are influenced more by retail investor enthusiasm and speculative trading than by traditional financial fundamentals (Yahya & Chiu, 2021). Meme stocks offer a natural experiment to study the impact of investors' discussion. expressed on social media. Their price movements are often directly influenced by the hype and discussions that take place on platforms like Twitter. Second, meme stocks typically showcase clear examples of herding behavior, where investors follow the crowd based on social media trends rather than traditional financial analysis. Moreover, the discussions surrounding meme stocks are usually robust and frequent, providing a substantial amount of data for feeding natural language processing (NLP) models. By focusing on meme stocks, it can be effectively explored how modern investment decisions are shaped by non-traditional influences such as social media, offering insights not only into market dynamics but also into the broader implications of technology on financial markets.

With the advancement of technology, the relationship between social media and the capital market has been a subject of academic inquiry for the past two decades. On the one hand, prior literature has documented the significance of social media sentiment and emotion on mainstream stock market movement (Awan et al., 2021; Khan et al., 2022; Fekrazad et al., 2022), trading volume and returns (Tan & Tas, 2021; McGurk et al., 2020), market liquidity (Agrawal et al., 2018), market volatility (Audrino et al., 2020), and market crash (Gider & Westheide, 2016; Guan, Liu, & Cheng, 2022; Hossain, Mammadov, & Vakilzadeh, 2022). Findings in those studies suggest social media's role in financial markets is multifaceted, influencing daily stock movements, impacting returns through the rapid spread of information and sentiment, and potentially contributing to market volatility during periods of financial stress. On the other hand, existing literature has found that the tone and the volume of discussions on social media and search engines during the short squeeze event had substantial impact on $GME return, volatility and put-call ratio (Anand & Pathak, 2021; Umar et al., 2021), price movement (Wang & Luo, 2021), and the profitability of investors (Vasileiou, Bartzou, & Tzanakis, 2021). Rather than focusing on the short squeeze event itself, our research question aims to explore whether changes in the topics discussed by retail investors on social media can serve as indicators of meme stock price movements following the short squeeze event.

Although literature pointed out the herding behavior of retail investors exists even after the GameStop short squeeze event (Lyócsa, Baumöhl, & Výrost, 2022; Klein, 2022; Aloosh, Choi, & Ouzan, 2021), there are limited studies examining whether the momentum amplified by social media lasts in the post-event period. In this research, we examine the association between retail investors' topics on Twitter and the meme stock price movement beyond the GameStop short squeeze. We argue that while the exact circumstances surrounding the GameStop short squeeze may not be easily replicable, the event has undeniably shifted some dynamics in the stock market, leading to a long-lasting momentum-driven market phenomenon, particularly in cases where stocks are heavily shorted. As a pivotal social media platform during the GameStop short squeeze, Twitter offers an invaluable context for exploring the discussions and topics of retail investors. The features of Twitter facilitate the mobilization of collective action, encourage information sharing, impel community building, and amplify the perception, allowing us to separate the discussion and timing of specific meme stocks. Our analysis of discussions on Twitter regarding meme stock activity reveals compelling evidence that momentum persists even after the initial event has concluded.

To gain insights into the predominant topics among retail investors concerning meme stocks, we selected the top 13 meme stocks and analyzed approximately 2.6 million tweets from March 2021 to May 2022. This is achieved by searching for tweets using the cashtag associated with each meme stock. Due to the nature of meme stocks, retail investors are inclined to express extreme sentiments on Twitter, such as bullish or bearish positions. To systematically analyze these topics, we employed a supervised topic modeling approach to train a customized BERT model. This model classifies each tweet into one of three categories: bullish, bearish, or neutral. Acknowledging that each tweet varies in its impact on Twitter, we use the sum of the number of retweets, replies, and likes as weights to construct daily bullish and bearish indices corresponding to each meme stock. Our analysis reveals that for meme stocks frequently discussed (highlighted), there is a positive correlation between the intraday Weighted Net Topic Densities (WNTD)—the difference between weighted bullish and bearish topic densities during trading hours—and their intraday price movements. Furthermore, the application of overnight WNTD during off-trading hours is also effective in predicting price movements between consecutive trading days. Interestingly, we do not find any significance for GME. For meme stocks that are less attractive (overlooked), our findings indicate a generally positive association between the overall WNTD and the intraday price movements. However, the results are mixed when we target the overnight price movements. In general, our findings provide support to our hypothesis that the momentum of the meme stock market lasts in the post-event period.

This research contributes to the rapidly expanding literature examining the impact of social media on financial markets. Diverging from previous research that primarily examines mainstream stocks (Agrawal, Azar, Lo, & Singh, 2018; (Awan et al., 2021)), our study delves into meme stocks, noted for their high volatility and the contrarian strategies they inspire. By focusing on real-time data from Twitter and its correlation with meme stock price movements, particularly after the high-visibility event of the GameStop short squeeze, this study provides valuable insights into how retail investors' topics and discussions can influence market dynamics. By incorporating investor attention as a weighting factor, we devise a novel methodology that enhances our ability to gauge the dissemination of tweets, offering a more precise measurement for tracking changes in the topics of interest to investors over time. Additionally, it explores the potential of social media as a predictive tool for financial analysts and investors, offering a new perspective on traditional market analysis techniques. This work not only bridges the gap between retail investors' herding behavior and momentum of financial markets but also suggests practical applications for market prediction strategies leveraging online discourse.

Our research introduces a significant advancement in the field of natural language processing (NLP) by customizing the BERT base model with a tailored tokenizer and domain adaptation techniques specifically for classifying tweets related to meme stocks. This customization has demonstrated superior performance compared to traditional machine learning methods, marking a pivotal contribution to the NLP literature. By adapting the BERT model to better understand

and process the unique lexicon and topics of social media discourse, our approach not only enhances accuracy but also offers deeper insights into the dynamic interplay between social media activity and financial market movements. This work underscores the potential of fine-tuned language models in specialized domains, providing a robust framework for future studies exploring the impact of online topics on the financial industry.

The structure of this research is as follows: The Literature Review discusses existing research on social media and financial markets, providing the foundation for our study and developing the hypotheses based on prior literature. The Data section describes the datasets used in this study, including data sources, collection methods, and data processing. In the Empirical Results section, we present and discuss our methodologies employed and primary empirical findings, followed by robustness tests. Finally, the Conclusion section summarizes key insights, implications, and potential directions for future research.

LITERATURE REVIEW

The burgeoning interest in leveraging online activity to predict financial market trends has become a prominent theme in recent academic and financial research. As platforms such as Twitter, Reddit, and Google become integral to daily communication, they also provide a wealth of real-time, user-generated data that reflects public attention and opinion. Li et al. (2015) explores the utility of Google search volume index (GSVI) in analyzing crude oil prices and trader behaviors from January 2004 to June 2014. The study finds that GSVI effectively captures the attention of noncommercial and nonreporting traders but not commercial traders, demonstrating that search data can predict price movements.

Existing literature explores the evolving methodologies and findings from studies that analyze how sentiments expressed on social media correlate with stock market movements and return. Ranco et al. (2015) analyzes the relationship between Twitter activity related to the Dow Jones Industrial Average (DJIA) companies and their stock market performance over 15 months. Results show that the direction of stock returns is influenced by the sentiment polarity during these Twitter peaks, with a modest but significant impact on stock returns in the days following special events such as earnings announcements.

Pagolu et al. (2016) examine the correlation between stock market movements and public sentiments on Twitter, focusing on how sentiments expressed in tweets about a company relate to its stock price changes. Utilizing sentiment analysis and supervised machine learning techniques, the research indicates that positive sentiments in tweets can lead to increased investment and higher stock prices, highlighting Twitter's utility as a predictive tool for stock market trends. However, existing research primarily examines mainstream stocks and industrial indices, which typically exhibit lower volatility. This focus often overlooks the unique dynamics and greater fluctuations seen in less conventional sectors or high-volatility assets like meme stocks, which can provide additional insights into market behavior under more extreme conditions.

Recent research has delved into the underlying factors and motivations behind the GameStop short squeeze, highlighting a range of interpretations that connect social media dynamics to the pricing behaviors of meme stocks. Vasileiou et al. (2021) utilizes intraday data from January 4, 2021, to March 26, 2021, to demonstrate that trading volume and Google search metrics are key indicators of GME's market performance. Findings suggest that rapid access to such data significantly benefits investors. Similarly, Lyócsa et al. (2022) investigates the price fluctuations of four major meme stocks during 2020 and early 2021, and find that part of the next day's price variation can be explained by an increase in activity on the WSB subreddit relative to Google searches. Moreover, Kim et al. (2023) examine the behavior of individual investors on social media platforms during the GameStop short squeeze in early 2021, specifically analyzing posts from the r/WallStreetBets subreddit. By conducting text-based sentiment analysis and comparing user informedness on different social media platforms, the study demonstrates that coordinated trading and strategy discussions among retail investors led to the short squeeze. Findings indicate that the tone and volume of social media posts are closely tied to GME's trading volumes and the emergence of potentially irrational trading patterns. While existing studies highlight the significant impact of retail investors' activities on social media during the GameStop short squeeze, research into the broader spectrum of meme stocks and the sustainability of their market momentum post-event remains sparse. Our research objective aims to fill this gap.

To decipher social media sentiments, researchers have leveraged advanced natural language processing (NLP) tools and machine learning techniques to mine and interpret textual data, employing a variety of approaches. Li et al. (2014) utilize the Harvard psychological dictionary and the Loughran-McDonald financial sentiment dictionary to create a sentiment space for stock price prediction. Ranco et al. (2015) train a Support Vector Machine (SVM) to classify 1.5 million tweets into positive, negative, and neutral sentiment for stock price return prediction. Mohan et al. (2019) enhanced the accuracy of stock price predictions by incorporating sentiment analysis derived from deep learning models trained on a substantial corpus of financial news articles. In various other works, other machine learning models such as Naïve Bayes (Kalyani et al., 2016; Khedr & Yaseen, 2017) and Random Forests (Kabbani & Usta, 2022; Kalyani, Bharathi, & Jyothi, 2016) have been used for sentiment extraction.

Although sentiment analysis offers a quick glimpse into the general mood around text, topic modeling reveals deeper insights by effectively capturing the rapid, sarcastic, or ironic nuances prevalent in social media language. Prior literature has demonstrated the effectiveness of topic modeling. Huang et al. (2018) employs textual analysis and topic modeling to investigate how analysts act as information intermediaries through their reports and corporate disclosures. Nguyen et al. (2015) propose a new topic model TSLDA to capture the topics on social media for predicting stock price movement. However, research on the attention-based topic modeling of highly shorted stocks that retail investors discuss on social media is limited. We intend to apply an advanced supervised topic modeling technique collaborated with a unique investors' attention factor to interpret meme stock tweets, offering fresh insights and expanding upon existing topic modeling literature.

The concept of the “wisdom of the crowd” refers to the idea that collective opinions or decisions of a group can be more accurate or effective than those of individual experts. It can sometimes manifest as herd behavior, which is proved to be noticeable among retail investors involved with meme stocks (Klein, 2022). By using the hourly return data, Aloosh et al. (2021) provide robust evidence that herding behavior exists in major meme stocks even beyond the GameStop short squeeze event. Consistent with the existing literature, Zhou et al. (2009) find that herding tends to be more prevalent with small stocks and in economic downturns.

When assessing the impact of social media on investor behavior and market trends, the use of engagement metrics such as the number of retweets, likes, and replies on Twitter can serve as effective weights for quantifying the importance and attention particular meme stocks or financial topics receive from investors. These metrics act as proxies for investor interest and sentiment, underpinning the assumption that higher engagement indicates greater relevance and influence among the speculator community. For instance, a meme stock tweet that garners a high number of retweets, likes, and replies suggests that the information is resonating widely with retail investors, potentially signaling market movements based on the topics expressed. Therefore, we expect that the difference between weighted bullish and bearish topic densities of each meme stock is positively associated with the corresponding price movement beyond the short squeeze for those that are highlighted. In our settings, those highlighted meme stocks are AMC, GME, NIO, PLTR, LCID, and SOFI, total tweets of which occupy 90% of the entire dataset. We state our first hypothesis as follows:

> Hypothesis 1: *There exists a positive relation between the weighted net topic density (WNTD) for each highlighted meme stock and the corresponding price movement beyond the GameStop short squeeze event.*

Conversely, meme stocks that receive minimal discussion on Twitter are often mentioned alongside more popular meme stocks and typically garner less attention from investors. More importantly, when the highly popular meme stocks such as GME and AMC experience significant hype and rallies to a peak, other less popular meme stocks may also see increased trading activity and price movements as investors anticipate similar trends or seek alternative opportunities within the same speculative environment. This phenomenon reflects the herd mentality and can lead to a ripple effect across related or similar stocks, driven by momentum rather than underlying fundamentals (Lyócsa et al., 2022). Therefore, we presume that the overall WNTD is positively related to the price movements beyond the short squeeze for less attractive (overlooked) meme stocks.[Those overlooked meme stocks are BB, ABNB, RIVN, MU, NOK, BBBY, and GOEV, total tweets of which account for 10% of the entire dataset.] Compared with the stock-specific WNTD, the overall WNTD offers a generalized view, potentially smoothing over spikes caused by events affecting any single stock and assessing broader trends across a sector influenced by similar trading behaviors and investor demographics. We conclude our second hypothesis as follows:

> Hypothesis 2: *There exists a positive relation between overall weighted net topic density (WNTD) and the overlooked meme stock price movement beyond the GameStop short squeeze event.*

DATA

As a dynamic and influential virtual community, Twitter is widely used by individuals and organizations to gain public visibility. The activities of individual investors on Twitter have become significantly impactful on the financial market and investment decisions. Twitter users have begun to rely on the platform as source of financial information, sharing their thoughts in real time. With its varied user base, Twitter offers an ideal setting for studying the behaviors of retail investors. Consequently, this thesis utilizes Twitter as the primary data source.
To access full text and structured data from Twitter, we use the Twitter Application Programming Interface (API) to connect to the platform's database. The API enables key words searches over designated date ranges, facilitating the creation of datasets tailored to our requirements.

Data retrieving

In order to thoroughly understand investors' opinion on Twitter, we consider 13 popular meme stocks [The list of the most mentioned meme stocks keeps changing over time, we refer to the list at the time we retrieve Twitter data. They are AMC Entertainment Holdings Inc. (AMC), GameStop Corp (GME), Nio Inc. (NIO), Palantir Technologies Inc. (PLTR), Lucid Group Inc. (LCID), SoFi Technologies Inc. (SOFI), BlackBerry Ltd. (BB), Airbnb Inc. (ABNB), Rivian Automotive Inc. (RIVN), Micron Technology Inc. (MU), Nokia Oyj (NOK), Bed Bath & Beyond Inc. (BBBY), and Canoo Inc. (GOEV).] that are most mentioned on */r/wallstreetbets* forum. The common characteristic of those meme stocks is that they are highly exposed and discussed on Twitter, resulting in plenty of data for retrieving.

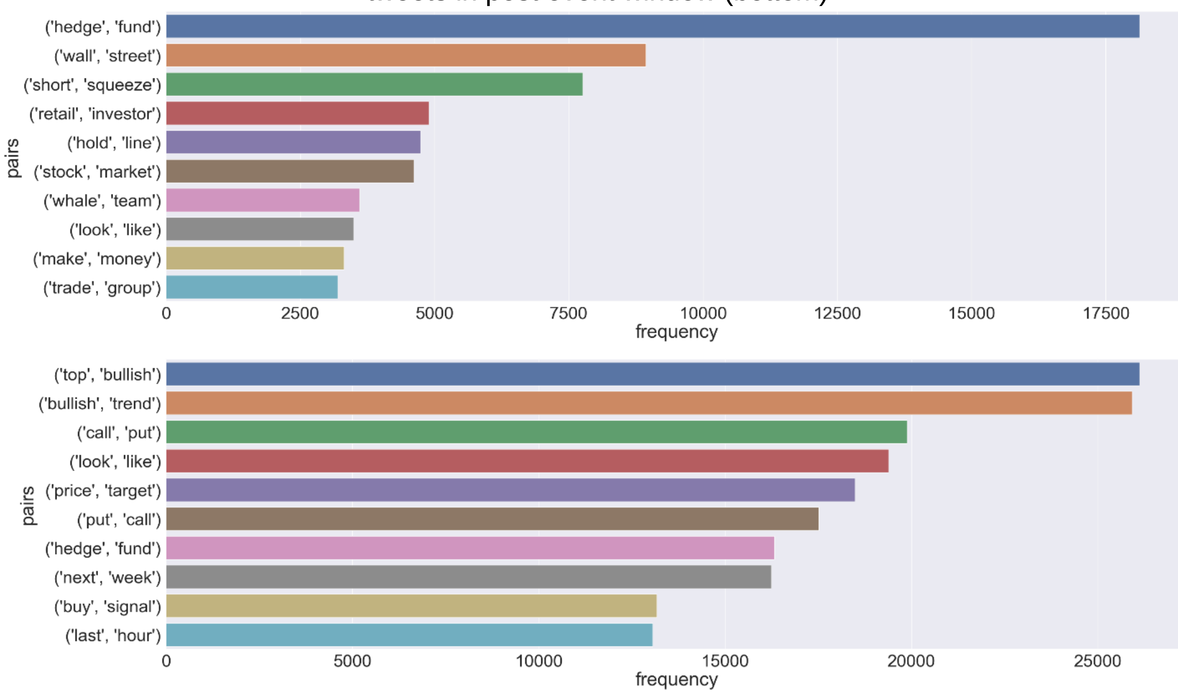
As $GME gained the most popularity and dramatic change among all the meme stocks discussed on social media, we use $GME as a benchmark to settle the time frame of this research. According to the definition on Wikipedia (Wikipedia contributors, 2023), GameStop short squeeze took place in January 2021 and reached its peak in late January. We set the date range from January 11th to February 5th in 2021 as the short squeeze window, which includes the initial growth, peak, and slump of the $GME trading volume, and the date range from March 1st, 2021, to May 31st, 2022, as the post-event window. We skip the rest of February to reduce the effect of this cooling period.

For the short squeeze window, we apply the Twitter API to search tweets that include keywords such as "$GME" and "@GameStop," and retrieve the tweets that contain information on both account level (i.e., number of followers) and tweet level (i.e., number of retweets). The account-level information includes the creation date of the account, whether it is verified, the total number of followers, and the number of tweets at the time of download, etc. The tweet-level information is meant for every single tweet message, including the date and time of posting, text content, hyperlink attached, picture or video embedded, hashtag and cashtag attached, and the number of replies, retweets, and likes received. For the post-event window, we search for tweets that are attached to 13 meme stock cashtags. As some tweets may contain one or more cashtags, we remove the duplicates and retain the distinct ones. In total, 481,793 tweets are retrieved in the short squeeze window, and 2,609,353 tweets are retrieved in the post-event window.

In our dataset, $AMC and $GME are the two meme stocks with the most popularity on Twitter, and a majority of the $AMC tweets have an exclusive cashtag. Conversely, about 63% of the $GME tweets have more than one cashtag.

In order to have an overall understanding of the topics included in the short squeeze window and post-event window, we investigate the words frequency in our dataset.

Figure 1: Top 10 word pairs of $GME tweets in short squeeze window (top) and meme stock tweets in post event window (bottom)



The words frequency shows that words "buy," "sell," "hold," "amp," etc. are strongly used to express investors' opinions. In addition, Figure 1 shows the top 10 word pairs in both windows, confirming that retail investors hype the short squeeze to push the market. Meaningful word pairs such as "short squeeze," "hold line," "top bullish," and "buy signal" are highly mentioned in those tweets in the short squeeze window as well as in the post-event window.

The BERT model

Although sentiment analysis can show us the emotional tone behind a body of text, typically categorized into positive, negative, or neutral sentiments, we are not able to discover hidden thematic structures in a large text corpus. The advance of machine learning techniques brings us state-of-the-art topic modeling algorithms to analyze the potential topics included in text content. Topic modeling is a tool used in natural language processing (NLP) to identify abstract topics in a collection of documents.

The core of traditional topic modeling algorithms is basically the application of statistical methods on the frequency and co-occurrence of words within the content of large documents such as tons of tweets. The results from sentiment topic analysis can vary substantially. Consider the below tweet about $AMC as an example:

> *"AMC's rollercoaster ride is nerve-wracking! The stock's unpredictability is a real concern, but I'm holding on because I believe in the 'apes' movement and expect a squeeze any day now. This could really shake up Wall Street!"*

In the example text above, the nuances of sentiment analysis versus topic analysis are particularly clear. Sentiment analysis focuses on determining the overall emotional tone of the text, which in this case is predominantly negative due to words like "nerve-wracking" and phrases such as "unpredictability is a real concern." These elements highlight the user's anxiety and discomfort regarding the current state of the stock, leading to an overall negative sentiment classification. Conversely, topic analysis, especially when analyzing investment intent, reveals a different aspect.

Despite the overall negative emotional tone, the topic analysis would uncover a bullish opinion towards the stock. The phrases "I'm holding on because I believe in the 'apes' movement" and "expect a squeeze any day now" indicate a strong belief in the potential for the stock's price to rise sharply due to collective actions by individual investors. This reflects a commitment to the stock based on specific market dynamics (the 'apes' movement and a potential short squeeze), which is indicative of a positive outlook on the stock's future performance.

There are a few popular traditional topic models, including Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Non-negative Matrix Factorization (NMF). In particular, these techniques are based on statistical modeling to extract topic patterns from a collection of documents. However, the efficacy of traditional topic modeling methods in analyzing social media data has been highly criticized (Egger & Yu, 2021; Egger & Yu, 2022). Prior literature shows that LDA is likely to ignore the co-occurrence relations of topics in tweets (Jaradat & Matskin, 2019) and social media texts are too noisy and sparse to provide sufficient information for LDA to conduct effective statistical learning (Chen et al., 2019).

In contrast, the BERT (Bidirectional Encoder Representations from Transformers) model (Devlin et al., 2018), developed by Google in 2018, represents a significant breakthrough in the field of natural language processing (NLP). BERT is built on the Transformer architecture, which was introduced in the paper "Attention is All You Need" by Vaswani et al. in 2017 (Vaswani et al., 2017). The Transformer model eschews traditional recurrent neural network (RNN) structures in favor of attention mechanisms. BERT specifically utilizes the Transformer's encoder mechanism. The core idea of the Transformer is to handle sequences of data in parallel and to capture contextual relationships between words in a sentence, regardless of their position. This bidirectional understanding allows BERT to grasp subtler nuances of language, making it highly effective for tasks that require a deep understanding of context, such as sentiment analysis, question answering, and language inference.

Another key advantage of BERT is its transferability across various NLP tasks with minimal task-specific tuning. Once pre-trained on a large corpus of text, BERT can be fine-tuned with just one additional output layer to adapt it for a wide range of applications, from text classification to entity recognition. This versatility reduces the need for extensive task-specific model training, saving time and computational resources.

In this research, instead of using BERT to discover previously unknown topics, we manually pass the pre-defined topics to the model and try to learn the probability of those topics corresponding to each meme stock tweet. In other words, we in fact perform a supervised topic modeling technique, which is transformed into a classification problem.

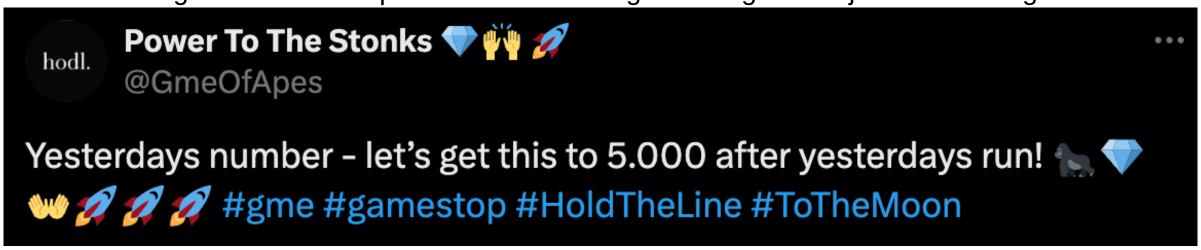
Customized tokenization

As one of the biggest pivotal players in the AI community, Hugging Face lowers the barrier to entry for using advanced natural language processing (NLP) models and fosters an open and collaborative ecosystem. Hugging Face provides multiple pre-trained BERT models, which are trained on various large corpora with different numbers of parameters. We choose the pre-trained "bert-base-uncased" model (12 layers, 768 hidden dimensions, 12 attention heads, and 110M parameters) for the following reasons.

On the one hand, in our tweets data, where case distinction might not affect the meaning of words, we use an uncased model to simplify processing and to prevent the model from attributing importance to case where none is intended. For example, we do not differentiate the meaning of words "buy" and "BUY". On the other hand, the "base" model of BERT, as opposed to the "large" model, has fewer parameters (110 million vs. 340 million). This makes the base model faster and less resource-intensive to train and deploy, which is beneficial for applications with limited computational resources. For many applications, BERT base provides a good balance between performance and computational efficiency. Additionally, with fewer parameters and no need to manage different cases, the BERT base uncased model can be easier to fine-tune on new datasets.

Meme stocks, like $GME and $AMC, often feature discussions filled with specialized jargon, internet slang, emojis, and community-specific references that standard language models trained on general text might not accurately interpret or understand. We start with a tokenizer from the BERT base model and extend the tokenizer vocabulary on the basis of unique and meaningful elements included in meme stock tweets. We therefore add tweet-specific tokens such as popular emojis or hashtags. For example, emojis and hashtags in Figure 2 show strong bullish sentiment related to $GME.

Figure 2: An example of tweet including meaningful emojis and hashtags



Domain adaptation

Domain adaptation of the BERT model is a powerful technique that tailors a general-purpose language model to specific contexts or industries, significantly enhancing its performance for targeted tasks. This adaptation is crucial because language usage varies widely across different fields such as legal, medical, financial, or technical domains, where specialized terminology and unique linguistic structures are prevalent. By training BERT on domain-specific texts, the model learns the peculiarities of language and context relevant to that field, which improves its understanding and predictive accuracy. This results in more reliable natural language processing applications, from improved document classification and more accurate information extraction to finer sentiment analysis and enhanced question-answering systems. Additionally, domain-adapted BERT models can handle domain-specific ambiguities and jargon better, reducing errors that occur due to misinterpretation of terms that have different meanings in general usage. Overall, domain adaptation makes BERT not just a versatile tool for broad NLP tasks but also a specialized instrument tuned to meet the exact needs of specific industries or applications.

In order to enhance the model's ability to understand and interpret the unique language, slang, and sentiment commonly found in meme stock-related tweets, we apply domain adaptation on the BERT base model using our entire meme stock market tweets. By adapting BERT to this specific domain—through unsupervised learning on large volumes of unlabeled meme stock tweets—the model can better grasp the nuances and colloquialisms unique to this investing subculture. This improved understanding enables more accurate sentiment analysis, better detection of trends, and more nuanced interpretation of the community's reactions, which are crucial for us trying to understand the relationship between investors' opinion and stock movements based on social media dynamics.

The primary technique used for further pre-training BERT on domain-specific data is Masked Language Modeling (MLM). Consistent with the original BERT model, we mask 15% of all customized tokens in each sequence during training. We configure our training environment with a batch size of 8, a learning rate of 5e-5, and 3 epochs. Because tweets are commonly shorter, we set the maximum sequence length to 128 for faster training without sacrificing performance. The objective for the base model is to predict the original token that was masked based solely on its context (the unmasked tokens). This task forces the base model to develop a deep contextual understanding of the meme stock tweets. Our domain-adapted model is saved for further fine-tuning.

Fine-tunning

Fine-tuning a BERT base uncased model after domain adaptation offers several significant advantages, particularly enhancing the model's effectiveness and efficiency for specialized tasks. Domain adaptation first tailors the model's embeddings and contextual understanding to meme stock market tweets. As we ultimately perform a sequence classification, following up with fine-tuning then allows this domain-adapted model to apply its finely tuned language capabilities to classify each meme stock tweet into our pre-defined categories: bullish, bearish, and neutral.

Training a fine-tuned model requires input texts and their corresponding labels. Due to the scarcity of labeled meme stock tweets, we use the existing financial-related tweets dataset "Twitter Financial News Sentiment" on Hugging Face as a main source of our training data.[The dataset contains 11,931 financial-related tweets that scrapped using Twitter API, with corresponding label of Bullish, Bearish, or Neutral. The version of the dataset is 1.0.0 and is released under the MIT License. https://huggingface.co/datasets/zeroshot/twitter-financial-news-sentiment] Additionally, we manually labeled 705 meme stock tweets as a supplement for the model to learn features beyond regular financial tweets. As a result, the data contains a total of 12,636 labelled tweets. The examples of labelled tweets are shown in Appendix. We next split the data into two sets: 70% for training and 15% for validation. The remaining 15% is the test set for the final evaluation.

After tokenizing the data, we use the sequence classification model from Hugging Face's Transformers library. this model uses the transformer architecture, primarily built around the concept of self-attention mechanisms, allowing them to process entire sequences of data (like sentences or paragraphs) simultaneously. It is, therefore, fundamentally different from traditional machine learning classification techniques. In comparison, we conduct conventional techniques include a variety of multi-classification algorithms such as Support Vector Classifier (SVC), Random Forests, and Multilayer Perceptron (MLP). In terms of the feature engineering, we consider two popular methods, Tf-idf and BERT embeddings. Although both of the methods can convert text into numerical format such as vectors, they are under different mechanisms. BERT starts by converting words into embeddings using an initial embedding layer that contains pro-trained representations, while Tf-idf is a statistical measure used to evaluate how important a word is to a document in a collection of documents. BERT base model typically outputs embeddings with a dimensionality of 768, which is expensive for the use in downstream classification tasks. We apply Principal Component Analysis (PCA), a statistical technique used for dimensionality reduction while preserving as much variance as possible, on the BERT embeddings, and keep 295 components (95%) to explain the variance. We evaluate the performance of models on our testing dataset.

Table 1: Performance of classifiers on different embedding methods

	Micro F1 Score			
	SVC	RF	MLP	Domain-Adapted BERT
Tf-idf	0.81	0.73	0.79	-
BERT Embeddings	0.74	0.7	0.77	-
BERT Embeddings with PCA	0.77	0.72	0.76	-
Customized Tokenizer	-	-	-	0.89

This table reports the micro F1 scores of classifiers on different embedding methods. The performance is evaluated using testing data set. SVC is the Support Vector Classifier, RF is the Random Forests Classifier (max depth=5), and MLP is the Multilayer Perceptron (64,2,8). Domain-Adapted BERT is the customized classifier from domain-adapted BERT base uncased model. Tf-idf and BERT Embeddings are two text embedding methods. BERT Embeddings with PCA reduces the dimension of BERT Embeddings to 295 (95%) principal components. Customized Tokenizer is the tokenizer adjusted based on our meme stock tweets, specifically for feeding domain-adapted BERT classifier.

Table 1 shows the micro F1 scores of running classifiers on different embedding methods. Compared with the traditional classifiers tested, our customized BERT model has a significant improvement on micro F1 score (0.89) on testing dataset, indicating that the features learned by BERT are highly relevant for the classification task.

Data processing

The next step is applying the fine-tuned model for detecting topics in our meme stock dataset. We use the customized tokenizer and feed the tokenized tweets in the fine-tuned model. At this stage, the BERT base model, which has been meticulously trained first on a general corpus and then domain-adapted on meme stock tweets, is fine-tuned to analyze and classify unseen tweets. In operation, the meme stock tweets data is preprocessed and tokenized in the same manner as the training data to ensure consistency. Each tweet entry is then fed into the fine-tuned BERT model, which uses its deep learned contextual embeddings to generate predictions. These predictions are the model's best guesses for the labels of each text based on the patterns and examples it learned during training. The results are typically probabilities assigned to each possible label, with the highest probability indicating the model's chosen label for each instance.

we processed 2.6 million tweets related to meme stocks, categorizing them into three distinct topics: bullish (53.6%), bearish (12.1%), and neutral (34.2%). Given the varying daily volume of tweets for each meme stock, we average the bullish and bearish topic densities to create a daily measure. These daily aggregates are further segmented by the specific cashtags associated with each meme stock, allowing for targeted analysis of topic trends within individual meme stocks.

Given the constraints of Twitter where users cannot modify the public's reaction to a tweet, we calculate the sum of the number of retweets, likes, and replies as the weight. This method serves as a proxy for gauging user attention and engagement with each tweet. Recognizing that some tweets may receive no likes, retweets, or replies, and thus have a weight of zero, it is important to note that such tweets can still attract views and make impressions, indicating that lack of recorded interactions does not necessarily mean absence of attention. Therefore, we add one to the weight to account for unseen engagements. We calculate the daily weighted topic density using the specified formula:

$$AVG_{\mathrm{ij}} = \frac{(\sum_{\mathrm{i}} (retweets_{\mathrm{ij}} + replies_{\mathrm{ij}} + likes_{\mathrm{ij}} + 1) * T_{\mathrm{ijt}})}{\sum_{\mathrm{i}} (retweets_{\mathrm{ij}} + replies_{\mathrm{ij}} + likes_{\mathrm{ij}} + 1)} \quad (1)$$

where $AVG_{i,j}$ is the weighted average for topic *t* on day *j*. $retweets_{i,j}$, $replies_{i,j}$, and $likes_{i,j}$ represent the number of retweets, replies, likes for the *i*-th tweet on the *j*-th day. $T_{i,j,t}$ is the dummy variable for the topic t for the *i*-th tweet on the *j*-th day. Due to the sparse and wide-ranging distribution of the weights (some tweets have extremely high attention from other users), we report the logarithm of daily aggregated weights in Table 2.

Table 2: Summary statistics of daily level Twitter variables

	P1	P25	Median	Mean	P75	P99	Std
Panel A: Twitter Variables in Trading Hours							
bullish	0.27	0.48	0.52	0.52	0.57	0.70	0.07
bearish	0.05	0.10	0.12	0.12	0.14	0.32	0.04
neutral	0.10	0.30	0.35	0.40	0.40	0.63	0.07
w_bullish	0.31	0.51	0.56	0.56	0.62	0.73	0.07
w_bearish	0.05	0.09	0.11	0.12	0.14	0.30	0.04
w_neutral	0.10	0.26	0.32	0.37	0.37	0.62	0.08
RRL	8.37	9.97	10.65	10.53	11.05	12.87	0.74
n_bullish	104	491	1336	1508	1898	17212	1428
n_bearish	35	103	310	357	456	4270	345
n_neutral	174	411	854	877	1148	6608	568
Panel B: Twitter Variables off Trading Hours							
bullish	0.35	0.46	0.51	0.51	0.55	0.69	0.06
bearish	0.06	0.09	0.10	0.11	0.13	0.29	0.03
neutral	0.18	0.34	0.38	0.38	0.42	0.53	0.06
w_bullish	0.26	0.52	0.57	0.57	0.62	0.77	0.07
w_bearish	0.01	0.07	0.10	0.11	0.14	0.55	0.06
w_neutral	0.11	0.27	0.32	0.32	0.36	0.63	0.08
RRL	9.21	10.30	10.67	10.67	11.01	12.45	0.54
n_bullish	290	884	1301	1484	1764	10371	1004
n_bearish	53	177	260	320	384	1951	221
n_neutral	307	694	951	1031	1243	4954	525

* For aggregated daily RRL, we report its natural logarithm in this table.
This table reports the summary statistics of 457 daily level Twitter variables. Panel A and B present the summary statistics of meme stock tweets in and off the trading hours, respectively. *bullish* is the daily average of bullish tweets, *bearish* is the daily average of bearish tweets, *neutral* is the daily average of neutral tweets, *w_bullish* is the weighted daily average of bullish tweets, *w_bearish* is the weighted daily average of bearish tweets, *w_neutral* is the weight daily average of neutral tweets, *RRL* is the daily log of 1 plus the sum of the number of retweets, replies, and likes, *n_bullish* is the daily number of bullish tweets, *n_bearish* is the daily number of bearish tweets, and *n_neutral* is the daily number of neutral tweets.

In our settings, we divide the dataset into two distinct subsamples based on the timing of the tweets. The first subsample consists of tweets posted during trading hours, specifically from 09:30:00 to 15:59:59. The second subsample includes tweets outside of trading hours, spanning from 16:00:00 to 09:29:59 of the next trading day. We report the summary statistics of aggregated daily variables in Table 2.

We plan to utilize the weighted topic density indices we've developed to examine their correlation with meme stock price movements both during a trading day and across consecutive trading days. Our historical stock data is extracted from Yahoo Finance and covers the post-event window from March 1st, 2021 to May 31st, 2022. For the analysis within a single trading day, we use an indicator variable DAY set to 1 if the closing price is higher than the opening price, and set to 0 otherwise. For the analysis between two trading days, we define another indicator variable ON set to 1 if the opening price on the second trading day exceeds the closing price from the previous day, and 0 otherwise. We next merge the weighted topic density indices for each meme stock to the corresponding price movement indicators, resulting in a total of 317 daily observations.

EMPIRICAL RESULTS

In this section, we initially explore whether the WNTD of highlighted meme stocks on Twitter is associated with intraday and overnight price movements. Subsequently, we examine the predictive capability of the overall weighted topic density for overlooked meme stock price fluctuations.

WNTDs and highlighted meme stocks

Our initial analysis investigates whether the WNTD for each highlighted meme stock correlates with their respective intraday and overnight price movements. We hypothesize that the herding behavior observed among retail investors during the GameStop short squeeze continues to influence the market. Consequently, we anticipate a positive relationship between WNTDs and those stock price changes. In other words, the more the discrepancies between bullish and bearish densities, the more likely the intraday stock return to be positive. Although our dataset includes 13 meme stocks, our focus is on the top six, as they account for 90% of the total tweets analyzed.

To test Hypothesis 1, we employ the following regression specifications:

$$DAY_{HL} \; or \; ON_{HL} = \beta_0 + \beta_1 WNTD_{HL} + \beta_2 SPY + \varepsilon \qquad (2)$$

where DAY_{HL} is an indicator set to 1 if the closing price of any of the six highlighted meme stocks exceeds their opening price within the same trading day, and set to 0 if it does not. ON_{HL} follows a similar logic, marked as 1 if the opening price of a stock on the following day is higher than the closing price from the previous day, again set to 0 otherwise. $WNTD_{HL}$ represents the WNTD associated with each respective meme stock. For example, $WNTD_{GME}$ is the WNTD from

Table 3: Weighted net topic densities (WNTDs) and price movements of highlighted meme stocks

Panel A: Intraday Price Movement Estimation

	DAY					
	GME	*AMC*	*NIO*	*PLTR*	*LCID*	*SOFI*
	(1)	(2)	(3)	(4)	(5)	(6)
$WNTD_{GME}$	1.01					
	(1.50)					
$WNTD_{AMC}$		8.15***				
		(6.74)				
$WNTD_{NIO}$			4.60***			
			(5.84)			
$WNTD_{PLTR}$				4.70***		
				(6.25)		
$WNTD_{LCID}$					1.37***	
					(3.44)	
$WNTD_{SOFI}$						1.64***
						(3.92)
SPY	0.35**	0.68***	0.75***	0.94***	0.65***	0.82***
	(2.54)	(4.39)	(4.90)	(5.99)	(4.44)	(5.26)
Constant	-1.07**	-5.33***	-1.97***	-2.08***	-0.84***	-1.13***
	(-2.43)	(-7.28)	(-7.12)	(-7.80)	(-5.66)	(-6.26)
No.Obs	317	317	317	317	317	317
Pseudo R^2	0.02	0.19	0.15	0.20	0.08	0.11

Panel B: Overnight Price Movement Estimation

	ON					
	GME	*AMC*	*NIO*	*PLTR*	*LCID*	*SOFI*
	(1)	(2)	(3)	(4)	(5)	(6)
$WNTD_{GME}$	-0.70					
	(-0.79)					
$WNTD_{AMC}$		5.98***				
		(5.54)				
$WNTD_{NIO}$			2.54***			
			(3.15)			
$WNTD_{PLTR}$				1.64**		
				(2.23)		
$WNTD_{LCID}$					1.08**	
					(2.35)	
$WNTD_{SOFI}$						0.98**
						(2.23)
SPY	0.80***	0.56***	0.92***	1.31***	0.91***	1.14***
	(5.33)	(3.70)	(6.08)	(8.28)	(6.14)	(7.25)
Constant	-0.13	-3.71***	-1.23***	-1.27***	-0.71***	-0.85***
	(-0.28)	(-5.97)	(-4.97)	(-5.25)	(-4.42)	(-5.13)
No.Obs	317	317	317	317	317	317
Pseudo R^2	0.07	0.12	0.13	0.19	0.10	0.15

This table reports the Probit estimation results from regressing the price movements of highlighted meme stock on weighted net topic densities (WNTDs). Panel A and B report the results for the intraday and overnight movements, respectively. $WNTD_{HL}$ is the difference of weighted *bullish* minus *bearish* for each highlighted meme stock. *DAY* is an indicator set to 1 if the closing price is higher than the opening price in a trading day, and sets to 0 otherwise. *ON* is an indicator set to 1 if the opening price second trading day is higher than previous closing price, and sets to 0 otherwise. Total of six highlighted meme stocks are estimated and reported. Cluster adjusted z-statistics are reported in parentheses. ***, **, * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

tweets attached with cashtag "$GME". Additionally, we incorporate *SPY*, an indicator for the S&P 500 index, as a control variable in our analysis. It is configured similarly to DAY_{HL} and ON_{HL} to help assess broader market trends and provide a benchmark against which to evaluate the movements of our selected meme stocks.

Our findings are presented in Table 3. In Panel A, the coefficients for the WNTD of each meme stock are significantly positive, with the exception of GameStop (GME). A similar trend is observed in Panel B, where all coefficients are significantly positive, except for GME. This pattern indicates that herding behavior is prevalent among highlighted meme stocks in and off trading hours even beyond the initial GME short squeeze event. However, this behavior is not notably significant with GME, the most iconic stock associated with the short squeeze phenomenon. It appears that retail investors may no longer be as motivated to influence GME's market as before, possibly shifting their interests to other meme stocks that present more profitable opportunities.

Overall WNTD and overlooked meme stocks

Next we examine the association between overall WNTD and the movements of the seven overlooked meme stocks within my dataset, which collectively account for just 10% of all tweets analyzed. Given their low volume, when aggregating these tweets to a daily level, there arises a significant challenge: the sparse data may not sufficiently represent the specific topics associated with these stocks. This sparsity complicates the relation derived from tweets tagged with specific cashtags, as these tweets are infrequent. Furthermore, these lesser-known meme stocks often appear in conjunction with cashtags of more popular meme stocks, complicating the isolation of their specific market topic density. Hence, we expect that the price movements of these overlooked meme stocks are positively correlated with the overall WNTD.

To test Hypothesis 2, we employ the following regression specifications:

$$DAY_{OL} \text{ or } ON_{OL} = \beta_0 + \beta_1 WNTD_{overall} + \beta_2 SPY + \varepsilon \tag{3}$$

where DAY_{OL} is an indicator set to 1 if the closing price of any of the seven overlooked meme stocks exceeds their opening price within the same trading day, and set to 0 if it does not. ON_{OL} follows a similar logic, marked as 1 if the opening price of a stock on the following day is higher than the closing price from the previous day, again set to 0 otherwise. $WNTD_{overall}$ represents the WNTD of the entire meme stock tweets. We incorporate *SPY*, an indicator for the S&P 500 index, as a control variable in our analysis following the same configuration of previous test.

We report our results in Table 4. In Panel A, the coefficients for $WNTD_{overall}$ are predominantly positive, except for Nokia Oyj (NOK), suggesting that the overall WNTD offers some predictive capability for overlooked meme stocks during trading hours. Conversely, Panel B displays mixed results, with more than half of the coefficients showing no significant impact. This indicates that the overall WNTD struggles to capture the price movements of overlooked meme stocks overnight. This may be attributed to the sparse discussion volume about these stocks outside of trading hours, which likely reduces the likelihood of investors engaging with this information the following day, thereby diminishing the herding effect on price movements.

Table 4: Overall weighted net topic density (WNTD) and price movements of overlooked meme stocks

Panel A: Intraday Price Movement Estimation							
	DAY						
	BB	*ABNB*	*RIVN*	*MU*	*BBBY*	*NOK*	*GOEV*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$WNTD_{overall}$	4.42***	2.33***	3.20***	1.89***	2.87***	1.18	2.99***
	(4.53)	(2.56)	(2.07)	(1.96)	(3.06)	(1.22)	(3.18)
SPY	0.81***	0.57***	1.03***	1.32***	0.70***	1.25***	0.67***
	(5.33)	(3.93)	(4.44)	(8.53)	(4.72)	(7.98)	(4.51)
Constant	-2.63***	-1.38***	-2.00***	-1.54***	-1.86***	-1.42***	-1.98***
	(-5.62)	(-3.26)	(-3.10)	(-3.38)	(-4.21)	(-3.12)	(-4.43)
No.Obs	317	317	138	317	317	317	317
Pseudo R^2	0.13	0.06	0.15	0.20	0.11	0.16	0.08
Panel B: Overnight Price Movement Estimation							
	ON						
	BB	*ABNB*	*RIVN*	*MU*	*BBBY*	*NOK*	*GOEV*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$WNTD_{overall}$	1.69	1.82*	1.32	2.86***	2.11**	1.11	-0.22
	(1.63)	(1.74)	(0.80)	(2.56)	(2.19)	(1.13)	(-0.21)
SPY	1.36***	1.37***	0.75***	1.79***	0.11	1.03***	1.32***
	(8.28)	(8.60)	(3.38)	(10.41)	(0.78)	(6.65)	(8.30)
Constant	-1.72***	-1.63***	-0.90	-2.29***	-1.09***	-1.13***	-0.67
	(-3.76)	(-3.52)	(-1.40)	(-4.59)	(-2.60)	(-2.63)	(-1.50)
No.Obs	317	317	138	317	317	317	317
Pseudo R^2	0.20	0.21	0.07	0.33	0.01	0.12	0.17

This table reports the Probit estimation results from regressing the price movement of overlooked meme stocks on overall weighted net topic density (WNTD). Panel A and B report the results for the intraday and overnight movements, respectively. $WNTD_{overall}$ is the difference of weighted bullish minus bearish for entire meme stock tweets. *DAY* is an indicator set to 1 if the closing price is higher than the opening price in a trading day, and sets to 0 otherwise. *ON* is an indicator set to 1 if the opening price on the second trading day is higher than the previous closing price, and sets to 0 otherwise. Total of seven overlooked meme stocks are estimated and reported. Cluster adjusted z-statistics are reported in parentheses. ***, **, * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Robustness tests

While our overall weighted net topic density (WNTD) effectively captures the price movements of overlooked meme stocks, it is anticipated to be more predictive for highlighted meme stocks. This expectation is based on the fact that highlighted meme stocks predominantly drive the trends observed in the WNTD. Therefore, we configure a robustness test with similar specification in Hypothesis 2, extending our examination to the price movements of highlighted meme stocks.

$$DAY_{HL}\ or\ ON_{HL} = \beta_0 + \beta_1 WNTD_{overall} + \beta_2 SPY + \varepsilon \quad (4)$$

where DAY_{HL} is an indicator set to 1 if the closing price of any of the six highlighted meme stocks exceeds their opening price within the same trading day, and set to 0 if it does not. ON_{HL} follows a similar logic, marked as 1 if the opening price of a stock on the following day is higher than the closing price from the previous day, again set to 0 otherwise. $WNTD_{overall}$ represents the WNTD of the entire meme stock tweets. We incorporate *SPY*, an indicator for the S&P 500 index, as a control variable in our analysis following the same configuration of previous test.

Table 5: Overall weighted net topic density (WNTD) and price movements of highlighted meme stocks

Panel A: Intraday Price Movement Estimation						
	DAY					
	GME	*AMC*	*NIO*	*PLTR*	*LCID*	*SOFI*
	(1)	(2)	(3)	(4)	(5)	(6)
$WNTD_{overall}$	1.01	5.99***	4.58***	5.58**	2.36***	3.22***
	(1.49)	(5.79)	(4.74)	(5.46)	(2.57)	(3.40)
SPY	1.32***	0.70***	0.71***	0.93***	0.60***	0.71***
	(8.52)	(4.62)	(4.70)	(6.04)	(4.12)	(4.79)
Constant	-0.87**	-3.27***	-2.63***	-3.21***	-1.57***	-2.07***
	(-1.96)	(-6.60)	(-5.72)	(-6.51)	(-3.63)	(-4.59)
No.Obs	317	317	317	317	317	317
Pseudo R^2	0.18	0.15	0.12	0.18	0.06	0.09
Panel B: Overnight Price Movement Estimation						
	ON					
	GME	*AMC*	*NIO*	*PLTR*	*LCID*	*SOFI*
	(1)	(2)	(3)	(4)	(5)	(6)
$WNTD_{overall}$	0.97	3.52***	1.88*	3.19***	3.05***	2.72***
	(1.03)	(3.51)	(1.90)	(3.02)	(3.04)	(2.68)
SPY	0.75***	0.51***	0.94***	1.28***	0.84***	1.07***
	(4.95)	(3.40)	(6.16)	(7.99)	(5.52)	(6.92)
Constant	-0.91**	-1.82***	-1.34***	-2.17***	-1.73***	-1.77
	(-2.17)	(-4.17)	(-3.11)	(-4.62)	(-3.93)	(-3.95)
No.Obs	317	317	317	317	317	317
Pseudo R^2	0.07	0.07	0.11	0.20	0.11	0.15

This table reports the Probit estimation results from regressing the price movements of highlighted meme stocks on overall weighted net topic density (WNTD). Panel A and B report the results for the intraday and overnight movements, respectively. $WNTD_{overall}$ is the difference of weighted *bullish* minus *bearish* for entire meme stock tweets. *DAY* is an indicator sets to 1 if the closing price is higher than the opening price in a trading day, and sets to 0 otherwise. *ON* is an indicator sets to 1 if the opening price second trading day is higher than previous closing price, and sets to 0 otherwise. Total of six highlighted meme stocks are estimated and reported. Cluster adjusted *z*-statistics are reported in parentheses. ***, **, * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Our findings are detailed in Table 5. In line with our expectations, the overall weighted net topic density (WNTD) shows a positive association with five of the six highlighted meme stocks, with significance levels generally surpassing those of the overlooked meme stocks. Notably, GameStop (GME) does not exhibit significant correlations with either intraday or overnight price movements, reinforcing our conclusion that retail investors have indeed shifted their focus away from GME as a target for short squeezing.

Additionally, we conduct an additional robustness test on a group of five stocks from S&P 500 firms that are not considered meme stocks due to their nature and minimal discussion on social media. These non-meme stocks belong to industries similar to those of our meme stocks. A key characteristic shared by these five non-meme stocks is that tweets containing their cashtags comprise less than 0.01% of the total tweets in our dataset. The stocks included in this test are Best Buy Co Inc (BBY), Motorola Solutions Inc (MSI), JPMorgan Chase & Co (JPM), Meta Platforms Inc (META), and Expedia Group Inc (EXPE). We expect there is no significant relation between overall WNTD and the price movements of those non-meme stocks. Our specification is similar to the previous test.

$$DAY_{SP} \ or \ ON_{SP} = \beta_0 + \beta_1 WNTD_{overall} + \beta_2 SPY + \varepsilon \tag{5}$$

where DAY_{SP} is an indicator set to 1 if the closing price of any of the five non-meme stocks exceeds their opening price within the same trading day, and set to 0 if it does not. ON_{SP} follows a similar logic, marked as 1 if the opening price of a stock on the following day is higher than the closing price from the previous day, again set to 0 otherwise. $WNTD_{overall}$ represents the WNTD of the entire meme stock tweets. We incorporate *SPY*, an indicator for the S&P 500 index, as a control variable in our analysis following the same configuration of previous test.

Table 6 presents the outcomes of our second robustness test. None of the five non-meme stocks demonstrated significant results when regressing their price movements against the WNTD, suggesting that non-meme stocks, which are seldom discussed on Twitter, are not influenced by discussions surrounding meme stocks. This finding supports the notion that the overall WNTD is particularly effective and relevant only for meme stocks.

Table 6: Overall weighted net topic density (WNTD) and price movements of non-meme stocks

Panel A: Intraday Price Movement Estimation					
	DAY				
	BBY	*MSI*	*JPM*	*META*	*EXPE*
	(1)	(2)	(3)	(4)	(5)
$WNTD_{overall}$	-0.03	1.00	0.42	1.11	0.74
	(-0.03)	(1.04)	(0.45)	(1.13)	(0.82)
SPY	0.86***	1.28***	1.01***	1.44***	0.62**
	(5.80)	(8.36)	(6.70)	(9.17)	(4.23)
Constant	-0.54	-1.05**	-0.8*	-1.29***	-0.75*
	(-1.28)	(-2.35)	(-1.88)	(-2.82)	(-1.81)
No.Obs	317	317	317	317	317
Pseudo R^2	0.08	0.18	0.11	0.22	0.05
Panel B: Overnight Price Movement Estimation					
	ON				
	BBY	*MSI*	*JPM*	*META*	*EXPE*
	(1)	(2)	(3)	(4)	(5)
$WNTD_{overall}$	1.63	1.65	0.37	1.24	1.60
	(1.58)	(1.59)	(0.37)	(1.14)	(1.55)
SPY	1.51***	1.35***	1.08***	1.73***	1.26***
	(9.31)	(8.50)	(7.02)	(10.27)	(8.01)
Constant	-1.44***	-1.20**	-0.66	-1.47***	-1.40***
	(-3.21)	(-2.65)	(-1.52)	(-3.09)	(-3.17)
No.Obs	317	317	317	317	317
Pseudo R^2	0.24	0.21	0.13	0.30	0.18

This table reports the Probit estimation results from regressing the price movements of non-meme stocks on overall weighted net topic density (WNTD). Panel A and B report the results for the intraday and overnight movements, respectively. $WNTD_{overall}$ is the difference of weighted *bullish* minus *bearish* for entire meme stock tweets. *DAY* is an indicator sets to 1 if the closing price is higher than the opening price in a trading day, and sets to 0 otherwise. *ON* is an indicator sets to 1 if the opening price second trading day is higher than previous closing price, and sets to 0 otherwise. Total of five non-meme stocks are estimated and reported. Cluster adjusted *z*-statistics are reported in parentheses. ***, **, * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

CONCLUSION

We examine the pre-defined topics related to 2.6 million tweets of 13 meme stocks from March 1st, 2021 to May 31st, 2022, a period beyond the GME short squeeze. Our results can be concluded in two main points. First, this research effectively leverages a customized BERT model to classify 2.6 million tweets related to meme stocks, highlighting the model's robust capability in understanding and processing large volumes of social media data. By tailoring the BERT model to specifically address the nuances of social media language, we are able to extract meaningful insights into the topics dominating discussions about meme stocks.

Second, this research further sheds light on the herding behavior exhibited by meme stock investors beyond GME short squeeze, which is a significant factor influencing market momentum. Herding behavior, often fueled by the echo chamber effect of social media platforms, contributes to rapid and substantial swings in stock prices. Our findings indicate that when a meme stock becomes a focal point of discussion on social media, particularly Twitter, investors tend to move collectively, either buying or selling en masse. This collective action is primarily driven by the topics expressed in these discussions.

By constructing the weighted net topic density (WNTD) of meme stock tweets, we document the effectiveness of how WNTD can be utilized in capturing the price movements of meme stocks, particularly those that are frequently highlighted on social media. For five out of the six highlighted meme stocks, there is a positive association with WNTD, suggesting that social media discussions have a significant impact on their market behaviors. While the WNTD can capture some movements of overlooked meme stocks, the predictive power is less pronounced compared to that for highlighted meme stocks. This difference underscores the influence of social media visibility and discussion volume on the predictive accuracy of the WNTD. Notably, GameStop (GME), despite its status as a meme stock, does not show significant correlations with WNTD in our analysis. This suggests a possible shift in retail investors' focus away from GME as a primary target for trading based on social media discussions.

Overall, this analysis not only reinforces the link between social media discussions and stock price movements but also highlights the critical role of herding behavior in driving the market momentum associated with meme stocks. This understanding is crucial for both investors and regulators as they navigate the complexities introduced by the digital age's influence on financial markets.

APPENDIX

Example of labelled tweets by topics

Topic	Tweets Related to Meme Stock Market
Bullish	Let’s buy into these and become partial owners and push for an #AMCTheatres and #GMEtothemoon merger. $AMC + $GME = Ready Player One experience.
Bearish	To all my friends holding $GME still, please don’t hold to make a point if you can’t afford to lose the money you put in. This bubble will pop. Get out if you can’t afford the loss, hold strong if you can!
Neutral	I have an active pro membership yet haven’t gotten the new $5 reward and also didn’t receive the 20% off a preowned game birthday coupon. Are these no longer given out?
Topic	**Tweets Related to General Stock Market**
Bullish	$ZYXI - Zynex reports 126% order growth and raises Q1 and FY2020 revenue outlook.
Bearish	Caterpillar reported lower sales across all three primary segments in the last quarter, adding to the industrial gl...
Neutral	(Free to read) Over 60% of Americans say that stock market performance has little or no impact on their personal life.

REFERENCES

Aloosh, A., Choi, H.-E., & Ouzan, S. (2021). Meme stocks and herd behavior. *Available at SSRN 3909945*.

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting, 36*(2), 334–357.

Awan, M. J., Rahim, M. S. M., Nobanee, H., Munawar, A., Yasin, A., & Zain, A. M. (2021). Social media and stock market prediction: A big data approach. *Computers, Materials and Continua, 67*(2), 2569–2583.

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF-based schemes. *Knowledge-Based Systems, 163*, 1–13.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Egger, R., & Yu, J. (2021). Identifying hidden semantic structures in Instagram data: A topic modelling comparison. *Tourism Review, 77*(4), 1234–1246.

Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2vec, and Bertopic to demystify Twitter posts. *Frontiers in Sociology, 7*, 886498.

Fekrazad, A., Harun, S. M., & Sardar, N. (2022). Social media sentiment and the stock market. *Journal of Economics and Finance, 46*(2), 397–419.

Gider, J., & Westheide, C. (2016). Relative idiosyncratic volatility and the timing of corporate insider trading. *Journal of Corporate Finance, 39*, 312–334.

Guan, C., Liu, W., & Cheng, J. Y.-C. (2022). Using social media to predict the stock market crash and rebound amid the pandemic: The digital 'haves' and 'have-mores'. *Annals of Data Science, 9*(1), 5–31.

Hossain, M. M., Mammadov, B., & Vakilzadeh, H. (2022). Wisdom of the crowd and stock price crash risk: Evidence from social media. *Review of Quantitative Finance and Accounting, 58*(2), 709–742.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science, 64*(6), 2833–2855.

Jaradat, S., & Matskin, M. (2019). On dynamic topic models for mining social media. *Emerging*

Research Challenges and Opportunities in Computational Social Network Analysis and Mining, 209–230.

Kabbani, T., & Usta, F. E. (2022). Predicting the stock trend using news sentiment analysis and technical indicators in spark. *arXiv preprint arXiv:2201.12283*.

Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications, 9*(7), 22.

Kim, K., Lee, S.-Y. T., & Kauffman, R. J. (2023). Social informedness and investor sentiment in the GameStop short squeeze. *Electronic Markets, 33*(1), 23.

Klein, T. (2022). A note on GameStop, short squeezes, and autodidactic herding: An evolution in financial literacy? *Finance Research Letters, 46*, 102229.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems, 69*, 14–23.

Li, X., Ma, J., Wang, S., & Zhang, X. (2015). How does Google search affect trade positions and crude oil prices? *Economic Modelling, 49*, 162–171.

Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 205–208). IEEE.

Nguyen, T. H., & Shirai, K. (2015). Topic modeling-based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1354–1364).

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS One, 10*(9), e0138441.

Umar, Z., Gubareva, M., Yousaf, I., & Ali, S. (2021). A tale of company fundamentals vs sentiment-driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance, 30*, 100501.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

Vasileiou, E., Bartzou, E., & Tzanakis, P. (2021). Explaining GameStop short squeeze using intraday data and Google searches. *Available at SSRN 3805630*.

Wang, C., & Luo, B. (2021). Predicting $GME stock price movement using sentiment from Reddit r/WallStreetBets. In *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing* (pp. 22–30).

Yahya, M. A., & Chiu, T. N. (2021). The meme stock paradox. *Corporate and Business Law Journal, 3*(1).

Zhou, R. T., & Lai, R. N. (2009). Herding and information-based trading. *Journal of Empirical Finance, 16*(3), 388–393.

Zheng, X., Tian, H., Wan, Z., Wang, X., Zeng, D. D., & Wang, F.-Y. (2021). Game starts at GameStop: Characterizing the collective behaviors and social dynamics in the short squeeze episode. *IEEE Transactions on Computational Social Systems, 9*(1), 45–58.

DECISION SCIENCES INSTITUTE

Navigating Sustainability Controversies: Implications for SSCM and Firm Performance in Turbulent Environments

Amir Naderpour
Eastern Kentucky University
amir.naderpour@eku.edu

Gregory Frazier
University of Texas at Arlington
frazier@uta.edu

ABSTRACT

Stakeholders today expect companies to take responsibility for sustainability throughout their supply chains and respond quickly to any issues. This study examines how sustainable supply chain management (SSCM) controversies impact sustainability practices and firm performance. Using stakeholder theory, we explore direct and indirect effects, while considering industry turbulence influence in this relationship. Our findings show that SSCM controversies do not directly harm firm performance but encourage better governance, social, and environmental practices. However, these benefits weaken in unstable industries. This research underscores the need for adaptive sustainability strategies that enable companies to maintain robust performance in dynamic business environments.

KEYWORDS: Sustainable supply chain management (SSCM); SSCM controversies; Firm performance; Environmental turbulence.

INTRODUCTION

In today's interconnected global business landscape, stakeholders expect companies not only to be accountable for their own operations but also for the actions of their suppliers and partners across the entire supply network. This heightened scrutiny emphasizes the importance of sustainable supply chain management (SSCM), which integrates environmental, social, and governance (ESG) considerations into supply chain operations. Consumers, investors, and regulators demand that firms address any sustainability-related incidents in their supply chains, thereby holding them accountable for broader operational impacts.

Despite increasing emphasis on sustainability, the relationship between SSCM and firm performance remains complex and contested. Previous studies have yielded contradictory findings—some report that sustainability enhances performance through improved reputation, customer loyalty, and operational efficiencies, while others caution that such practices may increase costs and operational risks. Aouadi and Marsat (2018) argue that these divergent outcomes stem from an incomplete understanding of the mechanisms underlying the sustainability–performance relationship. Investigating the indirect links between various sustainability dimensions and firm performance may resolve these contradictions (Surroca et al., 2010).

A critical dimension requiring further exploration is the role of SSCM controversies—incidents that compromise a firm's sustainability credentials and stakeholder trust. Such controversies may arise from environmental violations, labor abuses, or governance failures within the supply chain. According to stakeholder theory, firms that adhere to social norms and values secure long-term success by maintaining legitimacy and robust stakeholder relationships (Jiao, 2010; Aouadi & Marsat, 2018).

Implementing effective SSCM practices can bolster buyer–supplier relationships, ensuring financial prosperity and long-term viability (Parmigiani et al., 2011; Reuter et al., 2010). Conversely, when controversies arise, a firm's reputation and legitimacy may suffer, potentially leading to adverse financial outcomes (Du et al., 2011; Palazzo & Scherer, 2006). Proactive sustainability measures, however, can help firms mitigate these negative effects by building resilient reputations (Birindelli et al., 2015; Donaldson & Preston, 1995; Franceschelli et al., 2018).

This study addresses the gap by examining how SSCM controversies affect the implementation of sustainability practices—and, in turn, firm performance—while also investigating the moderating role of industry turbulence. Our research questions are:

Do SSCM controversies have any direct or indirect impact on firm performance? And how does industry turbulence modify this effect?

By answering these questions, our study contributes to a deeper understanding of SSCM and offers actionable insights for firms navigating sustainability challenges in varied industry contexts.

LITERATURE REVIEW

Sustainable Supply Chain Management (SSCM)

Sustainable supply chains and eco-friendly practices are increasingly seen as the future of supply chain management (Seuring, 2011). While many companies outsource various processes to reduce costs, they often overlook monitoring the environmental and social impacts of their external partners (Ramsay, 2001). This oversight has led to numerous incidents within the supply chains of major corporations. Scholars and practitioners alike have turned their attention to SSCM (Pagell & Wu, 2009; Svensson et al., 2018). Seuring and Müller (2008) define SSCM as managing the flow of materials, information, and capital while collaborating with multiple supply chain partners to address the three pillars of sustainability demanded by stakeholders.

However, many firms either neglect the importance of SSCM or find the required intercontinental leadership too costly (Damberg et al., 2022), leading to ongoing reports of unethical supplier practices, particularly in developing countries (Bradshaw et al., 2021). Researchers have explored ways to enhance SSCM through supplier engagement (Chen et al., 2017) and have highlighted the influence of external pressures—including controversies and negative media coverage—on driving better sustainability practices (Gualandris et al., 2015). SSCM can yield both environmental/social benefits and financial gains for supply chain participants by fostering collaboration and establishing new sustainability protocols (Golicic & Smith, 2013; Vachon & Klassen, 2006). Evaluating the impact of SSCM on the performance of focal firms is essential to demonstrate its economic viability (Golicic & Smith, 2013; Pagell & Wu, 2009; Svensson et al., 2018).

Supply Chain Incidents and Controversies

Globalization and outsourcing have increased the risk of social and environmental controversies, such as environmental law violations and labor abuses (Qiang, 2015). High-profile cases, including issues with Nike suppliers, have prompted public scrutiny of focal firms' roles in managing their supply chains (Locke et al., 2007; Anner et al., 2013). SSCM controversies—stemming from unethical treatment of workers, environmental degradation, and governance failures—have increasingly attracted stakeholder and media attention (Tamayo-Torres et al., 2019). Recent examples include lawsuits against fashion brands like Lululemon Athletica for alleged false claims about their environmental impact reduction plans (WSJ, July 2024). Such negative news not only reduces information asymmetry between firms and stakeholders (Dyck et al., 2008) but also provides opportunities for firms to improve their practices (Hartmann, 2020).

Damberg et al. (2022) show that adverse media coverage can spur firms to adopt better SSCM practices. To measure the efficacy of sustainability practices, we employ Tobin's Q—a forward-looking, market-based measure that captures long-term firm performance (Flammer, 2013; Servaes & Tamayo, 2013).

Stakeholder Theory

Stakeholder can affect the firm performance (Parmar et al., 2010) and has different needs, goals, expectations, and responsibilities (Clarkson, 1995). The externalities made by organizations and the conflicts with stakeholders will result in more risk in the operations, a bad reputation, a negative effect on the current and future sales of the firm, and higher costs (Klassen & Vereecke, 2012). The response of the organizations to these controversies includes actions like implementing new strategies and systems to improve their sustainability performance, and it is very important to have a good reaction to these controversies (Aouadi & Marsat, 2018; Klassen & Vereecke, 2012). Firms know the importance of their brand and reputation and not only try not to encounter any controversy with stakeholders but also try to have prompt reactive strategies in case of any conflicting incident with a level of potential to become controversial to improve their stigmatized reputation (Fombrun et al., 2000; Klassen & Vereecke, 2012).

Contingency View

Contingency theory emphasizes that management practices must be tailored to specific organizational contexts (Lawrence & Lorsch, 1967; Thompson, 1967). The theory suggests that there is no one "best" way to manage resources or organize a company that will work in all situations. Instead, the most effective management practices will depend on a variety of factors, such as the nature of the work being done, the size of the organization, the culture of the organization, and the external environment in which the organization operates (Thompson, 1967).

One key contingency variable is industry environmental turbulence. In turbulent environments, organizations face rapid changes and unpredictability, which can complicate the adoption of best practices and dampen the positive effects of sustainability efforts (Xue et al., 2012a; Miller, 1987). Firms in stable industries benefit from strong customer relationships and reputations, while those in turbulent sectors may need to prioritize short-term operational demands over long-term sustainability goals (Bansal, 2005; Pagell & Wu, 2009).

According to Pagell & Wu (2009), the effectiveness of SSCM practices is contingent on factors such as the firm's industry sector, its competitive position, and its internal organizational capabilities.

The mentioned studies show us the importance of the contingency approach in SSCM, which involves considering the specific contextual factors that influence their effectiveness in different organizational contexts.

Sustainability and Firm Performance

The literature presents mixed findings on the sustainability–performance relationship, with some studies showing a neutral (Garcia-Castro et al., 2010; Mcwilliams & Siegel, 2001), positive (Aouadi & Marsat, 2018), or even negative link (López et al., 2007; Marsat & Williams, 2012). To reconcile these differences, researchers have examined mediating and moderating factors (Surroca et al., 2010; Servaes & Tamayo, 2013). Tamayo-Torres et al. (2019) found that while SSCM controversies can spur improvements in sustainability practices (with varied effects across ESG dimensions), their direct impact on performance remains unclear. Pagell & Wu (2009) and Vachon & Klassen (2006) further illustrate that effective SSCM practices can enhance firm performance, particularly when moderated by factors such as R&D investment. These reviewed studies show that the relationship between SSCM controversies, sustainability practices, and firm performance is complex and multifaceted.

THEORETICAL DEVELOPMENT

Implementing sustainability through the supply chain not only yields economic benefits that safeguard long-term survival but also strengthens relationships with strategic suppliers (Parmigiani et al., 2011). Although previous studies present contradictory findings regarding the impact of sustainability controversies on firm performance, there is a recognized need to clarify the mechanisms by which such controversies might influence performance. Accordingly, this research posits hypotheses grounded in fundamental theories linking supply chain controversies and firm performance.

Organizations can attain long-term success by observing social norms and values and undertaking actions that confer a sense of legitimacy (Aouadi & Marsat, 2018). When organizations encounter various controversies, stakeholders become aware of these sustainability issues, potentially harming the organization’s reputation (Du et al., 2011). Consequently, we expect a negative effect of sustainability controversies on firm performance (Fombrun & Shanley, 1990), as empirical evidence indicates that socially irresponsible actions can drive down stock prices (Johnson, 2003). Klassen and McLaughlin (1996) similarly observed that “bad news” in sustainability is linked to a decline in market value. A firm with multiple stakeholder controversies may thus experience waning trust, a drop in sales, difficulties securing crucial resources from suppliers, and even governmental sanctions.

Stakeholder theory suggests that adhering to social norms and values is critical for legitimacy and long-term viability. Sustainability controversies challenge a firm’s legitimacy, with stakeholders responding negatively when controversies arise (Du et al., 2011). This adverse perception can depress sales, erode trust, and lead to other negative consequences. Sustainability controversies also trigger adverse investor reactions, reducing a firm’s market value. Evidence shows that organizations facing serious sustainability controversies frequently lose market share and see financial performance deteriorate (Gong et al., 2019; Klassen & McLaughlin, 1996). To avert these outcomes, firms are advised to minimize controversies and adopt sustainability practices throughout their supply chain to protect reputation and ensure long-term success. Based on this discussion, we propose:

H1: *SSCM controversies directly and negatively impact a firm’s financial performance.*

Parmigiani et al. (2011) argue that stakeholders hold the focal firm accountable for managing its supply chain responsibly. Implementing sustainability practices helps firms build reputation and trust, resulting in competitive advantage (Birindelli et al., 2015). Nevertheless, negative behaviors or unexpected incidents within the supply chain can generate stakeholder controversies. In such circumstances, firms may adopt a reactive strategy, deploying sustainability measures as tools to safeguard reputation and trust (Livesey & Kearins, 2002). Conversely, a history of proactive sustainability can moderate the reputational damage when controversies arise, since consumers may be more lenient (Klein & Dawar, 2004).
In recent years, stakeholders have become more vigilant about all facets of sustainability. Incidents involving high-profile brands in industries such as apparel and fashion—often under close scrutiny by consumers, journalists, lawyers, and regulators—can rapidly escalate to lawsuits or public outcry. For instance, Lululemon Athletica faced legal action alleging misleading advertising about its environmental commitments (WSJ, July 2024).
Stakeholders can address environmental controversies either directly or indirectly through allies (Frooman, 1999). NGOs and governments may also intervene to ensure firms comply with environmental standards (Ageron et al., 2012). Suppliers often face growing demands to produce ISO 14001-certified goods (Tamayo-Torres et al., 2019) or to establish environmental management systems (Mitra & Datta, 2014). By demonstrating responsible practices, the firm can reinforce relationships with customers, employees, suppliers, and investors (Parmar et al., 2010). Nonetheless, guiding supply chain members and stakeholders toward shared sustainability and profit goals requires targeted initiatives (Hall et al., 2012). This process is especially challenging when sustainability-related incidents—such as corruption or child labor—gain significant media attention. In global supply chains that span multiple continents, including developing regions where sustainability considerations may be lower, it becomes critical to assess how well the focal firm's leadership drives SSCM objectives. Thus, we propose:

H2: *Supply chain incidents and controversies positively influence sustainability practices.*

Global supply chains that operate across continents often encounter environmental turbulence—instability arising from sudden regulatory shifts, changing customer preferences, and rapid technological developments. Under such conditions, firms find it more difficult to implement sustainability practices, and controversies may be more frequent due to greater uncertainty. Costly adaptations to evolving regulations and markets also mean firms may have fewer resources to invest in sustainability initiatives, potentially weakening the link between SSCM controversies and sustainability practices. Furthermore, stakeholder attitudes differ in stable versus turbulent environments. In stable contexts, a proven record of sustainability can bolster a firm's reputation, whereas in highly volatile settings, stakeholders tend to react more swiftly and critically to controversies.
Firms operating in turbulent markets often face difficult trade-offs between short-term profitability and long-term sustainability goals. They may need to invest heavily in reputation management rather than in preventative or proactive sustainability solutions. Indeed, collaboration and maintaining sustainability integrity become more difficult in highly dynamic industries, thus slowing progress toward more sustainable supply chains and improved ESG performance (Silvestre, 2015). Hence, we propose:

H3: *The relationship between SSCM controversies and sustainability practices is moderated by environmental turbulence. Specifically, this relationship is weaker for firms operating in more turbulent industry environments.*

Heightened consumer and stakeholder expectations regarding sustainability have led to stricter demands not only for eco-friendly products but also for evidence of sustainable processes. Stakeholders now expect rapid and transparent responses to any environmental incidents in a company's supply chain. For example, the Swiss athletic apparel firm On introduced a circular subscription plan, On Cyclon, whereby customers return worn-out shoes in exchange for new ones under a promise to recycle the returned items. However, recycling did not commence by mid-2024 as planned, sparking criticism; the company subsequently announced an August 2024 start date (WSJ, July 2024). This trend illustrates that consumers are increasingly willing to pay premium prices for sustainable offerings (Altmann, 2015).
By implementing sustainability practices, firms not only comply with regulations but also enhance stakeholders' trust. Reducing waste and energy consumption can decrease production costs (Sarkis et al., 2010), and close collaboration with suppliers can bring long-term mutual benefits (Pagell & Wu, 2009). Although the empirical literature does not always yield consistent results on social sustainability and performance (Margolis & Walsh, 2003), many studies indicate that investments in social practices can improve financial outcomes (Tamayo-Torres et al., 2019). Like environmentally focused initiatives, social sustainability measures appeal to consumers who are willing to pay more for products with lower social impacts (Graff Zivin & Small, 2015). Taking social and environmental issues into account can unlock competitive advantages and access to innovative technologies (Porter & Van Der Linde, 1995). Moreover, firms with robust sustainability practices may mitigate reputational risks when tackling negative incidents in their supply chain.
By prioritizing sustainability, organizations differentiate themselves and cultivate a strong brand reputation (Eccles et al., 2014), leading to higher customer loyalty, greater sales, and improved financial performance. In an era of rapid information sharing, organizations can manage potential reputational damage by demonstrating genuine commitment to sustainability through their supply chains (Sarkis et al., 2010). Accordingly, we posit:

H4: *Implementing sustainability practices has a positive impact on firm performance.*
In the following section, we discuss the data sources, dataset, measures, methods, and models employed to test these hypotheses.

METHODS

Data and Sample

Our study's sample is derived by merging sustainability data from Sustainalytics and LSEG (previously Refinitiv) with financial data from COMPUSTAT. Specifically, we integrate SSCM controversies from Sustainalytics and sustainability ratings from LSEG for the years 2009 to 2019, augmenting these data with COMPUSTAT information from 2005 to 2019. The additional years capture rolling five-year windows needed for calculating industry sales volatility in order to measure environmental turbulence.
Following Servaes and Tamayo (2013), we restrict our sample to nonfinancial firms to avoid the distinct structural characteristics of financial institutions. We also exclude firms located outside the North American region and private firms, primarily due to lack of robust stock market data.

After removing observations with incomplete data in any of the required variables across all databases, the final sample consists of 610 firms and 5,625 firm-year observations.

Measures

To evaluate SSCM (Sustainable Supply Chain Management) controversies, we employ three indicators from Sustainalytics (as outlined in Tamayo-Torres et al., 2019):
Social Incidents – Controversies involving stakeholders on social issues in the supply chain (e.g., strikes at supplier facilities, inadequate labor standards, poor safety standards, or child labor).
Operational Incidents – Controversies related to operational, product, and service issues (e.g., poor waste management, air or water pollution linked to production activities).
Environmental Incidents – Controversies stemming from environmental problems (e.g., excessive emissions, deforestation).
These incidents are identified through publicly available news stories on problematic sustainability practices, collected from sources such as news agencies, NGOs, Human Rights Watch, trade unions, and other associations (Aouadi & Marsat, 2018). Following Aouadi and Marsat (2018), we use the raw values from Sustainalytics for each indicator and create a dummy variable (1 = presence of any SSCM controversies, 0 = no controversies).
We obtain environmental, social, and governance (ESG) sustainability indices from LSEG (Refinitiv).
Following Keats and Hitt (1988), we measure environmental turbulence through industry sales volatility. For each four-digit SIC industry, we calculate total industry sales, then regress the natural logarithm of these sales on a time index (year) using a rolling five-year period [t, t-4]. The antilog of the standard error of the regression coefficient serves as the measure of volatility. Higher standard errors signify greater difficulty in predicting sales growth rates for a given industry, reflecting higher turbulence and dynamism.
We measure firm size using the natural logarithm of the number of employees, expressed in millions.
Firm Performance is measured using Tobin's Q ratio (Aouadi & Marsat, 2018; Bharadwaj et al., 1999; Servaes & Tamayo, 2013). Based on (Chung & Pruitt, 1994) measure of Tobin's Q, Q = (market value of equity + book value of inventories + liquidating value of preferred stock + long-term debt + net short-term debt) / total assets. Net short-term debt is measured as the difference between current assets and current liabilities. Table 1 shows the significance and magnitude of the correlation between the variables of the study.

Table 1: Correlation Matrix

The numbers in the table are the Pearson correlation coefficients, and the asterisks show the significance of the correlation. Resource/Investment intensity measures are calculated as ratios of sales for each firm-year observation.

	1	2	3	4	5	6

1	Controversies	1					
2	Governance Sustainability	0.235***	1				
3	Social Sustainability	0.274***	0.433***	1			
4	Environmental Sustainability	0.349***	0.468***	0.752***	1		
5	Size	0.215***	0.166***	0.381***	0.350***	1	
6	Environmental Turbulence	0.029*	-0.040**	-0.133***	-0.089***	-0.148***	1

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

Model

To evaluate how SSCM controversies affect sustainability and how environmental turbulence moderates this relationship, we apply OLS regression to each dimension of sustainability. Consistent with prior research (Aouadi & Marsat, 2018; Soytas et al., 2019), firm size serves as a control variable. We also include year-fixed effects and firm and industry fixed effects by introducing dummy variables for each firm, industry (classified by initial SIC), and year. Following Soytas et al. (2019), we lag the predictors to address endogeneity concerns that can otherwise yield inconsistent results in sustainability research (Garcia-Castro et al., 2010; Luo & Bhattacharya, 2006). More specifically, the controversies variable is lagged to mitigate reverse causality. In this initial analysis, we examine the direct effect of SSCM controversies on sustainability while incorporating environmental turbulence as a moderator. The relationship is represented in Equation (1), with subscripts , representing firm, industry, and year, respectively.

(1)

To explore the indirect path from SSCM controversies to firm financial performance, we next consider the direct effect of sustainability practices on performance. We again employ OLS regression, treating each sustainability dimension as an independent variable. As before, we control for firm size (Aouadi & Marsat, 2018; Soytas et al., 2019) and include year-fixed effects and firm and industry fixed effects by adding dummy variables for firm, industry, and year., with subscripts , representing firm, industry, and year, respectively.

(2)

Table 2: Analysis Results

	Governance	Social	Environmental	Performance
SSCM Controversies	1.334** (1.007)	1.387* (.620)	.444** (.169)	-1.398 (1.218)
Environmental Turbulence	-3.564*** (.709)	-4.613*** (.537)	-3.791*** (.474)	-2.564 *** (.719)
Controversies * EnvTurb	1.650 (1.065)	-1.229 (.806)	-.314** (.711)	
Governance				.892*** (.336)
Social				.943** (.435)
Environmental				1.549*** (.339)
Size	0.022*** (0.001)	0.049** (0.022)	0.096*** (0.017)	0.002* (0.001)
Firm fixed effect	*	*	*	*
Industry fixed effect	*	*	*	*
Year fixed effect	*	*	*	*
Observations	2,688	2,688	2,688	5,000
	0.114	0.150	0.140	0.119

**p < 0.05; **p < 0.01; ***p < 0.001, EnvTurb = Environmental Turbulence*

RESULTS

As depicted in Table 2, there is no significant direct negative effect between SSCM controversies and firm performance, rejecting H1.

Our results show a significant positive effect of SSCM controversies on governance sustainability practices ($\beta_1=1.334$, $p<0.01$). Our results also show a positive and significant effect of SSCM controversies on future social sustainability practices ($\beta_1=1.387$, $p<0.05$). We find a positive and significant effect of SSCM controversies on future environmental sustainability practices ($\beta_1=0.444$, $p<0.01$). This result shows that SSCM controversies have positive effects on the movement towards better governance, social, and environmental sustainability practices, Thus, H2 is supported for all dimensions of sustainability practices.

The direct effect of environmental turbulence on environmental sustainability practices is negative and significant ($\beta_2=-3.791$, $p<0.001$). The moderating effect of environmental turbulence on the relationship between SSCM controversies and environmental sustainability practices is significant and negative ($\beta_3=-0.314$, $p<0.01$). This result shows that although SSCM controversies have positive effects on the movement towards better environmental sustainability practices, this positive effect would be lower in a turbulent environment or could even flip as the environmental dynamism increases. Thus, H3 is supported for environmental sustainability practices.

H4 is supported for all three dimensions of sustainability. A positive and significant effect of governance sustainability practices on performance ($\beta_1=0.892$, $p<0.001$) is shown in the results. Our results show a positive and significant effect of social sustainability practices on performance ($\beta\_1=0.943$, $p<0.01$). The results also show a positive and significant effect of environmental sustainability practices on performance ($\beta\_1=1.549$, $p<0.001$).

DISCUSSION

Our results show that SSCM controversies have a positive effect on governance, social, and environmental sustainability practices, indicating that firms are more likely to move towards and engage in sustainability after being faced with SSCM controversies. However, environmental turbulence has a negative effect on sustainability practices, implying that firms that work in turbulent industry environments may have difficulties implementing sustainability practices. The positive effect of SSCM controversies on environmental sustainability practices is reduced in turbulent environments.

The results of our study show that governance, social, and environmental sustainability practices have a positive effect on financial performance, indicating that firms that engage in sustainable practices are more likely to have better financial performance.

The findings of the present study have important implications for both researchers and practitioners. In line with the literature in the field, our study highlights the importance of SSCM controversies in increasing sustainability practices among firms. Policymakers and practitioners can use the controversies related to Supply Chain Sustainability Management (SSCM) as a tool to encourage firms to engage in sustainability practices. SSCM controversies may include issues such as social and environmental concerns related to the supply chain, ethical concerns about the sourcing of raw materials, labor practices, and human rights abuses.

By highlighting these controversies including social and environmental concerns related to the supply chain, ethical concerns about the sourcing of raw materials, labor practices, and human rights abuses, policymakers and practitioners can create awareness among firms and motivate them to move towards higher sustainability levels or persuade them by implementing regulations, guidelines, and standards that encourage sustainable supply chain practices to improve their reputation, meet stakeholders' expectations, and comply with regulatory requirements. They can help firms comprehend the importance of sustainability not only in the core company but also throughout their supply chain and encourage them to implement more sustainability practices in their supply chains all around the world.

Our research highlights the positive impact of sustainability practices on financial performance, indicating that sustainability practices should not be viewed as a cost but as an investment that can contribute to the long-term financial success of firms.

Environmental turbulence has a negative impact on sustainability and also a negative moderating effect on the relationship between SSCM controversies and environmental sustainability. This implies that firms need to be aware of the impact of environmental turbulence on their sustainability practices and financial performance. We emphasize the challenges that firms face in implementing sustainability practices in turbulent environments. Firms and policymakers need to consider these challenges when facilitating, developing, and implementing sustainability strategies.

Our findings of the study on the effect of SSCM controversies on sustainability practices and financial performance are mostly consistent with the results of previous research. As an example, a study by Sarkis et al. (2011) found that SSCM practices positively impact a firm's financial and operational performance. In line with our findings, the findings of these studies also suggest that firms can improve their financial performance by developing and implementing sustainable supply chain practices.

This study relied exclusively on archival data to generate and test its hypotheses, a choice that brings inherent limitations. To gain deeper insights, future research could adopt additional methods—such as interviews or surveys with senior supply chain and operations managers—to complement archival findings. Moreover, by focusing solely on publicly traded firms from Compustat and LSEG, this study excluded smaller firms, limiting the generalizability of the results. Future work can address these constraints by broadening both the range of data sources and the variety of data collection approaches.

In addition, rather than relying solely on general sustainability indicators, subsequent studies could use more specific SSCM performance metrics to examine the influence of SSCM controversies on social and environmental dimensions of sustainable supply chain management separately. This approach would offer more nuanced insights into how such controversies affect sustainability practices and overall firm performance.

REFERENCES

Ageron, B., Gunasekaran, A., & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *International Journal of Production Economics*, *140*(1), 168–182.

Altmann, M. (2015). A supply chain design approach considering environmentally sensitive customers: The case of a German manufacturing SME. *International Journal of Production Research*, *53*(21), 6534–6550.

Anner, M., Bair, J., & Blasi, J. (2013). Toward Joint Liability in Global Supply Chains: Addressing the Root Causes of Labor Violations in International Subcontracting Networks. *Comparative Labor Law & Policy Journal*, *35*.

Aouadi, A., & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics*, *151*(4), 1027–1047.

Bharadwaj, A. S., Bharadwaj, S. G., & Konsynski, B. R. (1999). Information technology effects on firm performance as measured by Tobin’s q. *Management Science*, *45*(7), 1008–1024.

Birindelli, G., Ferretti, P., Intonti, M., & Iannuzzi, A. P. (2015). On the drivers of corporate social responsibility in banks: evidence from an ethical rating model. *Journal of Management and Governance*, *19*(2), 303–340.

Bradshaw, C. J. A., Ehrlich, P. R., Beattie, A., Ceballos, G., Crist, E., Diamond, J., Dirzo, R., Ehrlich, A. H., Harte, J., Harte, M. E., Pyke, G., Raven, P. H., Ripple, W. J., Saltré, F., Turnbull, C., Wackernagel, M., & Blumstein, D. T. (2021). Underestimating the Challenges of Avoiding a Ghastly Future. *Frontiers in Conservation Science*, *1*, 9.

Chen, L., Zhao, X., Tang, O., Price, L., Zhang, S., & Zhu, W. (2017). Supply chain collaboration for sustainability: A literature review and future research agenda. *International Journal of Production Economics*, *194*, 73–87.

Chung, K. H., & Pruitt, S. W. (1994). A Simple Approximation of Tobin’s q. *Financial Management*, *23*(3), 70.

Clarkson, M. B. E. (1995). A Stakeholder Framework for Analyzing and Evaluating Corporate Social Performance. *The Academy of Management Review*, *20*(1), 92.

Damberg, S. V., Hartmann, J., & Heese, H. S. (2022). Does bad press help or hinder sustainable supply chain management? An empirical investigation of US-based corporations. *International Journal of Production Economics*, *249*, 108504.

Donaldson, T., & Preston, L. E. (1995). The Stakeholder Theory of The Corporation: Concepts, Evidence, And Implications. *Academy of Management Review*, *20*(1), 65–91.

Du, S., Bhattacharya, C. B., & Sen, S. (2011). Corporate social responsibility and competitive advantage: Overcoming the trust barrier. *Management Science*, *57*(9), 1528–1545.

Dyck, A., Volchkova, N., & Zingales, L. (2008). The Corporate Governance Role of the Media: Evidence from Russia. *The Journal of Finance*, *63*(3), 1093–1135.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835–2857.

Fashion Brands Face New Skepticism on Green Claims - WSJ. Retrieved July 28, 2024, from https://www.wsj.com/articles/fashion-brands-face-new-skepticism-on-green-claims-931c4806?mod=Searchresults_pos4&page=1

Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, *56*(3), 758–781.

Fombrun, C. J., Gardberg, N. A., & Sever, J. M. (2000). The Reputation QuotientSM: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management*, *7*(4), 241–255.

Fombrun, C., & Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *Academy of Management Journal*, *33*(2), 233–258.

Franceschelli, M. V., Santoro, G., & Candelo, E. (2018). Business model innovation for sustainability: a food start-up case study. *British Food Journal*, *120*(10), 2483–2494.

Frooman, J. (1999). Stakeholder influence strategies. *Academy of Management Review*, *24*(2), 191–205.

Garcia-Castro, R., Ariño, M. A., & Canela, M. A. (2010). Does social performance really lead to financial performance? Accounting for endogeneity. *Journal of Business Ethics*, *92*(1), 107–126.

Golicic, S. L., & Smith, C. D. (2013). A Meta-Analysis of Environmentally Sustainable Supply Chain Management Practices and Firm Performance. *Journal of Supply Chain Management*, *49*(2), 78–95.

Gong, M., Gao, Y., Koh, L., Sutcliffe, C., & Cullen, J. (2019). The role of customer awareness in promoting firm sustainability and sustainable supply chain management. *International Journal of Production Economics*, *217*, 88–96.

Graff Zivin, J., & Small, A. (2015). A Modigliani-Miller Theory of Altruistic Corporate Social Responsibility. *The B.E. Journal of Economic Analysis & Policy*, *5*(1).

Gualandris, J., Klassen, R. D., Vachon, S., & Kalchschmidt, M. (2015). Sustainable evaluation and verification in supply chains: Aligning and leveraging accountability to stakeholders. *Journal of Operations Management*, *38*, 1–13.

Hall, J., Matos, S., & Silvestre, B. (2012). Understanding why firms should invest in sustainable supply chains: a complexity approach. *International Journal of Production Research*, *50*(5), 1332–1348.

Hartmann, J. (2020). Toward a more complete theory of sustainable supply chain management: the role of media attention. *Supply Chain Management*, *26*(4), 532–547.

Jiao, Y. (2010). Stakeholder welfare and firm value. *Journal of Banking and Finance*, *34*(10), 2549–2561.

Johnson, H. H. (2003). Does it pay to be good? Social responsibility and financial performance. *Business Horizons*, *46*(6), 34–40.

Keats, B. W., & Hitt, M. A. (1988). A CAUSAL MODEL OF LINKAGES AMONG ENVIRONMENTAL DIMENSIONS, MACRO ORGANIZATIONAL CHARACTERISTICS, AND PERFORMANCE. *Academy of Management Journal*, *31*(3), 570–598.

Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science*, *42*(8), 1199–1214.

Klassen, R. D., & Vereecke, A. (2012). Social issues in supply chains: Capabilities link responsibility, risk (opportunity), and performance. *International Journal of Production Economics*, *140*(1), 103–115.

Klein, J., & Dawar, N. (2004). Corporate social responsibility and consumers' attributions and brand evaluations in a product-harm crisis. *International Journal of Research in Marketing*, *21*(3), 203–217.

Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, *12*(1), 1.
Livesey, S. M., & Kearins, K. (2002). Transparent and caring corporations? A study of sustainability reports by the body shop and Royal Dutch/Shell. *Organization and Environment*, *15*(3), 233–258.
Locke, R. M., Qin, F., & Brause, A. (2007). Does Monitoring Improve Labor Standards? Lessons from Nike. *Https://Doi.Org/10.1177/001979390706100101*, *61*(1), 3–31.
López, M. V., Garcia, A., & Rodriguez, L. (2007). Sustainable development and corporate performance: A study based on the Dow Jones sustainability index. *Journal of Business Ethics*, *75*(3), 285–300.
Margolis, J. D., & Walsh, J. P. (2003). Misery Loves Companies: Rethinking Social Initiatives by Business. *Administrative Science Quarterly*, *48*(2), 268.
Marsat, S., & Williams, B. (2012). CSR and Market Valuation: International Evidence. *SSRN Electronic Journal*.
Mcwilliams, A., & Siegel, D. (2001). Profit-Maximizing Corporate Social Responsibility. In *Source: The Academy of Management Review* (Vol. 26, Issue 4).
Mitra, S., & Datta, P. P. (2014). Adoption of green supply chain management practices and their impact on performance: An exploratory study of Indian manufacturing firms. *International Journal of Production Research*, *52*(7), 2085–2107.
Pagell, M., & Wu, Z. (2009). BUILDING A MORE COMPLETE THEORY OF SUSTAINABLE SUPPLY CHAIN MANAGEMENT USING CASE STUDIES OF 10 EXEMPLARS. *Journal of Supply Chain Management*, *45*(2), 37–56.
Palazzo, G., & Scherer, A. G. (2006). Corporate legitimacy as deliberation: A communicative framework. *Journal of Business Ethics*, *66*(1), 71–88.
Parmar, B. L., Freeman, R. E., Harrison, J. S., Wicks, A. C., Purnell, L., & de Colle, S. (2010). Stakeholder Theory: The State of the Art . *Academy of Management Annals*, *4*(1), 403–445.
Parmigiani, A., Klassen, R. D., & Russo, M. V. (2011). Efficiency meets accountability: Performance implications of supply chain configuration, control, and capabilities. *Journal of Operations Management*, *29*(3), 212–223.
Porter, M. E., & Van Der Linde, C. (1995). Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, *9*(4), 97–118.
Qiang, Q. (2015). The closed-loop supply chain network with competition and design for remanufactureability. *Journal of Cleaner Production*, *105*, 348–356.
Ramsay, J. (2001). The Resource Based Perspective, Rents, and Purchasing's Contribution to Sustainable Competitive Advantage. *Journal of Supply Chain Management*, *37*(2), 38–47.
Reuter, C., Foerstl, K., Hartmann, E., & Blome, C. (2010). Sustainable global supplier management: The role of dynamic capabilities in achieving competitive advantage. *Journal of Supply Chain Management*, *46*(2), 45–63.
Sarkis, J., Gonzalez-Torre, P., & Adenso-Diaz, B. (2010). Stakeholder pressure and the adoption of environmental practices: The mediating effect of training. *Journal of Operations Management*, *28*(2), 163–176.
Sarkis, J., Zhu, Q., & Lai, K. H. (2011). An organizational theoretic review of green supply chain management literature. *International Journal of Production Economics*, *130*(1), 1–15.
Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, *59*(5), 1045–1061.
Seuring, S. (2011). Supply chain management for sustainable products - insights from research applying mixed methodologies. *Business Strategy and the Environment*, *20*(7), 471–484.
Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, *16*(15), 1699–1710.

Silvestre, B. S. (2015). *Sustainable supply chain management in emerging economies: Environmental turbulence, institutional voids and sustainability trajectories*.

Soytas, M. A., Denizel, M., & Durak Usar, D. (2019). Addressing endogeneity in the causal relationship between sustainability and financial performance. *International Journal of Production Economics*, *210*(January 2013), 56–71.

Surroca, J., Tribó, J. A., & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal*, *31*(5), 463–490.

Svensson, G., Ferro, C., Hogevold, N., Padin, C., & Sosa Varela, J. C. (2018). Developing a theory of focal company business sustainability efforts in connection with supply chain stakeholders. *Supply Chain Management*, *23*(1), 16–32.

Tamayo-Torres, I., Gutierrez-Gutierrez, L., & Ruiz-Moreno, A. (2019). Boosting sustainability and financial performance: the role of supply chain controversies. *International Journal of Production Research*, *57*(11), 3719–3734.

Tobin, J. (1969). A General Equilibrium Approach To Monetary Theory. *Journal of Money, Credit and Banking*, *1*(1), 15.

Vachon, S., & Klassen, R. D. (2006). Extending green practices across the supply chain: The impact of upstream and downstream integration. *International Journal of Operations and Production Management*, *26*(7), 795–821.

Xue, L., Ray, G., & Sambamurthy, V. (2012). Efficiency or innovation: How do industry environments moderate the effects of firms' IT asset portfolios? *MIS Quarterly: Management Information Systems*, *36*(2), 509–528.

DECISION SCIENCES INSTITUTE

Age, Education, and Ethics: Reassessing CEO Influence on Global Supply Chain Violations

Amir Naderpour
Eastern Kentucky University
amir.naderpour@eku.edu

Ki-Jung Kim
Eastern Kentucky University
ki-jung.kim@eku.edu

ABSTRACT

Global supply chains face increasing scrutiny for social compliance, highlighting the importance of executive leadership. Drawing on Upper Echelons Theory, we investigate how CEO age and MBA education predict social compliance violations across supplier factories. Using Fair Labor Association audit data, our findings show that older CEOs, likely to rely on legacy practices, are associated with higher violation counts. However, an MBA moderates this effect by introducing contemporary management insights that reduce noncompliance, even for older executives. Our study advances theoretical understanding of leadership's impact on ethical operations and offers practical implications for fostering responsible management in global contexts ultimately.

KEYWORDS: Global supply chains, social compliance, CEO characteristics, upper echelons theory

INTRODUCTION

Global supply chains have become increasingly complex and scrutinized in recent years, particularly with respect to social compliance and ethical labor practices. Amid rising consumer awareness and regulatory pressures, firms are compelled to ensure that their supply chains adhere to internationally recognized standards of human rights and workplace conditions. However, despite these pressures, persistent social compliance violations continue to plague many organizations, often resulting in reputational damage, legal consequences, and adverse impacts on worker well-being (LeBaron, 2021).

The strategic decisions made at the top of a firm are widely recognized as pivotal in shaping organizational outcomes, including adherence to social and ethical standards. The Upper Echelons Theory (Hambrick & Mason, 1984) suggests that the characteristics of top executives—such as age and educational background—influence corporate strategies and, ultimately, performance outcomes. In this context, the present study investigates how CEO characteristics, specifically age and whether a CEO holds a graduate business degree (an MBA), relate to the incidence of social compliance violations within global supply chains. By linking executive demographic and educational attributes to labor practices in supplier factories, this research seeks to illuminate the extent to which personal attributes of leadership can serve as antecedents to ethical management practices in a globalized economic environment.

The importance of this study lies in its potential to inform both theory and practice. On the theoretical front, the research builds on established management theories to refine our understanding of how executive characteristics translate into strategic outcomes. From a practical perspective, the findings offer insights for firms and policymakers striving to enhance compliance monitoring and to mitigate risks associated with outdated managerial practices. Ultimately, this work contributes to a growing body of literature that seeks to reconcile the dynamics of leadership characteristics with the demands of modern corporate governance and social responsibility.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Despite the increasing scholarly attention on executive characteristics and organizational performance, research examining the nexus between CEO attributes and social compliance in global supply chains remains sparse (Borghesi et al., 2014; Slater & Dixon-Fowler, 2010; Bear et al., 2010; Byron & Post, 2016). By integrating insights from Upper Echelons Theory with empirical evidence on risk management and contemporary leadership education, the present study addresses this gap. It contributes to the understanding of the role of leadership in managing global supply chain ethics.

This study investigates how CEO characteristics—specifically age and educational background (MBA)—influence social compliance violations in global supply chains. Drawing on established theories such as Upper Echelons Theory (Hambrick & Mason, 1984) and incorporating recent empirical findings, the following hypotheses are proposed. Upper Echelons Theory posits that the personal attributes of top executives shape corporate strategies and outcomes. While some research suggests that older CEOs may be more risk averse (Custodio et al., 2019), there is also evidence that longer-tenured or older executives may rely on established, legacy practices. Such practices might be less adaptive to evolving regulatory and ethical standards in today's global supply chains. Older CEOs might, therefore, be less inclined to update policies or embrace innovative approaches to address contemporary social compliance challenges. This reliance on legacy systems could result in a higher number of social compliance violations when compared to firms led by younger executives who may be more open to change and modernization. The persistence of outdated practices and a potential resistance to change among older CEOs can lead to a higher incidence of compliance issues. Thus, it is hypothesized that:

> H1: There is a positive relationship between CEO age and the number of social compliance violations in the supply chain.

The educational background of a CEO significantly influences strategic decision-making and organizational practices. CEOs with an MBA are often exposed to advanced management concepts, including modern approaches to corporate governance, ethical leadership, and sustainability (Bertrand & Schoar, 2003; Lewis et al., 2014). Such training equips them with the skills necessary to integrate contemporary CSR and compliance practices into their strategic frameworks. As a result, CEOs with an MBA may be better positioned to implement and enforce robust social compliance systems, thereby reducing the number of violations within their supply chains.

Research indicates that a strong managerial education, particularly an MBA, is linked with improved monitoring practices and a more proactive approach to managing risks associated with non-compliance (Chin et al., 2013). Accordingly, firms led by CEOs with an MBA are expected to exhibit fewer social compliance violations.

> H2: Firms led by CEOs with an MBA degree experience a lower number of social compliance violations compared to those led by CEOs without an MBA.

Although older CEOs may be predisposed to adhering to legacy practices—which could lead to higher compliance violations—the presence of an MBA can introduce contemporary managerial insights and strategic rigor into their decision-making process. In this context, the MBA credential may serve as a moderating factor that mitigates the potential negative impact of age on social compliance. CEOs who combine extensive experience with modern management training are likely to be more adept at updating and refining compliance practices, thus offsetting the risks associated with the reliance on outdated systems. This interaction suggests that the positive relationship between CEO age and social compliance violations will be weaker among CEOs who possess an MBA. In other words, the adverse effects linked to older age—such as reliance on legacy practices—can be attenuated when the CEO also has advanced managerial training.

H3: The positive relationship between CEO age and social compliance violations is weaker for CEOs with an MBA degree than for those without an MBA.

METHODS

Data and Sample

We primarily use the Fair Labor Association's (FLA) audit data to assess labor conditions in supplier countries within global supply chains. The FLA is a non-profit collaborative effort involving universities, civil society organizations, and businesses. FLA member companies commit to subjecting their supply chains to independent assessments and monitoring as part of their organizational commitment to upholding fair labor standards through transparency. The FLA publishes these assessment results to foster open dialogue about worker conditions, ensure brand accountability, and help consumers make more informed purchasing decisions. We combined the FLA audit data with hand-collected CEO characteristics to investigate the relationship between CEO attributes and labor conditions in factories located in supplier countries.

Measures

The dependent variable in this study is the number of labor violations in each factory. This measure captures the total count of violations across six categories: harassment or abuse, forced labor, child labor, freedom of association and collective bargaining, health, safety, and environment, and hours of work and compensation. The primary explanatory variables include CEO age, measured in years; CEO MBA, a binary variable coded as 1 if the CEO holds an MBA as their highest degree and 0 otherwise; and CEO tenure, measured as the number of years the CEO has served in their role. To account for other factors that may influence labor violations, we include several control variables. CEO sex is a binary indicator coded as 1 for male and 0 for female. CEO nationality is included as a categorical variable. We also control for the parent company (buyer firm), which is encoded as a categorical variable, and the factory's location, represented as a categorical variable indicating the country in which the factory operates. Factory size, measured by the number of workers, is log-transformed to account for skewness. Finally, we control for the year in which the factory assessment took place to account for potential time-based effects.

Analytical Strategy

Given that the data structure is inherently hierarchical—with individual factories nested within larger buyer firms—we employ a multilevel mixed-effects modeling approach. This strategy is ideal for analyzing data where observations (i.e., factories) are grouped within higher-level units (i.e., buyer firms), as it allows us to account for both within-group (factory-level) and between-group (firm-level) variability.

RESULTS

Table 1 presents the estimates from three multilevel mixed-effects regression models examining the relationship between CEO characteristics and social compliance violations in the supply chain. Model 1 is the basest model without control variable. From Model 2, control variables were added. Consistent with Hypothesis 1 (H1), which predicted a positive relationship between CEO age and the number of social compliance violations, Model 2 revealed that increases in age were associated with more violations. Specifically, the coefficient for Age was .47 (SE = .16, $p < .01$), indicating that the expected number of violations increased significantly with CEO age. Hypothesis 2 (H2) posited that firms led by CEOs with an MBA would experience fewer social compliance violations compared to those led by CEOs without an MBA. Although the simple main effect of MBA in Models 1 and 2 was negative (–4.37 and –6.24, respectively), these effects were not statistically significant, failing to support H2 However, Model 3 included an interaction term between Age and MBA, which provides additional nuance. The significant interaction (Age × MBA: –2.07, SE = 0.40, $p < .001$) indicates that the effect of CEO age on social compliance violations is contingent upon MBA status. Specifically, the interaction suggests that while increasing CEO age is associated with a rise in violations among non-MBA CEOs, this positive association is significantly attenuated—and indeed reversed (Figure 1)—for CEOs with an MBA. For example, for MBA-holding CEOs the net effect of age on violations is 0.63 – 2.07 = –1.44, indicating that as these CEOs age, the number of violations tends to decrease. This finding provides robust support for Hypothesis 3 (H3), which predicted that the positive relationship between CEO age and social compliance violations would be weaker for CEOs with an MBA relative to those without an MBA.

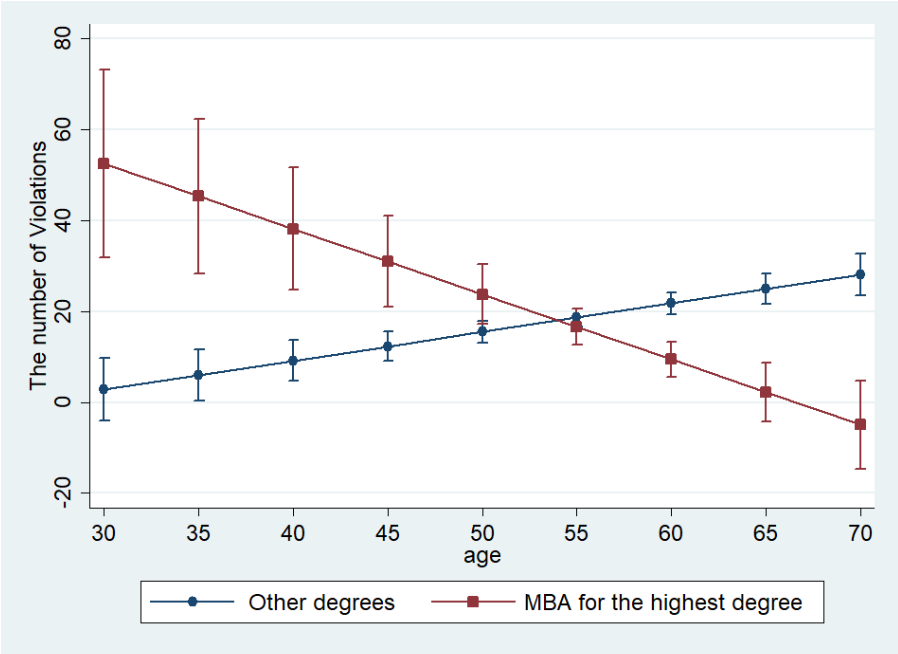
Table 1. The Results of Multilevel Mixed-Effects Regressions

	Model 1	**Model 2**	**Model 3**

Age	.18	.47**	.63***
	(.12)	(.16)	(.14)
MBA	-4.37	-6.24	111.59***
	(2.37)	(4.04)	(22.99)
Tenure	-.05	.05	.03
	(.10)	(.17)	(.13)
Age × MBA			-2.07***
			(.40)
Year		-2.34***	-2.16***
		(.00)	(.54)
Factory size (log)		.97	.89
		(.78)	(.76)
Factory nation		Yes	Yes
CEO sex		Yes	Yes
CEO nation		Yes	Yes
Random Effects			
Buyer Firm (Intercept) Variance	6.60	12.24	.00
	(13.87)	(7.86)	(.00)
Residual Variance	115.26	57.76	57.97
	(16.54)	(6.08)	(5.64)
ICC	.0542	.1757	.00
Model Fit			
AIC	1674.52	1557.85	1539.44
BIC	1694.83	1695.27	1680.22
Observations	218	211	211
Number of Groups	22	22	22

****p* < .001; ***p* < .01; **p* < .05; Standard errors in parentheses. ICC = Intraclass correlation coefficient; AIC = Akaike information criterion; BIC = Bayesian information criterion

Figure 1. The Interaction between CEO ages and MBA degrees



DISCUSSION

Drawing upon Upper Echelons Theory and recent empirical findings, our analysis provided insights into how executive attributes shape the ethical performance of firms in complex, globalized environments.

Our results indicate that CEO age is positively related to the frequency of social compliance violations, supporting the hypothesis that older CEOs, potentially due to a reliance on legacy practices and a higher degree of risk aversion, may be less inclined to adopt innovative compliance measures. However, our study also uncovered a nuanced interaction effect: the presence of an MBA significantly moderates the relationship between age and compliance outcomes. Specifically, while older CEOs without an MBA tend to experience a higher number of violations, older CEOs with an MBA demonstrate a reversal of this trend, suggesting that contemporary managerial training can offset the risks associated with age-related adherence to outdated practices. These findings not only reinforce the importance of executive characteristics in shaping corporate behavior but also highlight the potential for advanced education to foster adaptive, ethically sound decision-making even among more experienced leaders.

Limitations and Directions for Future Research

Despite these contributions, the study has several limitations that need consideration. First, the reliance on audit data from the Fair Labor Association (FLA) may introduce sample-specific biases, as the data is primarily drawn from firms committed to transparency and external review. This could limit the generalizability of the findings to firms or regions not covered by the FLA's monitoring framework. Future research employing longitudinal data could provide deeper insights into how changes in leadership attributes over time affect compliance outcomes.

Additionally, other unobserved factors—such as industry-specific regulatory environments, the role of board governance, or the influence of broader cultural norms—may also play a significant role in shaping compliance behavior. Expanding the scope of analysis to incorporate these variables could offer a more holistic view of the determinants of social compliance in global supply chains.

REFERENCES

Bear, S., Rahman, N., & Post, C. (2010). The impact of board diversity and gender composition on corporate social responsibility and firm reputation. *Journal of Business Ethics, 97*(2), 207–221.
Bertrand, M., & Schoar, A. (2003). *Managing with style: The effect of managers on firm policies. The Quarterly Journal of Economics, 118(4), 1169–1208.*

Borghesi, R., Houston, J. F., & Naranjo, A. (2014). Corporate socially responsible investments: CEO altruism, reputation, and shareholder interests. *Journal of Corporate Finance, 26*, 164–181.

Byron, K., & Post, C. (2016). Women on boards of directors and corporate social performance: A meta-analysis. *Corporate Governance: An International Review, 24*(4), 428–442.
Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: *The organization as a reflection of its top managers. Academy of Management Review, 9*(2), 193–206.

Chin, M. K., Hambrick, D. C., & Treviño, L. K. (2013). *Political ideologies of CEOs: The influence of executives' values on corporate social responsibility. Administrative Science Quarterly, 58(2), 197–232.*

Custodio, C., Ferreira, M. A., & Matos, P. (2019). *Do general managerial skills spur innovation? Management Science, 65*(2), 459–481.

Lewis, B. W., Walls, J. L., & Dowell, G. W. S. (2014). Difference in degrees: CEO characteristics and firm environmental disclosure. *Strategic Management Journal, 35(5), 712–722.*

LeBaron, G. (2021). *The role of supply chains in the global business of forced labour. Journal of Supply Chain Management, 57*(2), 29–42.

Slater, D. J., & Dixon-Fowler, H. R. (2010). The future of the planet in the hands of MBAs: An examination of CEO MBA education and corporate environmental performance. *Academy of Management Learning & Education, 9*(3), 429–441.

Using AI in Strategic Formulation to create Competitive Advantage

Abstract. Utilizing an RBV (Resource-based View) logic, this paper explores the possibility of whether the use of Artificial Intelligence (AI) can help generate strategic competitive advantage for a firm. The RBV perspective used is the VRIO model developed by Jay Barney, which can help one determine the potential competitive advantage a specific resource might give an organization. The analysis suggests that in and of itself AI cannot provide a specific company with a specific competitive advantage. However, by taking a complex systems approach that involves the innovation processes of an organization, AI can be used to develop an intangible competency. When combined with other resources such as specific industry knowledge and the innovation processes of the firm, this can have the potential for creating truly sustainable competitive advantage. It is concluded that, given the trajectory of AI development, no company can afford to ignore the impact and potential use of AI on its strategic business models but its effective use in creating a strategy for competitive advantage must still be discerned.

Keywords: Artificial Intelligence (AI), Resource-based View (RBV), VRIO.

The Resource-based View (RBV) perspective on Strategy

Strategic competitive advantage can be measured or operationalized as above average profits in for-profit businesses. Thus, to determine whether a company has a competitive advantage we can measure the profit of companies and compare them. However, this does explain why or where such profits materialize. Studies in strategic management have shown that 20% of a typical firm's profit is explained by industry structure, which can be determined using a Structure-Conduct-Performance (SCP) model like the Five Forces model [9]. However, what explains the other 80% of profit? The answer- as first delineated in the strategic scholarly works of Wernerfelt [16], Barney [1] and Hunt & Morgan [6]- is the internal resources of the company itself.

The Resource-based View (RBV) perspective on Strategy was developed to describe where the strengths of an organization exist and helped to differentiate between successful and less successful competitors within the same industry. The RBV perspective attempts to answer questions like- why does one company do so well in the same industry relative to a competitor? What are the fundamental forces behind the success of the former company?

The key, according to RBV, is in the resources and their proper integration into the company's processes. Successful companies had the required resources and less successful did not. Those successful companies also effectively deployed and utilized these important resources in ways less successful companies failed. As such, both assets as resources and capabilities as resources were identified as necessary for strategic success.

In the next two sections the importance of using the RBV perspective in analyzing the strengths and weaknesses of an organization is discussed. Then, the VRIO model is used to demonstrate how such a perspective can be used to determine the competitive advantage of a strategy for any organization. In the final two sections, that logical argument is extended to the use of AI in creating unique strategies that might lead a company towards creating a competitive advantage for itself.

Determining Strengths and Weaknesses- Basic Positions in the Internal Environment

SWOT analysis is often practiced in industry as a means of developing strategy. However, it suffers from a Garbage In-Garbage Out issue [7]. That is, the framework is a simple one, such that anyone can do it, but it is also easy to do poorly. Usually this is because of bad wording or formulation. If strategic management can be considered scientific, then like other scientific pursuits, we must use careful formulations to avoid ambiguity and error.

It is important to assign the designation of a strength to either an asset or capability that the company and to be careful about the wording. The reason it is important is because strengths are meant to be used by management to engage in strategic action that will produce a competitive advantage effect! A typical mistake by a careless strategist is to define a strength in such a way that it is, in effect, an outcome, and this is done even in some major textbooks on the topic. For example, someone might list profit as a strength of the firm. This is erroneous [7]. Why? Because profit is the result of a competitive advantage- it is an outcome, whereas a true strength must be defined as an input.

When analyzing strategic success, it is always helpful to ask “why” whenever given a fact or position. For example, if a manager says, “we have the best product in the industry!”- let us say it is the iPhone at Apple- the next question should be “Why?”. Let’s say that the manager answers … “Because it has the highest market share and sales”. But that would still not answer why. So, like Socrates might have done with his pupils, one must ask, “Yes, but why the highest market share and sales?” Now you might get into a discussion about how the product is the best designed and fits the market well- hence, these success measures. But that still avoids the issue of why the product is that way. After a while, you get to the core strengths that provide for the successful product in the first place. Quite often that core strength will be the people you have that developed and continually re-designed the “successful” product. Quite literally the strength of many companies is their people! In the case of the iPhone one important strength is the product development team and the underlying skills and knowledge that they have. Another is the market development team that promotes and distributes the product.

These are things that a manager can actually manage by either resourcing the teams or not, for example. At the strategic level, what is important is that the strength is recognized and strategically positioned to create competitive advantages. The important lesson here is that when one defines strategic strengths properly, they can then strategically manage it!

Resources (Assets & Capabilities) and Core Competencies

Resources, in general, can be categorized into the assets (both tangible and intangible) and the capabilities of which the organization has some ownership and control. In the literature, resources are often synonymous with the term “assets” but can also represent all aspects of internal strengths- like capabilities. We refer to the general concept of internal strengths as “resources” and specific items that give the organization its strength as “assets”. That is, resources incorporate assets, capabilities, and also higher-ordered competencies.

Hunt [5] defined resources as the tangible and intangible entities available to firms that enable them to produce market offerings that have value for some market segment(s). He categorized resources into seven distinct categories, as depicted in Table 1.

-Table 1 Here-

The resources illustrated in Table 1 include all types, both tangible and intangible, as well as capabilities. Describing assets in terms of tangibility is important because the more intangible an asset or resource is, the more likely it will provide protection against imitation. Tangible assets are “things” that can be touched and seen (although sometimes that might not be easy as in the example of capital resources, which are held in a bank deposit but could be eventually “grabbed” in the form of actual cash). Intangible assets are invisible to the eye and are more conceptual, but in many cases are even more important for competitive advantage. Often, tangible expressions of an intangible asset also exist, such as in the case of patent drawings. However, in the case of a patent, the asset itself is not the piece of paper describing the patented technology but actually the legal concept itself, which gets its value from the ability of the courts to protect the patent’s use and enforce judgment on misuse by others.

Table 2 shows general examples of each category of tangible and intangible assets for many organizations.

-Table 2 Here-

Capabilities are specific resources that refer to the skills at which the organization is adept. That is, they are things that the organization can do (not things that they own (i.e., assets)), which may be of value to the strategic goals of the organization. While some are more identifiable than others, capabilities as skills tend to be intangible. They are embedded in the routines and procedures of the organization- such as logistics skills or sales skills.

-Figure 1 Here-

As illustrated in Figure 1, core competencies are higher level manifestations of the enactment of a combination of an organization's assets and capabilities. Core competencies reinforce the activities that the organization enacts that gives it competitive advantage and leads to successful strategic goal achievement. An example of the core competency that underlay the emergent strategy of Honda's North American market entry in the 1970's is Honda's competency in designing, manufacturing, and marketing small engines. This led to Honda focusing on activities that leverage this core competency, which in turn is supported by the assets and capabilities in which Honda invests. When combined with the environmental fit (manifested as the need or value of such a core competency), this gives Honda a competitive advantage in the market for small engine offerings.

Determining Competitive Advantage Potential via the VRIO Model

As shown in Figure 1 competitive advantages (that is, being able to do better than your competitors and profit from that) can be explained by the unique combinations of assets, capabilities, and competencies that an organization has at its disposal [8]. However, that only tells half the story. It does not tell us, for instance, which assets, capabilities, and competencies are important for any particular situation. In fact, this initially led to criticisms of the RBV perspective as being tautological. Tautology is the repetition of an original idea using a different but comparable phrasing- the notion that something is something because it is something (i.e., a rose is a rose) ... Tautologies do not describe a true cause and effect relationship but merely say the same thing with different words. In the case of RBV's criticism, companies are successful because they have unique assets, capabilities, and competencies and because they have these unique assets, capabilities, and competencies they are successful. But we need to know why they are actually successful and what leads to competitive advantage, which is usually measured in terms of above-normal profits. The key is given by the clue in Figure 1- viz., the notion of environmental fit!

That is, strategic competitive advantage is possible when a company's strategic activities lead to it offering something different that is value-added. They must be different because if competitors offer similar strategic activities the most that one can hope for is competitive parity. Quite literally- being the same. They must be value-adding or fit with the needs of the environment's stakeholder- in this case, customers- because to generate profits the offerings must generate revenue beyond costs and, most importantly, for which the customer is willing to pay the company and not its competitors.

Indeed, assets, capabilities and competencies are important because they allow us to fulfill an environment need (for example, to fulfill the market needs of consumers for a communication technology, etc. one would require a set of skills and assets in smart phone and internet technologies). The key for an organization is that it must be the only one (ideally) that can fulfill these needs in order to profit most from it. These concepts extend from the RBV model on organizational competitive advantage: 1) resource heterogeneity and 2) resource immobility. Namely, that resources are different across organizations and that they may not transfer easily across organizations. The more that an asset, capability, or competency is unique and the harder it is to transfer across organizations, the more strategic value it may create.

The VRIO framework, depicted in Figure 2, was developed by Jay Barney to explain how one might go about analyzing the potential of an asset, capability, or competency to create a sustainable competitive advantage for an organization [1].

-Figure 2 Here-

First, one identifies an asset, capability, or competency to evaluate. The first question to ask is whether that asset, capability or competency is valuable. Here, as with much of strategic analysis, judgment is crucial and connection to the external environment “fit” is critical. In general, something is valuable if it allows the organization to exploit an opportunity or deflect a threat identified previously in our external analyses. An asset that either makes money or saves money would be valuable (note that there are also non-strategic assets that may not be valuable in this sense but nevertheless required- for example, to meet regulatory requirements like doing your taxes. In these cases, such assets etc. would not be strategic and, while necessary, do not add to the bottom line profitable (in fact, in most cases they cost more money than they save…)

Given this definition of valuable, any asset, capability or competency that does not give real value may lead to a competitive disadvantage because the asset, capability or competency will still cost the organization but does not bring in any revenue to cover the costs. In general, one would want to eliminate any such non-essential asset, capability, or competency (of course, the assets, capabilities and competencies needed for compliance issues mentioned in the last paragraph will remain).

If the asset, capability, or competency is considered valuable, the next question to ask is whether the asset, capability or competency is rare or unique to the organization. If it is, then the organization will be able to build on it towards superior performance. If the asset, capability, or competency is common then the best it can do for us is give us parity with our competitors as they will be able to utilize a similar asset, capability, or competency to do a similar thing as our organization.

Value and rarity relate to the notion of resource heterogeneity- or being different. In these cases, we may have a unique asset, capability, or competency. However, if a competitor is able to copy or substitute for the asset, capability, or competency (with another asset, capability, or competency), any resulting advantage would be worn away. Thus, only when an asset, capability or competency is costly to imitate can that asset, capability, or competency provide anything but a temporary competitive advantage. That is, it is only a matter of time before the asset, capability or competency that provided us with a competitive advantage is copied.

A few parameters have been identified that make it harder or more costly to imitate resources and competencies [2]. *Unique historic circumstances* may give a company a first mover advantage or access to raw materials or ideas/knowledge that others do not have. *Causal ambiguity* may also exist whereby the cause-and-effect between resources and competitive advantage is difficult to determine. This may be due to secrecy or other complex sets of interactions among resource conditions. *Social complexity*, whereby interactions are so complex that it is difficult to track them all, is another concept that can create intangible value that is hard to copy. In fact, it is important to reemphasize a point made earlier that most strategically valuable resources tend to be intangible. This is due to these ineffable-like traits. For example, most tangible assets are easily identified and thus half the battle towards imitation is completed by competitors. Indeed, recent research using Qualitative Comparative Analysis (QCA) has emerged to take a configurational approach to determining potential towards sustained competitive advantage [14]. We will return to this aspect of analysis later.

Finally, the last requirement of the VRIO model for determining a sustained competitive advantage is that one must be able to organize the asset, capability, or competency properly. If one owns an asset, capability or competency that is valuable, unique, and non-imitable but cannot truly use it properly competitive advantage can still be eluded[1]. Thus, the VRIO model suggests overall success also depends on an interconnected web of other assets, capabilities, and competencies.

Analyzing AI using the VRIO Model

We now have the tools to analyze whether artificial intelligence can create a competitive advantage for a specific firm. We do this by utilizing

[1] This is akin to having a beautiful piano (an asset) but not being able to play it because of lack of other skills (competence in piano playing). An organizational example similar to the piano might be a company that has a sophisticated accounting software system, but no one employed who knows how to use it properly.

the previously delineated VRIO model to determine whether AI can logically provide a competitive advantage for the company. Before we start, we need to first ask whether AI is indeed a resource at all. Otherwise, the VRIO model is irrelevant.

“AI is a machine’s ability to perform the cognitive functions we associate with human minds, such as perceiving, reasoning, learning, interacting with an environment, problem solving, and even exercising creativity” [10]. Chatbots are an often-utilized example of AI. This use is similar to the use of search engines and other research tools. As such, these can be defined as resources similar to computers.

The first question from the VRIO model is whether AI provides any value. This is always subjective and often suffers from the tautological problem mentioned earlier. However, the essential aspect of valuation is whether the resource can provide revenue generation or cost cutting abilities. It is also important to point out that this valuation is always a potentiality. A potentially valuable resource that is not utilized properly may not lead to any actual value. According to NIST, “new AI-enabled systems are revolutionizing and benefiting nearly all aspects of our society and economy” [11]. Thus, it is probably safe to argue that AI as a capability or tool is a potentially valuable resource for any company.

The next question from the VRIO model is about rarity. The singularity has not yet happened but that is a concept that AI practitioners suggest may not be too far off, demanding “conceptual clarity” on the issues it presents [4]. The singularity has been defined as “an event or phase that will radically change human civilization, and perhaps even human nature itself” [3:1]. It “describes the moment AI exceeds beyond human control”- essentially when all AI-based computers link into one thinking being (hence, singularity)- and some have predicted it is only 7 years away [12]!

We will leave the above exploration on the singularity and its effects on life as we know it on earth for the computer scientist and technological ethicist. Suffice it to state that for our purposes in determining the strategic implications of AI for firm competitive advantage the singularity provides an endgame. That is, if a singularity were to happen, it can be seen logically that anyone who is connected to AI (as a singularity) will be privy to the exact information that one's competitors are. (Leaving aside the issue of whether a singularity can result in loss of the ability for humans to control the power of the AI system...) As successful strategy is defined as the ability to create competitive advantage by being different in a value adding way, this suggests that AI cannot provide unique strategies because all actors will be exposed to the same set of information.

We probably do not need to continue the VRIO path argument from this point, as we have already discovered that AI resources are (like computers) not likely to provide competitive advantage for any firm, in and of itself. However, the computer metaphor is applicable here. Computers themselves were seen as important tools or resources for companies since IBM created its own competitive advantage by selling mainframes to larger corporations in the 1950's. As tools of business processing, etc. computers are/were indeed valuable and the first companies to invest in them could see some initial competitive advantage from those investments, particularly in terms of operational efficiency, which is not a source of competitive advantage [13]. However, as the computer became ubiquitous to all companies the rarity imperative of the resource failed the competitive advantage test. Computers then became a normal operating resource that was necessary but not sufficient for strategic success. Thus, AI is also likely to be considered a normal operating, and not a strategic, resource. Given the trajectory of AI technology, therefore, it is likely to quickly become a necessary but not sufficient resource for strategic success.

However, the key to utilizing such a technology to drive strategic competitive advantage can be seen in the historical case study of the use of computers and other logistics technologies in the success of Walmart over Kmart. An interesting aspect of the competition of these two companies that eventually lead to Kmart dissolution was that Kmart actually spent more money on IT systems than Walmart in the hopes that it might leapfrog Walmart in logistics success. The problem for Kmart was that they did not realize that it was an entire system of logistical competencies and not just IT systems that gave Walmart the edge in that competitive realm [15].

Thus, as we see, AI technologies, in and of themselves, cannot create competitive advantage. However, embedding them within a complex system with unique historic circumstances, causal ambiguity, and social complexity can provide the intangible strategic competencies that can lead to strategic competitive advantage. Despite the importance and impact of AI technology, our analysis demonstrates that it alone cannot give any one company a competitive advantage. The implications of these findings are discussed next.

Implications of the Analysis

The logical analysis of the impact of AI on a company's competitive advantage suggests that in the long run AI cannot create competitive advantage in strategic formulation. This is similar to the fact that products and services in and of themselves cannot provide competitive advantage. Rather it is likely that AI can help with competitive strategy by being used as a tool to generate options and ideas that can then be vetted by human intuition and cognition. In terms of strategic management, AI can be useful for generating lists of opportunities and threats as well as internal strengths and weaknesses. However, given the convergence of the AI technology, the tool is likely to have similar outputs for everyone using it. As such, as in many cases of strategic success, human intuition, and the interaction between the output of AI systems as well as human creativity will be necessary for developing truly unique and value-adding strategic initiatives that can lead to competitive advantage. This has major implications in terms of the use innovation processes and AI in creating new strategies.

Despite this conclusion, this does not suggest that companies should not invest in AI technologies with regard to their strategic activities. Specific use of AI in areas that are already proprietary in nature such as drug design, biogenetics, materials engineering, etc. are likely to remain paramount to success. This is because our analysis has shown that combining AI tools with the existing competencies of the firm is the best way to creating the kind of valuable and difficult to imitate intangible resources that may lead to competitive advantage. Certainly, companies that do not adequately invest in AI are likely to be at a competitive disadvantage. This, again, is akin to companies that refused to invest in computers when they were first introduced, only to fall behind to their more efficiently run competitors who did invest.

As such, the major implication of the analysis is to suggest that the way in which for-profit companies should engage in AI technologies is as resources or tools that augment other assets and capabilities within the company that can provide competitive advantage. Combining that with the existing innovation processes of the firm increases the probability of creating resources that follow the VRIO logic.

References

1. Barney, J.B. Firm resources and sustained competitive advantage. Journal of Management, 1, 99–120 (1991).
2. Barney, J.B. The resource-based theory of the firm, Organization Science, 7(5), 469–496 (1996).
3. Eden A.H., Steinhart E., Pearce D., Moor J.H. Singularity hypotheses: an overview. Introduction. In: Eden A.H., Moor J.H., Søraker J.H., Steinhart E., (eds) Singularity hypotheses: a scientific and philosophical assessment. Springer, New York, 1–12 (2012).
4. Hoffmann, C.H. A philosophical view on singularity and strong ai. AI & Soc 38, 1697–1714 (2023).
5. Hunt, S.D. A general theory of competition: resources, competences, productivity, economic growth. The review of Austrian Economics 16(4), 385–393 (2000).
6. Hunt, S.D. & Morgan, R. The comparative advantage theory of competition. Journal of Marketing 59(2), 1-15 (1995).
7. Johnson, W.H.A. How to do an effective swot analysis and tows matrix in strategic formulation in sage research methods business. Sage publications, ltd. (2023).
8. Johnson, W.H.A. Managing precision-driven strategy: developing an effective strategy for your organization. Amazon and iBooks. (2023).
9. McGahan, M. M. & Porter, M. E. How much does industry matter, really? Strategic Management Journal 18, 15–30 (1997).
10. Mckinsey Https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai#/ (2023).
11. Nist Https://www.nist.gov/artificial-intelligence. (2024).

12. Orf, D. By one major metric, artificial general intelligence is much closer than you think. Https://www.popularmechanics.com/technology/robots/a42612745/singularity-when-will-it-happen/.(2023).
13. Porter, M.E. "What is Strategy?" Harvard Business Review 74, No. 6 (November–December 1996): 61–78 (1996).
14. Rodríguez, J.C., Osuigwe, U.U.S., Mathibe, M, Calderón, E. Sustained competitive advantage and complexity: a configurational approach. Proceedings of the 22nd European conference on research methodology in business and management, ECRM 2023. (2023).
15. Sliwa, C. IT difficulties help take Kmart down | Computerworld. Https://www.computerworld.com/article/2585861/it-difficulties-help-take-kmart-down.html (2002).
16. Wernerfelt, B. A resource-based view of the firm. Strategic Management Journal 5(2), 171-180 (1984).

Table 1. Seven Resource Categories

RESOURCE	*DESCRIPTION/EXAMPLE*
Financial	Cash reserves and access to financial markets
Physical	Plant raw materials and equipment
Legal	Trademarks and licensees
Human	Skills and knowledge of individual employees including importantly their entrepreneurial skills
Relational	Relationship between competitors, suppliers and customers
Organizational	Controls, routines, cultures and competencies including importantly a competence for entrepreneurship
Informational	Knowledge about market segments competitors and technology

Table 2. Tangible and Intangible Assets

TANGIBLE	*INTANGIBLE*
Labor	Reputation and Brand equity
Capital	Proprietary knowledge (Trade secrets, R&D discoveries)
Buildings/plants	Culture and Organizational abilities
Land	Intellectual properties (Patents, designs, trademarks, copyrights)
Equipment and materials	

Figure 1. Core Competencies as Manifestations of Embedded Resources

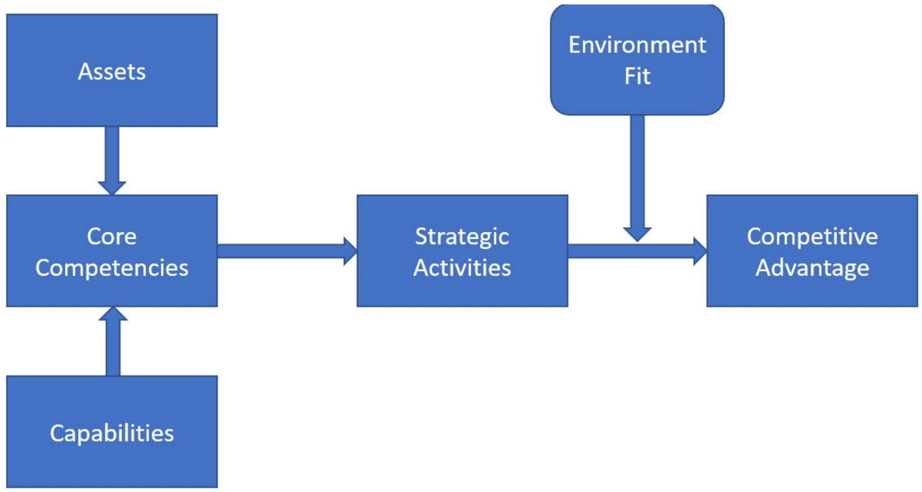
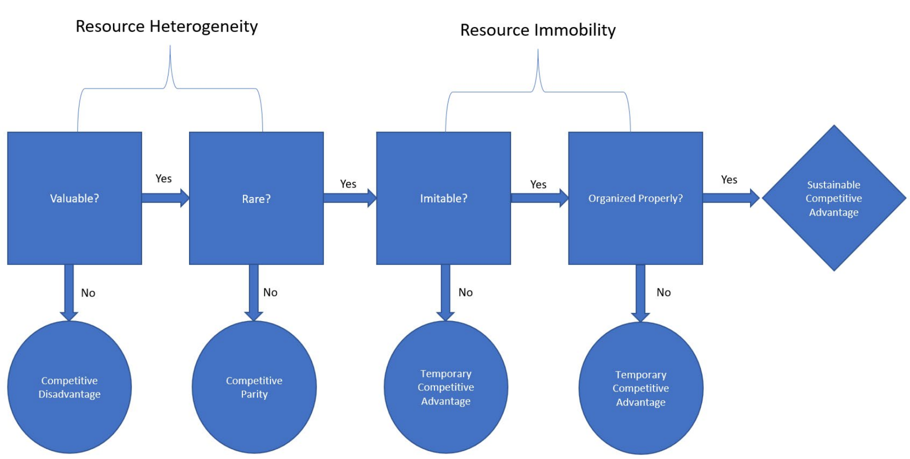


Figure 2. VRIO framework for Determining the Competitive Advantage of Resources



DECISION SCIENCES INSTITUTE

Simulating B2B Sales Process Using Agentic AI

Chhaya Tundwal, Kshitij Chauhan, Tirth Patel, Vishnu Anand, Yash Kothari, Yu-Yang Ho, Rohan Ajay, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
ctundwal@purdue.edu; chauha69@purdue.edu; pate2202@purdue.edu; anand173@purdue.edu; kothar37@purdue.edu; ho291@purdue.edu; rajay@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This study examines how companies can effectively balance AI automation and human oversight in B2B sales pipelines to maximize efficiency. We deploy Hermes-Llama-70B AI agents, leveraging Prediction Guard's API and LangChain, to automate the sales pipeline. Our analysis compares AI-driven pipelines against traditional human workflows using synthetically generated user personas and interactions. Results reveal key trade-offs across varying autonomy levels, offering practical insights. The project aims to enhances Prediction Guard's internal processes and provides a clear model for businesses seeking responsible AI integration alongside essential human roles.

KEYWORDS: AI Agents, AI Automation, B2B Sales Pipelines, LangChain, Process Automation

INTRODUCTION

The integration of artificial intelligence into enterprise sales operations represents both an unprecedented opportunity and a complex operational challenge. As organizations look to adopt more autonomous AI agents, critical questions emerge about the optimal balance between AI agent efficiency and human judgment. In this study we collaborate with Prediction Guard [https://predictionguard.com/], "a secure, scalable GenAI platform that can be self-hosted, safeguards sensitive data, prevents common AI malfunctions, and runs on affordable hardware." The research focuses on modern B2B sales pipelines, where businesses recognize AI's potential yet struggle with implementation risks ranging from data security concerns to inconsistent performance validation.

B2B sales pipelines are structured processes that guide potential customers (leads) through a series of predefined stages, including lead generation, qualification, meeting scheduling, proposal development, negotiation, and ultimately, deal closure. Each stage typically involves intensive information gathering, decision-making, and personalized communication, making it a complex environment where AI integration promises efficiency gains but also introduces new risks.

We address two core business challenges in AI-driven sales environments. First, we investigate how to replace the traditional sales pipeline stages with AI agents capable of performing tasks. For example, lead qualification, personalized outreach, follow-ups, and proposal development. Second, we develop a robust simulation framework to quantify the real-world impact of varying autonomy levels on metrics like cost per lead and time consumption throughout the sales pipeline. The project implements AI agents using Prediction Guard's LLM models and the LangChain framework, creating a testing environment that analyzes AI-driven performance against traditional human-driven processes.

An AI agent, in this context, is a software system powered by large language models (LLMs) that can autonomously perform complex sales tasks. These agents are designed to ingest large datasets, engage with potential customers through natural language interfaces (e.g., email or chat), and make real-time decisions based on pre-defined objectives and dynamic inputs. By leveraging machine learning and natural language processing (NLP) techniques, these AI agents aim to replicate—and in some cases, enhance—the activities traditionally performed by human sales representatives.

Methodologically, this investigation combines technical implementation with business impact analysis. Comparative testing of purely AI-driven workflows versus human-based workflows demonstrates nuanced performance trade-offs.

LITERATURE REVIEW

Introduction of AI in Sales

AI is reshaping sales by automating tasks and providing deeper insights into customer behaviors. According to Paschen et al. (2020), AI-driven automation has become integral to modern sales strategies, enabling companies to efficiently handle repetitive tasks. Yet, despite efficiency gains, “concerns persist regarding AI’s lack of emotional intelligence and nuanced human judgment,” highlighting the need to carefully manage autonomy levels (Umashankar & Geethanjali, 2024).

AI-Powered Sales Funnel: Opportunities and Challenges

AI significantly streamlines sales funnel stages, from lead qualification to proposal generation. Sharma et al. (2023) found AI-enhanced lead scoring improves sales conversion by effectively identifying high-value leads. However, Kumar et al. (2024) caution that “AI reliance on extensive customer data brings significant privacy concerns,” emphasizing a critical challenge in responsible AI adoption. Paschen et al. (2020) similarly underscore that although AI excels at automating routine interactions, “complex negotiations and relationship-building still necessitate human involvement.”

Comparative Studies of AI vs. Human Sales Teams

Multiple studies reveal AI’s strengths and limitations compared to traditional sales approaches. For example, Sharma et al. (2023) indicate AI delivers “higher lead conversion rates and operational efficiency” through automation. However, research consistently demonstrates human superiority in areas requiring emotional understanding and adaptability. Manoharan (2024) emphasizes that “AI systems fall short in personalization and trust-building,” suggesting an optimal sales pipeline integrates AI to handle routine processes while reserving complex interactions for humans (Paschen et al., 2020).

Future Directions in AI-Driven Sales

Future research must tackle challenges surrounding transparency and fairness in AI systems. Manoharan (2024) advocates strongly for “robust ethical frameworks to mitigate biases,” while Sharma et al. (2023) stress balancing automation with personalized customer engagement to achieve maximum effectiveness.

The foundational articles in this study are summarized in Table 1.

Table 1. Literature Review Summary

Study	Focus Area	Key Contributions	Relevance to Study
Kumar et al. (2024)	AI in marketing and its strategic applications	Explores AI's role in enhancing customer engagement and personalization.	Establishes a foundation for AI-driven sales optimization.
Paschen et al. (2020)	Collaborative intelligence in B2B sales	Analyses human-AI collaboration in sales and its impact on productivity.	Highlights AI's impact on efficiency in B2B sales processes.
Sharma et al. (2023)	AI-driven lead scoring and sales funnel efficiency	Evaluates machine learning techniques for optimizing lead scoring.	Offers a benchmark for AI applications in lead prioritization.
Manoharan (2024)	Ethical AI governance in sales automation	Discusses biases in AI-driven decision-making and ethical considerations.	Addresses risks and challenges in AI adoption for sales.
Umashankar & Geethanjali (2024)	Behavioral AI for customer engagement	Investigates AI's ability to decode consumer behavior for targeted marketing.	Supports AI-driven customer behavior analysis for sales strategies.

DATA

The efficacy of the proposed LLM-driven sales agent hinges on the availability of a robust and representative dataset for simulation, evaluation, and iterative refinement. To address this need, we constructed a synthetic dataset meticulously designed to mirror the complexities and nuances of real-world enterprise sales interactions. This dataset comprises structured customer profiles coupled with detailed meeting records, capturing the dynamic interplay between sales agents and prospective clients. This approach allows for controlled experimentation and analysis, circumventing the challenges associated with acquiring and utilizing sensitive real-world sales data.

While the customer profiles are synthetically generated, the distribution of professional designations (e.g., Engineers, CEOs, CIOs, AI Researchers) is calibrated to reflect common organizational structures and decision-making hierarchies relevant to the target industry. A core component of the dataset is the simulated meeting data. This includes a verbatim transcription of the interaction between the sales person and the prospect. These transcriptions represent the crux of the sales conversation, detailing the exchange of information, negotiation points, and expressions of interest.

Furthermore, the dataset incorporates critical sales performance metrics. Each lead is assigned a classification of cold, warm, or hot, reflecting its perceived potential based on engagement levels and indicators of purchasing intent gleaned from the meeting transcript. Finally, each meeting record is augmented with a post-meeting insights summary, outlining key discussion points and recommending actionable next steps. This information serves as a critical input for AI-driven follow-ups and strategic optimization of the sales process.

The synthetic dataset was generated using advanced prompt engineering techniques provided in the Appendix and summarized in Table 2 below. By carefully designing and iteratively refining the prompts used to guide the language model, we were able to produce structured and meaningful data that closely approximates real-world sales dynamics.

Table 2: Synthetic Data Generated

Variable	Type	Description
Name	String	Full name of the prospect.
Company	String	Name of the organization the prospect belongs to.
Designation	String	Job title of the prospect within the company.
Professional Email	String	Business email address used for communication
Meeting Date	Date	The scheduled date of the sales interaction.
Meeting Duration	Integer	The length of the meeting in minutes.
Meeting Transcript	Text	AI-generated transcript of the conversation.
Lead Score	Integer	Score indicating the likelihood of conversion.
Lead Classification	Categorical	Classification of the lead based on engagement and interest.
Customer Persona	String	Buyer behavior profile assigned to the prospect.
Key Discussion Points	Text	Summary of critical topics covered in the meeting.
Recommended Next Steps	Text	AI-suggested actions based on the meeting outcome.

METHODOLOGY

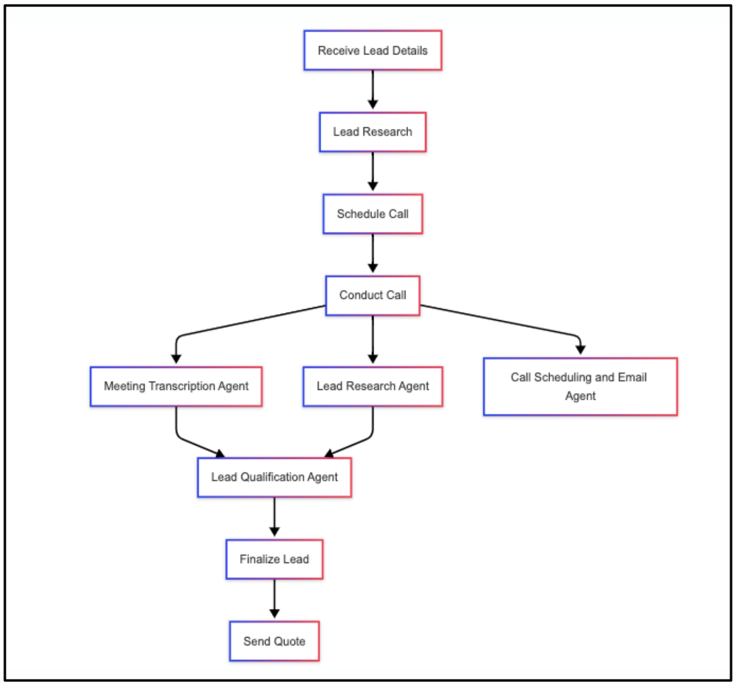
A structured, simulation-based experimental methodology in employed to evaluate the optimal integration of AI autonomy and human oversight within enterprise sales pipelines. The core of the methodology involves designing and simulating an AI-enabled sales pipeline, systematically comparing fully autonomous AI workflows with traditional human based pipeline, and analyzing the resultant performance metrics. The AI sales pipeline is structured around five distinct stages, each designed to automate a specific stage of the sales process:

Lead Research Agent - In the first stage as shown in Figure 1, prior to any direct engagement, this AI-powered module leverages publicly available data sources, including professional networking platforms (e.g., LinkedIn), and public channels, to gather and analyze prospective client data. This module aims to enhance lead quality by providing comprehensive prospect profiles, thereby reducing the time and resources expended on manual lead research.

Lead Qualification Agent - This agent analyzes the transcribed meeting data in conjunction with external insights to assess the potential of the lead. Prospects are scored based on predefined qualification criteria, providing a quantifiable measure of lead viability and enabling prioritization of sales efforts.

Call Scheduling and Email Agent - Upon identification of a promising lead, this agent automates the coordination of meetings. The agent proposes suitable meeting times, generates and sends calendar invitations, and manages follow-up reminders, thereby streamlining the scheduling process and improving sales team efficiency.

Figure 1. B2B Sales pipeline using AI Agent



Meeting Transcription Agent - During scheduled meetings, this agent records and transcribes key discussion points in real-time. The transcriptions are then summarized to provide structured insights for the sales team, mitigating the risk of overlooking critical information and facilitating comprehensive follow-up actions.

Quote Generation Agent - For leads deemed viable by the Lead Qualification Agent, this AI module generates customized quotes based on the information gathered during previous interactions and the identified needs of the prospect. This ensures timely and relevant proposal delivery, enhancing the likelihood of conversion

The experimental design centers on controlled simulations of the AI-driven sales pipeline. Key performance metrics are rigorously tracked and analyzed, including efficiency gains—measured by time saved per task and reductions in overall sales cycle length—and cost comparisons, specifically evaluating token utilization expenses versus traditional salesperson compensation.

RESULTS

Our analysis highlights the impact of integrating AI into sales pipelines, revealing advantages over traditional approaches. We discovered that AI-driven processes enhance outreach capabilities, increase the precision of lead qualification, and significantly reduce costs. Additionally, the adoption of AI streamlined scheduling tasks and freed sales teams from administrative burdens, enabling them to focus on higher-value interactions. We also observed there are limitations of AI like unavailability of nuanced negotiation scenarios for AI agents to tackle complicated scenarios making it crucial to include human in the loop.

To derive the metrics in our results, we integrated precise time-tracking mechanisms within our AI pipeline code, systematically capturing the duration for each step of the sales workflow. Additionally, we utilized Hugging Face's Token Counter tool to accurately determine token usage per interaction, enabling realistic cost estimations based on consumption. To validate accuracy metrics, we conducted simulations using a controlled sample of twenty leads, directly comparing predicted lead scores against known outcomes, resulting in the reported accuracy rate of 60%. These combined methodologies ensured a robust foundation for the performance metrics summarized in Table 3.

Table 3: Metrics Comparison

Metric	Description	AI Performance	Human Performance
Outreach Volume	Number of personalized customer interactions managed per day	100	30-50
Lead Qualification Accuracy	Ability to correctly identify high-quality leads likely to convert into customers	65-70%	70-75%
Lead Scoring Precision	Accuracy in prioritizing leads based on true sales potential	60%	65-75%
Cost per Qualified Lead	Average cost to generate and qualify each lead	Less than $0.03	$50-200
Sales Cycle time	Time required to complete lead qualification, send email & generate quote (per 100 lead)	4 hours	13 hours
Total Compensation Costs	Comparative total cost including human oversight, technology, and infrastructure	$5 for 100 leads	$450 for 100 leads
Scalability Costs	Cost growth pattern as sales operations expand	Linear	Exponential

Table 4: Metrics Sources and Assumptions

Metric	Sources for Human Benchmark	Assumption
Outreach Volume	HubSpot's 2023 State of AI in Sales report	-
Lead Qualification Accuracy	McKinsey's 2023 "AI in B2B Sales" report	-
Lead Scoring Precision	1. IBM Watson Marketing benchmark data (2023) 2. Gartner's 2023 "MarTech Survey"	Requires complete customer interaction history
Cost per Qualified Lead	HubSpot's 2023 "Inbound Marketing Benchmark Report"	Factors in technology costs but assumes scale
Sales Cycle time	Salesforce Research "State of Sales" report (2023)	Sales conversations are linear and non-friction.
Total Compensation Costs	Deloitte's 2023 "Future of Work in Sales" report	Includes tech infrastructure and human oversight
Scalability Costs	Gartner's 2023 "Cost of Sales Technology" study showing different scaling patterns.	At enterprise scale

These findings illustrate the practical advantages of adopting AI-driven sales pipelines. Organizations leveraging AI can achieve greater efficiency, reduce operational costs, and significantly improve lead quality and scalability. By integrating AI strategically with human oversight, enterprises can unlock new opportunities to elevate sales performance and drive sustainable growth.

CONCLUSIONS

This study explored how AI-driven agents can enhance B2B sales pipelines while maintaining a balance with human oversight. By automating key sales tasks - lead research, qualification, scheduling, transcription, and proposal generation - our AI-driven approach significantly improved efficiency, reducing costs and increasing outreach capacity. The AI agents outperformed traditional human-led processes in lead qualification accuracy, response times, and administrative workload reduction, making sales operations more scalable and cost-effective.

However, AI still has limitations. It lacks the human touch needed for complex negotiations, trust-building, and long-term relationship management - critical aspects of high-value sales. Additionally, while AI-driven decision-making is efficient, there's a risk of biases that require careful monitoring. Our study assumes AI models can generalize well across different industries, but real-world implementation may introduce challenges related to data privacy, compliance, and adaptability.

Looking ahead, businesses should focus on refining AI-human collaboration, improving AI's ability to handle nuanced sales interactions, and addressing potential biases in lead scoring. As AI adoption in sales continues to grow, finding the right balance between automation and human expertise will be key to building effective, scalable, and ethical sales processes.

For Prediction Guard, the dual-focused outcomes deliver immediate internal improvements and long-term strategic value. The simulation environment not only optimizes their sales processes but also serves as marketable case study demonstrating responsible AI implementation. By establishing measurable benchmarks for autonomy effectiveness across sales stages, this research provides enterprises with actionable frameworks to maximize ROI while preserving operational integrity—a crucial balance as global markets approach tipping points in AI adoption.

REFERENCES

Manoharan, J. (2024). Navigating the AI Wave in Martech: A Systematic Literature Review of Developments, Challenges, and Ethics., *SoutheastCon 2024*, Atlanta, GA, USA, pp. 616-622, https://doi.org/10.1109/southeastcon52093.2024.10500176

Umashankar, N. and Geethanjali, K. (2024). Decoding Consumer behavior: behavioral ai for smarter marketing analytics.. https://doi.org/10.31224/4155

Paschen, J., Wilson, M., & Ferreira, J. (2020). Collaborative Intelligence: How Human and Artificial Intelligence Create Value Along the B2B Sales Funnel. Business Horizons, 63(3), 403-414. https://doi.org/10.1016/j.bushor.2020.01.003

Sharma, K. K., Tomar, M., & Tadimarri, A. (2023). Optimizing Sales Funnel Efficiency: Deep Learning Techniques for Lead Scoring. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 261-274. https://doi.org/10.60087/jklst.vol2.n2.p274

Kumar, V., Ashraf, A. R., & Nadeem, W. (2024). AI-powered marketing: What, where, and how? International Journal of Information Management, 77, 102783. https://doi.org/10.1016/j.ijinfomgt.2024.102783

APPENDIX

Agent	Purpose	Exact Prompt
Lead Research Agent	Analyze company info and meeting notes	"You are a marketing assistant skilled in analyzing business information. Below is the company you have to analyze with information, meeting notes. COMPANY TO ANALYZE: {company} COMPANY INFORMATION: {company_info} MEETING NOTES: {meeting_notes} Based on the meeting notes and company information, provide a detailed summary that includes: 1. Key information about {company} 2. Important points from the meeting notes 3. Specific details about: {topics} Please organize your response in a clear, structured format."""
Lead Qualification Agent	Score lead quality from meeting transcript	""" You are a sales representative of a B2B company which provides a secure API to companies to access open-source LLMs. You are extremely skilled in analysing lead quality based on lead profile. Analyse the provided meeting transcription **without making assumptions**. - **Assign a lead score from 0 to 100.** - **Justify the score ONLY based on what was explicitly stated.** - **If the transcription lacks urgency, do not assume it.** - **Keep responses concise and avoid any extra details.** **Meeting Transcription:** {transcription} **Prediction Guard Offerings (Relevant Data for Comparison):** {offerings}"""
Call Scheduling Agent	Schedule calendar slots	"""Book slots according to {transcript}. Label them with the work provided to be done in that time period. Schedule it for the day mentioned in the todo. Today's date is {date} (YYYY-MM-DD format) and timezone is {timezone}. Create Google Meet links for all events. The output must be JSON: {{"meeting_date": "YYYY-MM-DD" "participants": ["John Doe", "Jane Smith"], "meeting_description": "Project discussion", "meeting_link": "https://meet.google.com/xyz-abc"}}"""
Email Agent	Email participants with meeting info	"""Take the meeting details from the previous task and send an email to the participants. The email should include the raw meeting details in the body. Use the Gmail API to send the email. Example JSON: { "user_id": "me", "recipient_email": "{participants}", # Only the first participant's email "subject": "Meeting Scheduled: {meeting_description}", "body": "Here are the meeting details:\n\n{raw_meeting_details}", "is_html": false # Plain text email format }"""
Quote Generation Agent	Recommend PG products	"Given the following business use case, recommend Prediction Guard products/services that best suit the company's needs." Business Use Case: {overall_summary}" "Format your answer as:"

		"<Product Name>:" <Reason for recommendation>" "Only suggest Prediction Guard products."

DECISION SCIENCES INSTITUTE

Enhancing Judicial Efficiency and Access to Justice Using AI

Shubham Gaddi, Saquib Hussain, Nandini Devalla, Hugo Bizcek, Ayush Gupta, Darshan Upadhyay, Rohan Ajay, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
sgaddi@purdue.edu; hussai71@purdue.edu; ndevalla@purdue.edu; hbizcek@purdue.edu; gupt1248@purdue.edu; upadhy33@purdue.edu; rajay@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This study explores how AI can enhance productivity in Indiana's legal system. The motivation is that Integrating AI into court workflows requires balancing efficiency gains with accountability and protecting sensitive data from AI feedback loops. We use survey data from over 100 Indiana judges, and apply Python-based NLP and Azure Language AI to extract sentiments and key concerns, offering guidance for responsible AI integration in judicial processes.

KEYWORDS: Artificial Intelligence, Judicial Productivity, Legal Technology, Natural Language Processing, Sentiment Analysis

INTRODUCTION

Artificial Intelligence (AI) has inevitably become a part of our daily lives in more ways than we realize, and the sooner we start accepting it, the better our lives are going to be. More and more industries and workplaces are incorporating AI into their daily workflows to optimize output, reduce processing time, and focus human resources on more complex tasks. One domain that struggles with AI incorporation is the legal system due to the nuances of how AI impacts its principles.

Not only do the courts have to set the tone on what is and what is not permissible, but they also need to set precedent by leading by example (Illinois Supreme Court, 2025). Given the sensitive nature of court data, security and safety of the information is also of unparalleled importance. These reasons make it tricky for the courts to adopt AI-enhanced operational workflows, which could make them lag behind in efficiency and productivity gains other domains are realizing.

This study explores two major problem statements—the incorporation of AI into everyday legal workflows to improve efficiency and facilitate easier access to justice, and the identification of AI-generated or AI-altered evidence during legal processes. The first problem deals with the aspect of ensuring that the courts are using AI ethically, optimally, and synergistically with human supervision, while the second concern ensures that access to justice is not deterred by modern-day AI capabilities. AI can generate and process a significant amount of information, and being able to flag errors in legal processes is important. It helps in raising reasonable doubt for further investigation and brings up the question of ethics and accountability around the use of AI.

Through this study, we explore some common challenges that stand in the way of incorporating AI into daily legal workflows, along with how these problems were tackled and the impact of these strategies. We also dive into the impact of introducing AI tools to help speed up processes like legal research, writing briefs, and courtroom transcriptions in terms of time saved, cost incurred, and overall efficiency gains. We then dive into a more specific question—how do we identify what information is AI-generated or AI-altered with a certain level of confidence and accuracy.

It is crucial that our legal system become appropriately aware and equipped of AI technologies as AI-generated images and videos, along with deepfakes, are becoming more common. Providing educational resources for these decision-makers can help them better identify, evaluate, and rule on these items with confidence (National Courts and Sciences Institute, 2024). This confidence could come from a process that creates reasonable doubt around the nature of such media, which in turn makes the decision-making process transparent by accounting for AI as a factor.

We structure the remainder of this paper by reviewing the academic literature, outline the methodology and models used which include the use of the Azure Language AI through Python. We find this approach could made it easier to identify and pick out key themes and concerns coming from the courts. We show how we built on these themes to optimize court workflows. Lastly, we provide a set of guidelines and best practices for potentially identifying AI-generated content and media based on our ongoing research, and currently available AI tools.

LITERATURE REVIEW

AI has been increasingly used in judicial systems to enhance decision-making, automate administrative tasks, and increase legal research capability. The Indiana National Courts and Sciences Institute conducted a survey of the awareness and use of AI among judicial officers and discovered that while some judges use AI for legal research and administrative tasks, 50% have never used generative AI tools (Indiana Judiciary Report, 2024). Likewise, the Illinois Supreme Court has crafted policies for AI adoption, focusing on the need to keep ethical and legal standards in place while recognizing the ability of AI to enhance access to justice. The National Center for State Courts has also issued advice regarding AI adoption, cautioning against threats like bias, accuracy problems, and ethics issues.

With the emergence of generative AI models such as ChatGPT, concerns over the authenticity of AI-generated content have grown. Some have experimented with the susceptibility of AI text detection techniques and established the susceptibility of existing systems under paraphrasing and spoofing attacks (Sadasivan et al., 2025). The authors found that watermarking techniques, retrieval-based detectors, and zero-shot classifiers all fall drastically if the attackers apply paraphrasing attacks. This reflects how difficult it is to ensure reliable AI text detection.

The massive expansion of deepfake technology has brought about colossal challenges to digital forensic examination and disinformation detection. Blümer et al. proposed a deepfake detection pipeline combining background-matching with CNN-based classification, whose efficiency was realized on unseen datasets. Traditional classifiers are very accurate on cleaned datasets but are not generalizable to real-world deepfake detection. Thus, making a need for robust and adaptive detection methods.

Despite advances in AI adoption in the judiciary and AI-generated content detection, there are several gaps:

- *Judicial AI Policies*: While there exist frameworks such as the Illinois Supreme Court AI Policy providing ethical guidelines, empirical research evaluating AI's impact on case outcomes and judicial biases is limited. The New Jersey Supreme Court authorized preliminary guidelines to be effective immediately, while also providing directions for lawyers to raise individual questions about specific AI ethics issues and provide comments and suggestions to inform the committee's ongoing work.

- *Robust AI-Generated Text Detection*: Today's AI text detectors do not work well with intelligent paraphrasing and spoofing attacks (Illinois Supreme Court Policy, 2025). Further research must be done to develop robust detection methods that can counter adversarial manipulations. One defense relies on retrieving semantically-similar generations and must be maintained by a language model API provider.
- *Scalable Deepfake Detection*: Semi-blind detection pipelines improve the performance on unseen datasets, yet real-world applications demand more scalable and computationally lighter models.

Some researchers have called for continual research in these avenues. The Illinois Supreme Court has committed to revisiting AI policies regularly as technologies emerge (Illinois Supreme Court Policy, 2025). Researchers in AI detection have stressed the importance of resilient watermarking and neural network-based approaches to address emerging threats (NCSC AI Guidelines, 2024). Additional advances in deepfake detection require integration of perceptual hashing techniques to create enhanced generalizability to datasets. We summarize these studies in Table 1 are compare it to this study.

Table 1: Literature review summary

Study By	Focus Area	AI Adoption	AI Detection	Policy Considerations	Empirical Evaluation	Limitations Identified
Indiana National Courts and Sciences Institute	Judicial AI Usage	✓	✗	✗	✓	✗
Illinois Supreme Court	AI Policy Guidelines	✓	✗	✓	✗	✓
National Center for State Courts	AI Advisory	✓	✗	✓	✗	✓
Sadasivan et al.	AI Text Detection	✗	✓	✗	✓	✓
Blümer et al.	Deepfake Detection	✗	✓	✗	✓	✓
Our Study	AI Use in the Judiciary	✓	✓	✓	✓	✓

In conclusion, reviewing the published literature on AI adoption in the judiciary, AI content detection, and deepfake detection has developed dramatically. Yet, policy implementation gaps, robust detection, and generalizability remain critical areas for continued research. Bridging these issues will advance knowledge on AI applications in legal domains could help shape more effective AI content detection technologies. This review motivates the relevance and need for our research against the backdrop of the wider scholarly debate.

DATA

The foundation of this research is anchored in the data collected from a comprehensive survey designed to elicit insights into the role, requirements, and concerns surrounding the integration of AI within the Indiana court system. Responses were obtained from a diverse cohort of over 100 judges, representing a wide spectrum of judicial experience and backgrounds. The participant pool encompassed judges from various court types, including criminal, tax, mental health, and appeals courts. This audience had variations in time served on the bench, thereby ensuring a comprehensive representation of the Indiana judiciary.

This dataset was instrumental in identifying overarching themes related to areas of concern regarding AI adoption, in addition to consolidating suggestions pertaining to preferred methods of instruction and training. By analyzing the responses, the study was able to discern common anxieties, ethical considerations, and practical challenges that judges perceive as potential impediments to the successful implementation of AI technologies. Furthermore, the survey provided valuable insights into the preferred modes of learning and professional development that would best facilitate the integration of AI into judicial workflows.

The survey instrument also gathered data on the types of legal tools and processes that would most benefit from the application of AI, providing the research team with a clear starting point for targeted intervention. This information enabled the identification of specific areas where AI could offer the greatest gains in terms of efficiency, accuracy, and access to justice. The findings from this section of the survey elucidated the courts' and judges' aspirations for AI, their ideal pathways toward achieving the desired state, and how this research could best support and facilitate their objectives.

The survey included three open-ended questions, participation in which was optional, designed to capture qualitative insights into key concerns, most valued resources, and miscellaneous comments regarding AI. These responses were subjected to qualitative analysis to explore ambiguous themes and nuanced perspectives that might not be fully captured through quantitative data alone. This approach allowed for a deeper understanding of the judges' attitudes, beliefs, and experiences with AI, enriching the overall findings of the study and providing a more holistic view of the challenges and opportunities associated with AI adoption in the Indiana court system.

We also conducted qualitative interviews with 12 judges spread across different jurisdictions courts across Indiana, all of whom had varying levels of experience being on the bench to get information about their daily workflows and their key concerns. The interview transcripts and noted were instrumental in recognizing potential throttle points that could be solved through AI and analytical solutions to speed of the overall efficiency and turnaround time for the court. Key workflows were identified based on their importance, impact and ease of resolution to provide pilot proof of concepts for initial testing.

Figure 1: Level of Familiarity that the judges have with using different AI tools.

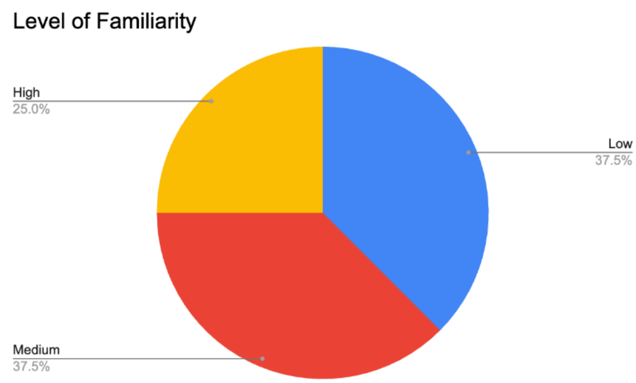
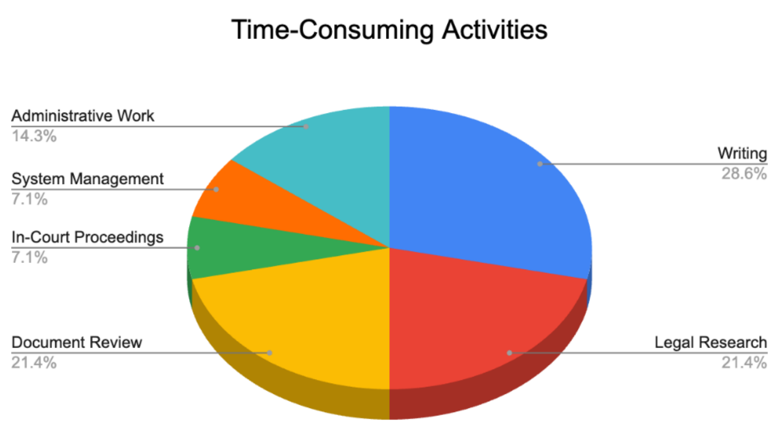


Figure 2: The most time-consuming activities according to Judges, some of which could be automated



METHODOLOGY

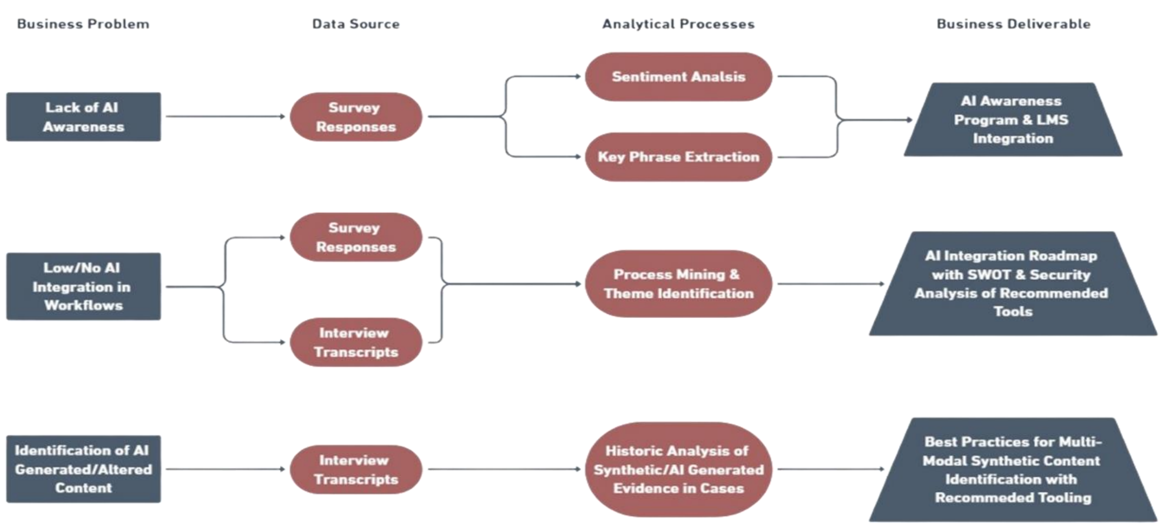
Our project operates on three major focus areas, each employing a blend of quantitative and qualitative modeling procedures.

The first area focuses on identifying gaps surrounding AI in the courts, which will inform the development and structure of an AI awareness packet for judges. The quantitative component of this area leverages survey data, wherein specific questions and responses are used to generate scores for key themes (e.g., Gen AI, Text to Speech, GPTs, Prompt Engineering). The most relevant themes, determined by these scores, serve as the foundation for the awareness packet. Additionally, sentiment analysis and key phrase extraction are performed on the open-ended survey responses to identify additional salient themes for inclusion in the awareness packet. The awareness packet will be created as an initial draft of documents and subsequently integrated into the Indiana Office of Court Services (IOCS) learning management system for regulation, control, and dissemination.

The second area of the project examines the daily workflows of judges to identify areas suitable for AI integration, with the aim of enhancing efficiency and productivity. This involves quantitatively scoring and identifying processes that judges would like to augment with AI. Qualitative interviews are then conducted with over ten judges across different counties, courts, and levels of seniority to gain deeper insights into these workflows, identify potential areas for improvement, and evaluate relevant AI tools for these specific needs.

The third area addresses multi-modal AI detection and is purely hypothetical. The goal is to identify tools that offer such services, identify datasets for testing these tools, and conduct a comparative market analysis to provide the IOCS with a list of costs and benefits associated with each tool. AI poses a wide range of challenges including bias, transparency and accountability2. A recent UNESCO survey revealed that 44% of judicial operators are utilizing AI tools1. A lack of guidance and training on AI usage is evident1. This highlights the need for scalable and computationally lighter models for real-world application. We summarize the three research focuses below in Figure 1.

Figure 3: Three focus areas and methodological approach



AI TECHNOLOGIES

The first phase of our project involved performing NLP analysis on survey responses, particularly focusing on the open-ended questions, to establish a clear starting point for the content of the AI awareness packet. These questions primarily revolved around gauging judges' concerns, requirements, openness toward adopting AI, current familiarity levels with AI technologies, and overall opinions regarding the use of AI within their courtrooms. Instead of relying on basic or naive NLP algorithms, we leveraged the advanced capabilities of Azure Language AI services for two main analytical purposes—sentiment analysis and key-phrase extraction.

Survey responses from over 100 Indiana judges served as the primary data source for this analytical effort. Sentiment analysis and key-phrase extraction performed using Azure Language AI were chosen due to their superior accuracy and reliability compared to simpler algorithms. These advanced techniques helped identify critical themes such as generative AI, text-to-speech tools, GPTs, and prompt engineering. The insights gained from this analysis were directly used to inform the development of an AI awareness packet, which serves as an educational resource specifically tailored for judges. This awareness packet was initially created as a draft document and subsequently integrated into the Indiana Office of Court Services (IOCS) Learning Management System (LMS) for controlled dissemination, regulation, continuous updating, and sharing among judicial officers.

The second part of the project aimed at identifying opportunities for integrating AI technologies into judicial workflows to enhance efficiency and productivity. Data sources included quantitative survey responses as well as qualitative interview transcripts from judges across different counties, courts, and seniority levels. Process mining techniques and thematic identification methods were employed to analyze this data comprehensively and pinpoint specific areas where AI could be effectively incorporated. The findings informed the creation of an actionable AI integration roadmap that included a detailed SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis and security evaluation of recommended tools. To ensure practical relevance and accuracy in understanding judicial needs, interviews were conducted with 12 judges representing diverse jurisdictions—including criminal courts, mental health courts, family courts—and varying levels of seniority. These interviews provided deeper insights into existing workflows and highlighted specific pain points faced by judges in their daily tasks.

The information collected through these interviews was combined with open-ended survey responses related to areas that judges believed would significantly benefit from AI integration. Some notable points of interest identified through this analysis included assistance in calendar tracking systems, opportunities for significant improvements in court transcription processes—particularly within mental health hearings—and the establishment of automated check-in and case-tagging systems during walk-in hours in family and misdemeanor courts. Although many of these processes are not inherently complex or overly convoluted, targeted AI integration could substantially boost efficiency by minimizing routine administrative burdens. This would allow judges more time to focus their efforts and decision-making capacities on more pressing judicial matters and complex processes that require nuanced human judgment.

The third component of our project addressed the growing challenge associated with identifying synthetic or AI-generated content within legal proceedings—a concern amplified by the rapid proliferation of generative AI models and deepfake technologies. Interview transcripts served as one primary qualitative data source for this prong. Additionally, historical analyses of synthetic or AI-generated evidence presented in legal cases were conducted to derive robust best practices for detection procedures. This involved systematically evaluating existing tools designed for multi-modal synthetic content identification and conducting a thorough comparative market analysis to assess each tool's costs, benefits, accuracy levels, scalability potential, transparency features, ease-of-use considerations, as well as overall suitability within a courtroom context.

To rigorously evaluate these detection tools' performance metrics—including precision rates and computational efficiency—we utilized publicly available benchmark datasets featuring diverse examples of AI-generated images, deepfakes videos, synthetic audio clips, and other forms of artificially generated media content. Each evaluated tool was documented comprehensively within a detailed proposal outlining their respective advantages and disadvantages across various critical metrics such as performance accuracy rates under realistic conditions, cost-effectiveness considerations for implementation at scale within judicial systems statewide or nationwide if necessary; transparency regarding underlying detection methodologies; ease-of-use by non-technical courtroom personnel; computational resource requirements; scalability; vendor support availability; training needs; integration complexity with existing judicial IT infrastructure; and potential ethical implications surrounding their deployment in sensitive legal environments. This comprehensive proposal was formally submitted to the IOCS leadership team for further deliberation regarding strategic planning decisions related to future adoption pathways.

RESULTS

The AI awareness packet was successfully integration within the IOCS learning management system allowing for it to be rolled out to judges on demand. The content can be controlled, managed, modified and monitored by the IOCS and make changes accordingly to keep the contents relevant over time.

Pilot programs for the workflows were setup and recommended to the IOCS and relevant owners to test and implements, subject to appropriate approvals being granted. If successful, the programs would be introduced to a wider set of judges with a choice to implement in their respective courts.

The overall comparison and proposal packet for AI tools to aid with multi-modal synthetic media detection was also prepared and provided to IOCS. The overall decision is subject to jurisdictional and approval process at multiple levels.

CONCLUSIONS

We were able to successfully obtain answers to our initial research questions and provide reasonable and practical suggestions to the Indiana Office of Court Services as a proposal for them to look at and decide on next steps.

We were able to successfully identify the key points of concern and the areas surrounding AI that the Indiana judges lacked understanding on. Some interesting things that came out of this analysis included – security concerns around sharing files with GenAI tools, lack of knowledge about AI tools and their use-cases, lack of understanding about other industrial use-cases for AI etc. The judges also seemed to agree on the fact that there were certain resources that would be more useful than others when it came to earning more about the current AI landscape. Most judges preferred videos as the preferred mode of instruction with a significant number of judges also giving a high priority to guest lectures. We also noticed a lot of judges being hesitant to use AI tools because there was no clear distinction between what tools classify as AI, a simple example would be ChatGPT v/s Grammarly.

We were able to tackle the question of what part of a judge's daily workflows can be optimized through the use of AI and analytical processes. The interviews with the judges proved to be immensely helpful in this regard and provided us with a lot of key insights. We learnt that a lot of judges would appreciate help in identifying useful or important parts of documents to speed up the process of doc review, allowing them to scan through larger documents in smaller amounts of time without having to spend a good chunk of their time on it. Calendar management was also a key concern that came up in some conversations. There were also court procedure related concerns such as the lack of a check-in process in the family and misdemeanor courts which makes it difficult for judges to figure out which defendants are in court, which ones have attorneys of their own etc. This slows down that overall process for everyone and therefore was something that the courts would appreciate. Expedition of transcription processes for mental health cases and other occasional cases was also a major concern and something that needed to be solved on the ready.

We also made progress on the front of identifying tools that can detect and flag synthetic media. We provided a comprehensive list of tools available for the courts to test and then choose to rely upon for identify AI generated texts, videos, images and audio.

REFERENCES

Artificial Intelligence Use in the Indiana Judiciary. (2024). National Courts and Sciences Institute

Illinois Supreme Court AI Policy. (2025). Illinois Supreme Court

National Center for State Courts. (2024). Artificial Intelligence: Guidance for Use of AI and Generative AI in Courts

Sadasivan, V., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2025). Can AI-Generated Text be Reliably Detected? University of Maryland, Harvard University. Retrieved from [AI Detection Study]

Blümer, S., Steinebach, M., Frick, R. A., & Bunzel, N. (2023). Detection of Deepfakes Using Background-Matching. TU Darmstadt, Fraunhofer SIT|ATHENE. Retrieved from [Deepfake Detection Study]

DECISION SCIENCES INSTITUTE

Leveraging Analytics & Modeling to Derive Insights for Point-based Loyalty Campaigns

Yifan Gu, Muskan Aggarwal, Sai Nithin Godi, Dhairya Dedhia, Harshal Amin, Nikitha Balaji, Rohan Ajay, Matthew A. Lanham
Purdue University Daniels School of Business
gu277@purdue.edu, aggar158@purdue.edu, sgodi@purdue.edu, ddedhia@purdue.edu, amin82@purdue.edu, balaji76@purdue.edu, rajay@purdue.edu, lanhamm@purdue.edu

ABSTRACT

This study examines how loyalty program design influences customer engagement and spending for a U.S. supercenter chain. We analyze the effectiveness of different offer types and delivery methods, controlling for pre-campaign behavior, to identify strategies driving incremental impact. Predictive models assess the ability of customer behavior to forecast campaign-period sales. Results indicate that per-visit incentives slightly outperform point multipliers, though differences are modest. A simple linear model performs comparably to more complex techniques, offering actionable insights for optimizing loyalty strategies. Future improvements could leverage granular segmentation and enriched data for enhanced program effectiveness.

KEYWORDS: loyalty programs, customer loyalty, predictive modeling, segmentation, retail marketing

INTRODUCTION

In the competitive retail landscape, loyalty programs are critical for customer retention and revenue growth. Shep Hyken (Forbes, 2025) notes that loyalty programs can boost customer spending by up to 20% through enhanced engagement. Deloitte (2025) further highlights that 72% of consumers are more likely to recommend brands with effective loyalty programs, emphasizing the need for personalized, value-driven experiences. Additionally, 65% of consumers prefer simple, accessible rewards, underscoring the importance of tailored benefits (WSJ, 2025).

Retailers globally leverage loyalty programs not only to retain customers but also to gather data for refining marketing strategies. Advanced analytics enable businesses to personalize offers based on consumer behavior, optimizing engagement strategies. This study explores the efficiency of a regional U.S. supercenter chain’s loyalty program, aiming to optimize offer characteristics to maximize engagement and drive behavioral change. The central research question is: How can loyalty program offers be optimized to enhance customer engagement and induce positive behavioral changes?

The analysis focuses on identifying which offer types and delivery methods most effectively drive engagement, distinguishing between transactional and incremental behavioral changes. Using data analysis and predictive modeling, we assess the impact of loyalty offers on sales, store visits, and department spending. Machine learning models are developed to forecast customer behavior, leveraging historical purchasing data.

This research integrates qualitative and quantitative analyses to evaluate the loyalty program's impact, offering actionable insights for program optimization. The findings contribute to retail marketing by demonstrating how data-driven decisions can improve business outcomes.

LITERATURE REVIEW

Customer loyalty programs (CLPs) are strategic tools for increasing retention and profitability. Point-based systems, prevalent in retail, hospitality, and financial services, reward customers with redeemable points for spending (Chen et al., 2021). Recent advancements in machine learning (ML) enable businesses to optimize CLPs by predicting customer responses and refining marketing strategies. This review examines the effectiveness of point-based loyalty programs, their impact on sales and profitability, and how predictive analytics can enhance loyalty strategies.

Effectiveness of Point-Based Loyalty Programs

Point-based programs encourage repeat purchases through delayed gratification, where customers accumulate points for rewards. Successful programs feature transparent redemption processes and meaningful rewards, driving engagement and spending (Gandomi & Zolfaghari, 2018). Conversely, complex structures, frequent point expirations, and high redemption thresholds can reduce effectiveness (Hofman-Kohlmeyer, 2016).

Loyalty programs also influence purchase behavior and price sensitivity. Enrolled customers become less price-sensitive, reducing the need for excessive discounting (Van Heerde & Bijmolt, 2005). However, their financial impact is mixed. While some studies show increased purchase frequency and lifetime value (Belli et al., 2022), others suggest they may not always yield sustainable profits (Daryanto et al., 2010; Meyer-Waarden et al., 2013).

Category-specific factors also play a role. For example, Lin & Bowman (2022) found that loyalty programs initially boost sales, but effects diminish over time. High-frequency categories like dairy sustain growth, while low-frequency categories like grooming items decline. This highlights the need for tailored strategies aligned with customer purchasing habits.

Behavioral mechanisms, such as the goal gradient effect (where customers spend more as they near a reward threshold), further drive engagement (Septianto et al., 2019). However, unredeemed points can lead to disengagement, emphasizing the need to balance aspirational rewards with attainable benefits (Gandomi & Zolfaghari, 2018).

Machine Learning for Predictive Loyalty Program Optimization

Traditional methods for evaluating loyalty programs, such as historical sales data and surveys, often lack predictive accuracy. ML models, however, enable data-driven optimization by forecasting customer responses and refining promotional strategies. For example, Usman et al. (2023) demonstrated ML's advantages in sales forecasting, with XGBoost outperforming other models like Linear Regression and ARIMA. ML allows businesses to segment customers based on purchasing behavior and tailor rewards, ensuring cost-effective and impactful promotions.

Sales Forecasting and Loyalty Program Performance

Sales forecasting is critical for loyalty program management, enabling businesses to anticipate demand fluctuations and adjust marketing efforts. ML models analyze transaction data to identify patterns traditional methods often miss (Usman et al., 2023). Predictive analytics can refine incentive structures by dynamically adjusting point accumulation and redemption values based on demand.

Lin & Bowman (2022) found that category penetration and purchase frequency are key to loyalty program success. Similarly, Verhoef (2003) emphasized targeting high-repeat-purchase segments. ML-driven price optimization further enhances loyalty programs by refining reward structures without unnecessary revenue loss (Van Heerde & Bijmolt, 2005). By integrating ML, businesses can move beyond traditional retention strategies, adopting data-driven approaches that maximize engagement and long-term profitability.

DATA

This study used a dataset comprising 400,662 distinct customer-level observations, including twelve variables capturing both pre-campaign and during-campaign customer behaviors. Each record is uniquely identified by a customer identifier and is assigned to one of five customer group categories. Behavioral features include total sales and store visits recorded separately for the pre-campaign (e.g., PreSales, PreTrips) and campaign periods (e.g., DuringSales, DuringTrips). Digital engagement is captured through the number of days customers visited the website, viewed offer pages, and claimed offers, also split into pre-campaign (e.g., PreWebVisitDays, PreOfferPageViewDays, PreClaimDays) and during-campaign (e.g., DuringWebVisitDays, DuringOfferPageViewDays, DuringClaimDays) metrics.

The dataset was constructed by merging three sources: customer group assignments, pre-campaign data, and during-campaign data, using the unique customer identifier. Initial validation confirmed the dataset was clean, with 4,259 rows of missing during-campaign values treated as zeros (i.e., no observed activity) after confirming with the retail company.

Customers were segmented into five groups based on combinations of offer delivery method, offer type, and campaign duration as shown in Table 1. Delivery methods included a newer, more informative format and a traditional coupon-style approach. Two groups received offers that rewarded additional points per qualifying visit, while another two received multipliers on base loyalty points earned per purchase. Four groups participated in a three-month campaign, while one group received a standard offer over a shorter, four-week period.

Table 1: Customer Treatment Groups

Group Number	Delivery Type	Offer Type	Duration
Group 1	New	PointsPerVisit	ThreeMonth
Group 2	New	PointsMultiplier	ThreeMonth
Group 3	Old	PointsPerVisit	ThreeMonth
Group 4	Old	PointsMultiplier	ThreeMonth
Group 5	Old	Normal Targeting	FourWeek

METHODOLOGY

Offer Effect Analysis

We first examined whether different combinations of offer types and delivery methods (captured through customer group assignments) impacted campaign-period spending. To isolate the effect of marketing treatments, we controlled pre-campaign sales (PreSales) using a multiple linear regression model. DuringSales was regressed on PreSales with one-hot encoded CustomerGroup indicators (Group 1 used as the baseline group). This allowed us to assess whether different offer delivery methods or offer types significantly influenced sales, holding historical purchasing behavior constant.

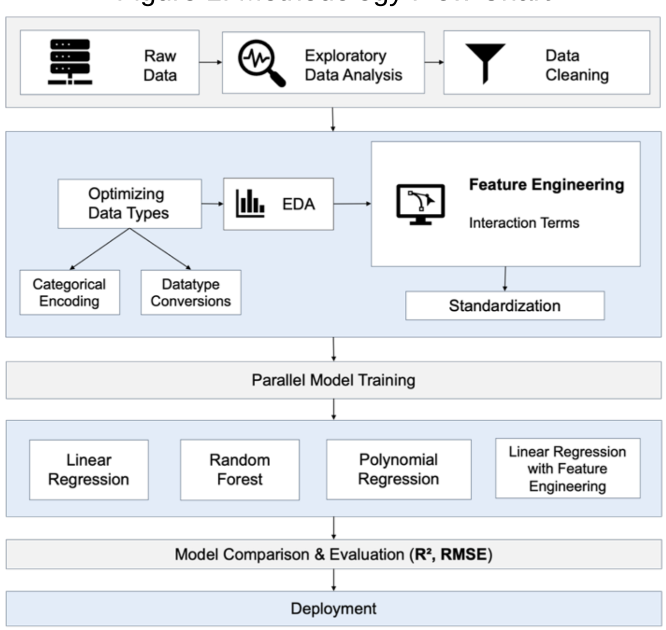
Modeling with All Pre-Campaign Features

We explored whether pre-campaign behavior alone could predict campaign-period sales using five pre-campaign predictors: PreSales, PreTrips, PreOfferPageViewDays, PreClaimDays, and PreWebVisitDays, reflecting in-store and digital engagement. Four modeling approaches were evaluated: linear regression, Random Forest, polynomial regression (Degree 2), and linear regression with interaction terms. Models were trained using an 80/20 train-test split, with performance assessed using RMSE and R^2 to balance complexity and predictive value.

Modeling with Best Selected Features

Beyond pre-campaign features, we expanded the feature set to include both pre- and during-campaign variables, identifying the most impactful predictors of campaign-period sales. The same four machine learning models are applied. Figure 1 represents the overall design of our explanatory analysis and predictive modeling:

Figure 1: Methodology Flow Chart



MODEL

We used a multiple linear regression (controlled offer group comparison) model as shown in Equation 1.

Equation 1: multiple linear regression used for our offer effect analysis:

$$DuringSales_i = \beta_0 + \beta_1 \cdot \boldsymbol{PreSales_i} + \Sigma k = 2^5 \boldsymbol{\beta_k} \cdot \boldsymbol{Group_k} + \epsilon_i$$

where:
- $\boldsymbol{PreSales_i}$ captures customer i's prior behavior
- $\boldsymbol{Group_k}$ are binary indicators for Customer Groups 2 through 5 (Group 1 as baseline)
- Coefficients $\boldsymbol{\beta_k}$ represent the adjusted difference in sales compared to Group 1

The variables were standardized for comparability and evaluated significance using p-values.

We developed the following additional models using pre-campaign features using the five selected pre-period variables:
- Linear Regression (OLS) to establish a simple, interpretable baseline.
- Random Forest (n = 100) to account for potential nonlinearities and interactions.
- Polynomial Regression (Degree = 2) to capture curvilinear relationships between predictors.
- Feature-Engineered Linear Regression, using second-degree interaction terms (without squared terms), to test for synergistic effects between variables.

The four models developed had the same structure to those with pre-period features, only that the model inputs were expanded to include pre- and during-campaign variables.

RESULTS

Offer Effect Analysis

The regression analysis shows in Table 2 that pre-campaign sales (PreSales) is the strongest predictor of campaign-period sales, while differences across customer groups were generally modest. Group 2 and Group 4 were found to have statistically significant negative coefficients, indicating lower sales compared to Group 1. Group 3 and Group 5 were not significantly different from Group 1. We also observed that per-visit offers (Groups 1 and 3) outperformed point multipliers (Groups 2 and 4), suggesting that rewarding specific shopping behaviors may be more effective than applying generalized point multipliers.

Table 2: Multiple Linear Regression Summary

Predictor	Coefficient	P-value	Interpretation
*constant	589.9122	0.000	-
Standardized PreSales	364.9537	0.000	-
Group 2	-2.8596	0.028	New Delivery: PointsPerVisit vs PointsMultiplier
Group 3	-1.1072	0.394	PointsPerVisit: New Delivery vs Old Delivery

Group 4	-3.1436	0.016	New Delivery PointsPerVisit vs Old Delivery PointsMultiplier
Group 5	-2.1214	0.081	New Delivery PointsPerVisit vs Old Delivery Normal Target

Model Comparison (Pre-Campaign Behavior)

Models using only pre-campaign behavioral features showed moderately strong predictive power (R^2 ~0.67, RMSE ~250) as shown in Table 3. The linear regression performed nearly as well as more complex models, stressing the effectiveness of simple, behavior-based predictors like store visits and digital engagement.

Table 3: Model Comparison with Pre-Campaign Features

Model	R^2	RMSE
Linear Regression	0.6761	251.1766
Random Forest (n=100)	0.6553	259.0880
Polynomial Regression (Degree=2)	0.6766	250.9664
Feature Engineered Linear Regression (2nd Degree Interactions)	0.6767	250.9433

Model Comparison (Best Selected Features)

The final models included three best features: pre-campaign sales, pre-campaign trips, and during-campaign trips. A summary of the model results is presented in Table 4. Polynomial Regression (Degree=2) was found to achieve the best R^2 and lowest RMSE, but improvements over Linear Regression were marginal. Increasing model complexity (e.g., higher-degree polynomials) was shown to degrade performance, suggesting linear relationships dominate.

Table 4: Model Comparison with Best Selected Features

Model	Predictors Used	R^2	RMSE
Best Incremental Linear Regression	DuringTrips, PreSales, PreTrips	0.8098	192.4681
Random Forest (n=100)	DuringTrips, PreSales, PreTrips	0.8102	192.2517
Polynomial Regression (Degree=2)	DuringTrips, PreSales, PreTrips	0.8245	184.8838
Feature Engineered Linear Regression (2nd Degree Interactions)	Same as LG + interactions	0.8131	190.8038

DISCUSSION

Measurable differences between groups were observed in the Offer Effect Analysis, but the effects were relatively small. Prior research suggests that loyalty program success depends on category-specific factors like purchase frequency and product type (Lin & Bowman, 2022), and not all customer segments respond equally to the same incentives (Verhoef, 2003). A more granular analysis at the retail department level is recommended to uncover where point multipliers or per-visit rewards are most effective. For example, high-frequency categories (e.g., grocery, produce) may sustain engagement better than impulse-driven ones. This aligns with the literature's call for predictive, category-aware program optimization.

Across both pre-campaign and final models, a simple linear regression model performed nearly as well as more complex alternatives. Despite testing nonlinear models (e.g., Random Forests, polynomial regression), no significant performance gains were observed. This suggests that, within the current dataset, the relationship between customer behavior and sales is largely linear, with past spending and store visits being the most predictive variables.

From a business perspective, the linear model offers a balance of accuracy, simplicity, and transparency, making it suitable for loyalty program planning, customer targeting, and internal forecasting. However, the analysis highlights a key limitation: model performance may be constrained by the available feature set. Advanced models like XGBoost and LSTM nerual networks may only outperform simpler methods when applied to richer datasets with granular behavioral and contextual variables (Usman et al., 2023).

Future research should expand the dataset to include additional attributes, such as offer interactions, temporal patterns, or external engagement indicators, to better leverage advanced machine learning techniques. For now, a well-tuned linear regression model remains a practical and effective solution, offering strong performance and interpretability for decision-makers.

CONCLUSION

This study explored how offer types, delivery methods, and customer behaviors influence sales during a loyalty campaign, with the goal of optimizing program design through data-driven insights. Our offer effect analysis showed that per-visit incentives slightly outperformed point multipliers, but the differences across groups were modest, highlighting that offer format alone does not drive substantial impact when prior behavior is accounted for.

We also built predictive models using both pre- and during-campaign behavioral data. Across all models tested, a simple linear regression model performed nearly as well as more complex alternatives, suggesting a largely linear relationship in the current dataset. Past spending and store visits emerged as the most influential predictors.

These findings are actionable for businesses aiming to enhance loyalty performance while keeping models interpretable and efficient. However, performance gains from more advanced models may require richer, more granular data, pointing to future opportunities in feature enrichment and segmentation at the category level to better tailor marketing strategies.

REFERENCES

Belli, A., Geyskens, I., Dekimpe, M. G., & Steenkamp, J.-B. E. M. (2022). 40 years of loyalty programs: How effective are they? *Journal of the Academy of Marketing Science, 50*(1), 147–173.

Chen, Y., Fay, S., & Wang, Q. (2021). Three decades of research on loyalty programs: A literature review and future research agenda. *Journal of Business Research, 124*, 179–197.

Daryanto, A., de Ruyter, K., Wetzels, M., & Patterson, P. G. (2010). Service firms and customer loyalty programs: A regulatory fit perspective of reward preferences in a health club setting. *Journal of the Academy of Marketing Science, 38*(5), 604–616.

Deloitte. (n.d.). *Transcending expectations in the new loyalty landscape.* Deloitte. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-transcending-expectations-in-the-new-loyalty-landscape.pdf

Deloitte. (2024, November 19). What loyalty program benefits do consumers want? *The Wall Street Journal.* https://deloitte.wsj.com/cmo/what-loyalty-program-benefits-do-consumers-want-4e0a6511

Gandomi, A., & Zolfaghari, S. (2018). To tier or not to tier: An analysis of multitier loyalty programs' optimality conditions. *Omega, 74*, 20–36.

Hofman-Kohlmeyer, M. (2016). Customer loyalty program as a tool of customer retention: Literature review. *CBU International Conference Proceedings, 4*, 202–206.

Hyken, S. (2025, February 9). The hidden revenue engine: Loyalty programs are more than just repeat business. *Forbes.* https://www.forbes.com/sites/shephyken/2025/02/09/the-hidden-revenue-engine-loyalty-programs-are-more-than-just-repeat-business

Lin, C., & Bowman, D. (2022). The impact of introducing a customer loyalty program on category sales and profitability. *Journal of Retailing and Consumer Services, 64*, 102769.

Meyer-Waarden, L., Benavent, C., & Castéran, H. (2013). The effects of purchase orientations on perceived loyalty programmes' benefits and loyalty. *International Journal of Retail & Distribution Management, 41*(3), 201–225.

Septianto, F., Sinha, R. K., & Sadarangani, P. (2019). The similar versus divergent effects of pride and happiness on the effectiveness of loyalty programs. *Journal of Business Research, 99*, 12–22.

Usman, S., Malik, K., Rauf, A., & Zahid, A. (2023). Predictive analysis of retail sales forecasting using machine learning techniques. *ResearchGate.*

Van Heerde, H. J., & Bijmolt, T. H. A. (2005). Decomposing the promotional revenue bump for loyalty program members versus nonmembers. *Journal of Marketing Research, 42*(4), 443–457.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing, 67*(4), 30–45.

DECISION SCIENCES INSTITUTE

Detecting and Predicting Phantom Inventory Using Random Forest and LSTM Models

Hung-Chen Hsu, Kratika Jain, Da Fang Lin, Yiran Liu, Kaushalya B. Naidu, Namitha Ramakrishnan, Rohan Ajay, Matthew A. Lanham
Purdue University Daniels School of Business
hsu367@purdue.edu, jain784@purdue.edu, lin1945@purdue.edu, liu387@purdue.edu, knaidu@purdue.edu, ramakr40@purdue.edu, rajay@purdue.edu, lanhamm@purdue.edu

ABSTRACT

Phantom inventory refers to discrepancies where inventory exists in records but not in reality, which can arise due to a variety of factors, such as misplacement, system glitches, theft, shrinkage, and inaccurate data entry. This paper will focus on exploring real business inventory data and modeling on detection and prediction. We aim to develop powerful machine learning models, random forest, time series, and clustering, by analyzing transaction data, stock level movement patterns, and sale anomalies to uncover hidden discrepancies and eventually manage to predict stockouts. Detecting and predicting phantom inventory case is crucial to optimize business performance and enhance profitability.

KEYWORDS: Phantom inventory, Detection, Prediction Random Forest, LSTM

INTRODUCTION

In the dynamic landscape of retail business, phantom inventory, referred to as ghost inventory or stockout, has been a problem for decades. Nowadays technological advances achieve customer convenience by offering online shopping and hence change shopping behavior. These changes result in new challenges for inventory management and supply chain operation. It is crucial to maintain accurate inventory records for both operational efficiency and customer satisfaction and is the driving motivation for our research.

In this paper, phantom inventory is defined as the merchandise that appears to be available in a company's records, but is not physically present on the shelves, indicating a discrepancy between the recorded inventory and actual stock. This issue can arise from various factors, including theft, misplacement, data entry errors, and system inaccuracies. When phantom inventory or stockout occurs, retailers will find that inventory records inaccurately reflect stock level, which lead to several challenges.

Loss of sales: Riaz (2021), points out that retailers worldwide lost more than $1.8 trillion annually due to inventory distortion issue. As one may imagine, offline retail accounts were the majority of this number. In the United States, a 2018 survey by the National Retail Federation (NRF) which is cited in by McCue (2019) reported a loss of $46.8 billion caused by inventory shrinkage.

Customer dissatisfaction: If a customer finds a product listed as available on a retailer's website, but encounters an empty shelf within store, this can lead to customer frustration and loyalty loss. According to Crowe (2023), after software provider, Retail Insight, surveyed more than 1000 U.S. shoppers, it was found that phantom inventory is negatively impacting sales and long-term loyalty.

Increased operational costs: Phantom inventory forces employees to take additional time searching for nonexistent merchandise and rectifying inventory records, largely and deeply diminishing business productivity and efficiency. As Paul Boyle, CEO of Retail Insight, pointed out, “Facing growing pressure on already razor-thin margins, phantom inventory risks becoming a profit-draining issue across every part of the retailer’s store operations, from wasted labor to out of stocks and increased shrink.” This comment was cited in the article by Mayer (2023) as well.

As we see from the past decade, phantom inventory continues to remain a persistent challenge in the retail industry. This issue leads to revenue loss, dissatisfied customers, and inefficiencies in operation and inventory management. Traditional inventory audits and manual checks fail to provide timely and accurate insights and forecasting into stock discrepancies, necessitating more advanced analytical approaches. With the rise of data-driven solutions, machine learning techniques such as time series forecasting, random forest, clustering, and other methods of anomaly detection offer potentials for addressing phantom inventory. By leveraging these techniques, it is now possible to accurately detect out-of-stock, effectively reduce revenue loss, and hence improve overall inventory accuracy.

In this study, we define and detect phantom inventory using key inventory features such as the daily balance on hand (BOH), which is recorded at the end of each day, and the ground truth label captured hourly by cameras. The primary objectives are to detect discrepancies between the BOH and the label data, and to predict the likelihood of out-of-stock in the future timeframe. By analyzing the discrepancies between inventory records and actual stock levels, we attempt to uncover patterns and trends that may signal phantom inventory. Time series models are explored to capture fluctuations in stock levels over time, allowing us to forecast potential stockouts before they occur. Additionally, we utilize random forest and clustering techniques for classification of phantom inventory merchandise, helping retailers understand which products or even locations are more prone to inaccuracies.

The integration of multiple modeling approaches enables a comprehensive analysis of phantom inventory. Ultimately, we aim to not only contribute to reducing loss of sales, improving customer experience and achieving operational excellence, but also provides a scalable solution that can also be adapted to various and broader industrial environments.

LITERATURE REVIEW

There are various definitions of academic phantom inventory in the academic literature. In retail operations, phantom inventory refers to inventory that appears in the records but is not actually available on the shelf. In other words, the system shows that an item is in stock when physically it is missing. Bassamboo et al. (2020) defines phantom inventory as “units of [a] good not available for sale” even though they are recorded in the inventory system. Farias et al. (2024) similarly describe phantom inventory events as instances where “inventory is recorded to be present on-shelf while in reality it is missing,” often due to causes like shrinkage or theft.

This concept is sometimes also called “ghost inventory” in practice, highlighting that the stock exists only in the computer records and not in the store’s actual stock. All these definitions emphasize a discrepancy in which recorded on-hand quantities exceed the true physical inventory, leading the system to falsely assume items are available. Such discrepancies are a well-known form of inventory record inaccuracy in retail (DeHoratius & Raman, 2008) and have been reported to be widespread. Studies have found that 50–65% of inventory records can contain errors, with phantom inventory (overstated stock) being a common outcome (Farias et al., 2024; Bassamboo et al., 2020). This phenomenon can significantly undermine operations, as it causes stockouts that are invisible to the inventory system – the shelf is empty but the system doesn’t trigger replenishment because it “thinks” products are still there (Rekik et al., 2019). In fact, undetected phantom inventory has been estimated to account for substantial lost sales (on the order of ~4% of annual revenue in the retail industry) because customers encounter empty shelves despite the system showing stock (Farias et al., 2024).

Since phantom inventory cannot be observed directly in systems until a correction is made, researchers have developed data-driven methods to detect or infer its occurrence from sales and inventory data. A key insight used in academic studies is that a phantom inventory often manifests as prolonged zero-sales periods for an item that is supposedly in stock. Tripathy (2019) explains that a “phantom inventory remains unnoticed unless an intervention (e.g., a shelf audit) occurs,” and proposes a detection approach based on consecutive days of zero sales while the inventory-on-hand (IOH) remains positive. In this Bayesian model, if a product shows no sales for an extended stretch of time despite the system indicating stock on hand, it likely signals a phantom inventory situation (Tripathy, 2019). The logic is that if the item were truly on the shelf, some sales would normally occur given underlying demand. This approach models daily demand probabilistically (e.g., using a negative binomial distribution) to estimate the likelihood that zero-sales over X days could happen if the inventory were actually there. A very low likelihood (i.e., unexpectedly long sales drought while stock is recorded) points to phantom inventory with high probability.

More recent data-driven methods cast the phantom inventory detection as an anomaly detection problem. Farias et al. (2024) treat the pattern of sales across stores as a matrix and look for anomalies indicating that a particular store-SKU combination has abnormally low sales compared to expectation, given that it should have stock. In essence, they look for when the “rate of transactions has slowed or stopped” for an item, which, coupled with a known positive inventory record, suggests a phantom inventory event (Farias et al., 2024). Their approach uses cross-sectional data (i.e., multiple stores) to detect these anomalies more robustly, framing it as identifying outliers in a low-rank Poisson sales matrix (Farias et al., 2024). This and similar AI or algorithmic methods allow retailers to flag likely phantom inventory instances without waiting for manual audits. The common thread in these data-driven approaches is leveraging point-of-sale (POS) data and inventory records over time: when POS data show zero sales for a SKU that the inventory system lists as in-stock, an alert for phantom inventory can be raised (Tripathy, 2019). Such techniques have been shown to significantly improve detection, enabling timely corrective actions (like triggering a physical count or restocking) to reconcile the records (Tripathy, 2019; Farias et al., 2024).

DATA

In collaboration with a national retailer, we were granted to access to four different dataset categories. Namely, "consumables", "dairy", "dry_grocery", and "frozen" from one of the store locations for the month of November 2024, structured into three CSV files for each category. Each category contains features-daily_2024-11.csv ("Features"), labels_2024-11.csv ("Labels"), and trxn_2024-11.csv ("Transaction"). The Features dataset (39 attributes) mainly captures product-level information, including various balance-on-hand (BOH) features and product sales data. We develop machine learning models to identify important features like product names, sales information, in-store location, and inventory level data (BOH). The Labels dataset (34 attributes), of which the stock level data is collected by the in-store shelf camera system on hourly basis, provides features such as out-of-shelf occurrences and their corresponding timestamps. Lastly, the Transaction dataset (17 attributes) provides transaction data that can help to identify potential demand level. These datasets offer a structured view of inventory flow, aiding in the detection and prediction of phantom inventory events.

METHODOLOGY

A structured approach is employed to detect and predict phantom inventory. We create two phases for the modeling. In Phase 1 we focus on detection of phantom inventory, while in Phase 2 we focus on prediction of product stockout day.

Phase 1: Detection of Phantom Inventory – Pre-processing and Exploratory Data Analysis

The dataset was given to us by one of the major players in U.S. grocery retailer. It consisted of three important tables that included Features, Labels and Transactions. To start EDA, we preprocessed the data by handling missing values. For feature engineering, K-means clustering was used to group the products on the basis of stock fluctuation patterns. Key variables such as SKU, BOH, transaction quantities and historical sales were selected.

Phase 1: Detection of Phantom Inventory – Model Building

After EDA, a random forest model was found to detect the phantom inventory by classifying the SKUs which have stock discrepancies. Many classification models were explored such random forest, CatBoost, XGBoost, and LightGBM. Random forest model was selected due to its robustness and interpretability. The cross-validation design used involved an 80-20% train-test split.

Phase 2: Prediction of Product Stockout Day – Pre=processing for LSTM-based Model

In Phase 2, we would adopt an LSTM-based forecasting approach for our predictive model. Since the dataset comprises daily inventory levels (DailyBOH) for each SKU, indexed by date, the preprocessing steps include the following two main steps: outlier capping and log transformation. We used outlier capping where the DailyBOH values were capped at the 99th percentile to limit the influence of extreme outliers, ensuring model stability. Log transformation was employed to cap the DailyBOH using a log1p transformation (np.log1p) to address skewness and zero-inflation. This created a more normalized distribution suitable for LSTM modeling.

2.2 Sliding Window Preparation

A sliding window of 7 days was used to create input sequences (X) and corresponding targets (y). For univariate forecasting, X contains only log-transformed DailyBOH; for multivariate forecasting, it includes both DailyBOH and transaction quantities. The data is scaled using either

MinMaxScaler (range 0-1) or RobustScaler, with the latter as an option for robustness against outliers.

MODELS

To build models for detecting phantom inventory, we required an implementation-based definition that will be carefully expounded in the remainder of this paper. We formulated a structured definition of phantom inventory using the following logic. Phantom inventory is identified by the coexistence of two conditions:

1. **Recorded On-Hand Stock > 0:** The inventory management system indicates that a given product is in stock (a positive quantity is recorded for that SKU).
2. **No Sales Over a Prolonged Period:** Despite the recorded stock, the point-of-sale data show zero sales for that product over a significant time frame (e.g., several consecutive days or weeks).

When both conditions hold, it implies the item is likely not actually present or available to customers, hence qualifying as phantom inventory. Essentially, the system "thinks" the item is on the shelf, but customers are not buying any – strongly suggesting the shelf is empty (the inventory exists only in phantom form). This definition reflects the operational logic used to flag phantom inventory in data-driven systems (Tripathy, 2019). It is consistent with academic characterizations: for instance, Bassamboo et al. notes that phantom inventory arises when products are "not available for sale" even though they appear in the records (Bassamboo et al., 2020), and Tripathy's detection rule captures exactly that scenario of zero sales with positive stock (Tripathy, 2019).

In summary, **phantom inventory = recorded stock – actual stock**, when the recorded stock is erroneously higher and items that are supposed to be "there" are not actually on the shelf (Tripathy, 2019; Farias et al., 2024). This explicit definition guided by the implementation logic is supported by the literature and helps ensure that our understanding of phantom inventory is both conceptually sound and practically measurable.

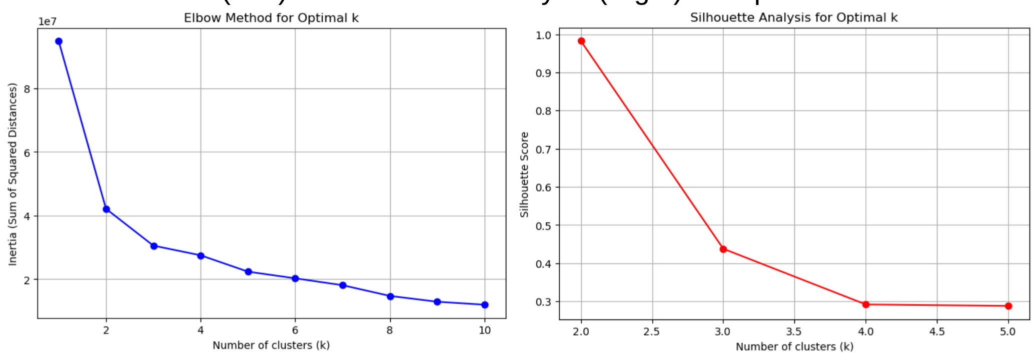
The key feature in the dataset is called "DailyBOH" (daily balance on hand), defined as "final recorded stock at the end of the day". Another relevant feature that will be considered for the models is "CurrentBOH" (current balance on hand) defined as a mid-day snapshot or system-generated number. As the daily final records, "DailyBOH" is more reflective of actual stock levels, and it will be prioritized in the modeling. On the other hand, "CurrentBOH" still contributes some value data to the model, so we are not removing it completely but only use it when it is needed.

We create a missingness indicator for the purpose of handling missing values that exist in the dataset. The idea is that we label the missing entry as "missing" (by missingness indicator) and the machine learning model will learn from this indicator. Our national retail collaborator, has conducted powerful demand forecasting analysis and the data was provided in the dataset. Hence, we create sales gap feature to measure the difference between the predicted demand and the actual sales quantity.

K-means Clustering Model

Our first model is a K-means clustering. The goal is to further group the products with similar behaviors in supply and demand dynamics. Using a naïve elbow plot approach, we find k=2 and k=3 are strong candidates. Moreover, with the help of Silhouette Analysis, we discover that k=2 has the highest silhouette score (near 1.0), which may indicate one super-small cluster (e.g., just a handful of points) and one giant cluster. This partition might not be very informative. Meanwhile, k=3 whose silhouette score is around 0.45 has a reasonable drop in inertia on the elbow plot, indicating that it is a natural choice of the number of clusters and that it is a possibly more balanced partition as shown in Figure 1. In many practical cases, a silhouette score of 0.4–0.6 is considered okay if it produces more interpretable clusters. And this makes sense since the target variable is binary, 0 or 1 (phantom inventory flag). Considering that we also want a bit more granularity (e.g., “low,” “medium,” “high”) for the model, we will use k=3 for clustering the products.

Figure 1: Elbow Method (Left) and Silhouette Analysis (Right) for Optimal k

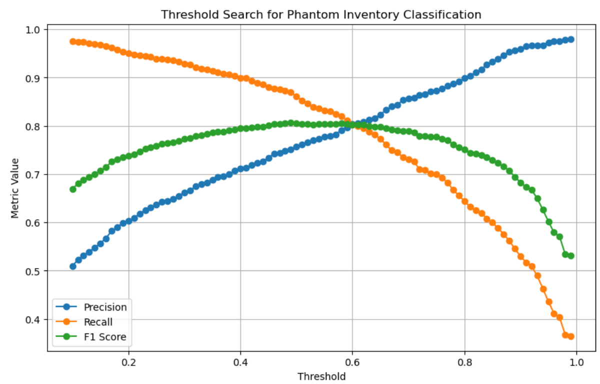


Random Forest Model

Based on the clusters, our second model was a Random Forest that will be used to classify phantom inventory and non-phantom inventory. Prior to building the random forest model, we tested a number of boosting algorithms which turned out to perform much worse than random forest. All of these algorithms resulted in precision, recall, and F1 scores strictly below 0.8, while random forest model outperformed all of them by giving all of the three evaluation metrics above 0.8.

We explored time sequential splitting techniques for our dataset which is inherently in time sequence. Counterintuitively, the random splitting turned to be a better choice. Thus, we randomly split the data into 80/20 for training-test set evaluation. We used the cross-validation process as an internal validation step. This approach saved us from having to explicitly carve out a separate validation set, especially when data provided was limited. The best threshold search for phantom inventory classification was found to have a threshold of 0.60 which provided a reasonable trade-off for balancing precision and recall. This is depicted in Figure 2.

Figure 2: Threshold Search for Phantom Inventory Classification



The results of random forest will be presented in the next section.

Long Short-Term Memory (LSTM) Networks

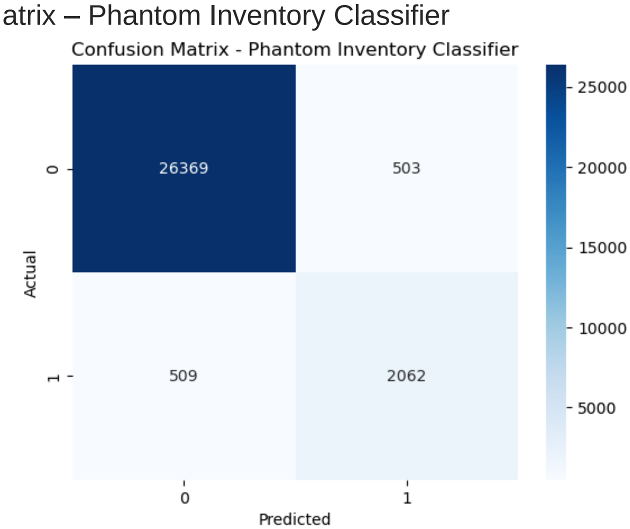
The LSTM time series model was developed and the configuration implemented was as follows. A single-layer LSTM comprised of 50 units with ReLU activation, followed by a dense layer outputting a single value. This was tested against a stacked LSTM configuration by adding a second LSTM layer (30 units) with return_sequences=True, enhancing the model's capacity to capture complex patterns. From our exploration we retained the single-layer LSTM as our final selection because it was more stable than stacked LSTM. Since this dataset is not that complicated for time series, the single layer performs well.

RESULTS

Phantom Inventory Detection

For phantom inventory detection, the result of random forest model is reasonable, as shown in the Figure 3 confusion matrix. The phantom inventory event is marked as Class 1. Precision is the proportion of detected phantom inventory events that are truly phantom inventory events, i.e., it tells how many of the phantom inventory alerts are correct, whereas recall is the proportion of actual phantom inventory events that are correctly identified.

Figure 3: Confusion Matrix – Phantom Inventory Classifier



Precision for phantom-inventory class (Class 1): 0.8039 while recall for Class 1: 0.8020. In phantom inventory detection, both precision and recall are important metrics, but their relative importance depends on the consequences of false positives and false negatives. If the retailer prefers to minimize the impact of false positives (e.g., to avoid unnecessary actions), then a high precision is needed. However, if failing to identify a phantom inventory is costly, then a high recall is required. We note that there is a trade-off between the two metrics, and hence we choose to put a result for both in a balanced way.

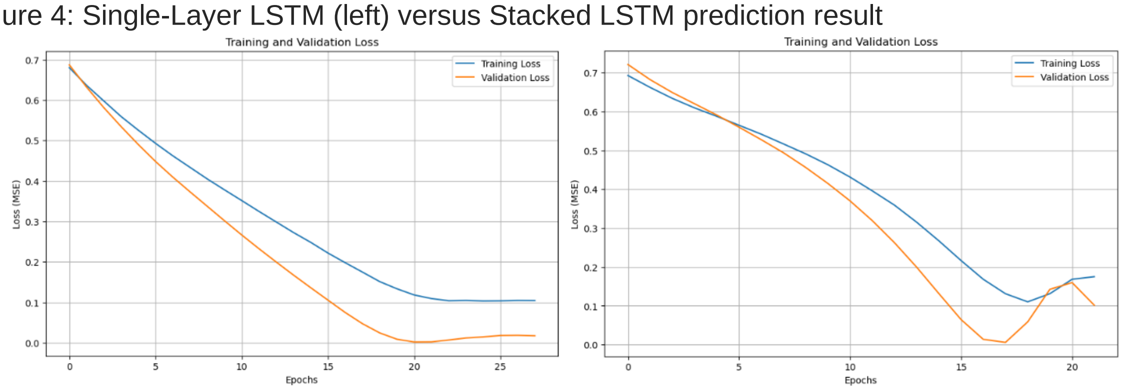
Phantom Inventory Prediction

The LSTM model predicts the next day's inventory iteratively over a 7-day horizon. Starting with the last observed window, each prediction updates the input sequence by shifting the window and appending the forecast. For multivariate inputs, additional features are held constant at their last observed values. Predictions are inverse transformed to return to the original DailyBOH scale.

We were able to predict OOS in the following way. The first day when forecasted DailyBOH falls to zero or below is identified as the out-of-stock date. If no such event occurs within the horizon, no run-out date is recorded. The predictive modeling performance is assessed using MSE on the original DailyBOH scale, comparing LSTM forecasts to a naive baseline (repeating the last observed value). The model's training loss (scaled MSE) is also monitored.

The performance of LSTM model is presented by the graph of training vs. validation loss in Figure 4. We include both Single-Layer LSTM and Stacked LSTM results for reference.

Figure 4: Single-Layer LSTM (left) versus Stacked LSTM prediction result



CONCLUSIONS

In this study, we defined and detected phantom inventory using key inventory features such as the daily balance on hand recorded at the end of each day, and the label data captured hourly by camera system. Furthermore, we were able to predict the likelihood of out-of-stock occurrences for the month of November 2024. In the end, we managed to detect and predict phantom inventory with the help of random forest and clustering techniques for categorizing merchandise, and time series models for forecasting the potential stockouts.

Beyond the used approaches in our study, other machine learning models may be investigated to improve the accuracy of phantom inventory detection. Supervised learning techniques may help to classify stock discrepancies based on historical data, while unsupervised learning techniques can help detect hidden patterns in inventory inaccuracies. We wish that our study not only contributes to reducing loss of sales, improving customer experience and achieving operational excellence, but also provides a scalable solution that may be adapted to various and broader industrial environments.

REFERENCES

Bassamboo, A., Moreno, A., & Stamatopoulos, I. (2020). *Inventory auditing and replenishment using point-of-sales data*. Production and Operations Management, 29(5), 1219–1231. DOI: 10.1111/poms.13153

Crowe, E. (2023, Oct 26). *Why Retailers Must Address Phantom Inventory Issues*. Progressive Grocer. https://progressivegrocer.com/why-retailers-must-address-phantom-inventory-issues

DeHoratius, N., & Raman, A. (2008). *Inventory record inaccuracy: An empirical analysis*. Management Science, 54(4), 627–641. DOI: 10.1287/mnsc.1070.0814

Farias, V. F., Li, A. A., & Peng, T. (2024). *Fixing inventory inaccuracies at scale*. Manufacturing & Service Operations Management, 26(3), 1102–1118. DOI: 10.1287/msom.2023.0146

Fleisch, E., & Tellkamp, C. (2005). *Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain*. International Journal of Production Economics, 95(3), 373–385. DOI: 10.1016/j.ijpe.2004.02.003

Mayer, M. (2023, Oct 26). *Phantom Inventory Haunts Retailers: Study*. Supply & Demand Chain Executive. https://www.sdcexec.com/warehousing/retail/news/22877459/retail-insight-phantom-inventory-haunts-retailers-study

McCue, T. (2019, Jan 31). *Inventory Shrink Cost The US Retail Industry $46.8 Billion*. Forbes. https://www.forbes.com/sites/tjmccue/2019/01/31/inventory-shrink-cost-the-us-retail-industry-46-8-billion/

Rekik, Y., Syntetos, A., & Glock C. (2019). *Inventory Inaccuracy in Retailing: Does it Matter?* ECR Community Shrinkage & On-shelf Availability Group. https://www.cardiff.ac.uk/__data/assets/pdf_file/0003/1729236/84486-ECR-Inventory-Inaccuracy-Report-v12.pdf

Riaz, O. (2021, Nov 15). *Amazon Inventory Management: Solutions To Common Problems*. Forbes. https://www.forbes.com/councils/forbesagencycouncil/2021/11/15/amazon-inventory-management-solutions-to-common-problems/

Tripathy, S. (2019). *Detection of phantom inventories at retail stores using a Bayesian approach* (Master's thesis, Massachusetts Institute of Technology). https://dspace.mit.edu/bitstream/handle/1721.1/122257/1119555264-MIT.pdf

DECISION SCIENCES INSTITUTE

Mapping the Path to Translations: Empowering SIL to Reach New Frontiers in Language Translations

Vandana Brahmasa, Amy Fischbach, Parvathy Nair, Gitanjali Nambiar, Aryan Ritesh, Huilin Xu, Rohan Ajay, Matthew A. Lanham
Purdue University Daniels School of Business
vbrahmas@purdue.edu; fischba@purdue.edu; nair221@purdue.edu; nambiarg@purdue.edu; aritesh@purdue.edu; xu1997@purdue.edu; rajay@purdue.edu; lanhamm@purdue.edu

ABSTRACT

We developed a mapping tool for SIL International, a global leader in language translation, to identify unmet translation needs and guide market expansion efforts. The motivation for this research is that despite global connectivity, many low-resource languages lack Bible translations; SIL seeks to address this gap using AI and data science. We design and build a mapping tool with Python and HTML which identifies underserved areas, integrates LLMs to help identify barriers, and offers insights applicable to both SIL and broader global outreach initiatives.

KEYWORDS: Language Translation, Low-Resource Languages, Market Expansion, Large Language Models, Data-Driven Mapping

INTRODUCTION

In today's rapidly evolving world, information flows smoothly, aided by advanced technologies and global communication networks. Despite these advancements, certain communities remain underserved due to the absence of resources in their native languages. This disparity becomes especially noticeable regarding vital documents such as legal guidelines, educational materials, and religious or faith-based texts. According to a survey done by Wycliff, only 10% of the languages have a fully translated Bible. Religion is an integral part of many individuals' day-to-day lives, highlighted in many of the decisions they make and comfort they feel. For Christians, connecting to the Word of God is an important aspect of their faith. A study published by Jonas S. Thinane highlights how important Bible translation is to the mission of the church and their community. When entire communities lack access to foundational content, they face barriers in literacy, cultural preservation, and participation in broader social dialogues. One significant challenge lies in prioritizing which languages should receive translated resources first, given the constraints of both financial and human capital. A major nonprofit organization, SIL International, seeks to maximize the impact of its translation efforts. SIL focuses on bridging the divide by bringing key texts, particularly foundational religious materials, to communities lacking translations.

The business problem in our study centers on identifying which language communities should be prioritized to ensure the most beneficial and timely return on investment. Many factors influence this prioritization. Geographic challenges, for instance, can make travel and on-site work prohibitively expensive. Similarly, languages that are closely related may allow for quicker translation adaptations, while those that are more distinct may require a full translation project from the ground up. Additionally, demographic indicators, such as the size of a language community or its socioeconomic conditions, play a role in determining both feasibility and long-term benefits. By assessing these interrelated variables, planners can arrive at more strategic decisions about where to focus immediate resources. Many companies, including SIL, seek a systematic, data-driven approach to guide their strategic planning. Decision-makers must also be able to visualize relationships between communities, calculate travel costs, evaluate linguistic distance, and weigh social or cultural impact. Graph-based analytics tools, such as the isochrone graph, offer a potential solution, as they can integrate disparate data sources into a single computational framework. In such a system, individual language communities become nodes in a network, while edges between them can represent everything from geographic proximity to linguistic similarity. This model allows for the calculation of "distances" in multiple senses—whether physical, linguistic, or socio-cultural—and thus enables sophisticated analyses.

Our primary research question is: How can a flexible, graph-based planning tool integrate these factors to help an organization allocate its limited resources more effectively? While the resulting strategic planning tool focuses on our industry collaborator, SIL, the design and technologies can be extended to other markets – retail, manufacturing, etc. All firms need the ability to make better decisions today but also adapt to future changes in travel routes, population shifts, and technological breakthroughs. The creation of this tool is not only to serve the specific needs of SIL, but also to be a framework for modifying it to address various other business needs as well. Isochrones maps are maps that depict the area accessible from a point within a certain time threshold, and are being utilized by companies in a variety of industries to address their geographic concerns. To address this business need, this study proposes the development and deployment of related geo-maps. Our tool aggregates information on travel time, cost estimates, linguistic distances, and demographic metrics in a way that is both dynamic and scalable. The goal is to present findings through intuitive data visualizations and dashboards that allow non-technical stakeholders to grasp complex networks and support data-driven decisions.

LITERATURE REVIEW

Bible Translation: History and Benefits

Bible translation (BT) into local languages and dialects has been a significant preoccupation of missionary work and has played a significant role in revitalizing distinct linguistic traditions. BT has seen various phases of evolution, from being an endeavor in word-for-word equivalence to an exercise in capturing dynamic equivalence with diverse languages (Naude, 2005). The Third Generation of Bible translations has expanded focus to incorporating oral traditions in its translations, which are seen to be more palatable to some communities. According to Naude, subsequent scripture translation should be motivated by balancing the need for amenability to chanting with literary quality and explainers on cultural context. Scripture translation has been accompanied by myriad socio-economic transformations in the communities it has been carried out, markedly in the spheres of education, literacy and health outcomes.

Barriers to Translation

Language translations are mediated by several factors, the most fundamental among them being the impact of geography on translation practices. Geographical contexts form the backdrop of development of different languages and linguistic traditions, which in turn pose a challenge in interpretability of translated works. Effective translation is thus an endeavor in balancing the source content's message with the target audience's interpretations, often through personalized metaphors rooted in the target audiences' cultural context (Ke, 2019).

Beyond the linguistic challenges in translation, there are practical socio-economic, cultural and political factors that mediate access to a region to allow translation efforts to flourish. The Skopos theory (Du, 2012) of language translation suggests that translation is a goal-driven activity primarily guided by the translator's intent.

Spatial Visualization Techniques

Isochrones can offer an intuitive way to visualize the regions that have available translations and the degrees to which translations are available. Isochrones are lines that connect points of equal travel time (Desai, 2008). In other words, it shows us the regions that can be reached in the same amount of time. Isochrones allow us to visualize the accessibility of a region and can be enriched with contextual inputs like mode of transport, speed, infrastructure available to glean insights into the accessibility of a location.

In addition to isochrone mapping, this paper utilizes Voronoi diagrams for visualizations of regions with and without scripture translations. Voronoi diagrams geometrically divide a plane into different regions based on their distance from 'seeds'- given points on the map. A Voronoi diagram divides the plane into several cells that enclose a region closest to each seed (DeLorenzo, 2021). For purposes of this study, the regions which have scripture translations available may be considered 'seeds' of the Voronoi cell. Each Voronoi cell would thus represent the wider region in which the language is spoken, a geospatial data point that can guide not only scripture translation efforts, but other region-determined endeavors like community building, business operations and expansion, tourism and aid work.

DATA

Bible translation data used in this data came from publicly available data at Progress.Bible (https://progress.bible/data/). As shown in Table 1, this dataset contained information on various regions and languages, along with the status and date of Bible translations available for each.

The ProgressBible dataset captures key details about Bible translation efforts across different languages and regions. It includes the country name (Country) and the language spoken (Language Name), along with the size of the language population (Population Group) and the stage of life for the language (Language Vitality). It also specifies which pieces of the Bible have been translated (Scripture), when it was published (Year Scripture Published), and whether there is an active Bible translation underway (Active Translation).

This data served as the foundation for our research and allowed us to assess the extent of Bible translation efforts across different linguistic and geographical contexts. By analyzing these attributes, we aimed to identify patterns in recent translation activity, gaps in coverage, and potential areas where additional translation efforts might be needed.

Companies will often use additional datasets to improve transparency from other angles. We integrated additional information from external sources, including geographic coordinates (latitude and longitude) for each region. For accurate location data, we utilized Glottolog (https://glottolog.org/), a comprehensive linguistic database that provides detailed information on the world's languages, including their geographic distribution. Glottolog is maintained by the Max Planck Institute for Evolutionary Anthropology and follows rigorous scholarly standards, ensuring the accuracy and reliability of its linguistic and geographic data. This expansion facilitated geo-based mapping, allowing for more precise visualization and assessment of translation distribution.

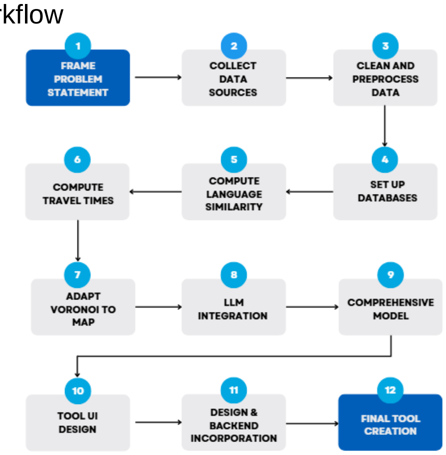
To estimate distance and travel time between locations, we incorporated the OSRM API (https://project-osrm.org/). Adding the Ease of Doing Business Index published by the World Bank Group added another intelligence layer (World Bank, 2020). This index evaluates the performance of different countries relative to each other on metrics like ease of starting a business, enforcing contracts, access to electricity, etc. By combining linguistic and geographical data, our tool has more potential to provide comprehensive insights about the accessibility and reach of Bible translations worldwide.

METHODOLOGY

To create a visualization tool for mapping locations with and without scripture translation, we first conducted market research to locate various sources of open-source data that could be utilized for this mapping tool. A rigorous review of literature was conducted to identify barriers to entry for Bible translations, as well as strategic and logistical challenges in accessing different linguistic regions. After narrowing down our sources of data, we proceeded to clean and preprocess the data to make it usable for analysis.

For language distance, we computed the similarity of languages with each other and crystallized it to a language similarity score. Our approach leverages lang2vec, which is built on the URIEL database, a large-scale linguistic resource that is structured compendium of information on language typology (https://aclanthology.org/E17-2002/). The Lang2vec library provides language embeddings based on these features allowing computational comparisons across languages. Travel times were computed, i.e. the time it takes to travel to a certain region using the OpenSky (https://openskynetwork.github.io/opensky-api/rest.html) and OSRM APIs. With these contours of our mapping tool in place, we created the necessary composite database to be utilized by the tool. Figure 1 below shows our step-by-step methodological workflow.

Figure 1: Methodological workflow



GEO-MAPPING TOOL DEVELOPMENT

To enhance the visualization of linguistic regions serviced by translation efforts, we employed a Voronoi-based geospatial mapping approach. This method partitions geographical areas based

on linguistic distribution, effectively delineating regions where translations are available and those that remain underserved. Given the global scope of translation expansion, we opted for a world map visualization, ensuring a comprehensive representation of translation accessibility across diverse regions. Additionally, we prioritized user accessibility, incorporating an interactive feature that allows users to input a language of interest, which serves as the reference point for analysis. The core functionality of the tool revolves around identifying the nearest hub language, defined as the geographically closest language to the input language that possesses an existing translation. To facilitate intuitive interpretation, we employed a color-coded classification system, highlighting gaps in translation coverage, as we aim to reveal regions with high linguistic accessibility versus those still in need of translation efforts.
We also sought to enhance the analytical capabilities of the tool by integrating external data sources to provide a more comprehensive assessment of translation accessibility. Apart from external APIs, we also used church location coordinates, sourced from OpenStreetMap (https://www.openstreetmap.org) (OSM), as religious institutions often play a pivotal role in translation dissemination and linguistic accessibility. The inclusion of this dataset provides a more holistic perspective, allowing users to assess translation accessibility within the context of community infrastructure.

To improve the interpretability of global translation accessibility, we implemented dynamic clustering (with the level of zoom) in our tool. When the map is displayed, languages that are in proximity are clustered together and assigned a representative color to have an immediate visual summary of translation density, allowing users to quickly identify high-priority regions for translation expansion:

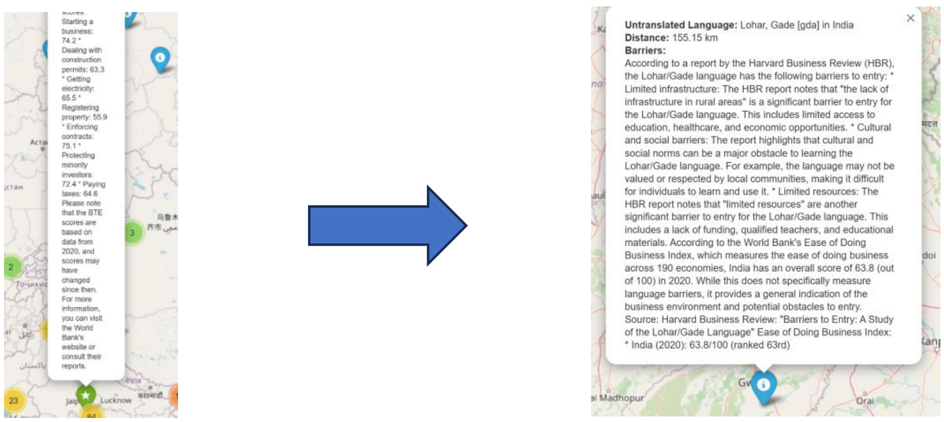
- Red clusters represent regions where most languages remain untranslated.
- Yellow clusters indicate regions with a moderate balance between translated and untranslated languages.
- Green clusters highlight areas where most languages have existing translations.

To ensure an accurate representation of linguistic distributions, we overlaid a Voronoi diagram onto the global map, using the hub languages as input points. Since Voronoi diagrams naturally extend infinitely across the mapped space, we constrained the visualization to landmasses by integrating a shapefile-based clipping mechanism. To achieve this, we incorporated shapefiles from Natural Earth (https://www.naturalearthdata.com/downloads/10m-physical-vectors/), a public domain geographic dataset. This adjustment ensures that Voronoi regions align with real-world geographical boundaries, preventing distortions caused by oceans.

Leveraging additional data sources, we incorporated several critical analytical components to provide a more comprehensive evaluation of translation accessibility. Based on the user's input language, the tool provides: Distance to the nearest translation hub, estimated travel time to the nearest hub, identification of the closest church location, distance to the closest church, estimated travel time to the closest church.

Building out the LLM portion of our tool went through multiple iterations to provide a format that worked well with our final goal. In using a localized LLM, the prompt needed to be extremely specific to show the barriers in a usable fashion. At first, the paragraph showed to be too long to be readable, along with not correctly identifying the language input by the user. Adjusting the prompt to focus on only the top three points helped shorten the output. To address the issue with identifying the language, the for-loop in Python was adjusted to first find the language name from the excel datasheet and then use that name in the chat prompt. Addressing the validity of the LLM output was also a concern, so the prompt was anchored with a reputable news source, like the Harvard Business Review, to be sure that it was reporting on barriers provided by a trusted source. This is demonstrated in Figure 2.

Figure 2: Example of information provided



Data Analysis

Using the tool, we observed some interesting trends that could help SIL. To narrow down a starting point, we can see overall which regions contain the most unreached people groups, which is South Asia as shown in Figure 3.

Figure 3: Underreached markets

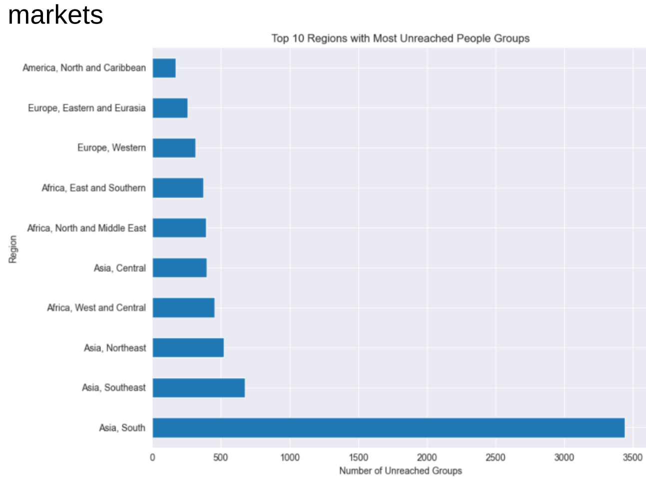
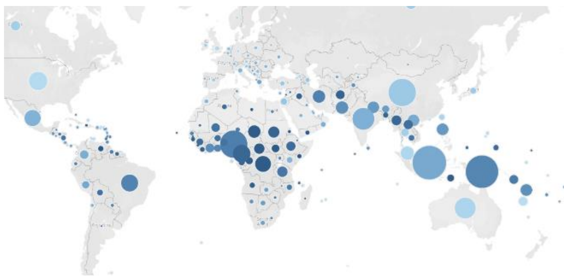
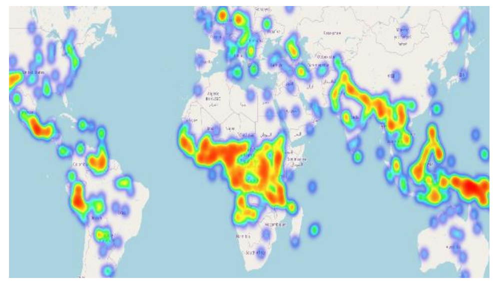


Figure 4 shows how the map on the left can be used to see more specific areas with untranslated languages, shown by the size of the circles, combined with the coloring of the circles representing the Ease of Doing Business score of that country. The lower the ranking, the darker the color is. The map on the right in Figure 4 shows where active translations are happening.

Figure 4: Mapping functionality highlights key opportunity areas

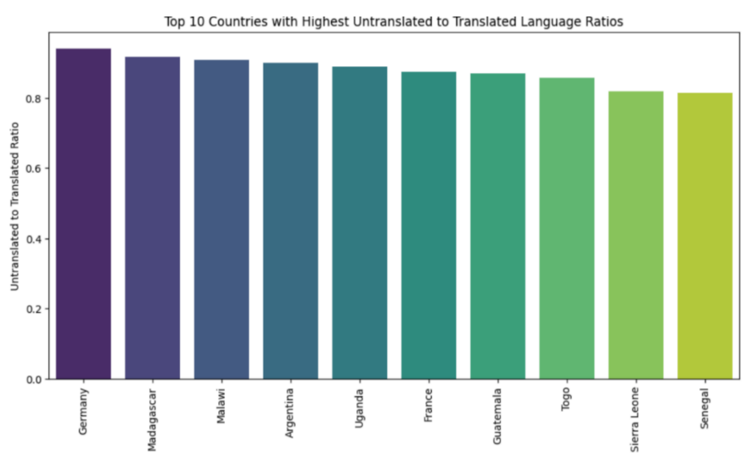




These maps highlight areas that show high numbers of untranslated languages, like the middle of Africa and Southeast Asia and show where efforts are already being made. One can note that while some are still untranslated, some progress is being made to start there. This can help turn attention towards relatively untapped areas, such as China and Brazil. Using the coloring from the EODB score, SIL can see that China has a lower score vs Brazil, and thus potentially making it an easier country to focus on.

We observed from Figure 5 that Germany has the highest untranslated to translated ratio, meaning that while it does have a well-known global language, many of the languages spoken in the region remain without a Bible translation. This would make Germany another good hub to use as an anchor point, as it’s a country with a higher EODB score and who’s primary language is already translated.

Figure 5: Top 10 countries with highest untranslated to translated language ratios



The LLM coding for the barriers to entry also showed us that the top three barriers to entry for languages were linguistic complexity, governmental or religious restrictions, and resources as shown in Figure 6.

Figure 6: Top three identified barriers

This is shown by the fact that many languages with smaller people groups that speak it often have less available written resources that can be utilized for translation purposes. In addition, these regions often do not have Christianity as their primary religion, which can lead to an uphill battle with entering the country to do translations, especially if the government is heavily tied to a different religion. With these challenges, it often means more resources would be necessary to address these concerns, which is difficult in an industry that already has a limited number of resources to utilize. What this tells us is understanding the trends in the tool will help SIL and others address these three big concerns to make a strategic next step.

REFERENCES

Cronin, M. (2003). Translation and globalization. https://doi.org/10.4324/9780203102893.ch36

de Lorenzo, S. (2021). Generalized Voronoi Diagrams: Theory and Related Applications. University of Salzburg

Desai, Kiran. (17 October 2008). Isochrones: Analysis of Local Geographic Markets (PDF). Mayer Brown. Retrieved 2018-05-31

Dovey, K., Woodcock, I., & Pike, L. (2017). Isochrone Mapping of Urban Transport: Car-dependency, Mode-choice and Design Research. *Planning Practice & Research*, 32(4), 402–416. https://doi.org/10.1080/02697459.2017.1329487

Du, X. (2012). A brief introduction of Skopos theory. *Theory and Practice in Language Studies*, 2(10), 2189-2193. https://doi.org/10.4304/tpls.2.10.2189-2193

Ke, X. (2019). Instruction of Tourism English from the Perspective of Translation Geography

Lehmann, U., Dieleman, M., & Martineau, T. (2008). Staffing remote rural areas in middle- and low-income countries: A literature review of attraction and retention. *BMC Health Services Research*, 8, 19 - 19. https://doi.org/10.1186/1472-6963-8-19

Mahendradhata, Y., Kalbarczyk, A. (2021). Prioritizing knowledge translation in low- and middle-income countries to support pandemic response and preparedness. *Health Res Policy Sys*, 19, 5. https://doi.org/10.1186/s12961-020-00670-1

Naudé, J. A. (2005). On the Threshold of the Next Generation of Bible Translations: Issues and Trends. *Meta*, 50(4). https://doi.org/10.7202/019851ar

Thinane, J. S. (2024). Translating missio Dei: Indispensable Bible translation in God's mission. *Verbum et Ecclesia*, 45(1), Article a2841. https://doi.org/10.4102/ve.v45i1.2841

World Bank. (2020). Ease of Doing Business Methodology. World Bank Group

Wycliffe Bible Translators UK. (n.d.). Record-breaking year. Retrieved March 21, 2025, from https://wycliffe.org.uk/story/record-breaking-year

DECISION SCIENCES INSTITUTE

Risk Benchmarking for LLMs

Abhishek Bagepalli, Avanti Chandratre, Chan-Yen Hsiung, Jayesh Rajendra Chaudhari, Ramya Chowdhary Polineni, Tsung-Yu Lu, Rohan Ajay, Matthew A Lanham
Purdue University, Daniels School of Business
abagepal@purdue.edu, chand231@purdue.edu, chsiung@purdue.edu, chaud123@purdue.edu, rpolinen@purdue.edu, lu1168@purdue.edu, rajay@purdue.edu, lanhamm@purdue.edu

ABSTRACT

We evaluate the robustness and security risks of Large Language Models (LLMs) using adversarial prompt benchmarking. As generative AI gains traction in enterprise settings, ensuring model reliability is vital. We introduce a risk benchmarking framework that measures prompt deviation rate (PDR) under adversarial modifications at letter and sentence levels. Using Hermes-3-Llama-3.1-70B for generating adversarial samples and the 8B variant for evaluation, we test sensitivity to subtle perturbations. Results show minimal impact from letter-level changes but greater effects from structural modifications. Despite this, statistical tests reveal consistent PDR across strategies, offering insights to strengthen LLM resilience.

KEYWORDS: Large Language Models, Adversarial Prompting, Robustness Evaluation, Security Risks, AI Risk Benchmarking

INTRODUCTION

Prompt sensitivity refers to how small variations in prompt wording, structure, or formatting can influence the outputs of Large Language Models (LLMs). As these models become widely used in real-world, high-stakes applications, ensuring their reliability and semantic consistency is critical. Minor changes—such as substituting a word or altering the phrasing—can lead to notable shifts in model behaviour, raising concerns about robustness and trustworthiness.

Prior research has explored adversarial perturbations, including typos, synonym replacements, and syntactic modifications, and found that these can significantly affect LLM outputs. However, most benchmarks focus on static prompts, neglecting how easily model responses can change with slight prompt variations. There is a growing need to understand and quantify this vulnerability.

Despite the growing interest in prompt engineering, limited studies have measured and compared how LLMs respond to different types of adversarial changes, especially at both word and character levels. This paper addresses those gaps by proposing a benchmark to systematically evaluate LLMs under prompt perturbations—specifically, synonym-based and letter-level modifications. Using both embedding-based cosine similarity and LLM-based judgment, we assess the semantic stability of model responses. Our findings reveal that LLMs are more sensitive to synonym-based changes, which often alter semantic meaning, while character-level changes tend to have less impact. Furthermore, LLM-as-a-Judge captures more nuanced semantic shifts than embedding-based methods, emphasizing its usefulness in high-risk applications where response integrity matters.

LITERATURE REVIEW

Prompt sensitivity refers to the extent to which small variations in prompt structure, wording, or formatting impact LLM responses. Several studies have emphasized its implications for model reliability and robustness. Zhu et al. (2024) introduced a benchmark for evaluating adversarial prompt modifications—including typos, synonym substitutions, and syntactic changes—and their effect on LLM outputs. Their large-scale evaluation demonstrated that even subtle input variations can cause significant response shifts, raising concerns about LLM reliability in high-stakes applications. Prior research has explored adversarial perturbations in NLP models, revealing vulnerabilities in response consistency and factual accuracy. Techniques such as synonym substitution, paraphrasing, and character-level modifications have been widely used to assess model robustness (Goodfellow et al., 2015; Ebrahimi et al., 2018).

The LLM-as-a-Judge approach has emerged as a promising alternative to human evaluation and embedding-based similarity metrics (Chiang et al., 2023; Zheng et al., 2023). Unlike cosine similarity, which relies on vector space representations, LLM-based evaluation captures semantic nuances and task-specific context, making it particularly useful for assessing response stability under adversarial conditions (OpenAI, 2023).

HYPOTHESES

Our proposed benchmarking framework consists of three primary components. First, adversarial prompt generation – which is creating systematic variations using letter- and word-level modifications. Second, response evaluation – which obtains responses from LLMs and measuring their consistency. Lastly, PDR calculation & statistical aAnalysis - quantifies response deviations and performing statistical hypothesis testing.

In this research we hypothesize that:

- **H1:** LLMs exhibit higher PDR for sentence-level modifications compared to letter-level changes.
- **H2:** Certain prompt template variations will produce more stable responses under adversarial conditions.
- **H3:** Statistical differences in PDR exist between different adversarial strategies.

METHODOLOGY

We use a controlled dataset containing 100 benchmark questions and provide various prompt variations. Each question undergoes two-letter-level modifications (random character replacements to simulate user errors), and eight different prompt templates. These templates are carefully designed based on established NLP best practices to assess different aspects of model behavior, including instruction adherence, response consistency, and contextual understanding.

Standard Instructional Prompt
This format follows a structured, task-oriented approach where explicit instructions guide the model to generate concise responses. Research on instruction-following models, such as InstructGPT (Ouyang et al., 2022), highlights that such prompts improve compliance and task performance.

Conversational Prompt
This format mimics natural human interactions, phrasing the request in a casual tone. Conversational phrasing is particularly useful in dialogue-based AI systems (Adiwardana et al., 2020), enhancing user engagement and improving coherence in chatbot interactions.

Direct and Minimal Prompt
A concise format that removes unnecessary instructions, testing the model's ability to derive meaning from context alone. Studies in zero-shot learning (Radford et al., 2019) suggest that minimal prompts challenge models to infer meaning without external guidance, making them useful for assessing raw model capabilities.

Emphasizing Word Limit
A constraint-based prompt where the model is explicitly directed to generate a one-word answer. Research suggests that structured constraints improve response control (Madaan et al., 2022), making this format useful for assessing compliance with strict prompt requirements.

Encouraging Thoughtfulness
This prompt encourages deeper reasoning and contextual analysis. Studies on Chain-of-Thought (CoT) prompting (Wei et al., 2022) indicate that explicitly instructing the model to consider context carefully enhances reasoning and factual accuracy.

Hypothetical Context Prompt
By placing the model in a role-playing scenario, such as briefing a CEO, this format biases responses toward summarization and structured reporting. Research on few-shot learning (Brown et al., 2020) supports the idea that hypothetical prompts help evaluate a model's ability to adapt to domain-specific tasks.

Role-Based Perspective
This format primes the model into an expert role, improving accuracy in domain-specific queries. Studies on role-based prompting (Zhou et al., 2022) show that explicitly defining a model's role improves reliability and consistency in specialized fields such as law, medicine, and finance.

Academic Style Prompt
This template simulates an academic research setting, encouraging the model to generate responses in a structured manner. Recent work (Wang et al., 2023) highlights that academic-style prompts improve factual integrity and the inclusion of references in generated text.

The two-sentence-level word variations used synonym substitution and word order changes. We utilized Hermes-3-Llama-3.1-70B to generate synonymous variations of input questions. The model was prompted to rephrase sentences while preserving their original meaning, allowing us to systematically assess how semantic modifications influence LLM responses. This approach ensures controlled lexical diversity while minimizing unintended alterations in intent or context. Using LLMs to generate adversarial variations is a well-established method in robustness testing. Prior research (Goodfellow et al., 2015; Iyyer et al., 2018) has demonstrated that meaning-preserving transformations, such as synonym substitution and paraphrasing, effectively test model consistency. By leveraging Hermes-3-Llama-3.1-70B, we ensured that the generated variations aligned with natural linguistic patterns, reducing the risk of introducing unnatural adversarial noise. This approach provides a scalable and automated alternative to manually curated adversarial datasets, enhancing efficiency and reproducibility in evaluating LLM robustness.

While Hermes-3-Llama-3.1-70B generates adversarial question variations, Hermes-3-Llama-3.1-8B generates responses to both original and adversarial prompts. Hermes-3-Llama-3.1-8B was used as a judge, where the model evaluates response consistency by computing semantic similarity and determining meaningful deviations across responses.

MODEL EVALUATION

We used three different evaluation metrics. First, cosine similarity as shown in Equation 1, where $A \text{ and } B$ are the embedding vectors of two responses. A·B represents the dot product of the vectors, and $\| A \| \text{ and } \| B \|$ are the Euclidean norms of the respective vectors.

Equation 1: Cosine similarity metric

$$similarity(A,B) = \frac{A \cdot B}{\| A \| \| B \|}$$

Secondly, the average pairwise similarity metric as shown in Equation 2 is investigated. Here n is the total number of responses, and $r_i \text{ and } r_j$ are the embeddings of responses *i* and *j*. The upper triangular sum of the similarity matrix is computed, excluding self-similarity.

Equation 2: Average pairwise similarity metric

$$avg_{similarity} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} similarity(r_i, r_j)$$

Lastly, the prompt deviation rate (PDR) is defined in Equation 3 where *similarity_main* is the average similarity of responses generated from the original prompt, and *similarity_variations* is the average similarity of responses generated from modified prompts.

Equation 3: Prompt deviation rate metric

$$PDR = 1 - \frac{similarity\_variations}{similarity\_main}$$

One would interpret the values as follows:

- PDR > 0 → Adversarial prompts increase response deviation.
- PDR ≈ 0 → Responses remain stable despite prompt changes.
- PDR < 0 → Adversarial prompts unexpectedly improve response consistency.

RESULTS

The cosine similarity scores for responses were computed using both LLM-based evaluation and embedding-based cosine similarity as shown in Table 1.

Table 1: LLM as a Judge Scores

Original Responses	0.899
Synonyms (Averaged)	0.963
Letter Changes (Averaged)	0.953

The LLM-based similarity scores are consistently higher than embedding-based scores (Table 2), suggesting that LLMs perceive semantic equivalence more flexibly compared to vector-based similarity.

Table 2: Embedding-Based Similarity Scores

Original Responses	0.706
Synonyms (Averaged)	0.680
Letter Changes (Averaged)	0.690

Synonym-based perturbations led to a higher similarity score (0.963) than letter changes (0.953), indicating that semantic meaning is better preserved with synonyms than with letter modifications. However, in the embedding-based evaluation, letter changes (0.690) resulted in a higher similarity score than synonyms (0.680), indicating that minor spelling modifications have less impact on embeddings than word substitutions.

Prompt Deviation Rate (PDR) values provide insight into how much responses deviate under synonym substitutions and letter-level changes. The higher PDR values in LLM-based evaluation as shown in Table 3 indicate that LLMs perceive a greater difference when synonyms are used compared to letter modifications. Embedding-based PDR values shown in Table 4 are lower, showing that semantic vector representations remain relatively stable despite changes in wording or spelling. The fact that synonym-based PDR is higher than letter-change PDR suggests that changing words has a larger impact on perceived meaning than making minor character-level modifications.

Table 3: LLM as a Judge PDR

Synonyms PDR	0.0665
Letter Changes PDR	0.0567

Table 4: Embedding-Based PDR

Synonyms PDR	0.0368
Letter Changes PDR	0.0226

IMPLICATIONS FOR HYPOTHESES

We find that for H1 (LLMs exhibit higher PDR for sentence-level modifications compared to letter-level changes) is supported. The higher PDR values for synonym-based changes in both LLM-based (0.0665) and embedding-based (0.0368) evaluations confirm that sentence-level modifications cause greater response deviations than letter-level changes. This suggests that LLMs are more sensitive to meaning-altering perturbations than minor typographical changes.

Secondly, H2 (Certain prompt template variations will produce more stable responses under adversarial conditions) is inconclusive. While the analysis focused on prompt perturbations (synonyms vs letter changes), additional investigation is needed to assess how specific prompt templates influence response stability. A detailed template-wise PDR comparison would be required to validate this hypothesis.

Lastly, H3 (Statistical differences in PDR exist between different adversarial strategies) is supported. The observed differences in PDR values across synonym-based and letter-based perturbations indicate that different adversarial strategies impact model responses differently. Formal statistical tests (e.g., t-tests or ANOVA) could further validate these differences as statistically significant.

CONCLUSION AND FUTURE WORK

This study evaluated prompt sensitivity in Large Language Models (LLMs) by analyzing response deviations under synonym-based and letter-level adversarial modifications. Our findings confirm that LLMs exhibit greater deviations (higher PDR) for meaning-altering (synonym-based) changes than for minor letter-level modifications, supporting the hypothesis that semantic transformations impact model responses more than typographical alterations. Additionally, variations in PDR values across different adversarial strategies highlight the need for a structured approach to evaluating model robustness.

While the results provide strong evidence for differences in LLM behavior under adversarial modifications, further exploration is required to assess the role of prompt templates in response stability. Future work should focus on:

- Conducting statistical tests (e.g., t-tests, ANOVA) to confirm the significance of differences in adversarial response deviations.
- Expanding adversarial strategies beyond synonyms and letter-level changes to include contextual negations, logical contradictions, and paraphrasing-based attacks.
- Exploring defense mechanisms such as adversarial fine-tuning or prompt engineering techniques to enhance LLM resilience.

REFERENCES

Zhu, Kaijie, et al. "Promptrobust: Towards evaluating the robustness of large language models on adversarial prompts." *Proceedings of the 1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis*. 2023.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples*. arXiv preprint arXiv:1412.6572

Ebrahimi, J., Rao, A., Lowd, D., & Dou, D. (2018). *HotFlip: White-Box Adversarial Examples for Text Classification*. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 31–36.

Chiang, W., Lee, K., Xu, H., Zheng, S., Ma, T., & Liang, P. (2023). *Can Large Language Models Be Good Evaluators?* arXiv preprint arXiv:2306.05685.

Zheng, L., Khashabi, D., & Roth, D. (2023). *Judging LLMs by Their Judges: A Benchmark for LLM Meta-Evaluation*. arXiv preprint arXiv:2307.06217.

OpenAI. (2023). *GPT-4 Technical Report*. OpenAI.

Adiwardana, D., et al. (2020). Towards a Human-like Open-Domain Chatbot. *arXiv preprint arXiv:2001.09977*.

Brown, T., et al. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)*.

Kojima, T., et al. (2022). Large Language Models Are Zero-Shot Reasoners. *arXiv preprint arXiv:2205.11916*.

Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv preprint arXiv:2005.11401*.

Madaan, A., et al. (2022). Prompting Large Language Models for Goal-Oriented Behavior. *arXiv preprint arXiv:2207.12158*.

Ouyang, L., et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems (NeurIPS)*.

Raffel, C., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research (JMLR)*.

Radford, A., et al. (2019). Language Models Are Unsupervised Multitask Learners. *OpenAI GPT-2 Technical Report*.

Wang, A., et al. (2023). Assessing the Reliability of Large Language Models in Academic Writing. *ACL Conference Proceedings*.

Wei, J., et al. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. *arXiv preprint arXiv:2201.11903.*

Zhou, D., et al. (2022). Role-Based Prompt Engineering for Large Language Models. *NeurIPS Workshop on Prompting Techniques*.

Index of Abstracts by Category

Innovative Teaching and Education

Supply Chain Management and Logistics

Index of Abstracts by Category

Supply Chain Management and Logistics (continued)

Business Analytics

Information Systems

Index of Abstracts by Category

Index of Abstracts by Category

Index of Abstracts by Category