The Evolving Retail Ecosystem: Spatiotemporal Data, Partnerships, and Crowd Wisdom for Strategic Advantage

Mingrui Zhang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Yong Tan, Chair

Wanning Chen

Shi Chen

Program Authorized to Offer Degree:

School of Business

University of Washington

Abstract

The Evolving Retail Ecosystem: Spatiotemporal Data, Partnerships, and Crowd Wisdom for Strategic Advantage

Mingrui Zhang

Chair of the Supervisory Committee:
Yong Tan
Department of Information Systems and Operations Management

The retail landscape is undergoing a profound transformation driven by technological advancements, changing consumer behaviors, and intensifying competition between online and offline channels. This dissertation explores three critical aspects of this evolving ecosystem: leveraging spatiotemporal data for strategic decision-making, forging partnerships between physical retailers and e-tailers for consumer returns, and harnessing crowd wisdom in the crowdfunding context. Chapter 3 employs tensor completion methods to estimate the treatment effect of deploying smart vending machines across various urban settings, demonstrating the superiority of this approach over traditional causal inference techniques. Chapter 4 investigates the impact of a returns partnership between a physical retailer and an e-tailer on their demand-side competition, revealing conditions under which both firms can benefit, but consumer surplus may decrease. Chapter 5 examines how atypical idea combinations on crowdfunding platforms influence project success, identifying an optimal balance between familiarity and novelty. Collectively, these chapters contribute to the literature on retail strategy, data-driven decision-making, and platform economies, offering valuable insights for firms navigating the complex and dynamic retail ecosystem.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the faculty, staff, and fellow students in the ISOM Department and Foster PhD Program for their unwavering support and mentorship throughout my graduate studies.

First and foremost, I extend my heartfelt appreciation to my primary faculty advisor, Professor Yong Tan, for being an exceptional mentor. Your guidance, wisdom, and unwavering support have been the pillars of my academic and personal growth.

Secondly, I want to express my profound gratitude to Professor Wanning Chen. Your mentorship has been instrumental in shaping me as a scholar and preparing me for the challenges and opportunities that lie ahead. I am deeply grateful for your on-hand guidance, not only in our research projects but also in life.

Next, I want to express my sincere gratitude to the rest of my committee members, Professor Shi Chen, Professor Kamran Moinzadeh, Professor Quan Wen, and Professor Jing Tao, for their invaluable contributions to my research. Your expertise, insightful feedback, and dedication to my success have greatly enriched my dissertation and helped me grow as a researcher. Your guidance and support have been crucial in navigating the complexities of my research topic and refining my ideas.

Lastly, I would like to thank my parents, Zongyi Zhang and Chunhong Li, and my in-laws, Minggang Kan and Chaohua Wang, for their unconditional love and support. Knowing that there is always a home behind me has given me the strength to reach this far. I also want to thank my classmate, friend, and wife, Yu Kan. We have explored and will continue to explore around the globe, but you are my greatest adventure.

Chapter 1

INTRODUCTION

The retail industry is in the midst of a profound transformation, shaped by the rapid growth of e-commerce, the proliferation of data, and the blurring of boundaries between online and offline channels. As firms strive to adapt and compete in this evolving landscape, they must harness the power of data, forge strategic partnerships, and leverage the wisdom of crowds to gain a competitive edge.

This dissertation comprises three essays that explore critical aspects of the modern retail ecosystem. Chapter 3, "Impacts of Smart Vending Machines in Different Urban Settings: Tensor Completion with Spatiotemporal Data," investigates how firms can leverage high-dimensional, spatiotemporal data to make strategic decisions. By applying advanced tensor completion methods to estimate the treatment effect of deploying smart vending machines across various urban settings, this chapter demonstrates the superiority of this approach over traditional causal inference techniques and highlights the importance of data-driven decision-making in optimizing retail operations.

Chapter 4, "Frenemies in the Retail Market: A Partnership Between a Physical Retailer and an E-tailer for Consumer Returns," examines the impact of a returns partnership between a brick-and-mortar retailer and an online retailer on their demand-side competition. By developing an analytical model, this chapter reveals the conditions under which both firms can benefit from such a partnership, but also uncovers a potential trade-off between retailer profitability and consumer surplus. This research contributes to the growing literature on omnichannel retailing and the strategic interactions between online and offline players.

Chapter 5, "Afraid of Niche, Tired of Mass: Atypical Idea Combination on Crowdfunding Platform," explores how the combination of familiar and novel elements in project ideas

influences success on crowdfunding platforms. By analyzing a large dataset from a leading crowdfunding website, this chapter identifies an optimal balance between familiarity and novelty that maximizes project funding. This research sheds light on the role of crowd wisdom in shaping the success of new ventures and contributes to the literature on platform economies and innovation.

Collectively, these three chapters provide valuable insights into the evolving retail ecosystem, highlighting the importance of spatiotemporal data, partnerships, and crowd wisdom for strategic advantage. The findings have significant implications for retailers, e-tailers, and platform operators seeking to navigate the complex and dynamic landscape of modern retail.

Chapter 2
LITERATURE REVIEW

This chapter provides an overview of the existing literature relevant to the methodologies and topics addressed in this paper. We begin by examining the intersection of econometric and machine learning techniques for causal inference in panel data settings. We then explore the dynamics of online-offline partnerships. And next we discuss atypicality in crowdfunding projects.

2.1 Casual Inference using Tensor Completion

The estimation of causal effects from observational data is a fundamental problem in many disciplines, including economics, statistics, and machine learning. In panel data settings, where units are observed across multiple time periods, a key challenge is to impute the missing potential outcomes for treated units had they not received treatment. This literature review explores several approaches to tackling this problem, focusing on the connections between econometric methods for causal inference and machine learning techniques for matrix and tensor completion. By examining these links, we lay the groundwork for the novel tensor completion methodology developed in this paper, which aims to unify and generalize existing approaches to estimating causal effects from panel data.

2.1.1 Unconfoundedness and Synthetic Controls

The unconfoundedness or selection on observables approach (Rosenbaum and Rubin 1983, Imbens and Rubin 2015) assumes that treatment assignment is independent of potential outcomes conditional on observed pre-treatment covariates and outcomes. In a panel data setting, this implies using control units with similar pre-treatment outcome trajectories to

impute the missing potential outcomes for treated units (Athey et al. 2021). In practice, this approach is often implemented by regressing post-treatment outcomes of control units on their pre-treatment outcomes, and using that model to predict counterfactuals for treated units.

In contrast, the synthetic control method (Abadie and Gardeazabal 2003, Alberto Abadie and Hainmueller 2010, Abadie et al. 2015) constructs a weighted average of control units designed to match the pre-treatment outcomes of a treated unit. The weights are chosen so that the weighted combination of control units closely tracks the treated unit's pre-treatment outcome trajectory. The post-treatment outcomes of this "synthetic control" unit serve as the counterfactual for the treated unit.

The two methods also tend to be applied in different data configurations - unconfoundedness with many treated and control units but few time periods, and synthetic controls with few treated units but many pre-treatment time periods.

2.1.2 Matrix Completion

An emerging literature connects these approaches to causal inference to matrix completion techniques. With panel data, one can view the causal inference problem as filling in the missing entries of the potential outcomes matrix, assuming it is low-rank plus noise (Athey et al. 2021).

The econometrics literature on interactive fixed effects (Bai and Ng 2002, Bai 2009, Xu 2017) models the outcome matrix as a combination of covariates, time-varying coefficients, unit and time fixed effects, and a low-rank matrix of unobserved factors. Estimation typically fixes the number of factors. Xu (2017) extends this to causal inference under restrictions on the treatment assignment mechanism.

In the machine learning and statistics literature, the matrix completion problem aims to impute missing entries of a low-rank matrix (Candès and Recht 2009, Candes and Plan 2010, Mazumder et al. 2010). Regularization, such as the nuclear norm penalty, is employed to encourage low-rankness without specifying the rank a priori. However, most work assumes

missing data is uniformly random.

Athey et al. (2021) draw connections between these literatures. They show that unconfoundedness, synthetic controls, and matrix completion methods can all be framed as minimizing the same least squares objective function, but differ in their restrictions or regularization on the underlying factor model. They propose a nuclear-norm regularized estimator that encompasses and generalizes these approaches. Their estimator adapts flexibly to different data configurations and outperforms unconfoundedness and synthetic controls in simulations.

2.1.3 Tensor Completion

Low-rank tensor completion has been widely used for estimating missing values in three-dimensional data, such as visual data (He and Atia 2021), seismic data (Kreimer et al. 2013), and traffic data (Chen et al. 2021). Tensor completion extends the matrix completion problem to higher-order arrays, aiming to impute missing entries of a low-rank tensor from a subset of its observed entries. Tensor factorization models, which decompose a tensor into a combination of lower-dimensional factors, are commonly used to capture the low-rank structure (Kolda and Bader 2009). Popular tensor decompositions include CANDECOMP/PARAFAC (CP), Tucker, and tensor train (TT) models.

Early tensor completion algorithms adopted convex relaxation approaches based on minimizing the tensor nuclear norm (Liu et al. 2013, Signoretto et al. 2011), generalizing techniques from compressed sensing and matrix completion. However, computing the tensor nuclear norm is NP-hard for general tensors (Friedland and Lim 2016).

More recently, many tensor completion algorithms optimize factorization models directly, often using alternating minimization schemes. For the CP model, Acar et al. (2011) proposed a weighted least squares approach called CP-WOPT. For the Tucker model, an alternating least-squares approach called Higher-Order Orthogonal Iteration (HOOI) has been widely used (Filipović and Jukić 2015, Karlsson et al. 2016). TT-based methods have also been developed (Yuan et al. 2019). While these techniques work well empirically, global convergence

guarantees are often lacking.

Another line of work establishes recovery guarantees for tensor completion under incoherence assumptions. Jain and Oh (2014) gave an alternating minimization method that provably recovers tensors whose factors are incoherent. Barak and Moitra (2022) showed the tensor nuclear norm can be approximated by a sum of matrix nuclear norms. Xia and Yuan (2020) proposed a Riemannian gradient descent method over the tensor train format and provided local convergence guarantees. However, these approaches still lack global optimality guarantees and can be computationally expensive.

Some recent works aim to bridge this gap between theory and practice. Cai et al. (2022) proposed a fast non-convex algorithm that provably completes tensors with near-orthogonal factors. Liu and Moitra (2020) introduced a provable alternating minimization method for tensor completion, analyzing its convergence under incoherence assumptions on the tensor factors. However, strong global guarantees, especially for tensors with highly correlated factors, largely remain an open problem.

2.2 Online-Offline Partnership

Our study is mainly related to three streams of literatures: (i) consumer return, (ii) consumer learning and product information disclosure, and (iii) supply chain contracts. In this section, we briefly discuss the relevant works in each stream and then highlight our contributions with respect to those studies.

2.2.1 Consumer Return

There is a wealth of literature that explores various aspects of customer returns in the retail industry. Shulman et al. (2009) develop analytical model that describes how consumer purchase and return decisions are affected by a seller's pricing and restocking fee policy. They find that it is impossible to eliminate returns costlessly through the provision of information but re-turns could still be part of the optimal product sales process. In the case of multi-channel retailers, Ofek et al. (2011) demonstrate that an online channel can help increase

investment in store assistance levels and decrease profit, when differentiation among competing retailers is not too high. Anderson et al. (2009) empirically show that there are large increases in demand when the option value is large using a structural model for consumer's decision to purchase and return a product.

2.2.2 Consumer Learning about Product Information

There is a large amount of literature regarding information disclosure and learning in omnichannel. On the modeling side, Hao and Tan (2019) investigate a retailer's and a supplier's incentive to facilitate information disclosure under the agency pricing model and the wholesale pricing model and find that information disclosure facilitation has a different interplay with the two models. Sun (2011) investigates how do multiple attributes of a product jointly determine a seller's disclosure incentives. He finds that monopolist in equilibrium does not always choose disclosure. Kwark et al. (2014) show the effect of online product reviews on different players in a channel structure. They find that quality information homogenizes consumers' perceived utility differences. It intensifies the upstream competition but benefits the retailers. Fit information heterogenizes consumers' estimated fits. It softens the upstream competition but hurts retailers. Moreover, Gao and Su (2017) show physical showrooming may cause retailers to reduce store inventory, and virtual showrooming may increase online returns and hurt profits. On the empirical side, Bell et al. (2018) demonstrate how showrooming benefit the two most basic retail objectives: demand generation and operational efficiency using quasi-experimental data. They find that show-rooming increase demand overall; generate operational spillovers to the other channels; improve overall operational efficiency; and amplify demand and operational benefits with customers who have acute need. Kumar et al. (2019) use customer-level data to estimate the effect of store presence on the online purchase behavior of its existing customers. They find that the retailer's store openings promote sales from existing customers.

Balakrishnan et al. (2014) characterize the conditions that consumers will do showrooming. They show that consumer showrooming intensifies the competition between a physical

retailer and an e-tailer, and eventually hurts both firms' profits. Mehra et al. (2018) demonstrate that showrooming is detrimental to the profits of the brick-and-mortar stores. They study price matching and find that with an increase in the fraction of customers who seek price matching, the store's profits initially decrease and then increase. For exclusivity of product assortments strategy, their analysis suggests that, when the product category has few digital attributes, implementing exclusivity through store brands is better than exclusivity through known brands. Also, showrooming intensifies competition and lowers both firms' profits, thus hurt the physical retailer when consumers' search cost is not sufficiently high (Jing 2018). When webrooming interacts with showrooming, the webrooming behavior might increase both the physical retailer's and the e-tailer's profit as it partially resolves consumers fit uncertainty.

2.2.3 Supply Chain Contracts

In the extensive body of research surrounding supply chain contracts, our study is particularly relevant to those that explore how shifts in consumer purchasing behavior impact firms' profits and strategies within distribution channels operating under either wholesale or agency contracts. Su (2009) demonstrates that allowing product returns can distort supplier and retailer strategies under common supply chain contracts, affecting channel coordination. To address this, the author proposes a new coordination strategy that accounts for consumer returns. Yue and Raghunathan (2007) investigate the impact of full returns policies, finding that while the retailer always benefits, the manufacturer and supply chain may sometimes suffer. Kwark et al. (2014) examine how the informativeness of online product reviews influences the profits of a retailer and two competing manufacturers in a wholesale contract distribution channel. Hao and Fan (2014) explore the effects of consumers' complementary consumption of e-books and e-readers on the profits of a publisher, who earns revenue from both e-book and physical book sales, and an e-tailer, who profits from e-book and e-reader sales, under both wholesale and agency contracts. Kwark et al. (2017) show that a retailer can benefit more from informative online product reviews by switching from an agency to

a wholesale contract with two competing manufacturers. Hao and Tan (2019) reveal that suppliers and retailers may have conflicting interests regarding facilitating consumer learning about product fit, with the nature of this misalignment depending on whether the supply chain operates under a wholesale or agency contract.

2.3 Atypicality in Crowdfunding

Multiple factors determine the success of a crowdfunding project. Researchers usually approach this question from two perspectives, intrinsic and extrinsic. Extrinsic factors include social network, geographic locations (Mollick 2014), scarcity strategy (Yang et al. 2020), campaign design (Chakraborty and Swinney 2021), information control level (Lin et al. 2021, Burtch et al. 2016), and founder race (Younkin and Kuppuswamy 2018), etc. Though many extrinsic factors have been observed and thoroughly investigated, the intrinsic factors remain underexplored. The intrinsic factors are associated with the quality of the projects. Researchers have suggested that *high-quality signals* can improve crowdfunding performance (Chakraborty and Swinney 2021). However, little work has sought to directly understand how the quality of crowdfunding projects affects their success due to the intangible nature of the quality feature. Creativity, one of the most important aspects of a crowdfunding project's quality, receives even less attention.

Therefore, the second stream of literature we contribute to is idea generation and creativity. Campbell (1960) first applies Darwinism to creative cognition. Idea combination as an essential part of "Blind variation" provides a cognitive view of idea generation. Later in cognitive psychology, the Geneplore model proposed by Finke et al. (1992) considers creativity iterative and alternating between the *generative* process and *exploratory* process. People prepare pre-inventive structures, which can be imagined as seeds of ideas in the first phase. Then these structures are analyzed, interpreted and combined to form new ideas. In some studies, researchers use creativity as a general concept without further definitions (Kornish and Jones 2021). In our context, we define creativity as a good balance between familiarity and atypicality (Toubia and Netzer 2017). In Table 2.1, we summarize the papers studying

this topic. We advance prior studies in multiple important ways.

Table 2.1: How the current paper advances prior studies

Context	Study	Content Analysis	Similarity Network	Interpretable Innovation Classification	Model Comparison	Across Product Categories	Underlying Mechanisms	Audience Sentiment Analysis	Audience Behavioral Sequence Analysis
Crowdfunding	**The Current Study**	✓	✓	✓	✓	✓	✓	✓	✓
	Chan and Parhankangas (2017)	✓							
	Taeuscher et al. (2021)	✓							
	Wei et al. (2022)	✓	✓						
Scientific Research	Uzzi et al. (2013)		✓						
	Foster et al. (2015)		✓						
	Wang et al. (2017)			✓					
Patent	Youn et al. (2015)								
	Fleming (2001)				✓		✓		
	Zhang et al. (2019)		✓		✓				
Ideation Tasks	Stephen et al. (2016)	✓	✓						
	Toubia and Netzer (2017)	✓	✓						
Music	Berger and Packard (2018)	✓							
	Askin and Mauskapf (2017)				✓				
Motion Pictures	Wei (2020)		✓		✓				
Social Identity	Chan et al. (2012)								✓
Product Design	Schnurr (2017)			✓			✓		

Though Wei et al. (2022) propose to use the proportion of singletons in the focal product's similarity network to simulate the atypicality of idea combination, the proportion itself is hard to interpret. Furthermore, it could be questioned why only singletons are considered atypical elements, given that small-sized clusters are also not commonly observed in mainstream ideas. Wang et al. (2017) employ cosine similarity to measure the novelty level of scientific research, then assigned them into three groups based on novelty scores. Using similarity scores can lead to ambiguity in terms of what the threshold of novelty is in a different scenario. Schnurr (2017) circumvent the question by using Amazon Turk to manually distin-

guish between typical and atypical design. In conclusion, prior studies have limitations from different perspectives. Both Wang et al. (2017) and Wei et al. (2022)'s approaches are valid in their respective contexts. However, there is still value in improving the interpretability of atypicality measures for broader applications. In addition, using only singletons instead of small-sized clusters, as Wei et al. (2022) do, can dilute the measurement of atypicality. Schnurr (2017)'s approach is static and hard to be automated in a large-scale setting. Consequently, there is a pressing need for a more robust and nuanced definition of atypicality. We propose a paradigm that follows the way how ideas are combined that can be easily interpreted and readily applicable to other contexts. Our proposed method also survives different robustness checks and shows a prominent effect with all outcome metrics.

Atypicality has usually been regarded as a one-side problem in prior research. Specifically, researchers learn from the perspective of initiators, but rarely from how the audience perceives the ideas. However, how well customers adopt atypicality is of great empirical implications. Yet the underlying mechanisms of atypicality in crowdfunding success from the audience's perspective have been understudied. To our knowledge, only two previous studies broadly relate to the topic of users' perception of atypicality. Fleming (2001) studies how recombination affects the usefulness and uncertainty of patents. He uses the dispersion parameter, i.e. variability of citations to measure uncertainty and mean of citations for usefulness. However, low variability does not necessarily implies low uncertainty. Customers may avoid products that are associated with high levels of uncertainty, leading to a combination of low mean and low variability in the demand for such risky products. Similarly, high citation indicates popularity rather than usefulness. This is especially true on crowdfunding platforms, where people fund projects about video games, food, or even movies. Popularity does not guarantee usefulness. Schnurr (2017) expand the study to how perceived interestingness and functionality are influenced by typical or atypical product designs. But their definition of atypicality is not based on idea combination. Instead, they artificially identify atypical designs. We fill this gap by implementing sentiment analysis on donors' comments and tracking donors' sequential behaviors after classifying projects based on their idea com-

bination types. We also conduct analysis across product types, which further testifies to the uncertainty theory.

2.3.1 Theoretical Background

Compared to typical products, atypical ones are usually more appealing. Maslow (1943) proposes the famous hierarchy of needs theory that describes a progression of needs that motivate individuals. Creativity and self-expression belong to the highest level of self-actualization. In other words, people have a fundamental need to be recognized for their unique qualities and talents. This need is met by niche products, which offer an opportunity to express oneself and stand out from the herd. Later, Berlyne (1970) further testify to the appealing effect of novelty by increasing people's evaluation. He demonstrates that people's evaluation is mediated by their cognition of *pleasingness* and *interestingness*. More empirical studies also find evidence for this claim. Yang and Smith (2009) argue that by increasing persuasive and emotional effects, creative ads exert more influence on consumers. Landwehr et al. (2013) show that the typical car design loses appeal in the long run, whereas atypical design gains more attention. Berger and Packard (2018) point out that songs with lyrics more differentiated from their genres are more popular.

Though psychological researchers support this favorable view of novelty, people have long been perplexed by the observation that organizations, scientific institutions, and decision-makers frequently reject new ideas (Staw and Cameron Ford 2004). Noseworthy and Trudel (2011) find that consumers struggle to make sense of incongruent products, essentially atypical products. As a result, atypicality negatively affects their adoption of the product. This negative effect is attributed to uncertainty by Fleming (2001). It is even the case that the more novel an idea is, the more uncertainty customers perceive (Amabile 2018). Mueller et al. (2012) also verify in their study that this 'bias against creativity' is mediated by uncertainty. People view creative ideas as novel and unpredictable, impractical and unreliable, which inevitably leads to uncertainty. Combining the positive and negative effects atypicality has on donors, we expect a dual pathways mechanism in our context.

In addition to that, time also plays an important role in the adoption of new ideas according to the theory of diffusion of innovation (Rogers et al. 2014). As Goldenberg et al. (1999) points out, generative learning process is an important property of creativity. The learning nature of innovation is not only shown by organizers borrowing from previous ideas, but also by donors adopting creative projects. Hence, we expect an atypical project in the later stage to be better received compared to the earlier stage. In the following chapters, we will further demonstrate these findings. In chapter 5.2, we construct the project similarity network. Then in chapter 5.3, we show four types of projects with different idea combination patterns. In chapter 5.4 and 5.5, we demonstrate our analysis and results. Then in chapter 5.6, we validate the results through robustness check. Lastly, chapter 5.7 concludes the paper.

Chapter 3

IMPACTS OF SMART VENDING MACHINES IN DIFFERENT URBAN SETTINGS - TENSOR COMPLETION WITH SPATIOTEMPORAL DATA

3.1 Introduction

Vending machines have become an integral part of our daily lives, providing convenient access to snacks, beverages, and various other products. With the rapid advancement of technology, the vending machine industry has undergone a significant transformation, giving rise to *smart* vending machines that offer enhanced features and services. In the United States, the vending machine market was valued at $15.21 billion in 2022, with an estimated 6.9 million machines in operation as of 2020. Similarly, in China, vending machines generated 28.91 billion yuan (approximately $4.05 billion USD) in retail revenue in 2022, accounting for nearly 97% of the unmanned retail market. The industry's robustness and potential for further expansion in both countries are evident, with approximately 300,000 vending machines now making up 80% of total unmanned retail sales in China.

Smart vending machines have revolutionized the retail landscape and significantly influenced consumer purchasing behavior. These machines not only offer a wider variety of products and services but also provide a seamless user experience through advanced technologies such as item recognition, computer vision, weight sensors, and RFID tags. The convenience and aesthetic appeal of smart vending machines has made them increasingly popular among consumers, as they can now easily access a broader range of items, including frozen goods and hot meals, thanks to the incorporation of heaters and coolers in some models.

Our research leverages a unique and comprehensive dataset from a leading manufac-

turer of technologically advanced vending machines in China. The dataset contains detailed transaction information from over 10,000 vending machines across 25 major Chinese cities, spanning the period from 2019 to 2022. These machines, which we refer to as smart vending machines, are equipped with advanced features such as digital payment acceptance through QR codes and facial recognition, item identification using camera sensors, weight sensors, and Radio Frequency Identification (RFID), as well as heaters and coolers for carrying a wider variety of products (Table 3.1). The dataset includes information on the specific item purchased, total order amount, anonymized user identification, vending machine identification, payment method, geographic location, machine type (smart or traditional), and the scene in which the machine is situated (e.g., school, hospital, train station). In total, the dataset encompasses 21 million transactions undertaken by more than 3.8 million distinct users, providing a rich and diverse source of information for our analysis.

Vending Machine Type	**Smart Machine**
Gravity	Yes
RFID	Yes
Dynamic Vision	Yes
Gravity + Vision	Yes
Closed Cabinet	No

Table 3.1: Types of Vending Machines and Their Classification as Smart Machines

Estimating the treatment effect of *smart* vending machines on sales across different scenarios presents several challenges due to the high-dimensional and sparse nature of the data. One of the primary challenges in understanding the effect of smart vending machines on sales is the lack of sufficient observations for accurate treatment effect estimation. In our research setting, we have data spanning 25 cities, 20 distinct scenes, and 81 weeks, leading to a substantial amount of missing data of over 70%. This sparsity in the data makes

it difficult to directly compare the performance of smart and traditional vending machines across different urban settings (region-scene pairs) and time periods. Moreover, the problem is further complicated by the presence of complex and heterogeneous interaction effects among regions, scenes, and time periods. The impact of smart vending machines on sales may vary significantly depending on the specific combination of these factors. For instance, the treatment effect of a smart vending machine in a hotel scene in New York City may differ from that of a similar machine in a school scene in Los Angeles.

Traditional approaches like difference-in-differences (DID) may struggle to capture the complex interactions in panel data, prompting the exploration of more capable methods such as matrix completion. DID is a common approach for estimating average treatment effects in panel data settings. It relies on the assumption that the counterfactual outcome matrix has a simple structure that captures unit and time-fixed effects. Under this assumption, DID can estimate the average treatment effect by comparing the pre-post treatment difference for the treated units to the same difference for the control units. However, DID may perform poorly if the assumption on the outcome matrix is violated, i.e., if there are more complex patterns over time and across units beyond just additive fixed effects. An alternative approach to estimating treatment effects in panel data is to view the unobserved counterfactual outcomes for treated units/periods as "missing" and use matrix completion techniques to impute them. This allows M* to have a more flexible low-rank structure. Athey et al. (2021) propose using a nuclear norm regularized estimator to impute the missing counterfactuals and estimate the average treatment effect, but they do not provide theoretical guarantees. Farias et al. (2023) propose an improved estimator that first computes a regularized estimate, then "debiases" it. They prove this achieves the optimal rate for recovering the average treatment effect under nearly minimal assumptions on the structure of the outcome matrix and the treatment assignment pattern. The matrix completion approach allows for complex, time-varying patterns in the outcome matrix and can handle very general treatment assignment mechanisms.

However, while matrix completion methods consider more complex effects compared to

traditional approaches, they still do not fully utilize the intrinsic data structure present in panel data with multiple dimensions such as regions, scenes, and time periods. In such settings, the data can be naturally represented as a tensor, which is a multi-dimensional array. Tensor completion methods, which extend matrix completion to higher-order arrays, have the potential to better capture the complex non-linear interactions among different dimensions, leading to more accurate predictions and treatment effect estimates.

Given these challenges, our research aims to address the problem of estimating the treatment effect of *smart* vending machines (treatment) compared to traditional vending machines (control) across various urban settings while accounting for the high-dimensional and sparse nature of the data. Our research aims to answer two primary questions: (1) How to predict vending machine sales in different urban settings (region-scene pairs)? (2) What is the treatment effect of deploying smart vending machines across these urban settings? By framing this problem as a tensor completion task, we seek to leverage the intrinsic structure of the data and employ advanced learning methods to predict missing sales outcomes and estimate the treatment effect accurately. Understanding the impact of smart vending machines on sales in specific urban settings can help companies make informed decisions regarding the installation and maintenance of these machines. By efficiently deploying smart and traditional vending machines based on their potential benefits in a given setting, firms can optimize their warehouse planning process and improve overall revenue management.

Our research makes several contributions to the field of causal analysis by applying advanced learning methods and value insights to the vending machine industry. First and foremost, we frame the treatment effect learning task as a high-dimensional machine learning problem. This approach deviates from traditional methods, which may struggle to handle the complex and large number of heterogeneous interaction effects present in our data. By formulating the problem in this manner, we open up the possibility of employing advanced learning methods, specifically tensor completion, to tackle the challenges associated with estimating treatment effects in high-dimensional settings. Secondly, we identify the intrinsic tensor structure of our data. By recognizing that our data naturally fits into a 3D tensor for-

mat, with dimensions corresponding to regions, scenes, and time, we can capture the complex interactions and dependencies among these dimensions. This tensor representation allows us to move beyond the limitations of traditional panel data analysis and matrix completion techniques, which may overlook the important role of the scene dimension in determining the treatment effect of smart vending machines. Thirdly, we provide a tensor completion-based method to estimate the treatment effect for smart vending machines, which outperforms state-of-the-art methods such as matrix completion. Our approach effectively handles the high-dimensional and sparse nature of the data, leveraging the low-rank structure of the tensor to predict missing sales outcomes and estimate the treatment effect across various urban settings.

Furthermore, our research highlights the importance of utilizing advanced learning methods in causal analysis. Our work showcases how machine learning techniques, specifically tensor completion, can be effectively applied to address these challenges and provide more accurate sales prediction and treatment effects estimation. The implications of our contributions extend beyond the vending machine industry. The framework and methodology we propose can be applied to various domains where high-dimensional and spatiotemporal data are prevalent, such as retail, healthcare, and transportation.

The rest of this paper is organized as follows. We describe the problem formulation in Section 3.2. Section 3.3 introduces the proposed tensor completion method, HQ-TCASD. The estimation procedure is presented in Section 3.4.1, followed by the benchmarks used for comparison in Section 3.4.2. Synthetic tests and their results are discussed in Section 3.4.3. Finally, Section 3.5 concludes the paper and provides directions for future research.

3.2 Problem Formulation

In this section, we explain how we shape the vending machine sales outcomes as tensors and introduce our causal problem: estimating the average treatment effect for the treated (ATT) of deploying smart vending machines compared to traditional vending machines across various urban settings.

We use calligraphic letters (e.g. $\mathcal{Y}$) for tensors, bold capital letters for matrices (e.g., $\mathbf{B}$), and non-bold capital letters for vectors (e.g., V). However, there are exceptions, we use non-bold capital letters $N1$, $N2$, and T for the dimensions of the tensor. For a positive integer m, we denote the set of integers $\{1, \ldots, m\}$ by $[m]$. For a tensor $\mathcal{Y}$, $\mathcal{Y}_{ijk}$ refers to its entry at (i, j, k), $\mathcal{Y}_{i::}$ refers to its i-th horizontal slice (fixing the first dimension), $\mathcal{Y}_{:j:}$ to its j-th lateral slice (fixing the second dimension), and $\mathcal{Y}_{::k}$ to its k-th frontal slice (fixing the third dimension). For a matrix $\mathbf{B}, \mathbf{B}_{ij}$ refers to its entry at (i, j), $\mathbf{B}_i$ refers to its i-th row, and $\mathbf{B}_j$ refers to its j-th column.

3.2.1 Vending Machine Sales Potential Outcome Tensors

There are $N_1 = 25$ regions corresponding to 25 major Chinese cities where the vending machines are installed; there are $N_2 = 20$ scenes corresponding to 20 scenarios that each machine is in (such as educational institutions or medical facilities); and there are $T = 81$ time periods corresponding to 81 consecutive weeks from October 2020 to May 2022. Figure 3.1 illustrates such $N_1 \times N_2 \times T$ tensor $\mathcal{Y}$ with T slices, where each slice is an $N_1 \times N_2$ matrix. For these two types of machines, we can represent the potential vending machine sales with two $N_1 \times N_2 \times T$ potential outcome tensors: a potential control tensor $\mathcal{Y}(0)$ representing the potential sales outcomes if all units were control (traditional) machines, and a potential treated tensor $\mathcal{Y}(1)$ representing the potential sales outcomes if all tuples were treated (smart) machines.

In this paper, we consider the deployment of smart vending machines as a treatment for traditional machines, and we focus on the treatment effect of deploying smart vending machines.

For the observed entries, we define $\mathcal{P}_\Omega(\mathcal{Y})$ such that:

$$P_\Omega(\mathcal{Y})(i,j) = \begin{cases} \mathcal{Y}_{ijt} & \text{if } (i,j,t) \in \Omega \\ 0 & \text{if } (i,j,t) \notin \Omega, \end{cases},$$

where $\Omega \subset \{1, \ldots, N_1\} \times \{1, \ldots, N_2\} \times \{1, \ldots, T\}$ denote the indices of observed sales out-

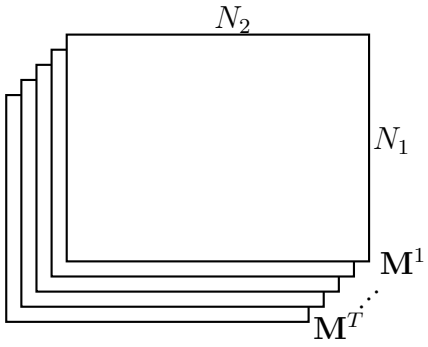


Figure 3.1: Illustration of a $N_1 \times N_2 \times T$ tensor with T slices of $N_1 \times N_2$ matrices $M^1 ... M^T$.

come. $P_\Omega(\mathcal{Y})$ is a projection of the tensor $\mathcal{Y}$ onto the observed entries. However, in reality, at each region-scene-time period tuple (i, j, t) we would either observe sales from smart machines, traditional machines, or no sales at all. So we define a treatment indicator tensor $\mathcal{W}$, where $\mathcal{W}_{ijt} = 1$ indicates the observed sales is from treated units (smart vending machines) for region-scene pair (i, j) at time period t, and $\mathcal{W}_{ijt} = 0$ indicates the observed sales is from control units (traditional vending machines). In addition, we let $\mathcal{W}_{ijt}$ be missing if the no sales observed.

With the definition of $\mathcal{W}$, we could further define the observed treated sales tensor $P_\Omega(\mathcal{Y}(1))$ and the observed control sales tensor $P_\Omega(\mathcal{Y}(0))$. For example, if we observe sales outcome $P_\Omega(\mathcal{Y})_{::t}$, a $N_1 \times N_2$ slice of tensors $\mathcal{Y}$ at time period t, then we could define $\mathcal{W}_{::t}$:

$$P_\Omega(\mathcal{Y})_{::t} = \begin{pmatrix} \mathcal{Y}_{11t} & \mathcal{Y}_{12t} & 0 & \dots & \mathcal{Y}_{1N_1t} \\ 0 & 0 & \mathcal{Y}_{23t} & \dots & 0 \\ \mathcal{Y}_{31t} & 0 & 0 & \dots & \mathcal{Y}_{3N_1t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{Y}_{N_21t} & 0 & \mathcal{Y}_{N_23t} & \dots & 0 \end{pmatrix}, \quad \text{and} \quad \mathcal{W}_{::t} = \begin{pmatrix} 0 & 1 & ? & \dots & 0 \\ ? & ? & 0 & \dots & ? \\ 0 & ? & ? & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & ? & 0 & \dots & ? \end{pmatrix};$$

then the sales outcome slice $P_\Omega(\mathcal{Y})_{::t}$ can be split into $P_\Omega(\mathcal{Y}(1))_{::t}$ and $P_\Omega(\mathcal{Y}(0))_{::t}$:

$$P_\Omega(\mathcal{Y}(1))_{::t} = \begin{pmatrix} 0 & \mathcal{Y}_{12t} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \mathcal{Y}_{3N_1t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{Y}_{N_21t} & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \text{and} \quad P_\Omega(\mathcal{Y}(0))_{::t} = \begin{pmatrix} \mathcal{Y}_{11t} & 0 & 0 & \dots & \mathcal{Y}_{1N_1t} \\ 0 & 0 & \mathcal{Y}_{23t} & \dots & 0 \\ \mathcal{Y}_{31t} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \mathcal{Y}_{N_23t} & \dots & 0 \end{pmatrix}.$$

Since we will mainly discuss the observed sales outcomes, for simplicity, for the rest of the paper we will use $\mathcal{Y}$ instead of $P_\Omega(\mathcal{Y})$ to discuss the observed sales outcome. And we refer to the potential sales outcome as *potential* $\mathcal{Y}$.

3.2.2 Causal Problem

We want to estimate the treatment effect of deploying smart vending machines (treatment) compared to traditional vending machines (control) in different urban settings across time, i.e., region-scene pairs, as aforementioned. Similar to Athey et al. (2021), we focus on estimating the ATT as defined below:

$$\tau_{ATT} = \sum_{(i,j,t):\mathcal{W}_{ijt}=1} \left[\mathcal{Y}_{ijt}(1) - \hat{\mathcal{Y}}_{ijt}(0)\right] / \sum_{(i,j,t):\mathcal{W}_{ijt}=1} \mathcal{W}_{ijt}.$$

Note that ATT only focuses on the treatment effect on the observed treated entries. The ATT formula suggests that we only need to impute the missing entries in $\mathcal{Y}(0)$ for tuple (i, j, t) since we observed the treated outcome in $\mathcal{Y}(1)$, i.e., $\mathcal{W}_{ijt} = 1$. Thus, we translate the causal inference problem into a missing data problem of imputing $\mathcal{Y}(0)$.

3.3 Tensor Completion as a Causal Inference Method

In this section, we introduce how we apply tensor completion to our causal inference task to estimate the ATT for deploying smart vending machines using the sales outcome tensor. We use the observed values in the $N_1 \times N_2 \times T$ control tensor to impute the missing entries in $\mathcal{Y}(0)$ for treated units. The imputed values represent the counterfactual outcomes for the treated units had they not received the treatment.

3.3.1 Low-Tubal-Rank Observation Model

Inspired by the literature on tensor completion, we model the observed control sales outcome $\mathcal{Y}$ with fixed effects and tensor factor model. We start with fixed effects the three dimensions:

$$\mathcal{Y}_{ijt}(0) := \beta_0 + a_i + b_j + c_t + \varepsilon_{ijt} \tag{3.1}$$

where a_i, b_j, c_t are region, scene, and time period specific fixed effects; β_0 is the intercept term, and ε_{ijt} is the independent error term, e.g., observation error.

Let $\mathcal{L}^* \in \mathbb{R}^{N_1 \times N_2 \times T}$ be a fixed low-tubal-rank r counterfactual tensor which indicates the expected sales outcomes absent treatment and noise, then $\mathcal{L}^*_{ijt}$ denotes the expected sales outcome for tuple (i, j, t). Kilmer et al. (2013) define the tensor tubal rank as follows: for any $\mathcal{L} \in \mathbb{R}^{N_1 \times N_2 \times T}$, the tensor tubal rank, $\text{rank}_t(\mathcal{L})$, is the number of non-zero singular tubes of $\mathcal{S}$ from the t-SVD (described in Appendix A.2.2), i.e.,

$$\text{rank}_t(\mathcal{L}) = \#\{i : \mathcal{S}(i, i, :) \neq 0\}.$$

Intuitively, the tubal rank captures the intrinsic dimension of tensors, which can be viewed as the counterpart of the rank of a matrix. It can be shown that, when $T = 1$, multiplications and factorizations based on the t-product will reduce to the standard matrix operations and factorizations. This allows researchers to generalize tools like singular value decomposition (SVD), QR decomposition and eigendecompostions to third-order tensors (Kilmer et al. 2013). Methods for estimating the rank R in matrix setups are discussed in Bai and Ng (2002) and Moon and Weidner (2015)). With the tubal rank definition, the low-tubal-rank assumption imposed on the sales outcome tensor $\mathcal{L}$ is as follows:

Assumption 1. We assume that tensor $\mathcal{L}$ is low-tubal-rank and

$$\text{rank}_t(\mathcal{L}) \ll \max\{N_1, N_2\}.$$

With the defined $\mathcal{L}^*$, we can model sales outcome tensor $\mathcal{Y}$, in the absence of covariates, as

$$\mathcal{Y}(0) := \beta_0 + \mathcal{L}^* + \varepsilon, \quad \text{where} \quad \mathbb{E}[\varepsilon \mid \mathcal{L}^*] = \mathbf{0}. \tag{3.2}$$

The goal is to estimate the tensor $\mathcal{L}^*$. With assumption 1, the tensor $\mathcal{L} \in \mathbb{R}^{N_1 \times N_2 \times T}$ can be factorized into the tensor-tensor product (t-product, denoted by operator $*$) of two tensors through low-tubal-rank factorization $\mathcal{L} = \mathcal{U} * \mathcal{V}$, where $\mathcal{U} \in \mathbb{R}^{N_1 \times r \times T}$ and $\mathcal{V} \in \mathbb{R}^{r \times N_2 \times T}$, and r is the tubal rank of $\mathcal{L}$ (Zhou et al. 2018). Note that since r is small, $\mathcal{U}$ and $\mathcal{V}$ are much smaller tensors than the original sales outcome tensor. The tensor-tensor product (t-product, introduced by Kilmer and Martin (2011)) of $\mathcal{U}$ and $\mathcal{V}$ is defined as

$$\mathcal{U} * \mathcal{V} = \text{fold}(\text{bcirc}(\mathcal{U}) \cdot \text{unfold}(\mathcal{V})).$$

The definition of bcirc(), fold(), and unfold() is defined in Appendix A.2.1. In the canonical domain, a three-way tensor of size $N_1 \times N_2 \times T$ can be considered as an $N_1 \times N_2$ matrix whose entries are $1 \times 1 \times T$ tubes lying in the third dimension. The t-product is then analogous to matrix-matrix multiplication, but where circular convolution replaces scalar multiplication between the matrix elements. Now we can rewrite 3.2 with the tensor factorization:

$$\mathcal{Y}(0) : \beta_0 + \mathcal{U} * \mathcal{V} + \varepsilon.$$

Analogous to matrix factor models, this setup can account for complex interaction effects between the three dimensions. Note that here the fixed effects are absorbed in $\mathcal{L}^*$ since they are rank 1 tenors and their addition does not affect our low-rank assumption on $\mathcal{L}^*$. But we separate fixed effects from the factorization, according to Athey et al. (2021), in practice, not regularizing the fixed effect terms will improve the quality of the imputations:

$$\mathcal{Y}_{ijt}(0) := \beta_0 + a_i + b_j + c_t + (\mathcal{U} * \mathcal{V})_{ijt} + \epsilon_{ijt} \tag{3.3}$$

3.3.2 Tensor Completion Algorithm

We adapt the low-rank tensor completion algorithm by He and Atia (2022) to estimate the ATT of deploying smart vending machines. To recover $\mathcal{Y}(0)$, our objective is to find a combination of $\widehat{\mathcal{U}}$, $\widehat{\mathcal{V}}$, $\hat{a}$, $\hat{b}$, and $\hat{c}$ that minimizes the sum squared error between the observed

value and the predicted value for all entries in the control tensor ($\mathcal{Y}_{ijt} = 0$):

$$\widehat{\mathcal{U}}, \widehat{\mathcal{V}}, \hat{a}, \hat{b}, \hat{c}, = \arg \min_{a,b,c,\mathcal{U},\mathcal{V}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} (1 - \mathcal{W}_{ijt}) \left(\mathcal{Y}_{ijt} - \beta_0 - a_i - b_j - c_t - (\mathcal{U} * \mathcal{V})_{ijt} \right)^2 . \quad (3.4)$$

The the low-tubal-rank constraint is employed through tensor factorization where $\mathcal{Y} \in \mathbb{R}^{N_1 \times N_2 \times T}$, $\mathcal{U} \in \mathbb{R}^{N_1 \times r \times T}$, and $\mathcal{V} \in \mathbb{R}^{r \times N_2 \times T}$. The objective is to ensure that the recovered values for the observed entries are not too far away from the actual observations and a low-rank constraint. Tensor factorization can avoid the high complexity associated with performing the SVD, which is reduced due to the inherent low-rank property. Low-tubal-rank-based tensor completion offers several advantages over tensor completion using other tensor rank models (e.g., CP rank, Tucker rank, and tensor ring rank). Other methods usually impose low-rank constraints through the nuclear norm minimization on the unfolded matrices of the tensor, which may destroy the original multi-dimensional structure of the tensor data. By contrast, based on tensor algebra, the tubal-rank-based methods directly impose a low-tubal-rank constraint on a tensor, and can well capture the inherent low-rank structure of a tensor.

When the observed entries $\mathcal{Y}_{ijt}$ are corrupted or contain large outliers, the ℓ_2 error measure can bias the optimization, which degrades the performance of tensor completion. To enhance robustness, He and Atia (2022) use the correntropy as the error measure. Correntropy is a local and nonlinear similarity measure between two random variables. Given a finite number of samples $\{x_i, y_i\}_{i=1}^N$ and using the Gaussian kernel $G_\sigma(x) = \exp(-\frac{x^2}{2\sigma^2})$, The Maximum Correntropy Criterion (MCC) can be formulated as minimizing the correntropy-induced loss (C-loss) function:

$$J_{C-\text{loss}} = \frac{1}{M} \sum_{i=1}^{M} \sigma^2 \left(1 - G_\sigma(e(i))\right). \quad (3.5)$$

By replacing the ℓ_2 error measure with correntropy, the objective function becomes

$$\min_{\mathcal{U},\mathcal{V}} J_{G_\sigma}(\mathcal{U}, \mathcal{V}) := \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} \mathcal{W}_{ijt} \sigma^2 (1 - G_\sigma \left(\mathcal{Y}_{ijt} - \beta_0 - a_i - b_j - c_t - (\mathcal{U} * \mathcal{V})_{ijt} \right)). \quad (3.6)$$

He and Atia (2022) proposed Half-Quadratic based Tensor Completion by Alternating Steepest Descent (HQ-TCASD) to solve the objective function. The formulation in (3.6) generalizes the correntropy-based formulation for matrix completion. HQ-TCASD adaptively selects the kernel width σ to improve convergence rate and performance. The algorithm is summarized in Algorithm 1 in the Appendix A.3.4, with the matrices $\boldsymbol{M}$ and $\hat{\boldsymbol{V}}$ updated alternately until convergence. Please see Appendix A.3 for more details regarding the tensor completion algorithm.

Investigating the asymptotic distribution of $\mathcal{L}^{-}\hat{\mathcal{L}}$ to develop confidence intervals is not within the scope of this paper, but it presents an intriguing opportunity for future research. However, re-sampling techniques (Alberto Abadie and Hainmueller 2010) can be utilized to examine the statistical fluctuations of the imputed tensor.

3.4 Empirical Results & Managerial Insights

We analyze the effects of smart vending machines using tensor completion on a real-world dataset. The data is structured as a three-dimensional tensor, and we estimate the average treatment effect on the treated (ATT) and the overall average treatment effect (ATE). Our proposed tensor completion algorithm, HQ-TCASD, is compared against the generalized Difference-in-Differences (DID) method and the matrix completion (MC) algorithm De-biased Convex Panel Regression (DC-PR). We evaluate their performance in predicting sales outcomes and estimating treatment effects using real-world data and synthetic datasets with known ground truth effects. We also examine the heterogeneity of treatment effects across different urban settings to identify contexts where smart vending machines have the most significant impact on sales.

3.4.1 Estimations

In this section, we present our analysis and findings on estimating the treatment effects of smart vending machines using tensor completion. We begin by introducing the vending machine dataset and discussing how we structured it into a three-dimensional tensor. We

then explain our methodology for estimating the average treatment effect on the treated (ATT).

Vending Machine Dataset

The vending machine sales dataset used in this study contains comprehensive and granular transaction data from various cities across China. The dataset encompasses itemized sales records from over 10,000 vending machines, serving more than 3 million users and generating over 21 million transactions between January 2019 and May 2022. The dataset includes several key variables: order amount, item details, machine locations, machine scenes, machine types, and timestamps. The machine locations are specified by coordinates, which can be used to identify the city in which the vending machine is located, such as Beijing, Shanghai, Chongqing, or Wuhan. The machine scenes variable indicates the surrounding settings of the vending machines, including schools, hotels, parks, and stations, providing valuable context for understanding the sales patterns in different environments. Furthermore, the dataset distinguishes between two machine types: smart vending machines, which are equipped with advanced technology and features, and traditional vending machines, which lack these enhancements. The timestamp variable records the time of each transaction in seconds.

The weekly sales data exhibits a long-tail distribution, with 95% of the sales not exceeding 10,000 CNY, despite the maximum recorded sales surpassing 130,000 CNY. This disparity poses challenges for analysis and modeling, as the presence of extreme values can skew results and make it difficult to identify underlying patterns. To address this issue and standardize the distribution, log transformation and Z-score normalization were applied. Figure 3.2 illustrates the distribution of the data after applying the log normalization to the weekly sales. The transformed data exhibits a symmetrical and bell-shaped distribution, albeit with a long left tail.

To enable tensor completion, the data was organized into a three-dimensional tensor $\mathcal{Y}$ with dimensions corresponding to Region, Scene, and Time Period. The Region dimension encompasses 25 major Chinese cities where the vending machines are deployed, while the

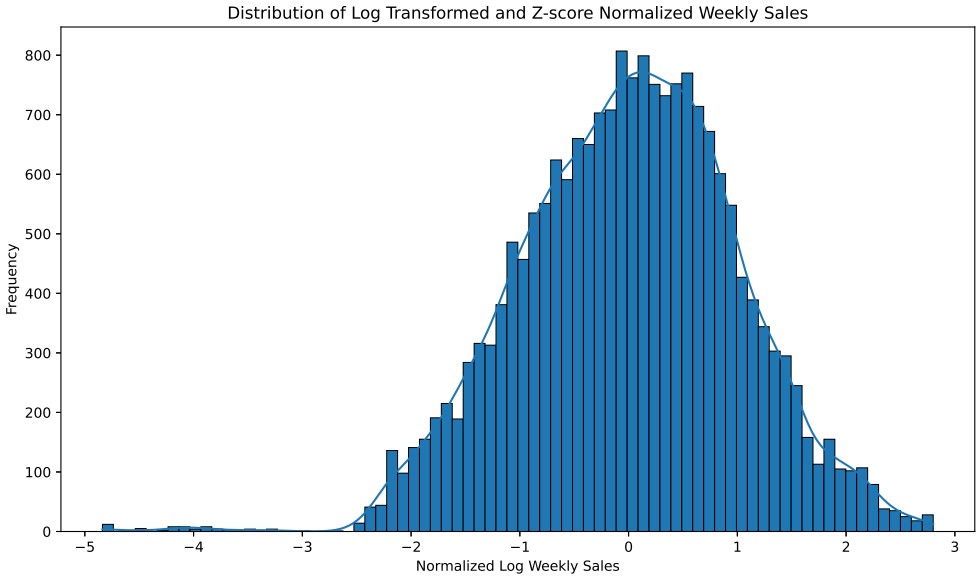


Figure 3.2: Histogram of log-normalized weekly sales

Scene dimension includes 20 categories describing the specific settings of the vending machines, such as educational institutions or hospitals. The Time Period dimension spans 81 continuous weeks from October 2020 to May 2022. Each tensor entry, denoted as Y_{ijt}, represents the cumulative sales across all machines in region i, under scene j, during time period t. The treatment variable $\mathcal{W}_{ijt}$ indicates whether vending machines at this region-scene-time period tuple are smart machines or traditional machines. Figures A1 and A2 show the log-normalized observed results for $\mathcal{Y}(0)$ and $\mathcal{Y}(1)$ by slice, respectively. These visualizations provide an overview of the observed data entries in both the control and treated tensor for each of the 81 time periods; the color, from dark blue to yellow, denotes the magnitude of the sales at each entry.

The pattern of missing entries and random treatment adoption in the dataset pose challenges for traditional causal inference methods. Approximately 58.57% of the entries in sales outcome tensor $\mathcal{Y}$ are missing, primarily due to the absence of vending machines in certain combinations of region and scene or the lack of recorded data for specific weeks. The control

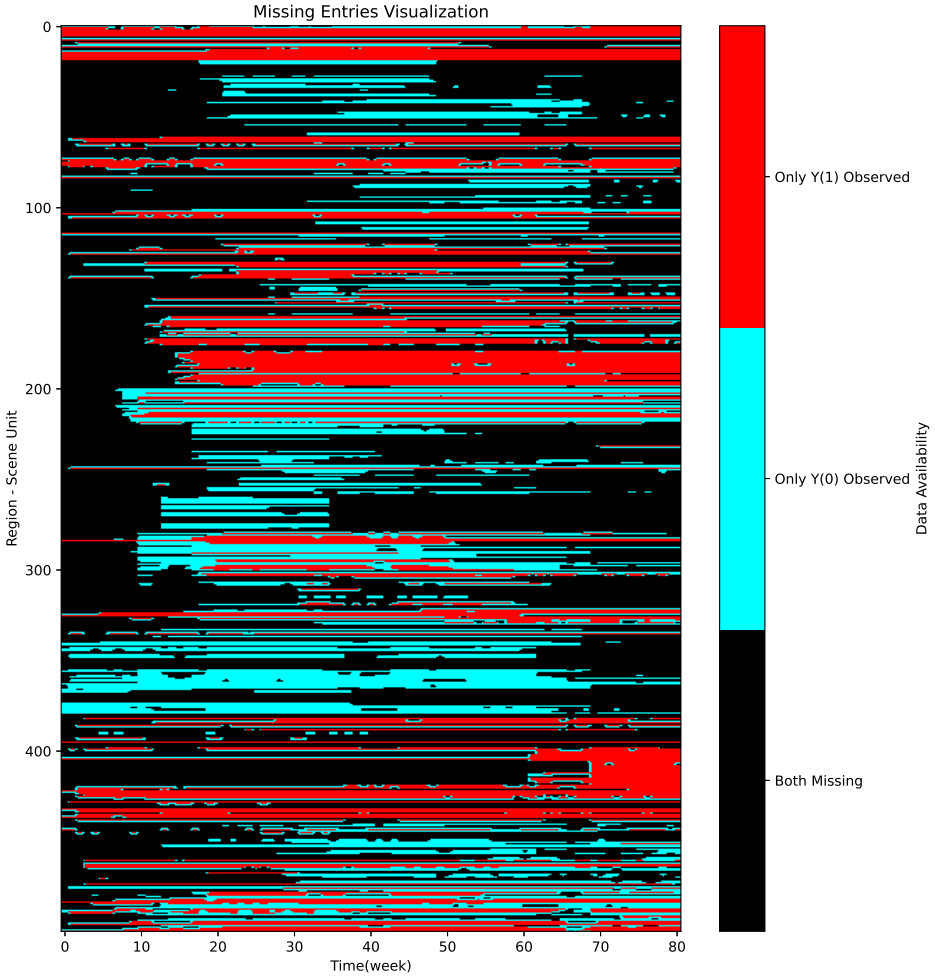


Figure 3.3: Visualization of Missing Entries. The color-coded matrix represents data availability, with different colors indicating the presence or absence of data across region-scene and time.

tensor $\mathcal{Y}(0)$ has a 31.26% non-missing rate, and the treated tensor $\mathcal{Y}(1)$ has a 26.74% non-missing rate, suggesting that smart vending machines have a slightly higher proportion of

missing data compared to traditional machines. Figure 3.3 visualizes the pattern of missing entries for $\mathcal{Y}$ in an unfolded matrix format. The x-axis represents 500 region-scene pairs, and the y-axis represents 81 time periods. Each row in the matrix corresponds to the sales from a specific region-scene pair at each time period. The color coding indicates the type of data observed at each region-scene-time period tuple: black for missing entries, blue for sales from control units (traditional vending machines), and red for sales from treated units (smart vending machines). Moreover, Figure 3.3 shows that the pattern of treatments is random, with no clear treatment adoption timeline across different region-scene pairs. Moreover, treated units do not necessarily remain treated after the initial treatment, with some reverting to being untreated in subsequent time periods. This random treatment pattern violates assumptions of traditional causal inference methods, such as the Difference-in-Differences (DID) approach, necessitating the use of more flexible and robust methods for estimating the treatment effect for smart vending machines.

Treatment Effect Estimation

Having prepared the tensor for the tensor completion algorithm, we now examine the empirical results from the vending machine dataset. After applying tensor completion on the $\mathcal{Y}(0)$ tensor, we obtain the estimated control outcomes for the treated units. Visualizations of the estimated log-normalized sales for $\hat{\mathcal{Y}}(0)$ and $\hat{\mathcal{Y}}(1)$ by slice are presented in Figure A3 and Figure A4, respectively. The color scale, ranging from dark blue to yellow, denotes the magnitude of the estimated sales at each entry.

Using formulation mentioned in Section 3.2.2, we are able to estimate the ATT. To interpret the results, we reverse the log-normalization process on the estimated tensors. After the transformation, the estimated ATT is 967.52 CNY, meaning on average, smart vending machines have 967.52 CNY higher weekly sales than traditional vending machines. The estimated $\hat{\tau}$ represents the average difference in outcomes between the treated and control groups after accounting for the fixed effects. A positive $\hat{\tau}$ suggests that the treatment had a beneficial impact on the outcome variable, while a negative $\hat{\tau}$ indicates an adverse effect.

3.4.2 Benchmarks

In this section, we implement Difference-in-Differences (DID) and matrix completion algorithm De-biased Convex Panel Regression (DC-PR) as benchmark methods for estimating treatment effects in panel data settings. We compare the performance of these methods with our proposed tensor completion algorithm, HQ-TCASD, by evaluating the estimated testing dataset root mean square error (RMSE). Finally, We report the statistics for the average treatment effect of the treated (ATT) from three algorithms.

To apply the benchmark methods, DID and DC-PR, we first transform the sales outcome tensor into a panel data format. We couple regions and scenes as region-scene pair dimension P, with index p, and keep the time period dimension T unchanged. The transformation unfolds the $25 \times 20 \times 81$ sales outcome tensor into a matrix $\mathbf{Y} \in \mathbb{R}^{P \times T}$ with dimensions 500×81. $\mathbf{Y}_{pt}$ represents the sales outcome for region-scene pair p for week t. In addition to the outcome matrix $\mathbf{Y}$, we also construct a treatment indicator matrix $\mathbf{Z} \in \mathbb{R}^{P \times T}$. $\mathbf{Z}_{pt}$ is a binary indicator that takes the value of 1 if we observed treated sales outcome (smart vending machine) for region-scene pair p at time period t, and 0 if observed control sales outcome (tradition vending machine). Similar to the setup to $\mathcal{W}$, we set $\mathbf{Z}_{pt}$ to missing if no outcome is observed.

Difference-in-Difference

Difference-in-difference (DID) is a widely used approach for analyzing panel data and estimating the causal effect of a treatment or intervention. The method compares the change in outcomes over time between a treated group and a control group, assuming that both groups would have followed parallel trends in the absence of the treatment. In this study, we employ a two-way fixed effects regression, a generalization of the standard DID estimator, to estimate the ATT.

The two-way fixed effects model extends the DID idea to settings with multiple time periods and multiple units. By including unit and time fixed effects, the model accounts for

any time-invariant differences between units and any common time trends affecting all units. They assume the observational model:

$$Y_{pt} = \beta_0 + \tau Z_{pt} + a_p + b_t + \epsilon_{pt} \tag{3.7}$$

The objective of a two-way fixed effects regression is to minimize the sum of squared residuals between the observed outcomes and the predicted values (Athey et al. 2021). Formally, we solve the following linear regression problem:

$$\hat{a}, \hat{b}, \hat{\tau} = \arg\min_{a,b,\tau} \sum_{pt} \left(\mathbf{Y}_{pt} - a_p - b_t - \tau \mathbf{Z}_{pt}\right)^2 \tag{3.8}$$

where $\hat{a} = (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p)$ is the vector of estimated unit fixed effects, $\hat{b} = (\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_T)$ is the vector of estimated time fixed effects, $\hat{\tau}$ is the estimated ATT. The objective function finds the combination of $\hat{a}$, $\hat{b}$, and $\hat{\tau}$ that solves the least square problem between the observed outcomes ($\mathbf{Y}_{pt}$) and the predicted outcomes ($a_p + b_t + \tau \mathbf{Z}_{pt}$).

De-biased Convex Panel Regression

We also consider the state-of-the-art matrix completion algorithm, De-biased Convex Panel Regression (DC-PR), introduced by Farias et al. (2023), extending the work by Athey et al. (2021) to a more generalized setting.

The limitation of the DID model is that the additive model of unit fixed effect and time fixed efffect $a_p + b_t$ is often too simplistic to capture the complex nature of the outcome variable. To address this, a low-rank factor model has been proposed as a generalization of $a_p + b_t$.

The observation model is given by:

$$\begin{gathered} \mathbf{Y}_{pt} = \beta_0 + \tau \mathbf{Z}_{pt} + a_p + b_t + \mathbf{U}_p \mathbf{V}_t^T + \epsilon_{pt} \\ \mathbf{U}_p \in \mathbb{R}^r, \mathbf{V}_t \in \mathbb{R}^{\mathrm{r}} \end{gathered} \tag{3.9}$$

where $\mathbf{Y}_{pt}$ is the outcome variable for unit p at time t, β_0 is the intercept, τ is the treatment effect, Z_{pt} is the binary treatment indicator, a_p and b_t are the unit and time fixed effects, and $\mathbf{U}_p$ and $\mathbf{V}_t$ are r-dimensional latent factors.

The objective function solves the following low-rank regression problem with a hard-rank constraint:

$$\hat{\mathbf{U}}\hat{\mathbf{V}}^T, \hat{\tau} = \arg \min_{\text{rank}(\mathbf{U}\mathbf{V}^T) \leq r} \sum_{pt} \left(\mathbf{Y}_{pt} - a_p - b_t - \mathbf{U}_p \mathbf{V}_t^T - \tau \mathbf{Z}_{pt} \right)^2 \tag{3.10}$$

where $\hat{\mathbf{U}}\hat{\mathbf{V}}^T$ is the estimated low-rank matrix (panel data), and $\hat{\tau}$ is the estimated treatment effect. The objective is to find the values of a, b, $\mathbf{U}$, $\mathbf{V}$, and τ that minimize the sum of squared residuals between the observed outcomes ($\mathbf{Y}_{pt}$) and the predicted outcomes ($a_p + b_t + \mathbf{U}_p \mathbf{V}_t^T + \tau \mathbf{Z}_{pt}$) across all region-scene pairs and time periods.

Then DC-PR mitigates the bias introduced by the regularization parameter λ:

$$\tau^d = \hat{\tau} - \lambda \frac{< \mathbf{Z}, \hat{\mathbf{U}}\hat{\mathbf{V}}^\top >}{\| P_{\hat{\mathbf{T}}^\perp}(\mathbf{Z}) \|_F^2}.$$

The orthogonal space of $\mathbf{T}^*$, denoted as $\mathbf{I}^{*\perp}$, is the subspace of $\mathbb{R}^{n \times n}$ with columns and rows that are orthogonal to the spaces $\mathbf{U}^*$ and $\mathbf{V}^*$, respectively. $P_{\mathbf{I}^{*\perp}}(\cdot)$ represents the projection operator onto $\mathbf{T}^{*\perp}$:

$$P_{T^{*\perp}}(A) = \left(\mathbf{I} - \mathbf{U}^* \mathbf{U}^{*\top} \right) A \left(\mathbf{I} - \mathbf{V}^* \mathbf{V}^{*\top} \right).$$

The De-biased Convex Panel Regression algorithm offers several advantages over existing approaches. It allows for the application to a wide range of settings with general treatment patterns.

Results

To obtain benchmark estimations, we generate 100 training and testing sets by randomly splitting the vending machine sales dataset, with 80% assigned to the training set and 20% to the testing set. We train three methods: tensor completion (TC) algorithm HQ-TCASD, generalized Difference-in-Differences (DID) method two-way fixed effects regression, and matrix completion (MC) algorithm DC-PR, and evaluate their performance in predicting the corresponding testing set.

Figure 3.4 compares the performance of the three estimation algorithms. The TC algorithm achieves the lowest root mean square error (RMSE) when predicting entries in the

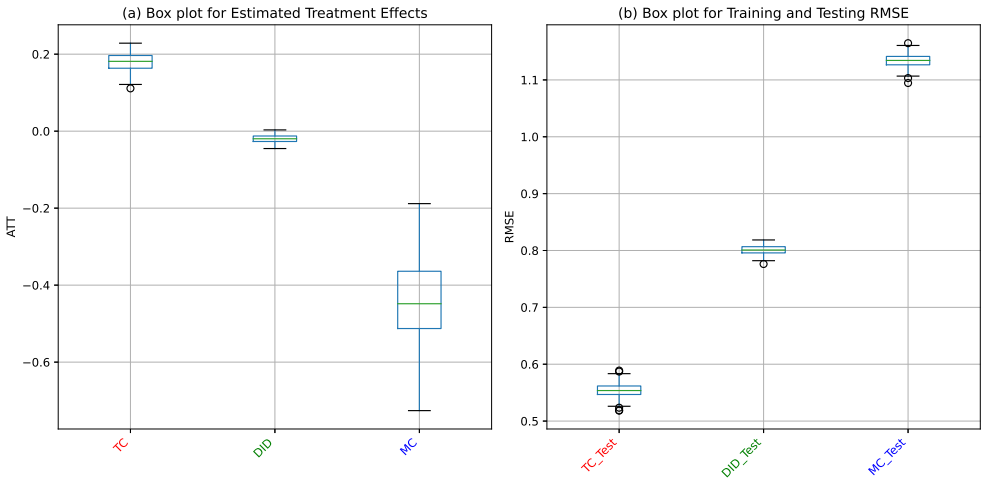


Figure 3.4: (a) Box plot of estimated treatment effects from TC, DID, and MC on normalized sales data (b) Box plot of root mean square error (RMSE) for both training and testing datasets from TC, DID, and MC.

testing dataset, suggesting it effectively identifies underlying patterns and makes accurate predictions. The DID method shows a higher testing RMSE compared to TC, possibly due to its assumptions or approach to handling temporal variations and control group comparisons. The MC algorithm has the highest testing RMSE, which could be the result of overfitting. Although the true treatment effect is unknown, TC estimates the highest average treatment effect on the treated (ATT), followed by DID, and then MC. The variation in the ATT estimates is small for TC and DID but large for MC, suggesting TC and DID's estimates are more consistent across different data subsets.

Given TC's higher accuracy in predicting vending machine sales, as evidenced by the low testing RMSE, we can conclude that its ATT estimates are more reliable. The low RMSE values indicate that the algorithm is learning the underlying relationships between variables, which is crucial for accurate treatment effect estimation. In contrast, MC's high variance in ATT estimates and poor performance on the testing set raises concerns about the reliability of its treatment effect estimates. DID's ATT estimates, while lower than TC's, show lower

variability, indicating consistency across data subsets. However, given DID's higher testing RMSE, its ATT estimates may be less accurate.

In summary, the benchmark results highlight the effectiveness of the tensor completion algorithm, TC, in estimating the treatment effect of deploying smart vending machines. Its low testing RMSE values suggest that TC performs better in handling the complexities of the data and provides more accurate treatment effect estimates compared to DID and MC.

Treatment Effects by Urban Settings

Building upon the estimations for the control tensor $\hat{\mathcal{Y}}(0)$, we conduct a comprehensive analysis of the average treatment effect of the treated (ATT) across different urban settings characterized by region-scene pairs. We calculate the ATT for each region-scene pair over time. Furthermore, we aggregate the results to single-dimension levels of region and scene to study the effects at a more consolidated level. This analysis allows us to identify the specific urban settings, scenes, and regions where smart vending machines have the greatest impact, both positively and negatively, on sales.

Table 3.2 presents the summary statistics for the top 20 region-scene pairs with the highest average ATT. The results show heterogeneous treatment effects across different regions. For instance, university dormitories in Beijing exhibit a highly positive ATT of 60,611 CNY, while those in Shaanxi Hanzhong and Shaanxi Shangluo have much lower, though still positive, ATT of 8,867 CNY and 5,834 CNY, respectively. Similarly, within the same region, the treatment effect differs across scene types. In Beijing, university dormitories, various stations, and university teaching buildings all show strong positive effects, while hospitals have a more modest positive effect. This suggests that both geographical location and scene play significant roles in determining the impact of smart vending machines on sales.

Table 3.3 presents the ATT aggregated by scene type. University dormitories exhibit the most positive effect, with an ATT of 4,131 CNY, indicating that college students are a key driving force behind the sales of smart vending machines. This can be attributed to several factors, including college students' greater familiarity with technology, newly acquired finan-

Table 3.2: Summary statistics for ATT by Region and Scene

Region-Scene	**Mean**	**Std**	**Min**	**Max**
Beijing_University Dormitory	60611.62	44933.41	1401.66	128889.62
Beijing_Various Stations	21176.14	15626.55	386.17	70328.22
Beijing_University Teaching Building	20063.19	18408.21	123.30	59232.43
Hubei Wuhan_Various Stations	18704.02	11890.12	2338.64	61030.05
Jiangsu Yancheng_Factory	17494.85	11655.71	-679.29	40885.56
Beijing_Administrative Office	9718.20	8311.33	-312.29	33500.81
Shaanxi Hanzhong_University Dormitory	8867.44	8139.32	-568.01	33169.17
Shandong Qingdao_Various Stations	6507.56	3603.83	168.63	19347.12
Jiangsu Yancheng_University Teaching Building	6461.61	8438.70	-5640.24	37574.00
Jiangsu Changzhou_Factory	6176.56	3583.14	168.77	13966.90
Ningxia Yinchuan_Hospital	6071.34	3990.96	-694.23	17583.25
Shaanxi Shangluo_University Dormitory	5834.27	4341.25	-291.86	16014.75
Beijing_Hospital	5128.39	3378.16	-2434.68	14052.22
Jiangsu Yancheng_Employee Dormitory	4735.76	2738.81	-253.09	9989.57
Chongqing_Various Stations	4182.05	2979.48	-64.37	13525.99
Xinjiang Urumqi_Hospital	4173.15	1755.85	-1479.31	8164.57
Ningxia Yinchuan_University Dormitory	3682.40	4046.70	-466.81	14662.80
Jiangsu Yancheng_Hospital	3330.74	1612.91	-258.15	6845.94
Chongqing_Factory	3199.99	599.95	2014.97	4486.38
Henan Zhengzhou_Various Stations	3007.63	2281.93	368.95	14790.05

cial independence, and willingness to experiment with new products. Various stations and administrative office areas also show positive effects, suggesting that smart vending machines perform well in business settings characterized by diverse demand profiles. In contrast, mid-

dle schools have the most negative effect, with an ATT of -7,851 CNY, despite being an educational setting similar to university dormitories. This discrepancy can be explained by several factors. First, high school students typically have more limited budgets compared to college students, which may restrict their spending on vending machine products. Second, high school students may have less access to digital payment methods, which are often required for using smart vending machines. Finally, parents may have concerns about their children overspending on vending machine products, leading to lower sales in middle school settings in China. University canteens and sports venues also exhibit negative effects, possibly because they have single-demand profiles. Smart vending machines, which often offer a wider variety of products, may not be as well-suited for these settings where consumers have more specific and limited needs.

Table 3.4 shows the ATT aggregated at the region level. The results show significant heterogeneity in treatment effects across regions. Beijing stands out with the highest positive effect, with an ATT of 8,217 CNY, followed by several regions in Shandong and Jiangsu provinces. On the other hand, Wuhan in Hubei province exhibits the most negative impact, with an ATT of -3,990 CNY. These regional variations could be attributed to differences in lifestyle, population demographics, and receptiveness to new technologies.

In conclusion, the tensor completion algorithm allows us to estimate the counterfactual control outcomes for the observed treated units in $\hat{\mathcal{Y}}(0)$. By calculating the ATT for each urban setting, we can pinpoint the specific contexts where smart vending machines have the most significant impact on sales and those where the effect is less favorable. This information can be invaluable for companies looking to optimize their deployment strategies and maximize the benefits of smart vending machines in the most promising urban settings. By considering the region and tailoring the setting accordingly, companies can drive sales growth and improve the overall effectiveness of their smart vending machine networks.

3.4.3 Synthetic Tests

This section introduces synthetic tests to evaluate the performance of the proposed tensor completion method for causal inference. These tests are designed to simulate the structure of real-world datasets, such as the vending machine dataset, while providing control over the underlying structure and treatment effects. By generating synthetic data with known properties, the effectiveness of the proposed method in estimating treatment effects under different scenarios can be assessed.

Synthetic Low-rank Tensors

Synthetic data is generated by creating a low-rank tensor with dimensions $N_1 \times N_2 \times N_3$, where N_1, N_2, and N_3 represent the number of regions, scenes, and time periods, respectively. Region, scene, and time period fixed effects are added to the tensor to introduce additional structure, and noise is incorporated to simulate real-world variability. Treatment effects are introduced by assigning a subset of region-scene-time period tuples to the treatment group and adding a corresponding treatment effect to their entries. Finally, entries are randomly masked out as missing values to simulate incomplete data.

In causal social science applications, missingness in data often arises from treatment assignments and the choices that lead to these assignments. As a result, there are often specific structures on the treatment pattern that depart substantially from complete randomness. Two main treatment patterns are considered: blocked and staggered, which are applied at the region-scene-time period tuple level.

Block Structure

A leading example of a treatment pattern is the block structure, where a subset of region-scene-time period tuples adopts an irreversible treatment at a particular point in time T_0+1. In the tensor setting, this can be represented in a unfolded panel format as:

$$\mathcal{Y}_{N_1N_2\times N_3} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & ? & \dots & ? \\ \checkmark & \checkmark & ? & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & ? & \dots & ? \end{pmatrix}$$

where ✓ marks indicate observed values and ? marks indicate treatment values.

There are two special cases of the block structure: 1. Single-treated-time-period block structure: This is an extension of the single-treated-period block structure to the tensor setting, where the only treated region-scene tuples are in the last time period of the tensor. 2. Single-treated-region-scene block structure: This is an extension of the single-treated-unit block structure to the tensor setting, where a single region-scene tuple is treated for a number of time periods from time period $T_0 + 1$ onwards.

A special case that fits both these settings is the one with a single treatment region-scene-time period tuple:

$$\mathcal{Y}_{N_1N_2\times N_3} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \dots & ? \end{pmatrix}$$

This specific setting is useful for contrasting methods developed for the single-treated time period (unconfoundedness) case with those developed for the single-treated region-scene tuple (synthetic control) case, as both sets of methods are potentially applicable.

Staggered Adoption

Another treatment pattern that has received attention in the literature is the staggered adoption design. In this setting, region-scene-time period tuples may differ in the time they are first exposed to the treatment, but the treatment is irreversible. In the tensor setting, the structure of the staggered adoption design can be represented as:

$$\mathcal{Y}_{N_1 N_2 \times N_3} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & ? & ? & \dots & ? \end{pmatrix}$$

where each row represents a region-scene tuple, and columns represent time periods. Region-scene tuples that adopt the treatment early will have more post-treatment observations than those that adopt later or never adopt at all. This structure presents unique challenges and opportunities for estimating treatment effects in the tensor setting, as the timing of adoption may be informative about the region-scene tuples' characteristics and potential outcomes.

Results

Synthetic test results from both blocked and staggered structures are presented in Figure 3.7 and Figure 3.8, respectively. The ground truth for the treatment effect in the synthetic data is known to be 1.5.

The synthetic test results reveal similar patterns in terms of RMSE across both blocked and staggered missing data structures. Tensor completion (HQTCASD) consistently outperforms the other methods, demonstrating its ability to accurately capture the underlying patterns and estimate the treatment effect. Although none of the methods precisely hit the ground truth value of 1.5, tensor completion provides the closest estimates, making it the most accurate among the three algorithms.

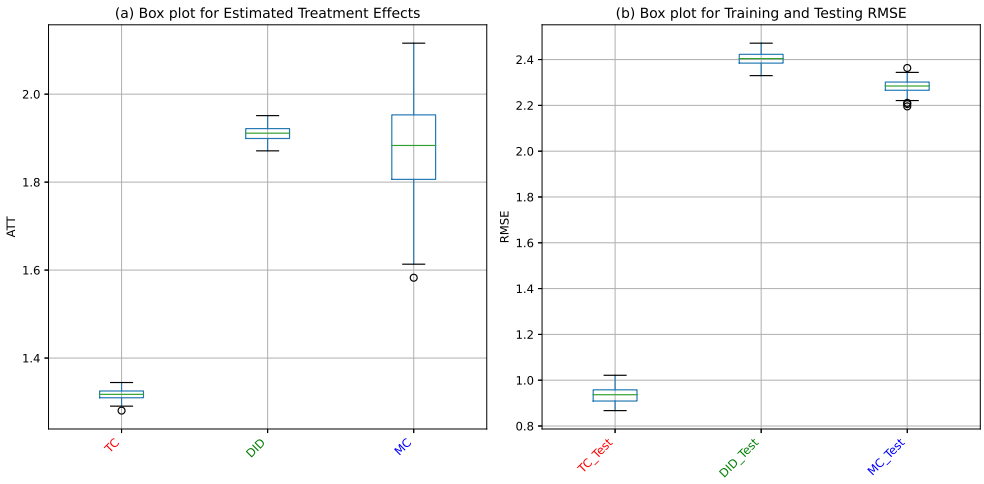


Figure 3.5: (a) Box plot of estimated treatment effects from TC, DID, and MC on synthesized data with blocked structure (b) Box plot of root mean square error (RMSE) for both training and testing datasets from TC, DID, and MC.

DID, in contrast, exhibits poor predictive performance in both missing data patterns, as evidenced by its higher RMSE values. This suggests that DID may not be the most suitable method for estimating the treatment effect in the context of deploying smart vending machines, regardless of the missing data structure.

MC suffers from an overfitting issue, performing well on the training set but poorly on the test set. This indicates that while matrix completion can capture patterns in the training data, it struggles to generalize effectively to unseen data, resulting in less accurate treatment effect estimates.

Notably, tensor completion demonstrates even stronger performance in the staggered treatment results compared to the blocked treatment effects. The estimated treatment effects from tensor completion are closer to the ground truth value of 1.5 in the staggered missing data pattern. This observation highlights tensor completion's robustness and adaptability to different missing data structures, solidifying its position as the best-performing algorithm among the three.

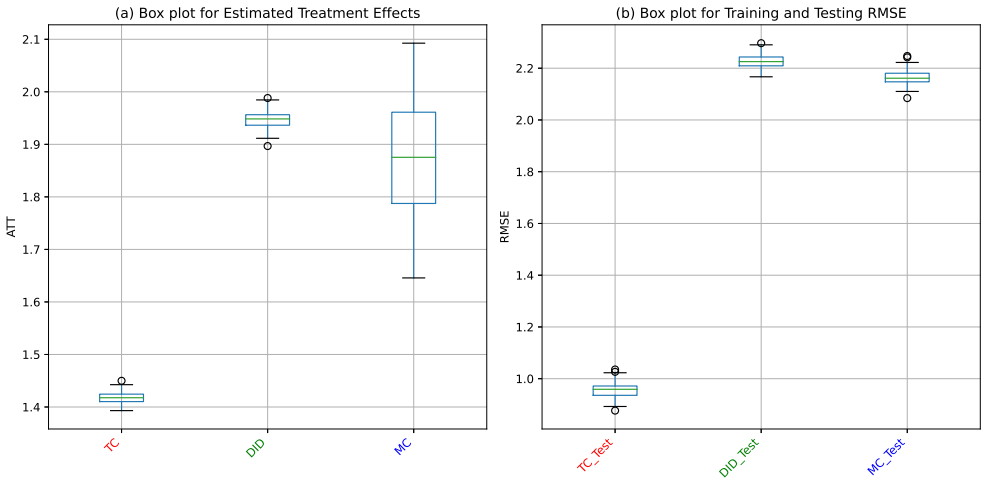


Figure 3.6: (a) Box plot of estimated treatment effects from TC, DID, and MC on synthesized data with staggered structure (b) Box plot of root mean square error (RMSE) for both training and testing datasets from TC, DID, and MC.

TC's superior performance can be attributed to its ability to leverage the multi-dimensional structure of the data, capturing the complex interactions and dependencies among various factors. By simultaneously considering the temporal and spatial dimensions, tensor completion effectively imputes missing values and estimates the treatment effect with greater accuracy.

However, it is crucial to recognize that while tensor completion surpasses the other methods, there remains room for improvement in precisely estimating the ground truth value. Further research and refinements to the tensor completion algorithm, such as incorporating domain-specific knowledge or exploring different regularization techniques, may help to enhance its accuracy and bring the estimates even closer to the true treatment effect.

In summary, the synthetic test results demonstrate the superiority of tensor completion algorithms in estimating the treatment effect of deploying smart vending machines, regardless of the missing data pattern. Its ability to provide accurate estimates closer to the ground truth, combined with its robustness and adaptability, positions it as the most promising

method among the three algorithms. Nonetheless, ongoing research and development efforts are essential to further improve its performance and minimize the discrepancy between the estimated and true treatment effects.

Figure 3.7(a) and Figure 3.8(a) show the box plots of the estimated treatment effects from three methods: TC, DID, and MC on synthesized data with a blocked structure. Figure 3.7(b) and Figure 3.8(b) displays the box plots of the root mean square error (RMSE) for both training and testing datasets from these methods. The ground truth treatment effect in the synthetic data is 1.5.

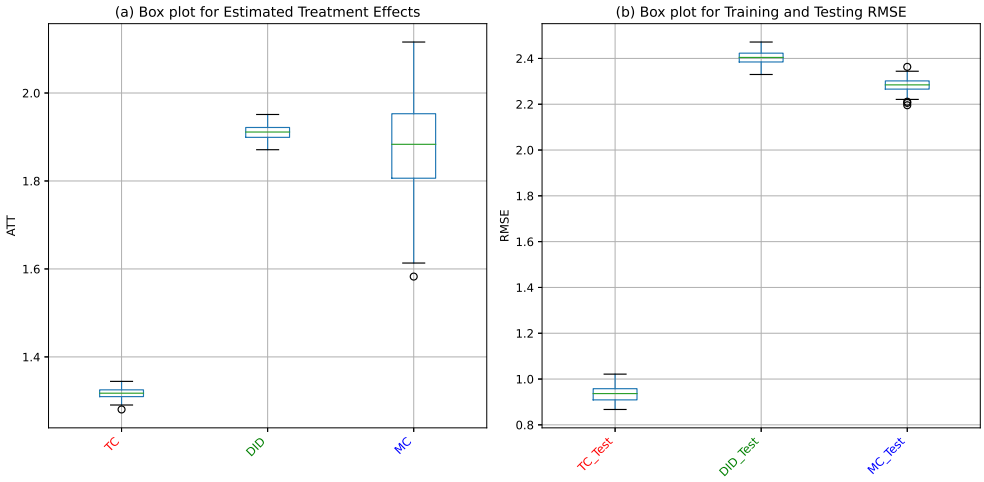


Figure 3.7: (a) Box plot of estimated treatment effects from TC, DID, and MC on synthesized data with blocked structure (b) Box plot of root mean square error (RMSE) for both training and testing datasets from TC, DID, and MC.

The synthetic test results reveal consistent patterns in terms of RMSE across both blocked and staggered missing data structures. TC consistently outperforms DID and MC, demonstrating its ability to accurately capture underlying patterns and estimate the treatment effect. Although none of the methods precisely match the ground truth value of 1.5, TC provides the closest estimates, making it the most accurate among the three algorithms. DID exhibits poor predictive performance in both missing data patterns, as evidenced by its

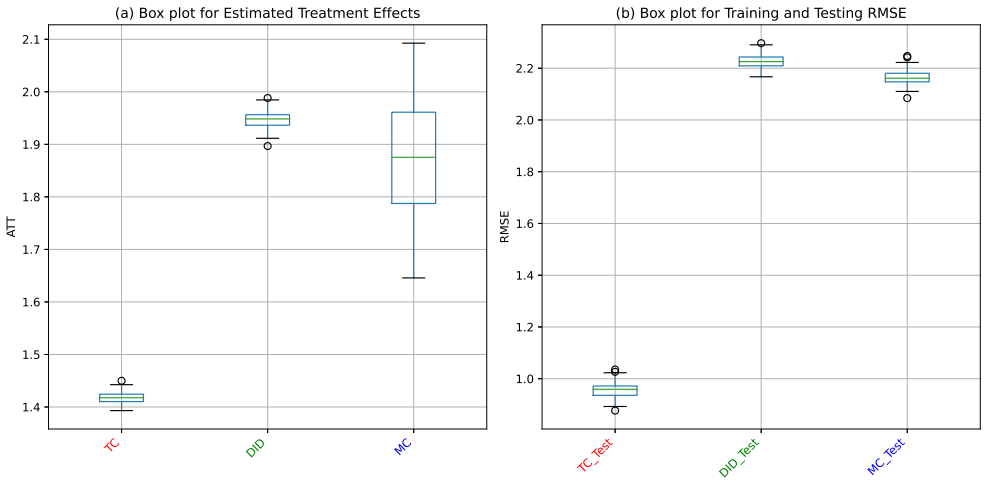


Figure 3.8: (a) Box plot of estimated treatment effects from TC, DID, and MC on synthesized data with staggered structure (b) Box plot of root mean square error (RMSE) for both training and testing datasets from TC, DID, and MC.

higher RMSE values, suggesting it may not be suitable for estimating the treatment effect in the context of deploying smart vending machines, regardless of the missing data structure. MC suffers from overfitting, performing well on the training set but poorly on the test set, indicating it struggles to generalize effectively to unseen data, resulting in less accurate treatment effect estimates. Notably, TC demonstrates even stronger performance in the staggered treatment results compared to the blocked treatment effects, with estimated treatment effects closer to the ground truth value of 1.5 in the staggered missing data pattern. This highlights TC's robustness and adaptability to different missing data structures.

TC's superior performance can be attributed to its ability to leverage the multi-dimensional structure of the data, capturing complex interactions and dependencies among various factors by simultaneously considering temporal and spatial dimensions. However, while TC surpasses the other methods, there remains room for improvement in precisely estimating the ground truth value. Further research and refinements to the TC algorithm may help enhance its accuracy and bring estimates closer to the true treatment effect.

The synthetic test results demonstrate the superiority of TC in estimating the treatment effect of deploying smart vending machines, regardless of the missing data pattern. Its accuracy, robustness, and adaptability position it as the most promising method among the three algorithms. Nonetheless, ongoing research and development efforts are essential to further improve its performance and minimize the discrepancy between estimated and true treatment effects.

3.5 Conclusions

This paper investigates the treatment effect of deploying smart vending machines compared to traditional vending machines across various urban settings. By framing this as a high-dimensional machine learning task and identifying the intrinsic tensor structure of the data, we developed a novel tensor completion-based method to accurately estimate the treatment effect.

Our research makes several key contributions. First, we demonstrate the value of formulating treatment effect estimation as a high-dimensional machine learning problem, enabling the application of advanced learning methods. Second, we recognize the inherent tensor structure of our data, with dimensions corresponding to regions, scenes, and time, to capture intricate dependencies. Third, we propose a tensor completion-based method that outperforms state-of-the-art approaches like matrix completion by leveraging the low-rank tensor structure.

The implications extend beyond the vending machine industry. Our work showcases how machine learning techniques like tensor completion can address challenges of high-dimensional, spatiotemporal data and provide more accurate treatment effect estimates. The framework can be adapted to various domains, including retail, healthcare, and transportation.

Future directions include incorporating instrumental variables and covariates, extending the methodology to handle time series components, and applying our approach to other domains like online advertising or personalized medicine.

In conclusion, our research underscores the power of leveraging advanced learning methods like tensor completion for causal analysis on high-dimensional, spatiotemporal data. Our novel tensor completion-based approach provides valuable insights for the vending machine industry and demonstrates the broader potential of these techniques for empowering informed decision-making across a wide range of applications.

Table 3.3: Summary statistics for ATT by Scene

Scene	**Mean**	**Std**	**Min**	**Max**
University Dormitory Area (Mixed)	4131.44	24512.98	-81287.44	128889.62
Various Stations	3177.21	8114.74	-22042.72	70328.22
Administrative Office Area	810.62	3181.14	-7669.85	33500.81
Employee Dormitory Area	703.85	1989.12	-3512.30	9989.57
Hospital	485.79	4557.20	-27231.72	17583.25
Business Office Area	360.83	1093.61	-2775.90	6943.97
University Teaching Buildings	277.18	12666.82	-77079.18	59232.43
Service Halls, Banks	193.79	1188.04	-5278.96	12580.93
Hotel	55.43	729.58	-5772.69	2644.35
Park	54.77	860.04	-7551.64	4133.19
Scenic Area	-12.66	843.77	-8024.84	2769.79
Exhibition Hall	-119.41	418.27	-2282.27	920.06
Library	-189.75	1303.96	-9838.74	1578.87
Shopping Center, Cinema	-205.12	1368.88	-9967.05	3109.82
Industrial Park	-212.96	2194.12	-13979.03	3037.74
Community	-379.20	3270.97	-35827.67	3928.81
Factory Workshop	-739.98	15047.11	-77765.59	40885.56
University Sports Venues, Gyms	-743.36	1852.18	-11907.98	2754.73
University Canteens	-810.65	4112.25	-33289.26	8527.04
Middle School	-7851.78	15708.38	-133677.29	2805.14

Table 3.4: Summary statistics for ATT by Region

Region	**Mean**	**Std**	**Min**	**Max**
Beijing	8217.22	21046.38	-13357.80	128889.62
Qingdao, Shandong	1830.44	3263.79	-2160.85	19347.12
Yancheng, Jiangsu	1823.26	5529.24	-11907.98	40885.56
Chongqing	1326.54	2234.66	-1152.40	13525.99
Yinchuan, Ningxia	1220.83	2734.54	-844.82	17583.25
Zhengzhou, Henan	839.34	1846.99	-4447.61	14790.05
Changzhou, Jiangsu	804.28	2608.03	-3756.99	13966.90
Nanjing, Jiangsu	794.87	1855.36	-7823.29	12508.41
Hanzhong, Shaanxi	424.36	3537.15	-10599.01	33169.17
Shangluo, Shaanxi	414.60	1919.23	-324.49	16014.75
Shenyang, Liaoning	406.53	1178.52	-323.26	8345.04
Lanzhou, Gansu	362.49	1149.69	-3399.70	7705.13
Taizhou, Jiangsu	152.11	1299.86	-8464.77	5230.55
Yichang, Hubei	68.19	304.99	-1993.03	321.16
Baoji, Shaanxi	-133.84	827.60	-5282.72	1807.26
Yibin, Sichuan	-179.92	1021.78	-5946.55	324.16
Tianjin	-180.13	1081.69	-8211.09	324.16
Suzhou, Jiangsu	-238.98	1154.92	-7551.64	3928.81
Urumqi, Xinjiang	-1153.06	5614.51	-31599.42	8164.57
Zhoukou, Henan	-1696.90	6414.27	-33401.75	483.07
Liupanshui, Guizhou	-1914.78	5899.49	-47915.67	324.16
Qianxinan, Guizhou	-2037.50	5785.15	-54866.08	323.16
Directly Governed Counties, Hubei	-3247.81	9623.37	-58299.20	3559.63
Xi'an, Shaanxi	-3339.46	10071.28	-81287.44	10063.99
Wuhan, Hubei	-3990.78	16634.01	-133677.29	61030.05

Chapter 4

FRENEMIES IN THE RETAIL MARKET: A PARTNERSHIP BETWEEN A PHYSICAL RETAILER AND AN E-TAILER FOR CONSUMER RETURNS

4.1 Introduction

The e-commerce market has experienced phenomenal growth, emerging as a significant force in the global retail landscape. As of 2023, the global e-commerce market has expanded to an impressive $3.183 trillion, marking a substantial increase from the previous year and reflecting a compound annual growth rate of 12.2% (Statista, 2023). This growth underscores the sheer size and influence of the e-commerce sector in contemporary commerce.

When there is a misfit between consumer and product, returns occur. Consumer returns have become a prevalent aspect of online shopping - in 2021, it was reported that 30% of online shoppers intentionally overbuy with the intention of returning items they do not want, highlighting the significant role returns play in consumer behavior. Moreover, the cost of e-commerce returns is estimated to be a staggering $400
billion annually in the United States alone (McKinsey & Company, 2020). E-tailers, in particular, face significant challenges as the returns process not only involves the direct costs associated with shipping and handling but also encompasses indirect expenses like processing and restocking. This complex scenario underscores the importance of efficient return management for maintaining profitability and customer satisfaction in the e-commerce domain.

4.1.1 Problem Background

Under this context, an intriguing partnership for consumer returns has been established between physical retailers and e-tailers. In this partnership, retailers like Kohl's have agreed to accept Amazon returns in exchange for store foot traffic. The partnership allows customers who purchased products online to return them at a physical store location, offering more convenient choices than the traditional returning method such as the parcel shipping stores or post offices.

Commentators have offered varied perspectives on who stands to benefit or suffer from such partnerships. On one hand, it seems beneficial to all the stakeholders. These collaborations are seen as a strategic move for retailers to increase foot traffic and potential sales. For e-tailers, this arrangement addresses the logistical and financial challenges of managing returns. For consumers, the partnership provides an added convenience for returns, potentially enhancing customer satisfaction and loyalty. On the other hand, retailers, by facilitating returns for their online competitors, might drive their customers who would visit the physical storefronts towards buying directly from e-tailers.

The situation is even more complex than it appears when we consider the potential impact on a particular shopping behavior, known as showrooming. While partnerships between e-tailers and retailers create foot traffic, their impacts on competition between e-tailers and retailers are far from clear. Will the increased foot traffic and potential sale opportunities for retailers outweigh the advantage given to etailers in terms of enhanced customer service and reduced return cost? Moreover, how will these partnerships impact consumer shopping behaviors? Of the consumers who return products through this partnership, how many of them would choose to showroom and purchase online; how many would make purchases at the retailer? To fully understand the effect of these partnerships on the demand-side competition between e-tailers and retailers, we look at their competitive dynamics in this paper.

4.1.2 Research Questions and Findings

Our paper aims to look at how the partnership on consumer returns will affect e-tailers' and physical retailers' profit by taking into account the indirect effect, i.e., the effect of cross-return on demand-side competition. Specifically, we ask the following research questions:

First, we ask the following question: *With the partnership on consumer returns, what is the equilibrium strategy for physical retailer's profit and the e-tailer?* To answer this question, we develop an analytical model to describe the consumer return process with or without a partnership between the physical retailer and the e-tailer. The physical retailer sells a product via the offline channel, and the etailer sells the identical product via the online channel. Without the partnership, the consumer can only return the product to the original seller, while with the partnership, the consumer has the option to return the product bought from the e-tailer to the physical retailer. With the presence of this retailer return partnership, when the offline shopping cost is very low and the physical retailer's marginal return cost is below a certain threshold, the e-tailer and physical retailer can set their compensation and prices such that they both achieve an equilibrium with higher profit compared to the scenario with the retailer return partnership. Conversely, where the offline shopping cost is high, and the physical retailer's marginal return cost is also high, the equilibrium prices will be higher than without the retailer return partnership, and the physical retailer's profit will depend on the e-tailer's marginal return and offline shopping costs. Under this setting, consumers no longer conduct showrooming at the physical store.

Next, we ask the second question of our paper: *What's the impact of the return partnership on the physical retailer's profit and the e-tailer's profit?* Our model reveals that forming a return partnership softens price competition between the e-tailer and physical retailer. The e-tailer can gain from the return partnership under two specific conditions. One condition is that when the offline shopping cost is below a certain threshold so that the inconvenience or additional costs associated with offline shopping are not excessively high, the e-tailer can benefit from the partnership. The e-tailer can utilize the physical retailer to facilitate re-

turns, thus increasing their profits by lowering return costs and providing more options to the consumer. The other condition is that when the offline shopping cost is above this threshold, as long as the physical retailer's marginal return cost is below another threshold, the e-tailer still benefits from the partnership. This situation suggests that if the physical retailer's cost of handling returns is relatively low, the e-tailer can effectively outsource the return service at a cost lower than managing returns independently, thereby improving e-tailer's profitability.

On the other hand, the physical retailer can also benefit from the consumer return partnership with two conditions. The first condition is when the offline shopping cost is above a threshold, implying that if consumers find it sufficiently inconvenient or costly to shop offline, they may be more inclined to purchase online. In this case, a partnership with an e-tailer can introduce more foot traffic to the physical retailer, thus capturing some of the online market. The second condition is when the physical retailer's marginal return cost is above a threshold, suggesting that if it is relatively expensive for the physical retailer to handle returns, partnering with the e-tailer for return service can alleviate some of the financial burden associated with returns, thus increasing the retailer's profitability. Overall, the return partnership can lead to a more collaborative relationship between e-tailers and physical retailers, reducing direct price competition and potentially leading to higher prices and profits for both retailers. This strategic collaboration could reposition physical retailers as valuable partners in the e-commerce ecosystem rather than merely competitors or showrooms for online purchases.

Finally, we ask the third question of our paper: *What's the impact of the return partnership on consumers? Would consumers treat physical retailers as showrooms?* The results indicate that consumer surplus can be lower in the cross-return scenario compared to the no-cross-return scenario under two conditions. The first condition is when the offline shopping cost is above a threshold, suggesting that if consumers find it relatively inconvenient or costly to shop offline, they may be more inclined to purchase online. The second condition is when the e-tailer's marginal compensation to the retailer for handling returns is above a threshold, implying that if the e-tailer pays a relatively high compensation to the retailer for

each returned item, it can lead to a decrease in consumer surplus. This finding highlights a potential trade-off between the benefits of the return partnership for retailers and its impact on consumers. While the partnership can lead to higher profits for both the e-tailer and the retailer under certain conditions, it may come at the expense of reduced consumer surplus. This reduction in consumer surplus could be attributed to the softening of price competition between the two retailers, leading to higher prices for consumers.

4.2 The Model

Our model considers two firms: a physical retailer and an e-tailer. Both firms offer the same product, but the physical retailer sells the product through an offline channel, while the e-tailer sells the product through an online channel. The price of the product in the online channel is denoted by p_O, and the price of the product in the offline channel is denoted by p_F.

4.2.1 Consumers' Choices of Shopping Methods

According to Bell et al. (2018), one of the main reasons consumers visit a physical showroom is to resolve uncertainty about how well a product fits their preferences by examining it in person. If consumers find that the product does not suit their tastes, they may choose not to purchase it. In light of this, our model considers consumers to have heterogeneous preferences regarding their individual product fit. Following the approach of Gao and Su (2017), we classify consumers into two categories: (i) fit consumers, for whom the supplier's product aligns with their tastes, and as a result, they derive a (gross) utility of v from the product; and (ii) misfit consumers, for whom the product does not suit their tastes, and consequently, they receive zero gross utility. In our model, we assume that θ proportion of the

consumers in the market are fit consumers, while the other $1-\theta$ are misfit consumers. Before the shopping process begins, each consumer is aware of θ but unaware of their true fit type.

Consumers have the option to directly purchase a product at a physical retailer's store,

a method referred to as "buy-offline." To buy offline, all consumers, regardless of their fit type, incur a cost related to shopping offline, such as the transportation cost of visiting the physical store. Forman et al. (2009) have empirically demonstrated the existence of these disutility costs associated with shopping offline. In this paper, we refer to these costs as the consumers' offline shopping cost, denoted by l. If the product fits the consumer's preferences, they can purchase it and gain a net utility of $v - p_F$, concluding their shopping experience. However, if the product does not fit, the consumer will end their shopping without making a purchase because their net utility of buying the misfit product at the store is negative, i.e., $-p_F$. Note that before visiting the store, a consumer only knows that the product has a proportion of θ fitting their taste. Therefore, the expected utility of choosing to buy offline is

$$U_F = -l + \theta\left(v - p_F\right).$$

Consumers also have the option to buy the product from the online retailer, which Balakrishnan et al. (2014) call "e-Direct" shopping. When choosing e-Direct, the consumer faces disutility of online purchasing irrespective of product fit, such as shipping costs and delivery wait times. Forman et al. (2009) provide empirical evidence that the disutility of buying online are significant. We refer to this disutility as the online shopping cost, denoted by the symbol h_O. If the delivered product matches the consumer's preferences, she will retain it, realizing a net utility of $v - p_0$, concluding her shopping journey.

If the product does not suit the consumer's taste, she will want to return it because it only brings a negative utility of $-p_0$. For analytical tractability, we consider that a consumer receives a full refund p_O during the return process (Gao and Su 2017), which is becoming an increasingly common practice in the online retail industry (Green 2018). Additionally, consumers must bear a return hassle cost ϕ (Anderson et al. 2009, Balakrishnan et al. 2014). When the physical retailer and the e-tailer do not form a partnership for consumer returns (referred to as the no-cross-return scenario, or N scenario), consumers must return the product through the channels designated by the e-tailer, such as local UPS stores. The return hassle cost under the N scenario is considered to be $\phi = h_r$.On the other hand, when both

firms form a partnership for consumer returns (referred to as the cross-return scenario, or C scenario), consumers have an additional option to return their e-Direct purchases by dropping off the misfit product at the physical retailer's store. Since cross-returning the product requires consumers to visit the physical retailer's store, the return hassle cost under the C scenario is considered to be approximately equal to the offline shopping cost, i.e., $\phi = l$.

Note that we consider $l \leq h_r$, which suggests that consumers will find it more convenient to return their e-Direct purchases through the physical retailer's store compared to other channels once this option becomes available. Consumers who feel differently are unlikely to be affected by the return partnership, as they will not consider the cross-return option in the first place. Therefore, these consumers are not considered in the model.

It is important to mention that when consumers return their Amazon.com purchases at Kohl's stores, they may make additional purchases, resulting in demand spillover. A strong demand spillover is undoubtedly favorable for both consumers and the physical retailer, making cross-returns a more attractive option for both parties. However, as mentioned in the introduction, our study aims to identify mechanisms other than demand spillover that may benefit both the physical retailer and consumers. Consequently, we normalize such additional sales to zero for the purpose of this analysis.

We assume that the return hassle cost in both the N and C scenarios is sufficiently low, such that a consumer will always return the product when a misfit occurs. This assumption has become much more acceptable than it was a decade ago due to the efforts of large e-tailers and partnering physical retailers to make product returns easy for consumers (Levy 2017, Ketzenberg et al. 2020, USPS 2022). As a result, the expected utility of choosing e-Direct is

$$U_E = -h_O + (\theta) \cdot (v - p_O) + (1 - \theta) \cdot (-\phi).$$

The third shopping method is showrooming. To engage in showrooming, a consumer must incur an offline shopping cost l to inspect the product at the store. If the product does not fit her taste, she will end her shopping without making a purchase. If the product fits, she can order it from the e-tailer at the online price p_O and bear an additional online

shopping cost h_O, such as waiting for delivery. It is important to note that before visiting the store, a consumer only knows that the product has a proportion of θ to fit her taste. Therefore, the expected utility of choosing showrooming is

$$U_S = -l + (\theta)(v - p_o - h_o).$$

Consumers are heterogeneous in their online shopping costs, denoted as h_O. For instance, the same two-business-day shipping time may incur different levels of disutility for different consumers. Consumers with relatively high online shopping costs may consider the price gap between offline and online purchases to be insufficient to justify the additional disutility they must bear when buying online, such as extra waiting time for delivery. As a result, these consumers may not switch to the e-tailer after examining the product at the store. In this model, we consider that the online shopping cost h_O is uniformly distributed in the range of $[0, 1]$.

Consumers will choose their shopping methods based on the highest expected utility among U_E, U_F, and U_S. Unless otherwise noted, we hereafter use the term "demand" to refer to the quantity of the product purchased and kept by consumers. In other words, it counts the purchases from consumers who find the product to be a good fit but not the purchases from consumers who find the product to be a misfit, as the latter will either not purchase or return the product. Specifically, e-Direct demand a_E counts all the e-Direct purchases from the fit consumers. Showrooming demand a_S and buy-offline demand a_F count all the showrooming and buy-offline purchases, respectively, from the same group of consumers. For the sake of simplicity, we assume that half of the consumers in the market are fit consumers, i.e., $\theta = 1/2$, while the other half are misfit consumers. Before the shopping process begins, each consumer is unaware of their true fit type. They only possess the knowledge that there is a 50% probability (1/2) that the product will match their taste and a 50% probability (1/2) that it will not. The derivation details for each demand can be found in Section A of the online appendix.

4.2.2 Cross-Return and Firms' Profits

This paper focuses on product returns due to misfit, which account for over 90% of total returns according to existing research (Shulman et al. 2010). When consumers choose showrooming or buy-offline, they do not incur returns since misfit consumers would not purchase the product. The average return rate for brick-and-mortar retailers is around 8.89%, while for e-tailers it is approximately 30% (Saleh 2016, Garcia 2018).

In the N scenario, the e-tailer bears the cost of handling e-Direct returns, incurring a marginal return cost of s. We hereafter refer to s as the e-tailer's marginal return cost. In the C scenario, the physical retailer handles e-Direct returns, and the e-tailer subsidizes them with a payment of s_F per unit returned. The physical retailer's cost for handling each unit return is f.

The firms' marginal cost of selling the product is normalized to zero, following previous studies (e.g., Kuksov and Liao 2018). Half of the consumers in the market are fit, while the other half are misfit. Only misfit e-Direct consumers will make returns. Unless otherwise noted, hereafter we use superscript $i \in \{N, C\}$ where N indicates variables under the N scenario, and C indicates variables under the C scenario. The physical retailer's and e-tailer's profit functions in the N scenario are the following, respectively. Note when $\theta = 1/2$, the unfit e-direct demand $\left(\frac{1-\theta}{\theta}\right) a_E$ equals a_E.

$$\pi_F^N = p_F \cdot a_F, \text{ and } \pi_O^N = p_O \cdot (a_E + a_S) - s \cdot (a_E)\,.$$

In the C scenario, their profit functions are the following, respectively,

$$\pi_F^C = p_F \cdot a_F + (f - s_F) \cdot a_E, \text{ and } \pi_O^C = p_O \cdot (a_E + a_S) - f \cdot a_E$$

The game sequence is as follows: In the N scenario, the physical retailer and e-tailer simultaneously decide offline price p_F and p_O in the first stage. In the C scenario, the e-tailer chooses marginal compensation f in the first stage, followed by price competition between the physical retailer and e-tailer in the second stage. Consumers then make decisions

based on p_O and p_F, and the firms' profits are realized. Each firm aims to maximize its own profit.

The analysis considers parameter values that satisfy the following conditions in market equilibrium: (i) the product value is large enough for full market coverage; (ii) at least one consumer chooses buy-offline; (iii) a decrease in offline shopping cost l triggers showrooming before eliminating e-Direct consumers; (iv) offline shopping cost is below threshold $\bar{l}$ and e-tailer's marginal return cost is below threshold $\bar{s}$ for showrooming consumers to exist in the N scenario equilibrium; and (v) physical retailer's marginal return cost is below threshold $\bar{s}_F$ for e-Direct consumers to exist in the C scenario equilibrium. Additional technical conditions for market equilibrium are provided in the online appendix (Section C). Table 1 summarizes the important notations.

4.3 Results and Managerial Insights

For the remainder of the paper, we use tilde accents for equilibrium variables and hat accents for variable thresholds. All proofs of lemmas and propositions are in the online appendix.

4.3.1 Equilibrium Analysis

Lemma 4.3.1. Consumers' choices of shopping methods in the no-cross-return (N) and the cross-return (C) scenarios are summarized in the following table based on their online shopping cost and the online price. Note that $i = N, C$.

The expressions of the thresholds in the table are $\hat{p}_{O1}^N = p_F + 2l - h_r$, $\hat{p}_{O1}^C = p_F + l$, $\hat{p}_{O2}^N = p_F - 2l + h_r$, $\hat{p}_{O2}^C = p_F - l$, $\hat{p}_{O3}^N = \hat{p}_{O3}^C = p_F - 1$, $\hat{h}_{OEF}^N = (p_F - p_O + 2l - h_r)/2$, $\hat{h}_{OEF}^C = (p_F - p_O + l)/2$, $\hat{h}_{OES}^N = 2l - h_r$, $\hat{h}_{OES}^C = l$, and $\hat{h}_{OSF}^N = \hat{h}_{OSF}^C = p_F - p_O$. ‘

Lemma 4.3.1 shows that consumers exhibit similar shopping patterns across the N and the C scenario. Specifically, when the online price is above an uppermost threshold (i.e., $p_O > \hat{p}_{O1}^i$ in Case F), all consumers will choose to buy offline. When the online price drops below the uppermost threshold (e.g., Case E-F), consumers with the low online shopping

Table 4.1: Key model parameters and variables

Parameters	**Description**
h_O	Consumers' online shopping cost
h_r	Consumers' return hassle cost through channels other than the physical retailer
l	Consumers' offline shopping cost
s	The e-retailer's marginal return cost
s_F	The physical retailer's marginal return cost
Variables	**Description**
p_F	Product's offline price
p_O	Product's online price
f	The e-tailer's compensation to the physical retailer for each consumer return handled
a_i	Consumer demand ($i = E$: e-Direct, $i = S$: showrooming, $i = F$: buy-offline)
π_O	The e-tailer's profit
π_F	The physical retailer's profit

cost (e.g., $h_O \leq \hat{h}^i_{OEF}$ in Case E-F) will start to choose e-Direct. When the online price becomes considerably low (e.g., Case E-S-F), consumers with medium online shopping cost (e.g., $\hat{h}^i_{OES} \leq h_O \leq \hat{h}^i_{OSF}$) will choose to do showrooming due to the considerable gap between the offline and the online price. Consumers with high online shopping cost (e.g., $h_o > \hat{h}^i_{\mathrm{OSF}}$) will still stay with buy-offline because they do not want to further incur the online shopping cost, for example, the disutility of waiting for delivery, after she has borne the offline shopping cost to visit the store. Lastly, in Case E-S where the online price is extremely low (i.e., $p_O \leq \hat{p}^i_{O3}$), no consumer will choose to buy offline.

Case	Online price p_O	Consumer segments based on online shopping cost h_O		
		e-Direct (E)	Showrooming (S)	Buy offline (F)
F	Extremely high ($p_O > \hat{p}^i_{O1}$)	—	—	$[0, 1]$
E-F	High ($\hat{p}^i_{O2} < p_O \leq \hat{p}^i_{O1}$)	$[0, \hat{h}^i_{OEF}]$	—	$(\hat{h}^i_{OEF}, 1]$
E-S-F	Low ($\hat{p}^i_{O3} < p_O \leq \hat{p}^i_{O2}$)	$[0, \hat{h}^i_{OES})$	$[\hat{h}^i_{OSS}, \hat{h}^i_{OSF}]$	$(\hat{h}^i_{OSF}, 1]$
E-S	Extremely low ($p_O \leq \hat{p}^i_{O3}$)	$[0, \hat{h}^i_{OES})$	$[\hat{h}^i_{OES}, 1]$	—

The main difference between the N and the C scenario in terms of consumers' choices of shopping methods is that the C scenario is a special case of the N scenario where consumers return hassle cost is equal to the offline shopping cost. To see this analytically, note that $\hat{p}^C_{o_j} = \hat{p}^N_{o_j}\big|_{h_r=l}$ for $j = 1, 2, 3$, $\hat{h}^C_{OEF} = \hat{h}^N_{OEF}\big|_{h_r=l}$, and $\hat{h}^C_{OES} = \hat{h}^N_{OES}\big|_{h_r=l}$. This is because consumers will return the product to the physical retailer's store in the C scenario.

It is also worth noting that in both the N and the C scenario, showrooming consumers will emerge only when the offline-online price gap reaches to a certain threshold. Specifically, according to the table of Lemma 1, it happens when $p_O \leq \hat{p}^i_{O2}$, which is equivaelnt to $p_F - p_O \geq 2l - h_r$ in the N scenario or $p_F - p_O \geq l$ in the C scenario. This will play an important role in formulating the market equilibrium later.

Lemma 4.3.2. Given the e-tailer's choice of marginal compensation f in the C scenario, the physical retailer's best response p^*_F to the e-tailer's choice of online price p_O is summarized in the following table. Note that the column titled "Seg" shows which case in Lemma l the best response p^*_F will lead to.

Case	Compensation f	Sub case	E-tailer's choice of p_O	p_F^*	Seg
F1	High $\left(f > \hat{f}_{F1}\right)$	1	$p_O \leq \hat{p}_{O11}^C = -1$	$\hat{p}_{F1}^C$	E-S
		2	$\hat{p}_{O11}^C < p_O \leq \hat{p}_{O12}^C = 1 - 2l$	$\hat{p}_{F2}^C$	E-S-F
		3	$\hat{p}_{O12}^C < p_O \leq \hat{p}_{O13}^C = f - s_F - 3l + 2$	$\hat{p}_{F3}^C$	E-F
		4	$\hat{p}_{O13}^C < p_O \leq \hat{p}_{O14}^C = f - s_F + l + 2$	$\hat{p}_{F4}^C$	E-F
		5	$p_O > \hat{p}_{O14}^C$	$\hat{p}_{F5}^C$	F
F2	Medium $\left(\hat{f}_{F2} < f \leq \hat{f}_{F1}\right)$	1	$p_O \leq \hat{p}_{O21}^C = \hat{p}_{O11}^C$	$\hat{p}_{F1}^C$	E-S
		2	$\hat{p}_{O21}^C < p_O \leq \hat{p}_{O22}^C = \sqrt{2}\left(f - s_F - l + 1\right) - f + s_F - l$	$\hat{p}_{F2}^C$	E-S-F
		3	$\hat{p}_{O22}^C < p_O \leq \hat{p}_{O23}^C = \hat{p}_{O14}^C$	$\hat{p}_{F4}^C$	E-F
		4	$p_O > \hat{p}_{O23}^C$	$\hat{p}_{F5}^C$	F
F3	Low $\left(\hat{f}_{F3} < f \leq \hat{f}_{F2}\right)$	1	$p_O \leq \hat{p}_{O31}^C = \hat{p}_{O11}^C$	$\hat{p}_{F1}^C$	E-S
		2	$\hat{p}_{O31}^C < p_O \leq \hat{p}_{O32}^C = 1 - 2\sqrt{\left(-f + s_F - 1\right) l}$	$\hat{p}_{F2}^C$	E-S-F
		3	$p_O > \hat{p}_{O32}^C$	$\hat{p}_{F5}^C$	F
F4	Extremely low $\left(f \leq \hat{f}_{F3}\right)$	1	$p_O \leq \hat{p}_{O41}^C = \left(f - s_F + 1\right) l$	$\hat{p}_{F1}^C$	E-S
		2	$p_O > \hat{p}_{O41}^C$	$\hat{p}_{F5}^C$	F

The expressions in the table above are: $\hat{f}_{F1} = s_F + l - 1, \hat{f}_{F2} = s_F - (3 + 2\sqrt{2})l - 1, \hat{f}_{F3} = s_F - (1/l) - 1$, $\hat{p}_{F1}^C = p_O + 1$, $\hat{p}_{F2}^C = (p_O + 1)/2$, $\hat{p}_{F3}^C = p_O + l$, $\hat{p}_{F4}^C = (p_O + f - s_F - l + 2)/2$, and $\hat{p}_{F5}^C = p_O - l$. Meanwhile, the e-retailer's best response p_O^*

to the physical retailer's choice of offline price p_F is summarized in the following table.

Case	Compensation f	Sub case	Physical retailer's choice of p_F	p_O^*	Seg
O1	Low $\left(f < \hat{f}_{O1}\right)$	1	$p_F < \hat{p}_{F11} = f - l$	$\hat{p}_{O1}^C$	F
		2	$\hat{p}_{F11} \le p_F < \hat{p}_{F12} = \sqrt{2}(f+l) - f + l$	$\hat{p}_{O2}^C$	E-F
		3	$\hat{p}_{F12} \le p_F < \hat{p}_{F13} = 2$	$\hat{p}_{O3}^C$	E-S-F
		4	$p_F \ge \hat{p}_{F13}$	$\hat{p}_{O4}^C$	E-S
O2	High $\left(\hat{f}_{O1} \le f < \hat{f}_{O2}\right)$	1	$p_F < \hat{p}_{F21} = 2\sqrt{fl}$	$\hat{p}_{O1}^C$	F
		2	$\hat{p}_{F21} \le p_F < \hat{p}_{F22} = \hat{p}_{F13}$	$\hat{p}_{O3}^C$	E-S-F
		3	$p_F \ge \hat{p}_{F22}$	$\hat{p}_{O4}^C$	E-S
O3	Extremely high $\left(f \ge \hat{f}_{O2}\right)$	1	$p_F < \hat{p}_{F31} = fl + 1$	$\hat{p}_{O1}^C$	F
		2	$p_F \ge \hat{p}_{F31}$	$\hat{p}_{O4}^C$	E-S

The expressions in the table above are: $\hat{f}_{O1} = (3+2\sqrt{2})l, \hat{f}_{O2} = 1/l, \hat{p}_{O1}^C = p_F + l, \hat{p}_{O2}^C = (p_F + f + l)/2$, $\hat{p}_{O3}^C = p_F/2$, and $\hat{p}_{O4}^C = p_F - 1$. The physical retailer's and the e-tailer's best responses in the N scenario are relegated to Section B of the online appendix for brevity.

Lemma 4.3.2 shows the physical retailer's (the e-tailer's) best response functions given its competitor's choice of the online price (the offline price) in the C scenario. As the tables show, the best response functions are step-wise and depend on the compensation f. Regarding the e-tailer's responses (i.e., Case O1-O3 in the second table), when compensation f is low (i.e., $f < \hat{f}_{O_1}$), as the offline price increases from low to high, the e-tailer would respond with an online price that induces market segmentation from F, E-F, E-S-F, and E-S. This is because when the offline price is extremely low, i.e., O1-1, the e-tailer would simply give up the entire market as it is not profitable to compete with such a low offline price. As the offline price

increases into the range of Case O1-2, the e-tailer would choose a online price $\hat{p}^C_{O2}$ that is relatively close to the online price

The main difference between the N and the C scenario in terms of consumers' choices of shopping methods is that the C scenario is a special case of the N scenario where consumers return hassle cost is equal to the offline shopping cost. To see this analytically, note that $\hat{p}^C_{Oj} = \hat{p}^N_{Oj}\big|_{h_r=l}$ for $j = 1, 2, 3$, $\hat{h}^C_{\text{OEF}} = \hat{h}^N_{\text{OEF}}\big|_{h_r=l}$, and $\hat{h}^C_{\text{OES}} = \hat{h}^N_{OES}\big|_{h_r=l}$. This is because consumers will return the product to the physical retailer's store in the C scenario.

Proposition 4.3.1. In the N scenario, the firms' equilibrium prices and equilibrium profits are summarized in the following table.

Case	Equilibrium profits	Equilibrium prices	Seg
N1	$\tilde{\pi}^N_{F1}, \tilde{\pi}^N_{O1}$	$\tilde{p}^N_{F1} = \frac{2}{3}, \tilde{p}^N_{O1} = \frac{1}{3}$	E-S-F

In the C scenario, depending on offline shopping cost l and physical retailer's marginal return cost s_F, the equilibrium compensation, the equilibrium prices, and the equilibrium profits are summarized in the following table.

Case	Offline shopping cost l & Marginal return cost s_F	Equilibrium compensation & profits	Equilibrium prices	Seg
C1	$l \le \hat{l}^C_{11}$ and $0 < s_F < \bar{s}_F$, or $\hat{l}^C_{11} < l \le \hat{l}^C_{12}$ and $0 < s_F < \hat{s}_{F11}$, or $\hat{l}^C_{12} < l \le \hat{l}^C_{14}$ and $0 < s_F < \hat{s}_{F12}$.	$\tilde{f}_1, \tilde{\pi}^C_{F1}, \tilde{\pi}^C_{O1}$	$\tilde{p}^C_{Fj} = \frac{2}{3}, \tilde{p}^C_{O_j} = \frac{1}{3}$ $(j = 1, 2, 3)$	E-S-F
C2	$\hat{l}^C_{11} < l \le \hat{l}^C_{12}$ and $\hat{s}_{F11} \le s_F < \bar{s}_F$.	$\tilde{f}_2, \tilde{\pi}^C_{F2}, \tilde{\pi}^C_{O2}$		
C3	$\hat{l}^C_{12} < l \le \hat{l}^C_{13}$ and $\hat{s}_{F12} \le s_F < \bar{s}_F$, or $\hat{l}^C_{13} < l \le \hat{l}^C_{14}$ and $\hat{s}_{F12} \le s_F < \hat{s}_{F13}$.	$\tilde{f}_3, \tilde{\pi}^C_{F3}, \tilde{\pi}^C_{O3}$		
C4	$\hat{l}^c_{13} < l \le \hat{l}^C_{14}$ and $\hat{s}_{F13} \le s_F < \bar{s}_F$, or $\hat{l}^C_{14} < l < \bar{l}$ and $\hat{s}_{F22} \le s_F < \bar{s}_F$.	$\tilde{f}_4$ can be any value in $\left[\hat{f}_{21}, \hat{f}_{22}\right]$, $\tilde{\pi}^C_{F4}, \tilde{\pi}^C_{O4}$	$\tilde{p}^C_{F4} = f + \frac{4-2s_F-l}{3}$, $\tilde{p}^C_{O4} = f + \frac{2-s_F+l}{3}$	E-F

Part of the expressions in the tables above are: $\hat{l}^C_{11} = (\sqrt{13} - 3)/6, \hat{l}^c_{12} = (\sqrt{2} - 1)/3$,

$\hat{l}^C_{13} = (24\sqrt{2}-25)/51, \hat{l}^C_{14} = (2\sqrt{2}-2)/3, \hat{s}_{F11} = 1+(1/9l), \hat{s}_{F12} = (5+2\sqrt{2}-(9+6\sqrt{2})l)/3,$ $\hat{s}_{F13} = (7+4\sqrt{2}-(18+12\sqrt{2})l)/3, \hat{s}_{F22} = (4-(4+3\sqrt{2})l)/2, \tilde{f}_1 = 0, \tilde{f}_2 = s_F - 1 - (1/9l)$ $\tilde{f}_3 = s_F + (3+2\sqrt{2})l - (5+2\sqrt{2})/3, \hat{f}_{21} = \left(1+2\sqrt{2}-(7+5\sqrt{2})l-(\sqrt{2}-1)s_F\right)/3$, and $\hat{f}_{22} = \left((7+5\sqrt{2})l+(2+\sqrt{2})s_F-(4+2\sqrt{2})\right)/3$. The rest of the expressions are relegated to Section C of the online appendix for brevity.

Proposition 4.3.1 shows the equilibrium prices, profits, and market segmentation under both the N and C scenarios. In the N scenario, where consumers can only return products to the original seller, there is a single equilibrium case (N1) with equilibrium prices $\tilde{p}^N_{F1} = \frac{2}{3}$ and $\tilde{p}^N_{01} = \frac{1}{3}$ for the physical retailer and e-tailer, respectively. The market is segmented into three groups: consumers who purchase from the e-tailer (E), those who showroom (S), and those who purchase from the physical retailer (F).

In the C scenario, the equilibrium outcomes depend on the offline shopping cost l and the physical retailer's marginal return cost s_F. The proposition identifies four distinct cases (C1, C2, C 3 , and C4) based on different ranges of these parameters. In cases C1, C2, and C 3 , the equilibrium prices for both the physical retailer and e-tailer remain the same as in the N scenario. The market segmentation also remains unchanged, with consumers divided into e-tailer purchasers, showroomers, and physical retailer purchasers (E-S-F). However, the equilibrium compensation paid by the e-tailer to the physical retailer for handling returns $\left(\tilde{f}_1, \tilde{f}_2, \tilde{f}_3\right)$ and the equilibrium profits ($\tilde{\pi}^C_{F1}, \tilde{\pi}^C_{O1}, \tilde{\pi}^C_{F2}, \tilde{\pi}^C_{O2}, \tilde{\pi}^C_{F3}, \tilde{\pi}^C_{O3}$) differ across these cases. In case $C4$, which occurs when the offline shopping cost and the physical retailer's marginal return cost are relatively high, the equilibrium prices are functions of the compensation (f), offline shopping cost(l), and marginal return cost (s_F). The physical retailer's equilibrium price is $\tilde{p}^C_{F4} = f + \frac{4-2s_F-l}{3}$, while the e-tailer's equilibrium price is $\tilde{p}^C_{O4} = f + \frac{2-s_F+l}{3}$. The equilibrium compensation ($\tilde{f}_4$) can take any value within the range $\left[\hat{f}_{21}, \hat{f}_{22}\right]$. In this case, the market is segmented into only two groups: e-tailer purchasers and physical retailer purchasers (E-F), with no showroomers.

This lays the foundation for understanding the impact of the cross-return partnership on

the equilibrium outcomes in the market. Figure 4.1 summarizes the different cases that arise in the C scenario based on the offline shopping cost l and the physical retailer's marginal return cost s_F. It highlights how the offline shopping cost and the physical retailer's marginal return cost influence the equilibrium prices, profits, and market segmentation under different scenarios.

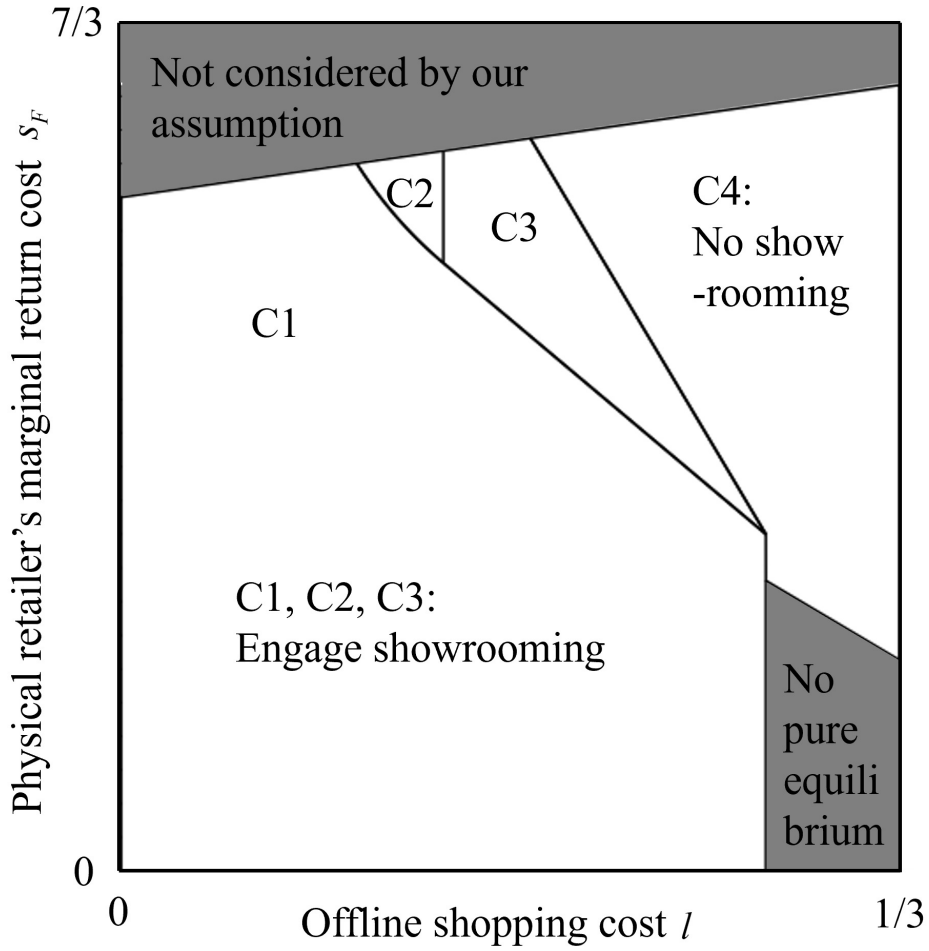


Figure 4.1: Summary of cases in C scenario from Proposition 1

4.3.2 Comparison of the **N** *and the* **C** *Scenario*

Based on the previous result, we next discuss the effects of the cross-return partnership on the etailer and physical retailer's corresponding profits and consumer surplus.

Proposition 4.3.2. The e-tailer is better off in the C scenario than in the N scenario:

1. When the offline shopping cost is below a threshold, i.e., $l < \hat{l}^c_{11}$,

2. or otherwise, when $l \geq \hat{l}^c_{11}$ and the physical retailer's marginal return cost is below a threshold, i.e., $s_F < \hat{s}_{F3}$.

The e-tailer cannot be better off in other market conditions.

The physical retailer can be better off in the C scenario than in the N scenario:

1. When the offline shopping cost is above a threshold, i.e., $l > \hat{l}^c_{13}$, and

2. the physical retailer's marginal return cost is above a threshold, i.e., $s_F > \hat{s}_{F4}$.

The physical retailer cannot be better off in other market conditions. The expressions of $\hat{l}^c_{11}$ and $\hat{l}^c_{13}$ are presented in Proposition 1. The remaining expressions are:

$$\hat{s}_{F3} = l+2-\sqrt{2-18s\,(2l-h_r)}, \quad \text{and} \quad \hat{s}_{F4} = \max\left\{\hat{s}_{F13}, l-1-3\sqrt{2}+\sqrt{(1+\sqrt{2})(1+17\sqrt{2}-36l)}\right\}$$

Proposition 4.3.2 states that the e-tailer can gain from the return partnership under two specific conditions. One condition is that when the offline shopping cost is below a certain threshold, i.e., $l < \hat{l}^C_{11}$, the e-tailer can benefit from the partnership. In this case, the inconvenience or additional costs associated with offline shopping are not excessively high. The e-tailer can utilize the physical retailer to facilitate returns, thus increasing their profits by lowering return costs and providing more options to the consumer. The other condition is that when the offline shopping cost is above this threshold, i.e., $l \geq \hat{l}^c_{11}$, as long as the physical retailer's marginal return cost is below another threshold, i.e., $s_F < \hat{s}_{F3}$, the e-tailer

still benefits from the partnership. This situation suggests that if the physical retailer's cost of handling returns is relatively low, the etailer can effectively outsource the return service at a cost lower than managing returns independently, thereby improving e-tailer's profitability.

On the other hand, the physical retailer can also benefit from the consumer return partnership under two conditions. The first condition is when the offline shopping cost is above a threshold, i.e., $l > \hat{l}^c_{13}$, implying that if consumers find it sufficiently inconvenient or costly to shop offline, they may be more inclined to purchase online. In this case, a partnership with an e-tailer can introduce more foot traffic to the physical retailer, thus capturing some of the online market. The second condition is when the physical retailer's marginal return cost is above a threshold, i.e., $s_F > \hat{s}_{F4}$, suggesting that if it is relatively expensive for the physical retailer to handle returns, partnering with the e-tailer for return service can alleviate some of the financial burden associated with returns, thus increasing the retailer's profitability.

Under certain market conditions, the return partnership can lead to a more collaborative relationship between e-tailers and physical retailers, reducing direct price competition and potentially leading to higher prices and profits for both retailers. This strategic collaboration could reposition physical retailers as friends in the e-commerce ecosystem rather than enemies, or merely showrooms for online purchases.

Proposition 4.3.3. Consumer surplus is lower in the C scenario compared to the N scenario when:

1. The offline shopping cost is above a threshold, i.e., $l > \hat{l}^c_{13}$, and

2. The e-tailer's marginal compensation f is above a threshold, i.e., $f > \hat{f}_1$.

Note that $\hat{f}_1 = (s_F - l)^2/36 - (2l - h_r)^2/2 + (4l/9) + (5s_F/9) - (11/18)$.

Proposition 4.3.3 reveals an intriguing insight into the impact of the return partnership on consumer surplus. The results indicate that consumer surplus can be lower in the cross-return (C) scenario compared to the no-cross-return (N) scenario under two conditions. The first

condition is when the offline shopping cost is above a threshold, i.e., \$1 > \hat{l}_{13} \wedge c\$, suggesting that if consumers find it relatively inconvenient or costly to shop offline, they may be more inclined to purchase online. The second condition is when the e-tailer's marginal compensation to the physical retailer for handling returns is above a threshold, i.e., \$f > \ hat {f}_1\$, implying that if the e-tailer pays a relatively high compensation to the physical retailer for each returned item, it can lead to a decrease in consumer surplus.

This finding is particularly interesting because it highlights a potential trade-off between the benefits of the return partnership for retailers and its impact on consumers. While the partnership can lead to higher profits for both the e-tailer and the physical retailer under certain conditions, as discussed in Proposition 2, it may come at the expense of reduced consumer surplus. This reduction in consumer surplus could be attributed to the softening of price competition between the two retailers, leading to higher prices for consumers.

Moreover, the result suggests that when consumers face high offline shopping costs, they may be more likely to purchase online. In this case, the return partnership could further encourage online purchases by providing a more convenient return option through the physical retailer. However, if the e-tailer's compensation to the physical retailer for handling returns is relatively high, it could lead to higher online prices, ultimately reducing consumer surplus.

This proposition highlights the importance of considering the impact on consumers when implementing strategic partnerships in the retail industry. While such collaborations can benefit retailers, it is crucial to ensure that they do not come at the expense of consumer welfare. Retailers should strive to find a balance that allows them to enhance their profitability while still providing value and competitive prices to consumers.

4.4 Conclusions and Future Research Directions

This paper investigates the impact of consumer return partnerships between e-tailers and retailers on their demand-side competition and profitability. Our analytical model reveals that such partnerships can soften price competition between the two types of retailers, potentially leading to a win-win situation for both parties under certain market conditions.

Specifically, when the offline shopping cost is below a threshold or when it is above the threshold but the retailer's marginal return cost is sufficiently low, the e-tailer can benefit from the partnership. On the other hand, the retailer can gain from the collaboration when the offline shopping cost is high, and its marginal return cost is above a threshold.

However, our findings also highlight a potential trade-off between the benefits of the return partnership for retailers and its impact on consumers. The softening of price competition may lead to higher prices, resulting in a decrease in consumer surplus when the offline shopping cost and the e-tailer's compensation to the retailer for handling returns are above certain thresholds. These insights contribute to the growing literature on consumer returns and the strategic interactions between e-tailers and traditional retailers. Our study underscores the importance of considering the indirect effects of return partnerships on demand-side competition and consumer welfare, in addition to the direct benefits of increased foot traffic and reduced return costs.

Future research could explore several extensions to our model. First, incorporating the impact of demand spillovers, where consumers returning products to the retailer make additional purchases, could provide a more comprehensive understanding of the partnership's effects. Second, considering heterogeneous consumer preferences for the online and offline channels could offer further insights into the market segmentation and competitive dynamics. Finally, empirical studies could test the predictions of our model and quantify the impact of return partnerships on retailer profitability and consumer surplus in real-world settings.

As e-commerce continues to grow and evolve, understanding the strategic implications of innovative partnerships between online and offline retailers will be crucial for firms to maintain their competitive edge and make informed decisions that benefit both their bottom line and their customers.

Chapter 5

AFRAID OF NICHE, TIRED OF MASS: ATYPICAL IDEA COMBINATION ON CROWDFUNDING PLATFORM

5.1 Introduction

Ideas are connected. Many models of creativity see new ideas as innovative combinations of existing ideas. The most well-known models among them should be the application of Darwinism in the behavioral sciences proposed by Campbell (1960). "Blind variation and selective retention in creative thought" (BVSR) illustrate the process of creativity, which also gained even more support during decades of research. As Simonton (1999) points out, the focus of this model is not the selection process but the variation process. The variations come from two means: recombining old genes and providing new genes through mutation. The metaphorical application of this theory describes the generation of new ideas. In this paper, we focus on crowdfunding ideas from an idea combination perspective, a major source of new variation, similar to the recombination of old genes in biology. Even though research on creativity has a long history, only recently have researchers found empirical evidence to support these theories. Wei (2020) argues that a new idea usually follows a stream of similar ideas yet simultaneously combines *atypical* elements from ideas outside this stream. Ideas that balance well familiarity and atypicality are considered as more creative (Toubia and Netzer 2017).

Hundreds of crowdfunding platforms have driven and witnessed the flourishing of crowdfunding during the past decade. The total volume increased rapidly across countries and gained more and more attention. In 2010, the global crowdfunding market was valued at 0.8 billion US dollars (Massolution 2013); in 2018, the number reached 10.2 billion US dollars and is expected to increase to 28.8 billion US dollars by the end of 2025 (QYResearch

2019). Creativity is essential to crowdfunding project success. If we take a closer look at the projects on *Kickstarter.com*, we notice that novelty is the key to success. One project from it, *terraplanter*, has already raised $4,502,569 by June 2020, wildly beyond the original goal of $20,000[1]. The main product is a vessel. But it takes only water and allows users to visibly follow the plant growth and flourishing process on the planter's exterior surface. No doubt, the novel design leads to its ultimate success.

Existing works on crowdfunding success have focused on the mechanism (Chemla and Tinn 2020), forms of crowdfunding (Belleflamme et al. 2014), and information revealing (Burtch et al. 2016). Extant research on atypicality innovation has found it to be important in scientific paper's impact (Uzzi et al. 2013), perceived creativity of ideas (Toubia and Netzer 2017), and motion picture popularity (Wei 2020). We are among the first to bridge the two streams of literature and investigate the business value and working mechanisms of atypicality innovation in crowdfunding. Specifically, our research questions are: (i) How to identify different idea combination patterns? (ii) What combination of ideas leads to high crowdfunding project outcomes? (iii) What are the underlying mechanisms that drive the prominent performance?

To address these questions, we choose our research context as *modian.com*, a crowdfunding website focusing on cultural and creative industries that was founded in 2014. Modian is one of the three biggest and most influential platforms in China and is regarded as "Kickstarter in China". For our work, we collect all the projects on this platform from 2014 until Feb 8, 2020. There are, in total, 1,826 projects with 1,062,182 donors and 1,087 organizers. We collect the data from three dimensions. Firstly, we retrieve all the user-level data, including all the projects they have donated to. Secondly, we look at it from the perspective of projects. We collect projects' text descriptions, user comments, the response of organizers, updates, locations, organizers' names, the categories they belong to, the amount that has been funded, the goal of the fundraising, number of users that have donated to the projects,

[1] https://www.kickstarter.com/profile/terraplanter

start date, and end date. Lastly, we collected the information of organizers, including the number of followers that organizers have (i.e., number of fans) and the number of projects they have created.

We first capture the different idea combination types by establishing a similarity network of crowdfunding projects. The nodes in the network are the projects, and the link between the two nodes represents that the two related projects are similar. We measure similarity between any pair of projects by calculating the semantic distance between the two projects' titles and descriptions. A project's atypicality or familiarity can be measured by the connectivity of its neighbors in the similarity network (Wei 2020). If the neighbors are less connected, it implies that the project is combining ideas from more distinct streams. If the neighbors are closely and extensively connected, it means the project is grounded in mainstream ideas. Then we leverage a cluster-robust double machine learning (DML) with Lasso (Chernozhukov et al. 2018) to investigate the relationship between atypicality innovation and project performance.

We find that (i) atypical combinations of mainstream and niche ideas greatly affect the success of crowdfunding projects by increasing the amount funded by almost five times; (ii) time affects how well atypical projects perform, with projects launched later exhibiting better performance; (iii) although donors are attracted by niche projects, the average amount donated by each donor is also lower; (iv) projects that are classified as experience goods have a dominant impact on the significant effect of niche projects; (v) the comments of mainstream projects are more negative, especially the idea related comment.

Based on the results we show above, our study provides important implications. We show that organizers should consider grounding their projects in the mainstream while also combining niche ideas to achieve success. We propose that the underperformance of niche projects is due to a duel pathways mechanism. The positive appealing effect cancels out the uncertainty inherent in niche projects. For mainstream projects, however, this poor performance originates from the audience's weariness. Though atypical projects stand out from the herd, we also show that it takes time before the audience adopts this new way of idea combination. Donors gradually become confident and convert to atypical projects. Lastly,

by comparing the DML method to other models, we find that crowdfunding data can suffer from endogeneity, nonlinearity and high dimensional issues. DML method requires the least model assumptions compared to traditional statistical models and resolves the regularization bias and regulation bias commonly seen in other machine learning models.

We contribute to the literature in several ways. First, although previous studies have made great efforts to identify atypicality, there is a lack of a universal interpretable paradigm that is readily applicable. Our study is among the first to model the process of idea combination and utilize it to differentiate various patterns of idea combination. The proposed method is not only validated by multiple robustness checks but also returns consistent results when using different outcome metrics. Second, we propose several underlying mechanisms from the donors' perspectives. We present the potential reasons why niche projects and mainstream projects are not well received, which is further validated by sentiment analysis, across product categories analysis, etc. We also highlight the important role of time in donors' adoption process of atypical idea combination.

5.2 Project Similarity Network Construction

We use the semantic similarity of project titles and descriptions to determine the similarity between two crowdfunding projects and construct a similarity network of these projects. In this section, we show how we measure semantic similarity and validate the semantic measurement in detail and how we construct and validate the network of projects.

5.2.1 Semantic Similarity

The title and description of a project summarize the main idea of the project. To measure the semantic similarity between two projects, we embed each project into a numerical vector, which contains the semantic information of the title and description. The cosine distance of the vectors indicates the semantic distance between the projects.

To do this, first, we extract key information from a project's title and description. We extract keywords and their corresponding weights from the title and description using the

term frequency–inverse document frequency (TF-IDF) algorithm (Rajaraman and Ullman 2011), a common term weighting scheme in information retrieval. The TF-IDF weight of a word increases proportionally to the number of times the word appears in the document and is offset by the number of documents in the corpus that contains the word. By its construction, such a weight indicates the informativeness of the corresponding keyword in the title or description. Formally, for each project i, we extract top K keywords $\{w_i^k\}_{k=1}^K$ with the highest TF-IDF weights from its title and description and use $\{v_i^k\}_{k=1}^K$ to denote their corresponding TF-IDF values. Here K is a hyper-parameter. In our dataset, we find that the title and description contain at most 104 words for all projects, with a mean of 72.62 and a standard deviation of 15.34. Considering these words contain stop words, we set K equal to 50[2]. If the number of words in the title and description is less than 50, we extract all the words.

Next, we use a pre-trained Chinese Word2Vec model[3] (Song et al. 2018) to map these keywords into 200-dimension vectors. The Word2Vec model is a commonly used statistical language model proposed by Mikolov et al. (2013), which maps a word to a high-dimensional vector that contains semantic information about the word. The model can exploit the semantic information of words and judge the semantic similarity between words by calculating the cosine distance between corresponding word vectors. With the model, we map each keyword w_i^k to a 200-dimension vector $\mathbf{w}_i^k$.

Third, we construct the semantic embedding vector $\mathbf{e}_i$ for each project i, by using the weighted average of its word vectors of the keywords with respect to their TF-IDF weights. It is empirically shown that averaging the embeddings of words in a sentence is a surprisingly successful and efficient way of obtaining sentence semantic information (Arora et al. 2019, Kenter et al. 2016). Formally we have $\mathbf{e}_i = \sum_k v_i^k \mathbf{w}_i^k$. By doing so, we give higher weights to keywords that are more informative. We then normalize all $\mathbf{e}_i$ to unit vectors so that the

[2]We changed K to 20 and 100 and found our semantic similarity measurements were not sensitive to this parameter.

[3]The model provides embeddings for over 8 million Chinese words and phrases, which are pre-trained on large-scale, high-quality data (Song et al. 2018).

inner product of two vectors will represent their cosine distance and range from -1.0 to 1.0.

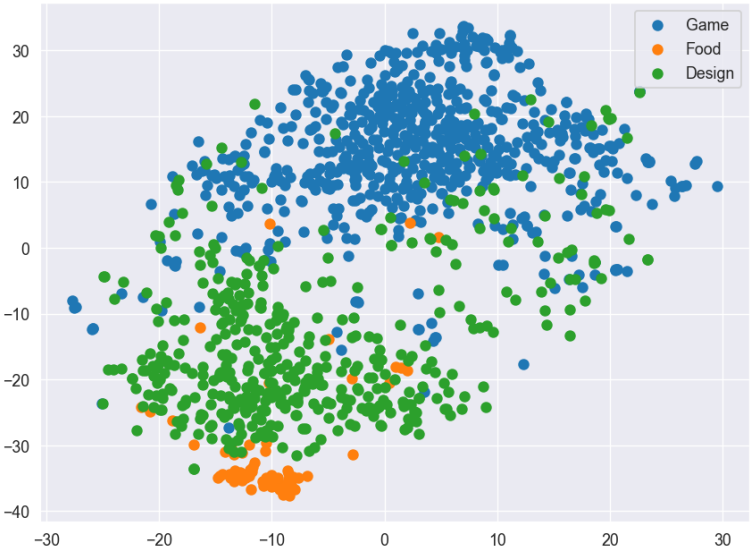
Finally, for each pair of projects (i, j), we compute their similarity distance s_{ij} by using the cosine distance of their embedding vectors, i.e., $s_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$.

We validate our embedding vectors as follows. If our vectors summarize the semantic information of the projects, they should be able to differentiate projects that belong to different categories. Such a practice is a common way to validate embedding vectors (Grbovic and Cheng 2018, Wang et al. 2018, Yu et al. 2022a). In our original dataset, there are 1,826 projects. *Game* (674 projects) and *Design* (455 projects) are the two categories that contain the highest number of projects. We also randomly select another category, *Food*, the projects which should be semantically different from those in *Game* and *Design*. For projects in the three categories, we use the T-distributed Stochastic Neighbor Embedding (t-SNE) algorithm (Maaten and Hinton 2008) to reduce our 200-dimension vector to 2 dimensions. It is a nonlinear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space of, for example, two or three dimensions. We find that projects in category *Game*, *Design* and *Food* clearly form three clusters (Figure 5.1). The results indicate our embedding vectors are valid in terms of measuring semantic distance.

5.2.2 Network Construction and Validation

We propose to use the similarity distance calculated by semantic analysis to build the projects' similarity network. We use a threshold α to discretize numerical distances to binary outcomes. If the semantic similarity of a pair of projects is above the threshold, we consider these two projects to be similar and build a link between the two projects. The optimal choice of α ($\alpha = 0.90$) is derived by considering both the validity of text analysis and atypicality, which we will further illustrate in the following section. As shown in Figure 5.2, each node represents a project, and an edge represents that two projects are similar to each other based on semantic distance. We choose three timestamps every two years and show the evolvement of the network. From the plots, we notice that there exists a core-

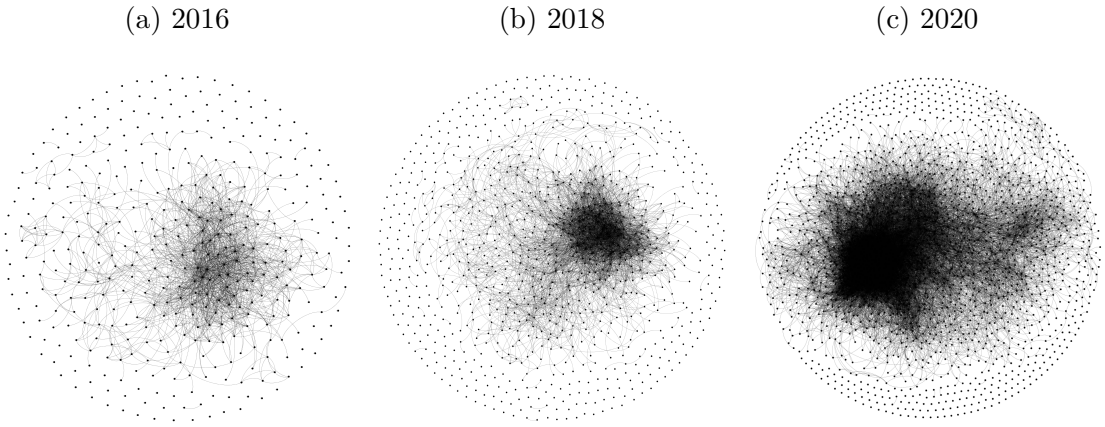
Figure 5.1: Visualization of three categories via t-SNE



peripheral structure. A large proportion of the projects are similar to each other, whereas others are peripheral and isolated, meaning that they are unique compared to others. There are also some projects which connect to both the core and some of the peripheral projects. Such projects are typically considered innovative by using "atypical combination" (Uzzi et al. 2013, Wei 2020). Further, as time goes by, the core grows bigger while, at the same time, more projects are connected to the network. This evolvement is in accord with the creation process of projects we assume. Organizers combine ideas from previous projects and they combine ideas differently. Some choose to ground the projects in the mainstream, while others choose to add atypical elements from peripheral projects.

To check the validity of the constructed network, Table 5.1 shows a logistic regression for all pairs of projects. The dependent variable is a dummy variable indicating whether there is a link between these two projects in the network. The independent variables are shown below. First, belonging to the same category is significantly associated with linkages. Apart from that, having the same organizers almost guarantees similarity. This echoes our result that sharing the same organizers is a significant indicator of the similarity between

Figure 5.2: Similarity network of crowdfunding projects



two projects. We also calculate the date difference between the release of two projects and include the first order and second order in the regression. The coefficients are negative and positive, respectively, indicating that the similarity decreases with longer release intervals within a period. After reaching a threshold, the similarity increases. This corresponds to the bandwagon effect that is commonly seen on crowdfunding platforms. Organizers learn from recent successful projects or hot topics on social media and follow the trend until new trends gain more attention. Therefore, we see a trend effect within a period. Lastly, we use the number of common donors to measure the similarity between the two projects. If two projects share common donors, the projects have latent characteristics that meet donors' joint preferences. With more sharing donors, the joint preferences become more specific (Wei 2020). Therefore, if two projects share many common donors, they must have particular characteristics in common, indicating high similarity. This agrees with our model that the projects are more likely linked if they share more common donors. In sum, the similarity measure obtained by our machine learning model is verified by the four interpretable features.

Through this logistic regression, we can check the robustness of our project similarity network by choosing different thresholds α. We also choose α around 0.90 and perform the

logistic regression again. We find that the variables in these regressions are statistically significant and have expected signs, indicating that our similarity network is not sensitive to the choices of α.

Table 5.1: Logit Regression Predicting Links

Variable	**Coefficient**	
Intercept	-4.60***	(0.0645)
Same category	1.49***	(0.0146)
Sharing organizer(s)	1.49***	(0.0531)
log(Date difference)	-0.321***	(0.0257)
log(Date difference)^2	0.0435***	(0.00256)
log(Number of sharing donors)	0.282***	(0.00670)
N	1,666,225	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

5.3 Descriptive Analysis

In this section, we will talk about how we distinguish between niche and mainstream ideas. Based on that, we also show how we define different types of idea combinations.

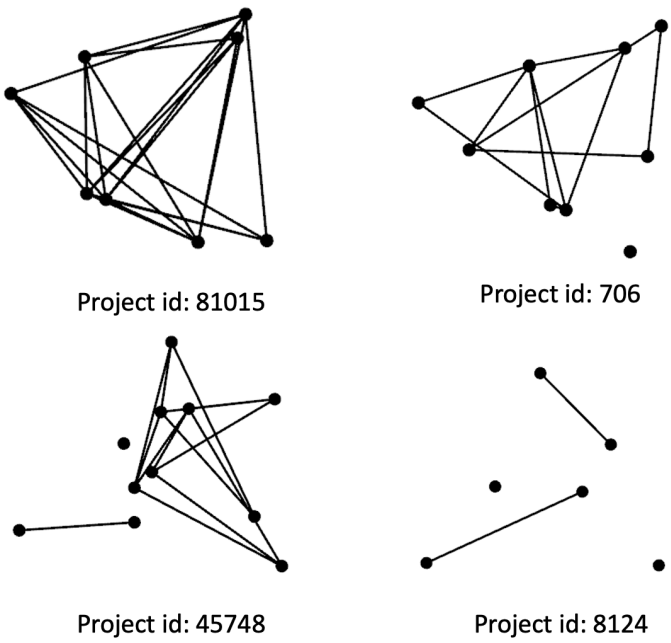
5.3.1 Niche Ideas and Mainstream Ideas

Previous studies (Uzzi et al. 2013, Toubia and Netzer 2017) have theorized and applied the idea of combinatorial creativity to a different context. In our case, we see that more than 70% (1,341 out of 1,823) projects are connected in the project similarity network. So it is reasonable to think of one project as a combination of previous ideas. This general point is also especially true because of the nature of the crowdfunding platform in our study. Modian is a crowdfunding platform focusing on creative cultural industries such as board games, animation, books, etc. The targeted area and audience determine the fact that ideas are concentrated and closely connected.

As Uzzi et al. (2013) pointed out, creativity results from the optimal balance between novelty and familiarity. Novel ideas are usually less known, more eye-catching, and also riskier. Ideas that people are familiar with are less risky but also less attractive. Therefore, we can correspondingly classify the ideas that projects grounded themselves in into two categories, niche ideas and mainstream ideas. For example, the anime garage kit is one stream of projects on the platform. Organizers produce creative figurines depicting characters from various anime. The bigger idea behind these projects can be generalized as the garage kit. The difference is that they center around different anime or characters for each project. So we can identify garage kit as a mainstream idea and anime as either a niche or mainstream idea depending on its popularity. So following Toubia and Netzer (2017), we construct a semantic similarity network of all projects. For any given project j, there is a local network, where nodes are j's prior similar projects (not including j), and links indicate whether any pair of these projects are similar. Hereafter, we refer to this network as j's local network. Figure 5.3 shows four typical local network structures. Project 81015 has seven interconnected nodes in its local network, which form one big cluster. We can conclude that project 81015 follows a mainstream idea. Similarly, project 706 has one large cluster but one singleton disconnected from the cluster. So, project 706 combines a mainstream idea and a niche idea. Project 45748, however, has three clusters, a major one and two minor ones. Building on

the mainstream idea, this project combines more niche ideas. The last project has four small clusters in its local network, which means combining four niche ideas and no mainstream ideas.

Figure 5.3: Local Network



We can easily identify project 81015 as a project grounded in mainstream ideas. The challenge is to classify the other three projects properly. Therefore, we focus on the clusters from local networks with more than one cluster. There are 1044 clusters, and 70.9% of them have a size less than or equal to 2. Considering originality and creativity are usually viewed as crucial features of niche ideas, and the general classification rules between the niche and the popular products, we set the threshold between niche and mainstream ideas as 2.

5.3.2 Types of Projects

Using the niche-mainstream idea paradigm, we identify four types of projects on this platform, i.e., mainstream projects, atypical mainstream projects, niche projects, and atypical niche projects.

Table 5.2: Summary of Different Types of Projects

Type of Projects	**Count**	**Avg(Idea Count)**	**Avg(Idea Size)**
Mainstream	617	1	26.512
Atypical Mainstream	288	2.441	7.486
Niche	291	1	1.265
Atypical Niche	137	2.431	1.210

Mainstream projects correspond to project 81015 in Figure 5.3. In its local network, prior similar projects form only one large cluster, i.e., one mainstream idea. We can conclude that project 81015 only learns from this mainstream idea, and we classify it as a mainstream project. In contrast, we categorize projects 706 and 45748 as atypical mainstream projects. They not only gain inspiration from mainstream ideas but also niche ideas. This non-conventional combination of previous ideas makes them both related and atypical of the mainstream. Similar to mainstream ideas, if a project only has one small cluster in its local network, i.e., it only learns from a niche idea, we call it a niche project. However, if a project combines multiple niche ideas as project 8124 in Figure 5.3, it is atypical of any of these niche ideas. Hence, we categorize it as an atypical niche project. For nodes that are singletons in the whole project similarity network and nodes that are the first among their similar projects, they do not borrow any ideas from previous similar projects. They are essentially

different from projects with small clusters in their local network since they are outside the framework of idea combination. So instead of defining them as niche or atypical niche ideas, we remove them from the model.

If we lower the linkage threshold α in the project similarity network, there will be more linkages. With more linkages, the number of small clusters will decrease. Given the definition of atypicality, we can easily conclude that lowering the linkage threshold dilutes atypicality. If the threshold is excessively low, all nodes in the similarity network will connect to each other. Then, no meaningful pattern can be found. On the other hand, if we increase the linkage threshold to a sufficiently large value, there will not be enough linkages to build the network, and the number of atypical projects will also decrease. Again, no meaningful pattern can be detected. Thus, it is essential to set an appropriate threshold value such that we can understand the atypicality of each project in a meaningful way. As we detail later, choosing α around 0.90 produces a robust similarity network that is neither too sparse nor too dense, from which consistent patterns are detected.

5.4 Empirical Strategy

There are several main challenges in the empirical analysis of causal inference of idea combination on projects' outcomes. First, the endogeneity issue in constructing the similarity network can make the estimates based on ordinary least squares (OLS) unreliable. Second, since organizers can create multiple projects, there also exists within-organizer correlations. Third, multiple control variables may exert nonlinear effects on the outcomes. For example, the interaction between multiple variables can significantly influence the outcome. Organizers who have created more projects previously are also likely to have more fans. Hence, to solve the aforementioned concerns, we compare the results using different models. Our baseline model is OLS. Then, we construct two-stage least squares by including seven instrumental variables. Further, to improve the accuracy and reduce model dependence, we employ double machine learning (DML) with Lasso (Chernozhukov et al. 2018), which enables us to add high-dimensional, nonlinear terms and resolves the curse of dimensionality. Finally, we

implement a cluster-robust DML (Chiang et al. 2022) to account for the within-organizer correlations in the DML procedure.

5.4.1 OLS Model

Equation 5.1 motivates baseline OLS model specifications that relate the observed project's outcome to project's type (constants omitted for brevity). Each project has a preset fundraising goal, and we use the percentage funded of this goal, $PctFunded$, to measure its success. $AtypicalMainstream_i$, $Niche_i$, and $AtypicalNiche_i$ are all dummy variables, indicating the type of projects. We use $Mainstream_i$ as the baseline, so the coefficient β is compared to it. In addition, x_i is a vector of control variables, as shown in Table 5.3. There are mainly two groups of variables, project-level controls and organizer-level controls. For project-level controls, we use the fundraising duration, the category it belongs to, count of updates, etc. For organizer-level controls, we employ the number of fans and the number of projects organizers have initiated. To account for within-organizer correlation, we allow the standard errors to be clustered at the organizer level.

$$log(PctFunded_i) = \beta_0 AtypicalMainstream_i + \beta_1 Niche_i + \beta_2 AtypicalNiche_i + \gamma x_i + \epsilon_i \tag{5.1}$$

5.4.2 Two-Stage Least Squares

We then estimate our data with two-stage least squares. Our instrumental variable specifications address confounding due to unobserved organizers' characteristics and simultaneity issues rising from the similarity network. Our instruments are motivated by the learning process of organizers. As Weitzman (1998) points out, "knowledge can build upon itself in a combinatoric feedback process." One channel of knowledge comes from previous projects posted on the platforms. Therefore, we collect the number of projects organizers have liked, favored, and backed. If an organizer like a project, he will be more likely to learn from it, which further influences his decision about project type. Additionally, organizers are

Table 5.3: Variables and Summary Statistics

Variables	Description	Mean	Std.	Max
Outcome variables				
$PctFunded$	Percentage funded of the preset fundraising goal	443.0	1275.9	19542.1
$DonorsCt$	Number of donors	536.6	2010.8	31866
$AvgAmount$	Average amount backed by donors	280.7	450.4	9733.6
$IdeaRatio$	Idea related comments ratio	0.26	0.15	1
$NegativeRatio$	Negative comments ratio	0.041	0.047	0.50
Control variables				
$TimeInterval$	Duration of fundraising of a project	36.8	15.9	141
$IdeaCt$	Number of ideas combined in a project	1.46	0.80	7
$CatTelevision$	Category television	0.16	0.37	1
$CatFood$	Category food	0.028	0.16	1
$CatGame$	Category game	0.56	0.50	1
$CatCreativity$	Category creativity	0.25	0.43	1
$UpdateCt$	Number of updates released by the organizer	9.18	11.4	98
$FanCt$	Number of fans each organizer has	1031.4	1696.0	9624
$CommentCt$	Number of comments each project has	385.0	1606.9	40452
$ProjectCreatedCt$	Number of projects each organizer has initiated	4.01	5.50	27
$FolSimilarProjectCt$	Number of similar projects after the release of this project	15.7	25.2	259
Projects Type (PT)				
$AtypicalMainstream$	Dummy variable, indicating a project is an atypical mainstream project	0.22	0.41	1
$Mainstream$	Dummy variable, indicating a project is a mainstream project	0.46	0.50	1
$Niche$	Dummy variable, indicating a project is a niche project	0.22	0.41	1
$AtypicalNiche$	Dummy variable, indicating a project is an atypical niche project	0.10	0.30	1
IVs				
$AtypicalNicheLN$	Average $AtypicalNiche$ in the local network	0.11	0.21	1

also very likely to learn from prior similar projects. Hence, we also employ the average of $AtypicalMainstream_i$, $Niche_i$, $AtypicalNiche_i$, and $Mainstream_i$ from its local network.

This learning process can also validate the exclusion conditions. The projects an organizer likes or are similar to are not directly connected to the project he creates. Additionally, the learning process of organizers is unobservable to donors. Thus, the correlation between the project's performance and the projects that organizers potentially learn from originates from how organizers construct the project, not other causal processes. Therefore, the instruments are exogenous to donors' decisions. To construct two-stage least squares, we build the following model:

$$\begin{aligned} PT_i &= \delta z_i + \theta x_i + e_i \\ log(PctFunded_i) &= \beta \widehat{PT_i} + \gamma x_i + \epsilon_i \end{aligned} \tag{5.2}$$

The first equation is the first-stage regression, where z_i is a vector of instrumental variables, x_i is a vector of controls, and PT_i is the project type. The F-statistic on the instruments greatly exceeds the rule-of-thumb threshold, which validates the relevance of the instruments. We obtain the estimated $\widehat{PT_i}$ from the first-stage regression and include it in the second-stage equation, as the second equation shows. The Sargan statistic is 3.492 ($p = 0.479$). The p-value is much higher than 0.05, indicating that the instrument variables pass the exogeneity test.

5.4.3 Debiased (Double) Machine Learning

To resolve the complex issues suffered by our dataset, we employ the DML method to estimate the effect of different idea combinations on the performance of each project. We first describe the basic setup that takes the form of the following partially linear IV regression model:

$$\begin{aligned} log(PctFunded_i) - \beta PT_i = g(X_i) + \zeta_i, \quad \mathrm{E}[\zeta_i \mid z_i, X_i] = 0, \\ z_i = m(X_i) + V_i, \quad \mathrm{E}[V_i \mid X_i] = 0, \end{aligned} \tag{5.3}$$

where $PctFunded_i$ is the project outcome, X_i is the controls and, z_i represents organizers' learning-related instrument variables. $g(\cdot)$ and $m(\cdot)$ are unknown nonparametric functions to be estimated. To consistently estimate β, it involves valid inference on $g(\cdot)$ and $m(\cdot)$.

However, a naive application of machine learning will result in inconsistent estimates $\hat{\beta}$ (Chernozhukov et al. 2018). Thus we alternatively construct the DML model that overcomes the "regularization biases" using orthogonalization.

$$\begin{aligned} W_i = U_i\beta + \zeta_i, \quad & \mathrm{E}[\zeta_i \mid z_i, X_i] = 0, \\ W_i = log(PctFunded_i) - \ell(X_i), \quad & \ell(X_i) = \mathrm{E}[log(PctFunded_i) \mid X_i], \\ U_i = PT_i - h(X_i), \quad & h(X_i) = \mathrm{E}[PT_i \mid X_i], \\ V_i = z_i - m(X_i), \quad & m(X_i) = \mathrm{E}[z_i \mid X_i]. \end{aligned} \tag{5.4}$$

We can summarize the procedure as follows: First, we use machine learning methods and cross-validation to estimate ℓ, h, and m. We divide the data into D equal-sized folds, and then for each fold $d = 1, \ldots, D$, we run a lasso regression on the other $D-1$ folds. According to Chernozhukov et al. (2018), the selection of D is usually 5 or 10. To get a robust estimator, we tried 10, 15, and 20, and the results remain consistent through different choices. The estimated results are $\widehat{\ell}^{(-d(i))}$, $\widehat{h}^{(-d(i))}$ and $\widehat{m}^{(-d(i))}$. Then from Equation 5.4, we can obtain the estimated residuals as $\widehat{W}_i$, $\widehat{U}_i$, and $\widehat{V}_i$. Using these estimations, by two-stage least squares, we can get our estimator $\hat{\beta}$.

We notice that in our context, the data is indexed one-way by organizer i. For the same organizer, their projects are likely to share some similarities. Therefore, our data entail a one-way dependence. The clustered structure implies a correlation of errors from train and test samples in the cross-fitting procedure if we employ the DML method. Therefore, we further propose a cluster-robust DML model (Chiang et al. 2022). The key difference of cluster-robust DML is a special sample splitting method used for the cross-fitting. To achieve independent data splits, Chiang et al. (2022) propose to incorporate the clusters into the partitioning process. For one-way clustering, we assume N clusters in the first dimension, and a K fold crossing-fitting. Then we randomly assign N clusters into K subsets $\{I_1, \cdots, I_K\}$. We train the model on a specific fold and use the other folds for prediction. In this way, we ensure the independence between train and test samples.

5.4.4 Results

To address the causal inference between idea combination and projects fundraising performance, we compare including and not including instruments in columns (1) and (2) in Table 5.4. Columns (3) and (4) show the DML model results without and with clustered Std. of organizers. The OLS estimates $AtypicalNiche$ projects perform significantly better than the others. After including instrumental variables, we observe the estimate of $AtypicalNiche$ becomes insignificant. This indicates that the positive effect of $AtypicalNiche$ suffers endogeneity bias. We construct the DML Lasso model by considering the nonlinearity virtue shared by multiple variables and the possible omitted interaction terms. In particular, we add the interaction terms between each pair of control variables into the model, which becomes high-dimensional and nonlinear. Then the Lasso model serves as a screening device to select the meaningful interaction terms and compress the coefficients of the meaningless terms to zero. As shown in column (3), the estimate of $AtypicalMainstream$ is 2.332. This implies that omitting certain interaction terms of control variables may induce over-estimation for the parameter of interest. We also note that the coefficient becomes significant. We deduce that this is because adding meaningful controls reduces the variance of the outcome variable and thus makes the relationship of interest more significant. If control for organizer clustered std, the estimate remains robust and decreases to 1.790. This indicates that omitting the within-organizer correlation may make the parameter of interest over-estimated. Since we use log transformation on the dependent variable, this estimate corresponds to a 498.9% increase in our dependent variables. Simply put, if the organizers adopt an atypical mainstream combination of ideas, their project will be five times more successful than projects with other idea combination strategies.

5.5 Mechanisms

In this section, we explore the underlying mechanisms that drive the success of atypical mainstream projects from multiple perspectives.

Table 5.4: Model Comparison

	(1)		(2)		(3)		(4)	
	OLS		2SLS		DML		One-Way DML	
$AtypicalMainstream$	0.0941	(0.0719)	3.08	(3.06)	2.33**	(1.11)	1.79**	(0.855)
$Niche$	0.0559	(0.0491)	0.239	(1.05)	0.224	(0.159)	0.0840	(0.252)
$AtypicalNiche$	0.204**	(0.0864)	2.67	(4.53)	0.490	(0.406)	0.348	(0.543)
$TimeInterval$	-0.00372***	(0.00134)	-0.00392**	(0.00190)	Yes		Yes	
$IdeaCt$	-0.0419	(0.0424)	-1.39	(1.87)	Yes		Yes	
$CatTelevision$	-0.158**	(0.0662)	-0.145	(0.140)	Yes		Yes	
$CatFood$	0.00803	(0.145)	-0.0106	(0.250)	Yes		Yes	
$CatGame$	-0.173***	(0.0582)	-0.135	(0.0849)	Yes		Yes	
$log(UpdateCt)$	0.207***	(0.0260)	0.234**	(0.0929)	Yes		Yes	
$log(FanCt)$	0.00326	(0.0146)	0.00352	(0.0372)	Yes		Yes	
$log(CommentCt)$	0.283***	(0.0305)	0.286***	(0.0401)	Yes		Yes	
$ProjectCreatedCt$	0.0170***	(0.00488)	0.0136*	(0.00721)	Yes		Yes	
$log(SimilarProjectCt)$	0.0272*	(0.0149)	0.0287	(0.0707)	Yes		Yes	
$Constant$	-0.510***	(0.123)	0.443	(2.02)	Yes		Yes	
$ClusteredStd$	Yes		Yes		No		Yes	
IV	No		Yes		Yes		Yes	
$InteractionTerm$	No		No		Yes		Yes	
Observations	1337		1337		1337		1337	
R2	0.428		.		.		.	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

Donors' comments on a project reveal some of the key mechanisms through which ideas of projects affect their crowdfunding performance. We conduct a deep-learning-based content analysis on comments to understand donors' opinions and sentiments toward different types of ideas.

To begin, we identify comments that are idea-related. First, we randomly sampled 4,100 comments and manually sorted them into two categories: idea-related or not. We find that 33.9% of the annotated comments (1,389) are idea-related. Second, we randomly select 90% of the annotated comments (3,690) as the training set and select the rest of them as the test set. We implement a Bidirectional Encoder Representations from Transformers (BERT) model pre-trained for Chinese text (Devlin et al. 2018) to map each of the comments into a 768-dimensional vector, which is also referred to as an embedding vector. Developed by Google, the BERT model is one of the most popular language models and achieves state-of-the-art performance on a number of natural language understanding tasks (Devlin et al. 2018). The number of dimensions is a preset parameter by the model. An embedding vector produced by the BERT model preserves the semantic information of a Chinese document and is commonly used as a semantic feature for document classification tasks.

Third, based on the embedding vectors and their corresponding labels, we select the best machine-learning-based classification model from the three candidate models, that is, the penalized logistic regression (PLR) (Bishop and Nasrabadi 2006), the random forest (RF) (Breiman 2001), and the extreme gradient boosting (XGB) (Chen and Guestrin 2016). These models are either commonly used or considered state-of-the-art for document classification tasks. Moreover, the dimension of our training data is a matrix of 3,690 rows and 768 columns, which is relatively high-dimensional. PLR, RF, and XGB are well-known to fit high-dimensional data in the machine learning literature because they either use a penalization strategy to automatically select the most important features or use a tree-based boosting strategy to mitigate the overfitting issue. We use five-fold cross-validation and the randomized parameter search approach (Bergstra and Bengio 2012) to determine the optimal parameters for each model using the validation set. After training each model with its opti-

mal parameters, we compare their classification performance using the test set. Specifically, we calculate the weighted average precision, recall, and F1-score based on the out-of-sample predictions generated by each model. Using the weighted average scores allows us to account for the unbalance between the two classes (i.e., idea-related or not). The classification results are presented in Table 5.5 panel (a). PLR outperforms RF and XGB regarding precision, recall, and F1-score. Training a PLR model with BERT-generated embedding vectors resembles the canonical BERT in which embedding vectors are forwarded to a fully-connected neural network layer with penalization and when the cross-entropy loss function is used to evaluate the prediction errors. Therefore, although RF and XGB are more flexible models that can incorporate nonlinear relationships, PLR fits the relationship between embedding vectors and content labels better than RF and XGB in our task. We deploy the trained PLR model to sort all 281,763 comments in our sample into 109,109 idea-related ones and 172,654 idea-unrelated ones.

Table 5.5: Classification Model Performance

	(a) Content Classification			(b) Sentiment Classification		
Model	Precision	Recall	F1-score	Precision	Recall	F1-score
PLR	**0.83**	**0.82**	**0.82**	**0.82**	0.77	**0.78**
RF	0.79	0.80	0.79	0.75	0.75	0.75
XGB	0.79	0.79	0.79	0.81	**0.78**	0.76

Next, we analyze the sentiments expressed in the idea-related comments. In our annotated set, we identify 1,389 idea-related comments. We further manually sort them into one of the three categories, i.e., positive (532 comments), negative (125 comments), and neutral (732 comments). We follow the same procedure to map comments to embedding vectors and

train and test the three classification models. As shown in Table 5.5 panel (b), we find that PLR outperforms RF and XGB in precision and F1-score, whereas XGB slightly outperforms PLR in the recall. Based on the result, we use PLR to sort all idea-related comments identified in the last step into positive, negative, and neutral ones. Out of 109,109 idea-related comments, there are 16,475 negative ones, 51,608 neutral ones, and 41,026 positive ones, which also follows the distribution of our annotation set.

5.5.1 Niche Projects Dual Pathways

As discussed previously, users are more attracted by niche projects (Berlyne 1970), while people also hold a 'bias against creativity' due to inherent uncertainty (Mueller et al. 2012). Hence, apart from the appealing effect, we propose that niche projects repel donors through a cognitive pathway, such that they increase the magnitude of uncertainty. In conclusion, we hypothesize niche projects impact donors through these dual pathways. In addition, the negative effect through uncertainty can cancel out the positive appealing effect. We testify to this theory in the rest of this section.

Following the analysis of how well each project fulfills its fundraising goal, we also measure the performance of a project in a comprehensive way. Besides the aforementioned perspective of donors' comments, we also examine the number of donors and the average amount of money backed by each donor. We construct a similar DML model as Equation 5.4 and change the outcome variables y_i to $log(DonorsCt)$, $log(AvgAmount)$, $IdeaRatio$, and $NegativeRatio$. Here, $IdeaRatio$ refers to the ratio of idea-related comments to all comments, and $NegativeRatio$ refers to the proportion of negative comments to all comments.

Results are presented in Table 5.6. Column (1) indicates that niche projects have a positive and significant estimate of 0.872. This suggests a 139.2% increase in the number of donors. Compared to mainstream projects, niche projects attract more donors to raise money. This result reveals the positive effect of niche projects and provides evidence of the first pathway. Then, we estimate the average amount of money backed by each donor in column (2). The estimate of $Niche$ is -1.083, which means the average amount decreases by

66.1% compared to mainstream projects. Therefore, we can draw the conclusion that even though more donors are fascinated by the novelty of niche projects, their payment choices remain relatively conservative. This provides vitally important implications to organizers. Niche ideas can be an eyecatcher and even increase donors' conversion rate, but donors are cautious and unwilling to spend more on niche projects due to the symbiosis of high uncertainty and novelty. This dual-pathways mechanism also explains the question of why the overall performance of niche projects is not improved with more donors.

To check the robustness of the proposed dual pathways, we adopt the concept of search and experience goods that is first introduced by Nelson (1970, 1974). He divides products into search and experience goods depending on how easily customers can obtain product quality information prior to purchase. Customers can easily investigate and evaluate search goods' attributes before buying. On the other hand, it is more difficult for customers to determine experience goods' attributes before making a purchase. Due to this intrinsic nature of experience goods, they are usually considered more novel and unpredictable. Hence, experience goods better satisfy donors' need to be recognized from the crowd, which in turn attracts more attention from donors. Also, the magnitude of uncertainty for experience goods is also higher. By dividing niche projects into search and experience goods, we further improve the threshold of niche projects. Thus, we should observe a more pronounced effect for niche projects about experience goods.

There are 12 categories of projects on the platform. Among them, we identify 'Publication', 'Activity', 'Media', 'Music', 'Food', 'Video game', and 'Board game' as experience goods, whereas 'Creative product', 'Toys', 'Anime memorabilia', 'Tech product' and 'Donation' as search goods. Table 5.7 columns (1) and (2) display the results of the DML models, which are comparable to those in Table 5.6 columns (1) and (2). The only distinction is that *Niche* has been substituted with *NicheExperience* and *NicheSearch*, respectively. First, when using donors' count as the outcome variable in column (1), *Niche* has a positive and significant estimate of 0.872, while this number increases to 1.31 for *NicheExperience*. *NicheSearch* becomes insignificant, which indicates that the effect of niche projects is driven

by niche projects about experience goods. Second, when the average amount backed by each donor is the outcome variable in column (2), *Niche* has a negative and significant estimate of -1.08, while this number drops to -1.35 for *NicheExperience* and *NicheSearch* becomes insignificant. This result further tests that niche projects about experience goods dominate the effect.

In conclusion, though niche projects attract more users, each donor's low engagement greatly impedes crowdfunding performance. Since crowdfunding projects entail uncertainty by nature, we can rationalize the underlying cause of low engagement of donors lies in the high risk associated with niche projects. For mainstream projects, countless prior similar projects provide a clear benchmark for donors. As previous studies show that consumers experience higher learning costs when adopting innovations, donors are more concerned about niche projects due to limited prior knowledge. This further results in a conservative choice made by donors.

5.5.2 Mainstream Projects Weariness

Donors are more hesitant about niche projects, but do mainstream projects outperform them? In column (3), we summarize the estimate of the ratio of idea-related comments to the total number of comments. *AtypicalMainstream* is negative and significant, while *Niche* is positive. This suggests that the idea-related comments ratio of atypical mainstream projects decreases by 43.5% compared to mainstream projects, and that of niche projects increases by 9.1%. In column (4), we change the outcome variable to the ratio of negative comments. All three types of projects have significantly fewer negative comments compared to mainstream projects. Among them, the ratio of *AtypicalMainstream* witnesses the highest decrease of 10.6%. Keep in mind that out of 109,109 idea-related comments, there are only 16,475 negative comments. Therefore, a ten-percent drop in the negative comments ratio is a huge difference. Combining these two columns, we can come to the following conclusions. Even though more idea-related comments are seen under mainstream projects, significantly more of them are negative comments. Niche projects inspire slightly more idea-related comments

Table 5.6: Measuring Project's Other Outcomes

	(1)		(2)		(3)		(4)	
	$log(DonorsCt)$		$log(AvgAmount)$		$IdeaRatio$		$NegativeRatio$	
$AtypicalMainstream$	1.77	(1.68)	2.86	(2.08)	-0.435***	(0.152)	-0.106**	(0.0525)
$Niche$	0.872**	(0.403)	-1.08**	(0.515)	0.0910*	(0.0502)	-0.0270*	(0.0143)
$AtypicalNiche$	1.58	(1.06)	-0.875	(1.29)	0.0467	(0.109)	-0.0790**	(0.0357)
$Controls$	Yes		Yes		Yes		Yes	
$Constant$	Yes		Yes		Yes		Yes	
$ClusteredStd$	Yes		Yes		Yes		Yes	
IV	Yes		Yes		Yes		Yes	
$InteractionTerm$	Yes		Yes		Yes		Yes	
Observations	1337		1337		1337		1337	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

Table 5.7: Interaction with Project's Type

	(1)		(2)		(3)	
	$log(DonorsCt)$		$log(AvgAmount)$		$log(PctFunded)$	
$AtypicalMainstream$	0.557	(1.40)	3.84*	(2.03)	-	
, $AtypicalMainstreamEarly$	-		-		-0.0351	(2.41)
$AtypicalMainstreamLater$	-		-		1.97**	(0.940)
$Niche$	-		-		0.0775	(0.249)
$NicheExperience$	1.31**	(0.553)	-1.35**	(0.661)	-	
$NicheSearch$	1.17	(0.922)	-1.40	(1.11)	-	
$AtypicalNiche$	1.89*	(1.11)	-0.439	(1.36)	0.389	(0.480)
$Controls$	Yes		Yes		Yes	
$Constant$	Yes		Yes		Yes	
$ClusteredStd$	Yes		Yes		Yes	
IV	Yes		Yes		Yes	
$InteractionTerm$	Yes		Yes		Yes	
Observations	1337		1337		1337	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

than mainstream ideas, and most importantly, donors are also more positive about them. Atypical mainstream projects have the most stable performance compared to the other two. It has relatively fewer idea-related comments and the lowest negative ratio. The idea-related comments ratio of $AtypicalNiche$ is not significantly different from $Mainstream$, but its negative comments ratio decreases by 7.9% compared to $Mainstream$. This results in a significant drop in the absolute amount of negative comments on atypical niche projects.

We further validate these results by looking from donors' perspectives. We observe 126,207 out of 679,211 donors have backed more than 1 project during the data collection period. This allows us to compare the performance of different types of projects vertically within each donor. When the platform was first founded in 2014, customer retention was low, which results from the fewer projects on the platform back then. Hence, to exclude this exogenous influence, we focus on the period after 2018, when the platform was already the leading company in the field. We fit the following model (Equation 5.5) to our data and the results are shown in Table 5.8.

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1 AtipicalNiche_{ij} + \beta_2 AtypicalMainstream_{ij} \\ + \beta_3 Mainstream_{ij} + \beta_4 Niche_{ij} + \mu_i + \theta_t + \epsilon_{ij} \end{aligned} \tag{5.5}$$

In column (1), the dependent variable is $NotActiveSince$, indicating whether we observe any future purchasing behavior after the donor backs on the focal project. Considering the majority of donors are one-time users, it is not surprising to see significant results for all types of projects. However, it is still worthwhile comparing the magnitude of the coefficient across different types. $Mainstream$ reaches the highest as 0.0340 and $AtypicalMainstream$ has the lowest as 0.00626. Therefore, users are most likely to churn after they back mainstream projects, while atypical mainstream projects significantly decrease this possibility. In columns (2) and (3), we use donors' comments' sentiment as dependent variables. $Mainstream$ has a significantly lower sentiment score compared to $AtypcialMainstream$ and $Niche$. After we filter the comment and keep only idea-related comments, we observe the negative effect of $Mainstream$ declines even more drastically.

Table 5.8: Donor's Engagement Analysis

	(1)		(2)		(3)	
	NotActiveSince		SentimentAvg		Idea Related SentimentAvg	
AtypicalNiche	0.00988***	(0.00285)	-0.0383***	(0.0147)	-0.0593**	(0.0231)
AtypicalMainstream	0.00626***	(0.00243)	-0.0105	(0.0118)	-0.0248	(0.0181)
Mainstream	0.0340***	(0.00201)	-0.0311***	(0.00996)	-0.0666***	(0.0155)
Niche	0.0252***	(0.00244)	0.00784	(0.0123)	-0.0254	(0.0192)
Constant	0.226***	(0.00150)	0.126***	(0.00757)	0.221***	(0.0118)
year	Yes		Yes		Yes	
month	Yes		Yes		Yes	
backer_id	Yes		Yes		Yes	
Observations	413333		37190		15606	
R2	0.388		0.315		0.352	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

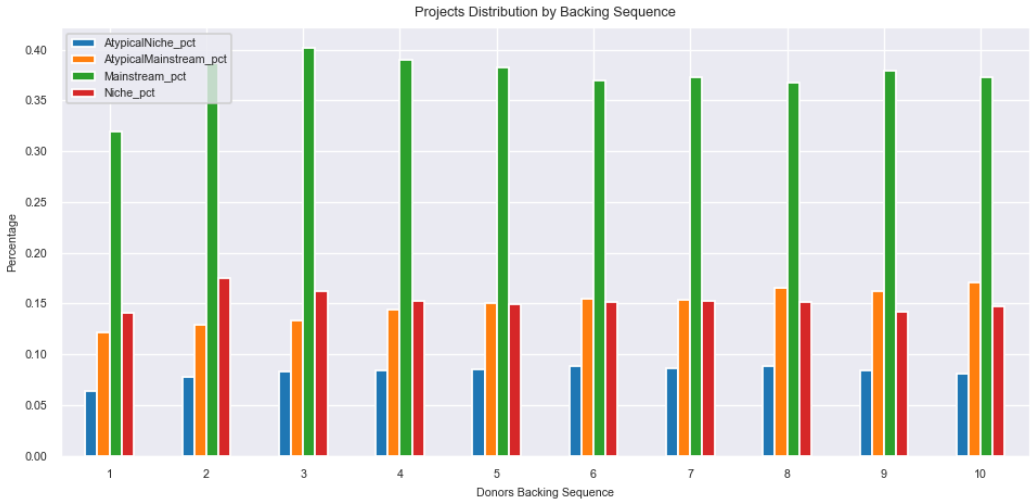
These results suggest that donors are dissatisfied with the idea proposed by organizers of mainstream projects. Too many similar projects presented on the platform bore the audiences with their repetition. Donors want the excitement of new ideas. In conclusion, the lack of freshness in mainstream projects induces weariness among donors, which further leads to donors leaving this platform.

5.5.3 Donors' Sequential Behavior Analysis

Previous studies have demonstrated that novel products are considered unpredictable and thus often rejected by decision makers(Staw and Cameron Ford 2004, Mueller et al. 2012). However, with donors devoting more time on the platform and familiarizing themselves with the context, do the 'bias against novelty' still remains true?

To test the effect of timing, we first plot the distribution of projects type based on donors' backing sequence in Figure 5.4. The x-axis indicates the sequence of projects backed by donors, and the y-axis is the percentage. Because singletons are out of the scope of our study, we do not include them in the plot. Figure 5.4 show evidence of an ever-increasing trend of atypical mainstream projects - the percentage of donors adopting this type of project becomes larger. To further test this observation, We divide projects into early-initiated and later-initiated ones. We define projects launched prior to 2018 as early projects (411 out of 1337); after and including 2018 as later projects (926 out of 1337). Then we identify early atypical mainstream projects and later atypical mainstream projects. We replace $AtypicalMainstream$ in our main model (Table 5.4) with the $AtypicalMainstreamEarly$ and $AtypicalMainstreamLater$. In Table 5.7 column (3), we observe $AtypicalMainstreamLater$ has an effect of 1.97 on the outcome $PctFunded$, whereas the effect of $AtypicalMainstreamEarly$ becomes insignificant. This validates our assumption that it takes some time before atypical mainstream projects become popular and recognized by donors.

Figure 5.4: Projects Distribution by Backing Sequence

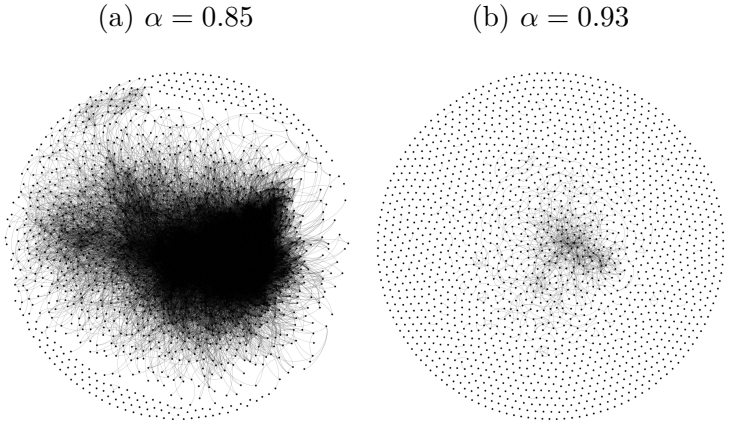


5.6 Robustness Check

To further assess the stability of our main results, we rebuild the semantic similarity network. In our prior work, we set the threshold of similarity network as $\alpha = 0.9$ to generate a network that is not too densely or loosely connected. In this section, we test our models on different values of α to prove our results are robust to the choice of the threshold value. As discussed above, if we set the threshold too low, all projects are connected. As a result, we dilute the proportion of niche ideas in such case. Figure 5.5 (a) demonstrates a densely connected network when $\alpha = 0.85$. On the other hand, if we set the threshold too high, no projects are connected, and we deduce the proportion of mainstream idea. Figure 5.5 (b) demonstrates a loosely connected network when $\alpha = 0.93$.

Hence we can conclude from Figure 5.5 that when $\alpha <= 0.85$ or $\alpha >= 0.93$, the network does not capture the similarity between projects. Also, the majority of semantic similarity scores s_{ij} between a pair of projects fall in the range of 0.8 to 1 (the 25% percentile is 0.787).

Figure 5.5: Extreme cases of similarity network



For ease of reading, we rescale α from 0.8 to 1 as 0 to 2 and denote it as $\hat{\alpha}$. Therefore, the rescaled $\alpha = 0.9$ is $\hat{\alpha} = 1$. We pick 10 values from $\hat{\alpha} = 0.95$ to $\hat{\alpha} = 1.05$. In table 5.9 we summarize the model results when we pick threshold values smaller than the original threshold, and table 5.10 shows the results when threshold values are larger than it. The variable of interest $AtypicalMainstream$ remains positive and significant from 0.97 to 1.04. As shown in Figure 5.6, though the networks with $\hat{\alpha} = 0.97$ and $\hat{\alpha} = 1.04$ are not well-balanced as $\hat{\alpha} = 1.00$, they are still within a reasonable range. It is also worth noticing that the count of projects included in the model drops with the increase of $\hat{\alpha}$. The higher the threshold value, the less the number of linkages in the network, and as a result, there are more singletons that are not connected to the network. In conclusion, the change of $\hat{\alpha}$ changes the network structure, but our findings are still valid.

5.7 Conclusion

Our work makes several significant contributions. First, for studies on crowdfunding projects' success, we introduce the intrinsic factor of projects' ideas quality (i.e., atypicality), while previous researchers primarily focus on extrinsic factors. Second, we enrich the literature on

Table 5.9: Model Comparison

	(1)		(2)		(3)		(4)		(5)	
	$\hat{\alpha} = 0.95$		$\hat{\alpha} = 0.96$		$\hat{\alpha} = 0.97$		$\hat{\alpha} = 0.98$		$\hat{\alpha} = 0.99$	
Controls	Yes		Yes		Yes		Yes		Yes	
AtypicalMainstream	0.512	(0.716)	1.80	(1.11)	1.55**	(0.786)	1.40**	(0.786)	1.40**	(0.694)
Niche	-0.125	(0.172)	-0.134	(0.166)	-0.0124	(0.132)	-0.00438	(0.125)	0.0656	(0.119)
AtypicalNiche	0.0382	(0.376)	-0.0502	(0.362)	-0.0453	(0.353)	0.172	(0.370)	0.413	(0.398)
Constant	Yes		Yes		Yes		Yes		Yes	
ClusteredStd	Yes		Yes		Yes		Yes		Yes	
IV	Yes		Yes		Yes		Yes		Yes	
InteractionTerm	Yes		Yes		Yes		Yes		Yes	
Observations	1430		1416		1393		1381		1365	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

Figure 5.6: Similarity network with different threshold

(a) $\hat{\alpha} = 0.97$ (b) $\hat{\alpha} = 1.00$ (c) $\hat{\alpha} = 1.04$

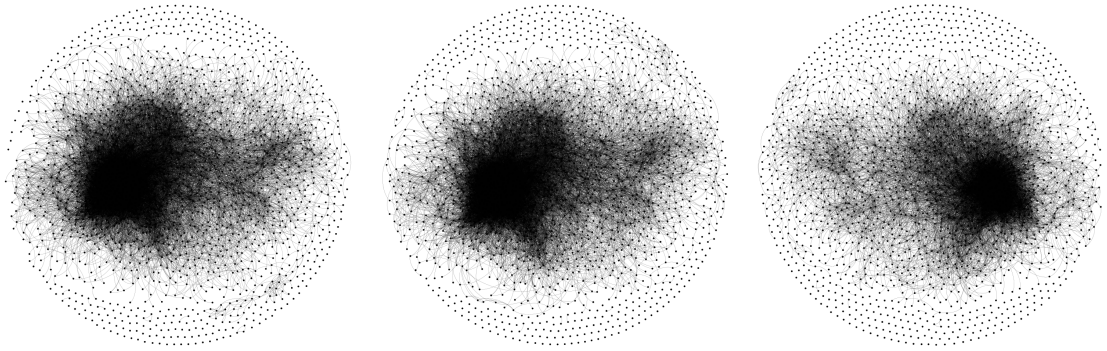


Table 5.10: Model Comparison

	(1)		(2)		(3)		(4)		(5)	
	$\hat{\alpha} = 1.01$		$\hat{\alpha} = 1.02$		$\hat{\alpha} = 1.03$		$\hat{\alpha} = 1.04$		$\hat{\alpha} = 1.05$	
Controls	Yes		Yes		Yes		Yes		Yes	
AtypicalMainstream	1.36*	(0.695)	3.04*	(1.78)	2.34**	(1.01)	1.90*	(1.14)	1.79	(1.16)
Niche	0.0502	(0.118)	0.0523	(0.122)	0.0816	(0.109)	0.131	(0.113)	0.166	(0.118)
AtypicalNiche	0.335	(0.411)	-0.289	(0.533)	0.150	(0.482)	0.122	(0.569)	0.249	(0.542)
Constant	Yes		Yes		Yes		Yes		Yes	
ClusteredStd	Yes		Yes		Yes		Yes		Yes	
IV	Yes		Yes		Yes		Yes		Yes	
InteractionTerm	Yes		Yes		Yes		Yes		Yes	
Observations	1319		1295		1266		1250		1221	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors in parentheses

idea generation and creativity in an empirical context. We employ a unique network construction approach and multiple machine learning methods to identify different idea combination patterns on crowdfunding platforms. Previous studies measure atypicality according to their own context without a universal and interpretable solution. We propose an interpretable approach to detect and classify different types of idea combinations. We demonstrate that the success of a project can be maximized by creatively combining mainstream and niche ideas. Thirdly, we bring cluster-robust double machine learning to practice. It resolves multiple issues that are commonly seen in the crowdfunding context. Traditional statistical models require many assumptions, which can be hard to satisfy, while machine learning models often suffer regularization bias and overfitting bias. DML, however, incorporates machine learning to deduce model assumptions and solves regularization bias through orthogonalization and overfitting bias through cross-fitting. Using this method, we can accurately and efficiently estimate the models. Fourthly, by investigating the underlying mechanism, we also add to the stream of psychological studies on novelty. We empirically show that the negative feedback ratio of novel projects is significantly lower than mainstream projects, and more users back niche projects. These results validate the point that niche projects appeal more to customers than mainstream projects. But on the other hand, we also find that users are more cautious about niche projects, and the average amount they back are lower. The result confirms that the value of the atypical combination is from reducing uncertainty and increasing novelty. Fifthly, we point out the importance of time in atypicality adoption by the audience. Though the atypical idea combination promises great success, it still takes time until the audience realizes their values. Lastly, we are among the first to bridge the literature on crowdfunding success and atypicality innovation. Despite its importance in scientific paper impact (Uzzi et al. 2013) and idea generation (Toubia and Netzer 2017), the business value of atypicality innovation has been understudied, with one remarkable exception in the movie industry (Wei 2020). We hear anecdotally that creativity has its business value, but few empirical works investigate what is the optimal balance between familiarity and atypicality and more importantly, provide a robust and interpretable approach to identify such balance. Our pa-

per contributes to this literature by proposing different patterns of idea combination, and to what extent an optimal pattern can lead to unexpected crowdfunding success in multiple dimensions.

In addition to that, our work also provides important managerial implications. First, we observe significantly more negative emotions in the comments of mainstream projects, and we empirically show they are less successful. Hence, both the platform and organizers should put more focus on atypical or niche projects. Second, although niche projects attract more attention and donors are more positive, the high uncertainty hinders their donation decisions. This is especially true for experience goods. Therefore, reducing the perceived uncertainty of donors is of great importance when organizers start crowdfunding for niche projects. Organizers can provide more visual information or have a trial period. Third, it takes time before the audience adopts atypical projects. Both the platform and organizers should not be deterred if they do not observe the success of atypical projects in the beginning as anticipated. Instead, the platform should gradually cultivate users' confidence toward such idea combination projects. Lastly, our approach can help the platform distinguish different types of projects by how organizers combine ideas, thereby helping the platform identify promising projects.

Our work also motivates future research in several directions. First, due to the limitation of data, we do not observe enough projects with the combination of mainstream ideas in our dataset. It is valuable to see how such type of projects performs compared to other combination patterns. Secondly, we hardly observe sequential data of organizers. Therefore, the learning effect from the organizers' viewpoint remains to be investigated.

Chapter 6

CONCLUDING REMARKS

This dissertation has explored three critical aspects of the evolving retail ecosystem: leveraging spatiotemporal data for strategic decision-making, forging partnerships between physical retailers and e-tailers for consumer returns, and harnessing crowd wisdom in the crowdfunding context. The findings from these essays contribute to the growing literature on retail strategy, data-driven decision-making, and platform economies, offering valuable insights for firms navigating the complex and dynamic landscape of modern retail.

Chapter 3 demonstrates the power of tensor completion methods in leveraging high-dimensional, spatiotemporal data to estimate the treatment effect of deploying smart vending machines across various urban settings. By showcasing the superiority of this approach over traditional causal inference techniques, this research highlights the importance of data-driven decision-making in optimizing retail operations and underscores the potential of advanced machine learning techniques in tackling complex business problems.

Chapter 4 reveals the nuanced impact of a returns partnership between a physical retailer and an e-tailer on their demand-side competition. By identifying the conditions under which both firms can benefit but consumer surplus may decrease, this research contributes to the growing literature on omnichannel retailing and the strategic interactions between online and offline players. The findings underscore the importance of carefully evaluating the implications of strategic partnerships on both firm profitability and consumer welfare.

Chapter 5 sheds light on the role of atypical idea combinations in driving success on crowdfunding platforms. By identifying an optimal balance between familiarity and novelty that maximizes project funding, this research highlights the power of crowd wisdom in shaping the success of new ventures and contributes to the literature on platform economies and

innovation. The findings offer valuable insights for entrepreneurs and platform operators seeking to leverage the potential of crowdfunding to bring innovative ideas to life.

Together, these chapters paint a rich picture of the opportunities and challenges facing firms in the evolving retail ecosystem. As the boundaries between online and offline channels continue to blur, and the importance of data, partnerships, and crowd wisdom grows, firms that can effectively harness these forces will be well-positioned to gain a strategic advantage. Future research could build on the findings of this dissertation by exploring additional dimensions of the retail ecosystem, such as the impact of emerging technologies, the role of social influence, and the dynamics of multi-sided platforms.

As the retail landscape continues to evolve at a rapid pace, it is imperative for firms to remain agile, adaptable, and open to new ideas. By embracing data-driven decision-making, forging strategic partnerships, and leveraging the wisdom of crowds, retailers and e-tailers can navigate the complexities of the modern marketplace and chart a course towards sustainable growth and success. The insights and frameworks provided in this dissertation offer a foundation for future research and practice, as firms seek to thrive in the ever-changing world of retail.

BIBLIOGRAPHY

Abadie A, Diamond A, Hainmueller J (2015) Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2):495–510, URL http://dx.doi.org/https://doi.org/10.1111/ajps.12116.

Abadie A, Gardeazabal J (2003) The economic costs of conflict: A case study of the basque country. *American Economic Review* 93(1):113–132, URL http://dx.doi.org/10.1257/000282803321455188.

Acar E, Dunlavy DM, Kolda TG, Mørup M (2011) Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems* 106(1):41–56, ISSN 0169-7439, URL http://dx.doi.org/https://doi.org/10.1016/j.chemolab.2010.08.004, multiway and Multiset Data Analysis.

Agrawal A, Catalini C, Goldfarb A (2014) Some simple economics of crowdfunding. *Innovation policy and the economy* 14(1):63–97.

Ahlers GK, Cumming D, Günther C, Schweizer D (2015) Signaling in equity crowdfunding. *Entrepreneurship theory and practice* 39(4):955–980.

Alberto Abadie AD, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association* 105(490):493–505, URL http://dx.doi.org/10.1198/jasa.2009.ap08746.

Amabile TM (2018) *Creativity in context: Update to the social psychology of creativity* (Routledge).

Anderson ET, Hansen K, Simester D (2009) The option value of returns: Theory and empirical evidence. *Marketing Science* 28(3):405–423, ISSN 07322399, 1526548X, URL http://www.jstor.org/stable/23884196.

Arora S, Liang Y, Ma T (2019) A simple but tough-to-beat baseline for sentence embeddings. *5th International Conference on Learning Representations, ICLR 2017*.

Askin N, Mauskapf M (2017) What makes popular culture popular? product features and optimal differentiation in music. *American Sociological Review* 82(5):910–944.

Athey S, Bayati M, Doudchenko N, Imbens G, Khosravi K (2021) Matrix completion methods for causal panel data models. *Journal of the American Statistical Association* 116(536):1716–1730.

Bai J (2009) Panel data models with interactive fixed effects. *Econometrica* 77(4):1229–1279, URL http://dx.doi.org/https://doi.org/10.3982/ECTA6135.

Bai J, Ng S (2002) Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221, URL http://dx.doi.org/https://doi.org/10.1111/1468-0262.00273.

Balakrishnan A, Sundaresan S, Zhang B (2014) Browse-and-switch: Retail-online competition under value uncertainty. *Production and Operations Management* 23(7):1129–1145, ISSN 1059-1478, URL http://dx.doi.org/https://doi.org/10.1111/poms.12165.

Barak B, Moitra A (2022) Noisy tensor completion via the sum-of-squares hierarchy. *Mathematical Programming* 193(2):513–548, ISSN 1436-4646, URL http://dx.doi.org/10.1007/s10107-022-01793-9.

Bayati M, Cao J, Chen W (2022) Speed up the cold-start learning in two-sided bandits with many arms. *arXiv preprint arXiv:2210.00340* .

Bell DR, Gallino S, Moreno A (2018) Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science* 64(4):1629–1651, URL http://dx.doi.org/10.1287/mnsc.2016.2684.

Belleflamme P, Sahm M, Lambert T, Schwienbacher A (2014) Crowdfunding: Tapping the right crowd. *Journal of Business Venturing* 29(5):585–609.

Berger J, Packard G (2018) Are atypical things more popular? *Psychological science* 29(7):1178–1184.

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *Journal of machine learning research* 13(2).

Berlyne DE (1970) Novelty, complexity, and hedonic value. *Perception & psychophysics* 8(5):279–286.

Bishop CM, Nasrabadi NM (2006) *Pattern recognition and machine learning*, volume 4 (Springer).

Breiman L (2001) Random forests. *Machine learning* 45(1):5–32.

Burtch G, Ghose A, Wattal S (2016) Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research* 27(3):iii–v, 471–464.

Cai C, Li G, Poor HV, Chen Y (2022) Nonconvex low-rank tensor completion from noisy data. *Operations Research* 70(2):1219–1237, URL http://dx.doi.org/10.1287/opre.2021.2106.

Campbell DT (1960) Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review* 67(6):380.

Candes EJ, Plan Y (2010) Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936, URL http://dx.doi.org/10.1109/JPROC.2009.2035722.

Candès EJ, Recht B (2009) Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9(6):717–772, ISSN 1615-3383, URL http://dx.doi.org/10.1007/s10208-009-9045-5.

Chakraborty S, Swinney R (2021) Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management* 23(1):155–169.

Chan C, Berger J, Van Boven L (2012) Identifiable but not identical: Combining social identity and uniqueness motives in choice. *Journal of Consumer research* 39(3):561–573.

Chan CR, Parhankangas A (2017) Crowdfunding innovative ideas: How incremental and radical innovativeness influence funding outcomes. *Entrepreneurship Theory and practice* 41(2):237–263.

Chemla G, Tinn K (2020) Learning through crowdfunding. *Management Science* 66(5):1783–2290, iii–iv.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen X, Lei M, Saunier N, Sun L (2021) Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *CoRR* abs/2104.14936, URL https://arxiv.org/abs/2104.14936.

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters.

Chiang HD, Kato K, Ma Y, Sasaki Y (2022) Multiway cluster robust double/debiased machine learning. *Journal of Business & Economic Statistics* 40(3):1046–1056.

Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Farias VF, Li AA (2019) Learning preferences with side information. *Management Science* 65(7):3131–3149.

Farias VF, Li AA, Peng T (2023) Learning treatment effects in panels with general intervention patterns .

Filipović M, Jukić A (2015) Tucker factorization with missing data with application to low-

n

n-rank tensor completion. *Multidimensional Syst. Signal Process.* 26(3):677–692, ISSN 0923-6082, URL http://dx.doi.org/10.1007/s11045-013-0269-9.

Finke RA, Ward TB, Smith SM (1992) Creative cognition: Theory, research, and applications .

Fleming L (2001) Recombinant uncertainty in technological search. *Management science* 47(1):117–132.

Forman C, Ghose A, Goldfarb A (2009) Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Science* 55(1):47–57, ISSN 00251909, 15265501, URL http://www.jstor.org/stable/40539126.

Foster JG, Rzhetsky A, Evans JA (2015) Tradition and innovation in scientists' research strategies. *American Sociological Review* 80(5):875–908.

Friedland S, Lim LH (2016) Nuclear norm of higher-order tensors. URL https://arxiv.org/abs/1410.6072.

Gao F, Su X (2017) Online and offline information for omnichannel retailing. *Manufacturing & Service Operations Management* 19(1):84–98, URL http://dx.doi.org/10.1287/msom.2016.0593.

Goldenberg J, Mazursky D, Solomon S (1999) Creative sparks. *Science* 285(5433):1495–1496.

Grbovic M, Cheng H (2018) Real-time personalization using embeddings for search ranking at airbnb. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 311–320.

Hao L, Fan M (2014) An analysis of pricing models in the electronic book market. *MIS Quarterly* 38(4):1017–1032, ISSN 02767783, 21629730, URL https://www.jstor.org/stable/26627960.

Hao L, Kumar S (2024) Benefit of consumer showrooming for a physical retailer: A distribution channel perspective. *Management Science* 70(8):5208–5225, URL http://dx.doi.org/10.1287/mnsc.2020.01990.

Hao L, Tan Y (2019) Who wants consumers to be informed? facilitating information disclosure in a distribution channel. *Information Systems Research* 30(1):34–49, URL http://dx.doi.org/10.1287/isre.2017.0770.

He Y, Atia GK (2021) Robust m-estimation-based tensor ring completion: a half-quadratic minimization approach. *CoRR* abs/2106.10422, URL https://arxiv.org/abs/2106.10422.

He Y, Atia GK (2022) Robust low-tubal-rank tensor completion based on tensor factorization and maximum correntopy criterion.

Hox JJ, Moerbeek M, Van de Schoot R (2010) *Multilevel analysis: Techniques and applications* (Routledge).

Hwang EH, Nageswaran L, Cho SH (2022) Value of online–off-line return partnership to off-line retailers. *Manufacturing & Service Operations Management* 24(3):1630–1649, URL http://dx.doi.org/10.1287/msom.2021.1026.

Imbens GW, Rubin DB (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge University Press).

Jain P, Oh S (2014) Provable tensor factorization with missing data. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1*, 1431–1439, NIPS'14 (Cambridge, MA, USA: MIT Press).

Jain P, Pal S (2022) Online low rank matrix completion. *arXiv preprint arXiv:2209.03997* .

Jing B (2018) Showrooming and webrooming: Information externalities between online and offline sellers. *Marketing Science* 37(3):469–483, URL http://dx.doi.org/10.1287/mksc.2018.1084.

Karlsson L, Kressner D, Uschmajew A (2016) Parallel algorithms for tensor completion in the cp format. *Parallel Comput.* 57(C):222–234, ISSN 0167-8191, URL http://dx.doi.org/10.1016/j.parco.2015.10.002.

Kenter T, Borisov A, De Rijke M (2016) Siamese cbow: Optimizing word embeddings for sentence representations. *arXiv preprint arXiv:1606.04640* .

Kilmer ME, Braman K, Hao N, Hoover RC (2013) Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications* 34(1):148–172, URL http://dx.doi.org/10.1137/110837711.

Kilmer ME, Martin CD (2011) Factorization strategies for third-order tensors. *Linear Algebra and its Applications* 435(3):641–658, ISSN 0024-3795, URL http://dx.doi.org/https://doi.org/10.1016/j.laa.2010.09.020, special Issue: Dedication to Pete Stewart on the occasion of his 70th birthday.

Kim K, Park J, Pan Y, Zhang K, Zhang X (2022) Risk disclosure in crowdfunding. *Information Systems Research* .

Koch JA, Siering M (2019) The recipe of successful crowdfunding campaigns. *Electronic Markets* 29(4):661–679.

Kolda TG, Bader BW (2009) Tensor decompositions and applications. *SIAM Review* 51(3):455–500, URL http://dx.doi.org/10.1137/07070111X.

Kornish LJ, Jones SM (2021) Raw ideas in the fuzzy front end: Verbosity increases perceived creativity. *Marketing Science* 40(6):1106–1122.

Kreimer N, Stanton A, Sacchi MD (2013) Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *GEOPHYSICS* 78(6):V273–V284, URL http://dx.doi.org/10.1190/geo2013-0022.1.

Kumar A, Mehra A, Kumar S (2019) Why do stores drive online sales? evidence of underlying

mechanisms from a multichannel retailer. *Information Systems Research* 30(1):319–338, URL http://dx.doi.org/10.1287/isre.2018.0814.

Kwark Y, Chen J, Raghunathan S (2014) Online product reviews: Implications for retailers and competing manufacturers. *Information Systems Research* 25(1):93–110, ISSN 10477047, 15265536, URL http://www.jstor.org/stable/24700107.

Kwark Y, Chen J, Raghunathan S (2017) Platform or wholesale? a strategic tool for online retailers to benefit from third-party information. *MIS Q.* 41(3):763–785, ISSN 0276-7783.

Lal R, Matutes C (1989) Price competition in multimarket duopolies. *The RAND Journal of Economics* 20(4):516–537, ISSN 07416261, URL http://dx.doi.org/10.2307/2555731.

Landwehr JR, Wentzel D, Herrmann A (2013) Product design for the long run: Consumer responses to typical and atypical designs at different stages of exposure. *Journal of Marketing* 77(5):92–107.

Lin YK, Rai A, Yang Y (2021) Information control for creator brand management in subscription-based crowdfunding. *Information Systems Research* .

Liu A, Moitra A (2020) Tensor completion made practical. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20 (Red Hook, NY, USA: Curran Associates Inc.), ISBN 9781713829546.

Liu J, Musialski P, Wonka P, Ye J (2013) Tensor completion for estimating missing values in visual data. *IEEE Trans Pattern Anal Mach Intell* 35(1):208–220.

Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9(Nov):2579–2605.

Maslow AH (1943) A theory of human motivation. *Psychological review* 50(4):370.

Massolution (2013) The crowdfunding industry report, 2013cf .

Mazumder R, Hastie T, Tibshirani R (2010) Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* 11:2287–2322, ISSN 1532-4435.

Mehra A, Kumar S, Raju JS (2018) Competitive strategies for brick-and-mortar stores to counter "showrooming". *Management Science* 64(7):3076–3090, URL http://dx.doi.org/10.1287/mnsc.2017.2764.

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Mollick E (2014) The dynamics of crowdfunding: An exploratory study. *Journal of business venturing* 29(1):1–16.

Montanari A, Sun N (2018) Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics* 71(11):2381–2425.

Moon HR, Weidner M (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83(4):1543–1579, URL http://dx.doi.org/https://doi.org/10.3982/ECTA9382.

Mueller JS, Melwani S, Goncalo JA (2012) The bias against creativity: Why people desire but reject creative ideas. *Psychological science* 23(1):13–17.

Mugge R, Dahl DW (2013) Seeking the ideal level of design newness: Consumer response to radical and incremental product design. *Journal of Product Innovation Management* 30:34–47.

Nageswaran L, Hwang EH, Cho SH (2024) Offline returns for online retailers via partnership. *Management Science* 0(0):null, URL http://dx.doi.org/10.1287/mnsc.2023.01291.

Nelson P (1970) Information and consumer behavior. *Journal of political economy* 78(2):311–329.

Nelson P (1974) Advertising as information. *Journal of political economy* 82(4):729–754.

Nguyen H, Calantone R, Krishnan R (2020) Influence of social media emotional word of mouth on institutional investors' decisions and firm value. *Management Science* 66(2):887–910.

Noseworthy TJ, Trudel R (2011) Looks interesting, but what does it do? evaluation of incongruent product form depends on positioning. *Journal of Marketing Research* 48(6):1008–1019.

Ofek E, Katona Z, Sarvary M (2011) "bricks and clicks": The impact of product returns on the strategies of multichannel retailers. *Marketing Science* 30(1):42–60, ISSN 07322399, 1526548X, URL http://www.jstor.org/stable/23012321.

QYResearch (2019) Global crowdfunding market size, status and forecast 2019-2025 .

Rajaraman A, Ullman JD (2011) *Mining of massive datasets* (Cambridge University Press).

Rogers EM, Singhal A, Quinlan MM (2014) Diffusion of innovations. *An integrated approach to communication theory and research*, 432–448 (Routledge).

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55, ISSN 0006-3444, URL http://dx.doi.org/10.1093/biomet/70.1.41.

Schnurr B (2017) The impact of atypical product design on consumer product and brand perception. *Journal of Brand Management* 24(6):609–621.

Schumpeter JA (1934) *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle.* (Harvard University Press.).

Shulman JD, Coughlan AT, Savaskan RC (2009) Optimal restocking fees and information provision in an integrated demand-supply model of product returns. *Manufacturing & Service Operations Management* 11(4):577–594, URL http://dx.doi.org/10.1287/msom.1090.0256.

Shulman JD, Coughlan AT, Savaskan RC (2010) Optimal reverse channel structure for consumer product returns. *Marketing Science* 29(6):1071–1085, URL http://dx.doi.org/10.1287/mksc.1100.0578.

Signoretto M, Van de Plas R, De Moor B, Suykens JAK (2011) Tensor versus matrix completion: A comparison with application to spectral data. *IEEE Signal Processing Letters* 18(7):403–406, URL http://dx.doi.org/10.1109/LSP.2011.2151856.

Simonton DK (1999) Creativity as blind variation and selective retention: Is the creative process darwinian? *Psychological Inquiry* 309–328.

Song W, Chen J, Li W (2021) Spillover effect of consumer awareness on third parties' selling strategies and retailers' platform openness. *Information Systems Research* 32(1):172–193, URL http://dx.doi.org/10.1287/isre.2020.0952.

Song Y, Shi S, Li J, Zhang H (2018) Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 175–180.

Staw BM, Cameron Ford D (2004) 37. why no one really wants creativity. *Psychological Dimensions of Organizational Behavior* 476.

Stephen AT, Zubcsek PP, Goldenberg J (2016) Lower connectivity is better: The effects of network

structure on redundancy of ideas and customer innovativeness in interdependent ideation tasks. *Journal of Marketing Research* 53(2):263–279.

Su X (2009) Consumer returns policies and supply chain performance. *Manufacturing & Service Operations Management* 11(4):595–612, URL http://dx.doi.org/10.1287/msom.1080.0240.

Sun M (2011) Disclosing multiple product attributes. *Journal of Economics & Management Strategy* 20(1):195–224, ISSN 1058-6407, URL http://dx.doi.org/https://doi.org/10.1111/j.1530-9134.2010.00287.x.

Taeuscher K, Bouncken R, Pesch R (2021) Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Academy of Management Journal* 64(1):149–179.

Toubia O, Netzer O (2017) Idea generation, creativity, and prototypicality. *Marketing science* 36(1):1–20.

Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342(6157):468–472.

Wang J, Huang P, Zhao H, Zhang Z, Zhao B, Lee DL (2018) Billion-scale commodity embedding for e-commerce recommendation in alibaba. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 839–848.

Wang J, Veugelers R, Stephan P (2017) Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* 46(8):1416–1436.

Wei Y (2020) The similarity network of motion pictures. *Management Science* 66(4):1647–1671.

Wei YM, Hong J, Tellis GJ (2022) Machine learning for creativity: Using similarity networks to design better crowdfunding projects. *Journal of Marketing* 86(2):87–104.

Weitzman ML (1998) Recombinant growth. *The Quarterly Journal of Economics* 113(2):331–360.

Xia D, Yuan M (2020) Statistical Inferences of Linear Forms for Noisy Matrix Completion. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83(1):58–77, ISSN 1369-7412, URL http://dx.doi.org/10.1111/rssb.12400.

Xia D, Yuan M, Zhang CH (2021) Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *The Annals of Statistics* 49(1).

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1):57–76, URL http://dx.doi.org/10.1017/pan.2016.2.

Yang L, Wang Z, Hahn J (2020) Scarcity strategy in crowdfunding: An empirical exploration of reward limits. *Information Systems Research* 31(4):1107–1131.

Yang X, Smith RE (2009) Beyond attention effects: Modeling the persuasive and emotional effects of advertising creativity. *Marketing Science* 28(5):935–949.

Youn H, Strumsky D, Bettencourt LM, Lobo J (2015) Invention as a combinatorial process: evidence from us patents. *Journal of the Royal Society interface* 12(106):20150272.

Younkin P, Kuppuswamy V (2018) The colorblind crowd? founder race and performance in crowdfunding. *Management Science* 64(7):3269–3287.

Yu Y, Wang Y, Zhang G, Zhang Z, Wang C, Tan Y (2022a) Outstream video advertisement effectiveness. *Available at SSRN 4098246* .

Yu Y, Yang Y, Huang J, Tan Y (2022b) Unifying algorithmic and theoretical perspectives: Emotions in online reviews and sales. *MIS Quarterly* .

Yuan L, Zhao Q, Gui L, Cao J (2019) High-order tensor completion via gradient-based optimization under tensor train format. *Signal Processing: Image Communication* 73:53–61, ISSN 0923-5965, URL http://dx.doi.org/https://doi.org/10.1016/j.image.2018.11.012, tensor Image Processing.

Yuan M, Zhang CH (2016) On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics* 16(4):1031–1068.

Yue X, Raghunathan S (2007) The impacts of the full returns policy on a supply chain with information asymmetry. *European Journal of Operational Research* 180(2):630–647, ISSN 0377-2217, URL http://dx.doi.org/https://doi.org/10.1016/j.ejor.2006.04.032.

Zhang J, Yan Y, Guan J (2019) Recombinant distance, network governance and recombinant innovation. *Technological Forecasting and Social Change* 143:260–272.

Zhou J, Hao B, Wen Z, Zhang J, Sun WW (2024) Stochastic low-rank tensor bandits for multi-dimensional online decision making.

Zhou P, Lu C, Lin Z, Zhang C (2018) Tensor factorization for low-rank tensor completion. *IEEE*

Transactions on Image Processing 27(3):1152–1163, URL http://dx.doi.org/10.1109/TIP.2017.2762595.

Appendix A

ONLINE APPENDIX TO IMPACTS OF SMART VENDING MACHINES IN DIFFERENT URBAN SETTINGS - TENSOR COMPLETION WITH SPATIOTEMPORAL DATA

A.1 Additional Graphs and Plots

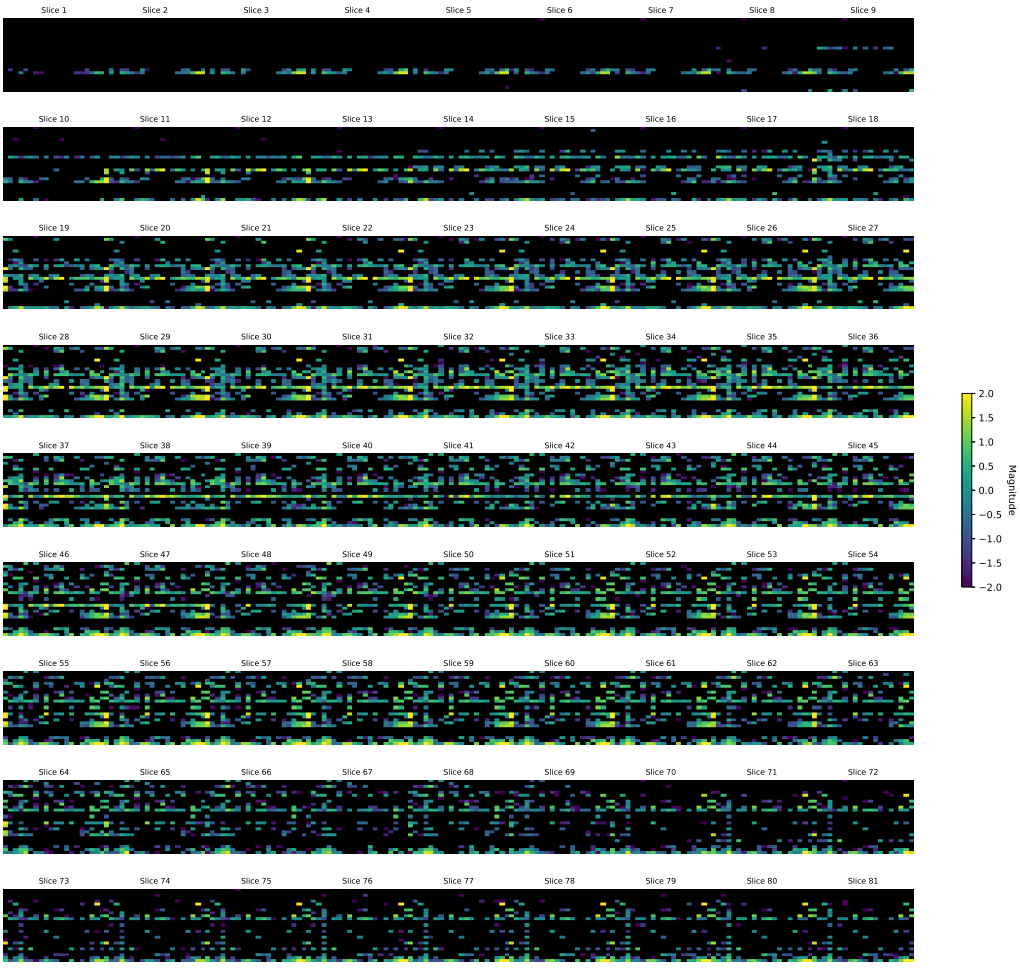


Figure A1: Visualization of the (25,20) slices of the observed $Y(0)$ tensor, with color denoting the log normalized magnitude of the entries. Each subplot represents one of the 81 slices.

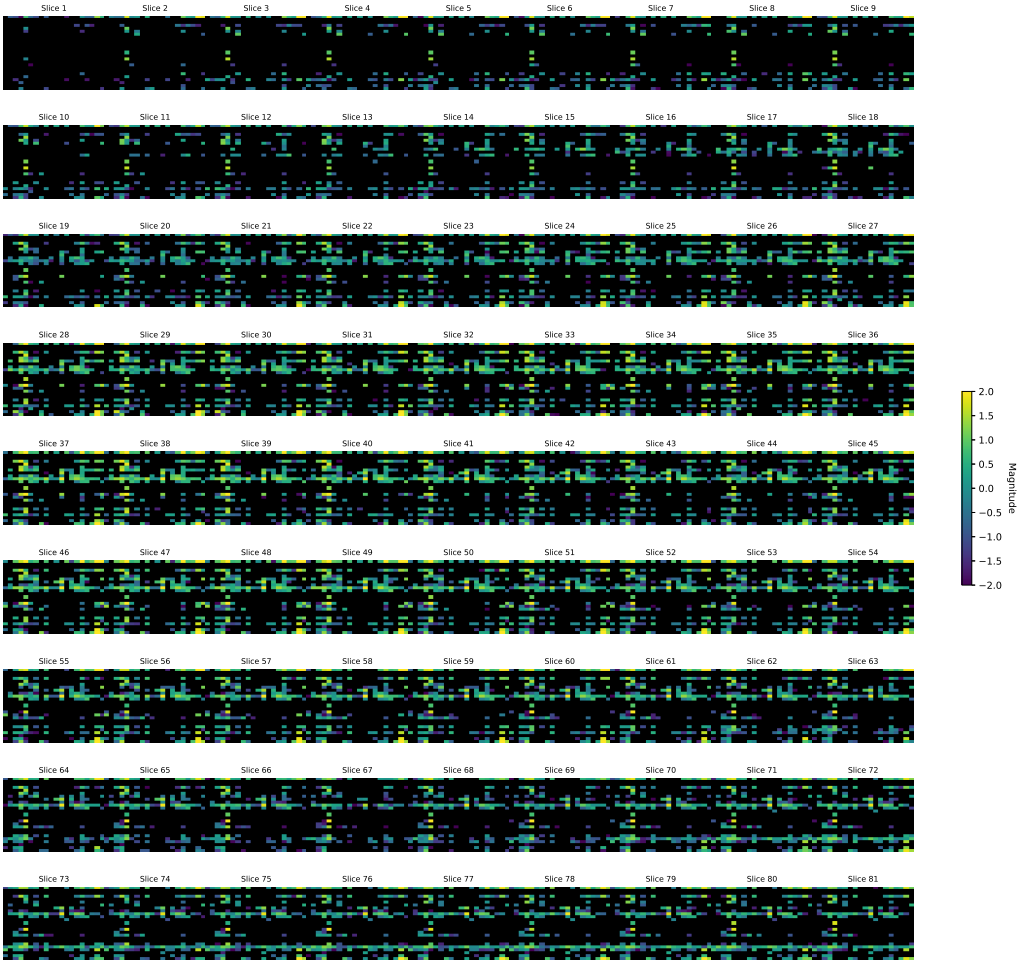


Figure A2: Visualization of the (25,20) slices of the observed $Y(1)$ tensor, with color denoting the log normalized magnitude of the entries. Each subplot represents one of the 81 slices.

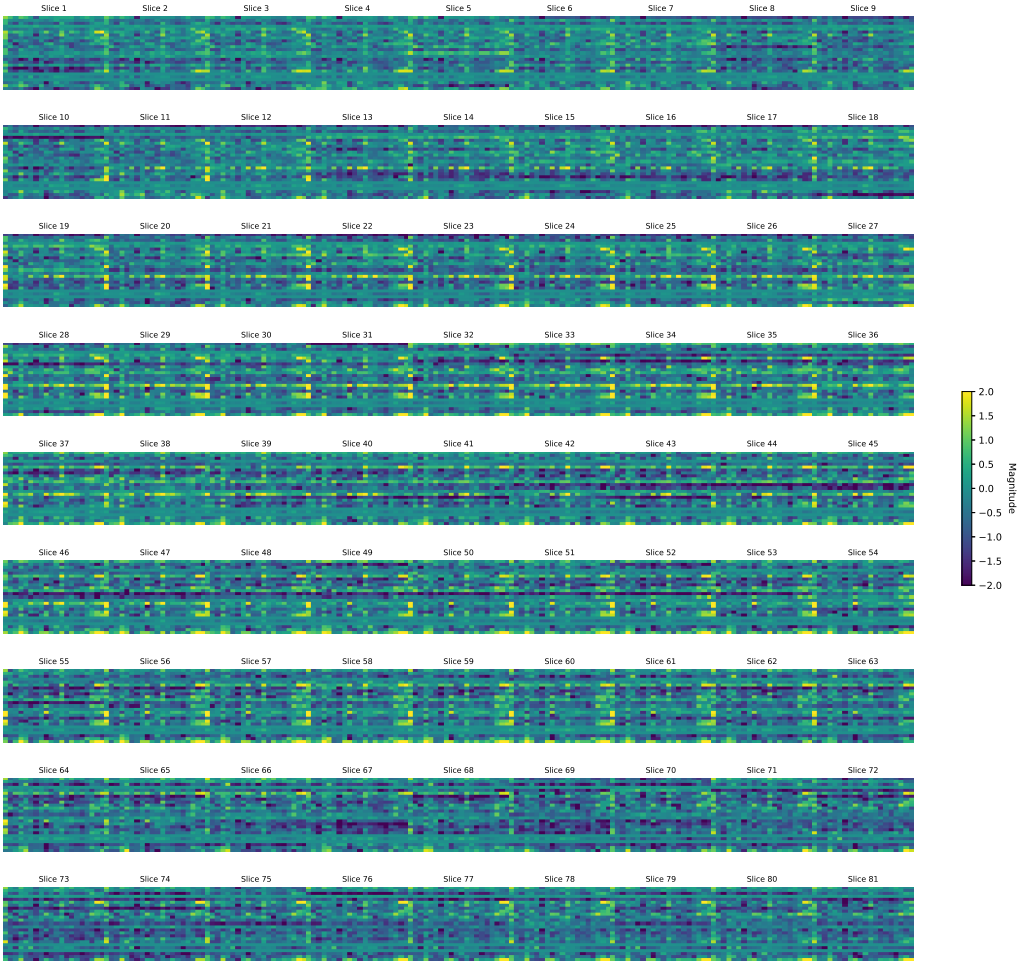


Figure A3: Visualization of the (25,20) slices of the estimated $Y(0)$ tensor, with color denoting the log normalized magnitude of the entries. Each subplot represents one of the 81 slices.

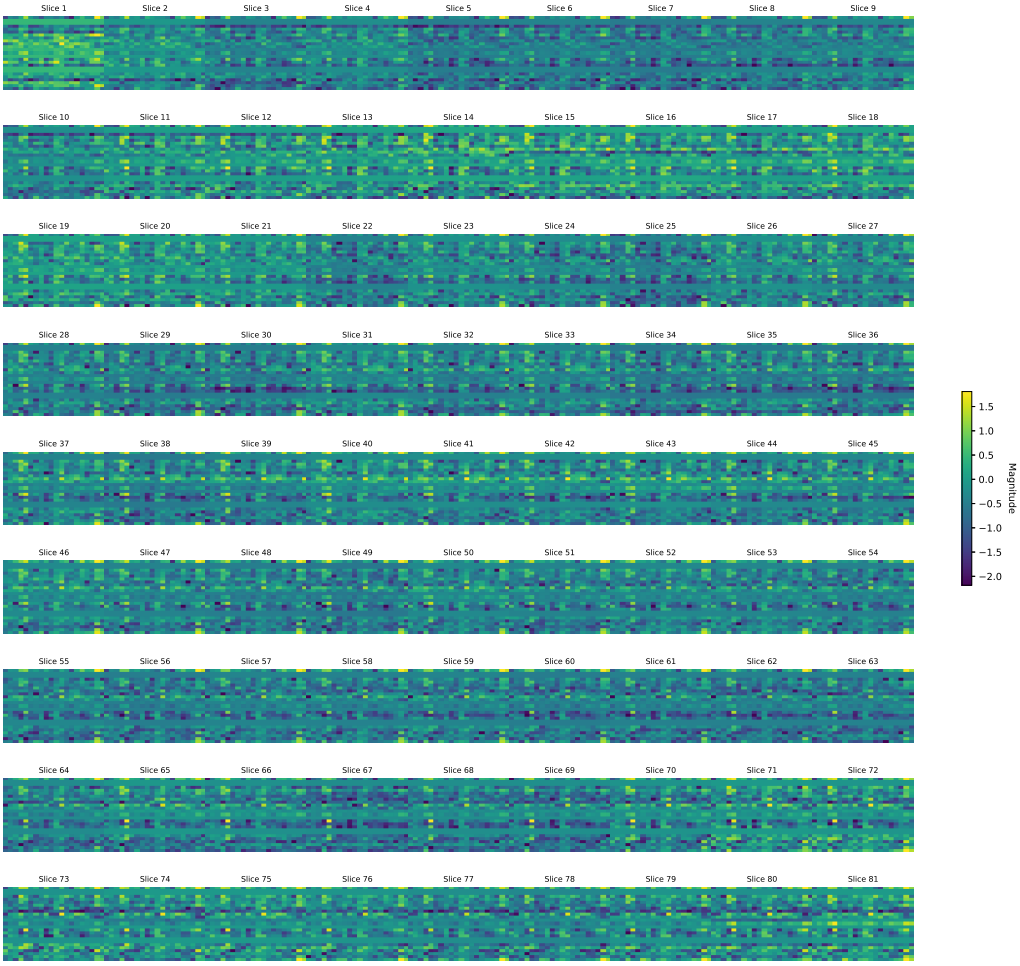


Figure A4: Visualization of the (25,20) slices of the estimated $Y(1)$ tensor, with color denoting the log normalized magnitude of the entries. Each subplot represents one of the 81 slices.

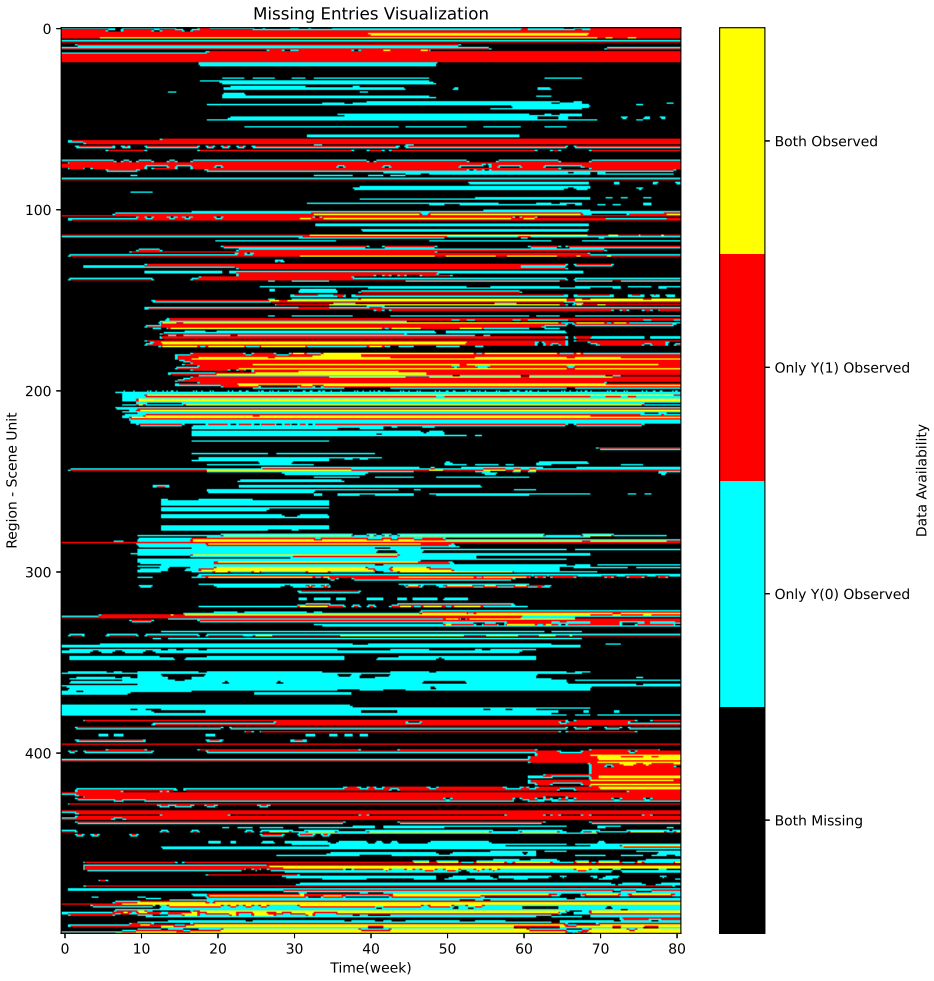


Figure A5: Visualization of Missing Entries. The color-coded matrix represents data availability, with different colors indicating the presence or absence of data across region-scene and time.

A.2 Definitions and Preliminaries

A.2.1 Tensor-tensor Product

In order to discuss multiplication between two tensors and to understand the basics of the algorithms we consider here, we first must introduce the concept of converting $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ into a block circulant matrix. If $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ with $\ell \times m$ frontal slices denoted $A^{(i)}$, then

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(n)} & A^{(n-1)} & \dots & A^{(2)} \\ A^{(2)} & A^{(1)} & A^{(n)} & \dots & A^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A^{(n)} & A^{(n-1)} & \ddots & A^{(2)} & A^{(1)} \end{bmatrix}$$

is a block circulant matrix of size $\ell n \times mn$. We anchor the unfold command to the frontal slices of the tensor. That is, unfold $(\mathcal{A})$ takes an $\ell \times m \times n$ tensor and returns a block $\ell n \times m$ matrix, whereas the fold command undoes the operation:

$$\text{unfold}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{bmatrix}, \quad \text{fold } (\text{unfold}(\mathcal{A})) = \mathcal{A}.$$

Now with the definitions:, Let $\mathcal{A}$ be $\ell \times p \times n$ and $\mathcal{B}$ be $p \times m \times n$. Then the t-product $\mathcal{A} * \mathcal{B}$ is the $\ell \times m \times n$ tensor

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})).$$

A.2.2 t-SVD

The tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be factorized as $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, i.e., each of the frontal slices of $\mathcal{S}$ is a diagonal matrix. The diagonal entries in $\mathcal{S}(:,:,1)$ are called the singular values of $\mathcal{A}$.

A.3 Tensor Completion

A.3.1 Low-Tubal-Rank Tensor Completion

Tensor completion is the task of recovering a tensor $\mathcal{Y} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ from a subset of its entries by leveraging the low-rank property of the tensor. The low-tubal-rank tensor completion problem can be formulated as:

$$\min_{\mathcal{Z} \in \mathbb{R}^{N_1 \times N_2 \times N_3}} \text{rank}_t(\mathcal{Z}), \text{ s.t. } \mathcal{W} \circ (\mathcal{Z} - \mathcal{Y}) = \mathbf{0},$$

where $\mathcal{W}$ is the indicator tensor of the observed entries, and $\circ$ denotes the Hadamard product. This problem is known to be NP-hard.

To address this issue, tensor factorization is used. The recovered tensor $\mathcal{Z}$ can be factorized into the t-product of two smaller tensors $\mathcal{U} \in \mathbb{R}^{N_1 \times r \times N_3}$ and $\mathcal{V} \in \mathbb{R}^{r \times N_2 \times N_3}$, where r is the tubal rank of $\mathcal{Y}$. The tensor factorization solves the completion problem by minimizing:

$$\min_{\mathcal{U},\mathcal{V}} J(\mathcal{U}, \mathcal{V}) := \|\mathcal{W} \circ (\mathcal{U} * \mathcal{V} - \mathcal{M})\|_F^2.$$

Low-tubal-rank tensor completion offers several advantages over other tensor rank models, such as directly imposing a low-rank constraint on the tensor and providing an optimal approximation with truncated decomposition.

A.3.2 Maximum Correntropy Criterion (MCC)

Correntropy is a local and nonlinear similarity measure between two random variables within a "window" in the joint space determined by the kernel width. Given two random variables U and V, the correntropy is defined as

$$V(U, V) = \mathbb{E}\left[\kappa_\sigma(U, V)\right] = \int \kappa_\sigma(u, v) dF_{UV}(u, v),$$

where κ_σ is a shift-invariant Mercer kernel with kernel width σ, $F_{UV}(u, v)$ denotes the joint probability distribution function of U and V, and $\mathbb{E}[\cdot]$ is the expectation operator. Given a

finite number of samples $\{u_i, v_i\}_{i=1}^N$, and using the Gaussian kernel, $G_\sigma(u) = \exp\left(-\frac{u^2}{2\sigma^2}\right)$, as the kernel function, the correntropy can be approximated by

$$\hat{V}(U,V) = \frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{e_i^2}{2\sigma^2}\right),$$

where $e_i = u_i - v_i$. Compared with the ℓ_2-norm based second-order statistic of the error, the correntropy involves all the even moments of the difference between U and V and is insensitive to outliers. Replacing the second-order measure with the correntropy measure leads to the maximum correntropy criterion (MCC). The MCC solution is obtained by maximizing the following utility function

$$J_{MCC} = \mathbb{E}\left[G_\sigma(e(i))\right].$$

Moreover, in practice, the MCC can also be formulated as minimizing the following correntropy-induced loss (C-loss) function

$$J_{C-loss} = \frac{1}{Y}\sum_{i=1}^{Y} \sigma^2\left(1 - G_\sigma(e(i))\right).$$

A.3.3 Optimization via Half-Quadratic Minimization

The correntropy-based objective function $J_{G_\sigma}(\mathcal{U}, \mathcal{V})$ can be expressed as:

Here is the modified text with the requested substitutions:

$$J_{G_\sigma}(\mathcal{U}, \mathcal{V}) = \min_{\mathcal{P}} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} \left(\mathcal{P}_{ijk}\mathcal{W}_{ijk}\left(\mathcal{Y}_{ijk} - (\mathcal{U} * \mathcal{V})_{ijk}\right)^2 + \mathcal{W}_{ijk}\phi\left(\mathcal{P}_{ijk}\right)\right).$$

By defining the augmented cost function:

$$J_{HQ}(\mathcal{U}, \mathcal{V}, \mathcal{P}) = \|\sqrt{\mathcal{P}} \circ \mathcal{W} \circ (\mathcal{Y} - \mathcal{U} * \mathcal{V})\|_F^2 + \psi_{\mathbf{\Omega}}(\mathcal{P}),$$

the correntropy-based optimization problem is formulated as a half-quadratic based optimization:

$$\min_{\mathcal{U},\mathcal{V}} J_{G_\sigma}(\mathcal{U}, \mathcal{V}) = \min_{\mathcal{U},\mathcal{V},\mathcal{P}} J_{HQ}(\mathcal{U}, \mathcal{V}, \mathcal{P}).$$

An alternating minimization procedure is proposed to solve the optimization problem by optimizing $\mathcal{U}$ and $\mathcal{V}$. Given a fixed $\mathcal{P}$, it becomes a weighted tensor completion problem:

$$\min_{\mathcal{U},\mathcal{V}} \|\sqrt{\mathcal{P}} \circ \mathcal{W} \circ (\mathcal{Y} - \mathcal{U} * \mathcal{V})\|_F^2.$$

The weighting tensor $\mathcal{P}$ assigns different weights to each observed entry based on error residuals, reducing the impact of large outliers.

A.3.4 Alternating Steepest Descent-based Algorithm

The alternating steepest descent (ASD) method is introduced for tensor completion. By optimizing $\mathcal{P}$ first and then gradually updating $\mathcal{U}$ and $\mathcal{V}$ using gradient descent, the optimization problem can be rewritten as:

$$\frac{1}{2}\min_{\mathcal{U},\mathcal{V}} \|\sqrt{\tilde{\boldsymbol{P}}} \circ \tilde{\boldsymbol{W}} \circ (\tilde{\boldsymbol{Y}} - \text{bcirc}(\boldsymbol{\mathcal{U}})\tilde{\boldsymbol{V}})\|_F^2.$$

Based on block-circulant diagonalization, it becomes:

$$\min J(\boldsymbol{M}, \hat{\boldsymbol{V}}) := \frac{1}{2}\|\sqrt{\tilde{\boldsymbol{P}}} \circ \tilde{\boldsymbol{W}} \circ (\tilde{\boldsymbol{Y}} - \boldsymbol{M}\hat{\boldsymbol{V}})\|_F^2.$$

The partial derivatives and update rules for $\boldsymbol{M}$ and $\hat{\boldsymbol{V}}$ are provided, along with a scaling method for the gradient descent direction of $\hat{\boldsymbol{V}}$ to improve convergence rate.

Algorithm 1: HQ-TCASD for robust tensor completion

Input: $\mathcal{W}, \mathcal{W} \circ \mathcal{Y}, r$ and λ

Initial: matrices $\boldsymbol{M}^0$ and $\hat{\boldsymbol{V}}^0$, $t = 0$

repeat

- compute σ^{t+1} and $\mathcal{P}^{t+1}$ using $\mathcal{P}_{ijk}^t = G_{\sigma^t}\left(\mathcal{Y}_{ijk} - (\boldsymbol{M}^t * \hat{\boldsymbol{V}}^t)_{ijk}\right)$
- compute $\boldsymbol{M}^{t+1}$ using $\boldsymbol{M}^{t+1} = \boldsymbol{M}^t - \mu_{\boldsymbol{M}}^{\prime t}\boldsymbol{g}_{\boldsymbol{M}}^{\prime t}$.
- compute $\hat{\boldsymbol{V}}^{t+1}$ using $\hat{\boldsymbol{V}}^{t+1} = \hat{\boldsymbol{V}}^t - (1-\lambda)\mu_{\hat{\boldsymbol{V}}}^t \boldsymbol{g}_{\hat{\boldsymbol{V}}}^t - \lambda\mu_{\hat{\boldsymbol{V}}}^{\prime t}\boldsymbol{g}_{\hat{\boldsymbol{V}}}^{\prime t}$.
- $t = t + 1$

until *stopping criterion is satisfied*

Output: $\boldsymbol{M}^t * \hat{\boldsymbol{V}}^t$

The proposed Half-Quadratic based Tensor Completion by Alternating Steepest Descent (HQ-TCASD) algorithm adaptively selects the kernel width σ to improve convergence rate and performance. The algorithm is summarized in Algorithm 1, with the matrices $\boldsymbol{M}$ and $\hat{\boldsymbol{V}}$ updated alternately until convergence.

Appendix B

ONLINE APPENDIX TO FRENEMIES IN THE RETAIL MARKET: A PARTNERSHIP BETWEEN A PHYSICAL RETAILER AND AN E-TAILER FOR CONSUMER RETURNS

The online appendix is available on my personal website: https://mingruirayzhang.github.io/research/.

VITA

Mingrui Ray Zhang is a researcher specializing in Information Systems. Her research interests revolve around Consumer Behavior, Causal Inference, and Fin-Tech. He employs a range of methodologies, such as Econometrics, Game-theoretical Models, and Machine Learning, to promote and explore consumer behaviors in emerging markets. He received his Ph.D. degree from the Michael G. Foster School of Business at the University of Washington (UW). Prior to joining UW, he completed his Master's degree at Columbia University and his bachelor's degree from University of Illinois at Urbana-Champaign. As of September 2024, he will be joining the Department of Business Information and Analytics at the Daniels College of Business, University of Denver, as an Assistant Professor. He welcomes your comments to `mingrui.zhang@du.edu`.