UNIVERSITY OF CALIFORNIA,
IRVINE

Choices in Economics: From Movie Videos, Leisure to Migration Decisions

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Hao-Che Hsu

Dissertation Committee:
Professor Matthew Harding, Chair
Associate Professor Jiawei Chen
Professor Yingying Dong
Associate Professor Ying-Ying Lee

2023

DEDICATION

To my parents, for their unwavering love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to take this opportunity to express my profound gratitude and appreciation to those who have played pivotal roles in the successful completion of my dissertation. This represents a significant milestone in my academic journey, and I am deeply thankful to everyone who has supported me throughout this process.

First and foremost, I extend my gratitude to my advisor, Matthew Harding, for his exceptional guidance and support. His mentorship has been invaluable to me throughout my research journey. Our shared passion for deep learning enabled us to engage in stimulating discussions. My time at UC-Irvine was enriched starting from when he welcomed me into his Deep Data Lab. He provided me with abundant proprietary data, computational resources, and funding. The numerous opportunities he gave me to collaborate on various projects, ranging from inter-department and inter-university partnerships to collaborations with government institutes and industry, have shaped my ideas and granted me precious experiences.

I would also like to thank the other members of my committee: Jiawei Chen, Yingying Dong, and Ying-Ying Lee. Their insightful comments and suggestions have consistently enhanced my research. Their encouragement during our regular check-ins was vital to my progress. The direction they provided during our discussions and consultations was always beneficial. I am further indebted to them for their support when I initiated reading groups and brown bag organizations.

My heartfelt thanks go to my colleagues at UC-Irvine. Their friendship, kindness, and encouragement have been integral throughout my academic journey. Their stimulating discussions, research chats, and invaluable feedback have been a source of inspiration and motivation.

Lastly, my deepest gratitude goes to my parents for their consistent love and support over the years. To my wife, who has been by my side from the beginning to the end of this journey. Her faith in me has been a cornerstone of my perseverance.

VITA

Hao-Che Hsu

EDUCATION

Doctor of Philosophy in Economics — **2023**
University of California, Irvine — *Irvine, CA*

Masters of Arts in Economics — **2019**
University of California, Irvine — *Irvine, CA*

Masters of Science in Economics — **2017**
University of Wisconsin, Madison — *Madison, WI*

Bachelor of Arts in Economics — **2014**
National Chung Cheng University — *Chiayi, Taiwan*

RESEARCH EXPERIENCE

Graduate Student Researcher — **2021, 2023**
University of California, Irvine — *Irvine, CA*

TEACHING EXPERIENCE

Teaching Assistant — **2018-2020, 2023**
University of California, Irvine — *Irvine, CA*

Teaching Assistant — **2013**
National Chung Cheng University — *Chiayi, Taiwan*

FIELDS OF STUDY

Econometrics, Machine Learning, Industrial Organization

ABSTRACT OF THE DISSERTATION

Choices in Economics: From Movie Videos, Leisure to Migration Decisions

By

Hao-Che Hsu

Doctor of Philosophy in Economics

University of California, Irvine, 2023

Professor Matthew Harding, Chair

The chapters of this dissertation analyze the choice decisions of various groups of individuals. Chapter 1 examines the choices of consumers regarding movie videos. We utilize weekly video sales data, encompassing digital videos from online platforms and physical discs from retail stores, to investigate the director's backward spillover effect, factors influencing consumer preferences, and the market structure. Our findings confirm the presence of backward spillover effects on video sales and provide an estimation of the movie video market's cost structure. Additionally, we analyze the impact of structural changes in the market using a merger simulation. Chapter 2 investigates household choices related to leisure activities. We consider both the time spent and the associated expenditure measurements to offer a comprehensive analysis of time allocations and costs for leisure pursuits. By combining a time-use survey with mobile app scanner data, we categorize activities and products into 14 leisure categories. Then we uncover the geography of leisure, constructing leisure price indexes and estimating the time spent and expenditure variations in income. Our study further probes choice sensitivity, employing a causal inference framework to estimate heterogeneous elasticity and leisure substitution patterns. Chapter 3 focuses on the choices made by refugees regarding their destination cities during migration. This analysis focuses on those escaping the Russia-Ukraine war that erupted in February 2022. We use data from Hornet, a Queer Social Network, to classify refugee users into three groups: Ukrainians, Russians, and

Foreigners. An interactive map has been developed to visualize the daily aggregated movements of the refugees. Subsequently, we investigate the factors that influence refugee preferences when choosing destination cities. We also analyze the elements that impact both the number of refugees in these cities and the migration movements between them due to the conflict. Lastly, we estimate the average migration patterns of the refugees.

All analyses in this dissertation were conducted at a 5% significance level.

Chapter 1

Choice and Backward Spillover Effects in the Movie Video Market

There are significant discrepancies in consumer preferences between choosing theater movies and movie videos. By using weekly movie video sales data instead of box office revenues and a demand model, we examine the factors influencing consumer purchasing decisions and uncover the video cost, profit margins, and market structure. Specifically, we present evidence supporting the existence of the director's backward spillover effect. A counterfactual merger simulation illustrates how the market would respond to structural adjustments.

1.1 Introduction

Accurately capturing consumer choice behaviors has a strong implication for marketing strategies. Movie-watching plays an important role in leisure activities. Exploring the choices and the preferred types of movies of a general audience benefits the video platforms in carrying out a much more efficient promotion program. To gain insights into the movie market video, it is crucial for the sellers to investigate and identify the market demand and learn the factors that affect consumers'

choices in movie videos. Hence, demand estimation has become an important and straightforward approach to probing customer behavior.

Movies offer a dynamic method of storytelling. Enhanced by computer graphics and special effects, they deliver a vibrant and immersive viewing experience to the audience. With easy accessibility and virtually no entry barriers, there's a wealth of creative content available, presenting audiences with countless films to choose from. High-budget movies with sophisticated production designs often premiere in theaters, while others are available on digital platforms or physical video discs. Additionally, there are informal productions, like fan films or home videos captured with phones or mirrorless cameras, shared on social media. It's this diversity and creativity that underpins the multi-billion-dollar movie industry.

Much of the analysis surrounding the movie industry is rooted in box office performances. In 2019, the global box office revenue reached $42.5 billion. North America's box office revenue accounted for over a quarter of this amount at $11.32 billion[1]. The discourse often revolves around movies that shatter box office records or sustain high revenues over time. The performance of box office revenues directly reflects the choice behavior of moviegoers. But do these metrics truly reflect audience preferences? A glance at box office rankings reveals that most top-grossing movies belong to the *Action* or *Adventure* genres. However, is this consistent with consumers' movie video choices when they browse movies on YouTube Movies, Amazon, or shop at Target?

Off-screen video sales, which include both digital platforms and physical disc sales, provide another perspective. For instance, Amazon's Prime Video generated $1.7 billion in revenue in 2018[2]. On the physical media front, DVD sales peaked at $16.3 billion in 2005, capturing 64% of the U.S. home video market. Meanwhile, Blu-ray disc sales, having launched in 2006, reached $2.37 billion by 2013. By 2018, however, DVD sales had declined to $2.2 billion, and Blu-ray sales to $1.8 billion[3]. Given these considerations, while 2018 data from Statista indicates an annual box office

[1] Source: billboard.com, *"2019 Global Box Office Revenue Hit Record $42.5B"* and statista.com.
[2] Source: variety.com, *"Amazon's Prime Video Channels Biz to Generate $1.7 Billion in 2018."*
[3] Source: cnbc.com, *"The death of the DVD: Why sales dropped more than 86% in 13 years."*

gross of $11.89 billion in the U.S. and Canada, drawing conclusions based solely on these figures without considering various online platforms and video discs might be misleading. The preferences for off-screen movies can differ significantly from on-screen selections, suggesting that box office metrics alone might not provide a comprehensive view of consumers' movie preferences.

We will investigate consumers' preferences for off-screen movie selections in both online platforms and retail stores. Delving deeper into consumer choice, we will explore the *backward spillover effects* that might affect these decisions. Treating movies as differentiated products, this study will examine the characteristics of films that sway consumer choices and evaluate the existence of choice-stimulated backward spillover effects on sales attributable to directors in the film industry.

In the fast-paced world of film production, not every movie reaches the broader public due to information gaps. Even when films premiere in theaters, they often remain undiscovered by the non-theatergoers. When a movie hits the big screen, it benefits from advertisements, press coverage, and ongoing discussions, which bring it to the forefront of public attention. Following a film's theatrical release, its in-theater experience spreads across the public. Stellar performances by actors or actresses are noticed, prompting fans to seek out their earlier works. For example, after watching "The Revenant," for which Leonardo DiCaprio won an Oscar for Best Actor, viewers might be inclined to explore other films like "The Wolf of Wall Street" or "Catch Me If You Can," both starring DiCaprio in lead roles. But does this same phenomenon hold for directors? We will delve into the potential backward spillover effect associated with directors.

The impact of backward spillovers stems from movie discovery. In the promotional phase leading up to a film's theatrical release, trailers often spotlight the director as part of the marketing strategy. For instance, films like "Avatar" and "Interstellar" prominently feature their directors in their trailers. After a film's release, if audiences appreciate it, they may attribute its success to strong direction. This admiration can lead them to seek out earlier films helmed by the same director.

We utilize weekly movie video sales data spanning 12 years to investigate the backward spillover effect. As illustrated in Figure 1.1, out of 1,492 directors in the dataset, 31.83% have directed

multiple movies. Yet, only 14.45% of these movies garnered over 100 million in box office revenue, as shown in Figure 1.2. Given that the likelihood of movie discovery is positively correlated with its popularity and considering that only a limited number of films achieve blockbuster status, identifying the backward spillover effect necessitates a careful approach rather than a straightforward examination of sales data fluctuations.

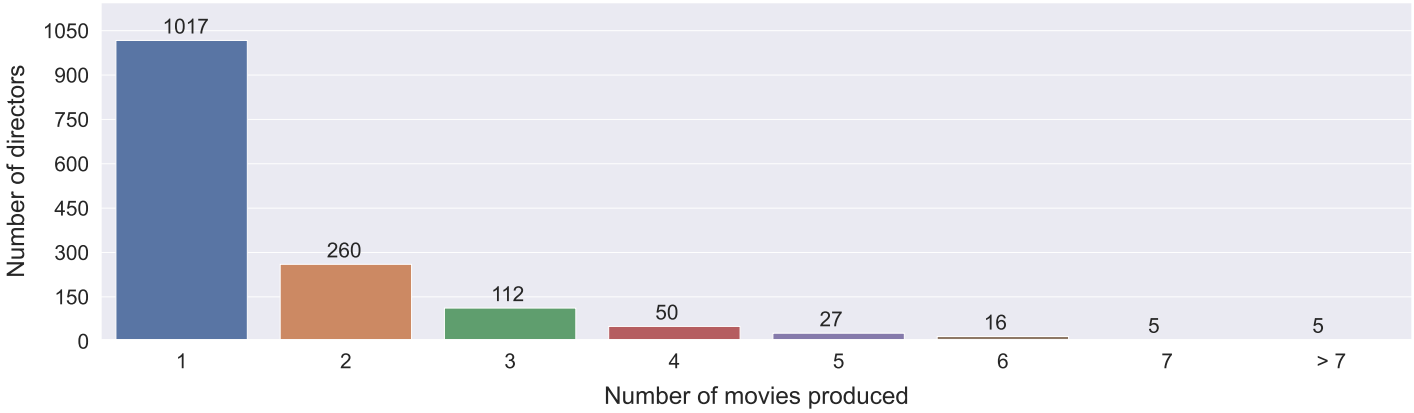

Figure 1.1: Number of Films Directed by a Director

Note: Out of the 1,492 directors listed in the data (from 2006 to 2017), approximately 30% directed more than one film. Of those with multiple credits, nearly half directed more than three films.

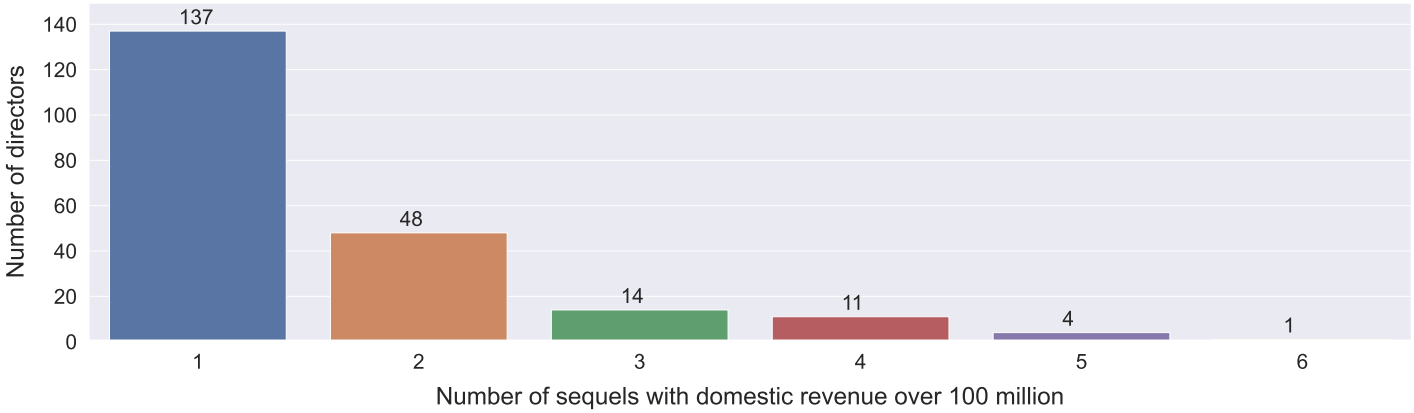

Figure 1.2: Number of Blockbusters Directed

Note: In Figure 1.1, only 13.5% of directors with just one film to their credit managed to produce blockbuster movies that grossed over $100 million at the domestic box office in the United States. Of the 2,388 films examined, 345 are classified as blockbusters.

Sales typically fluctuate, but they tend to trend downwards over time. However, when affected by a backward spillover shock, there is a temporary surge in sales, resulting in a pronounced spike

or jump on the release date of the subsequent product. Sometimes, this surge might manifest with a delay. To accommodate these situations, We've introduced a *backward spillover impact window*, which we will detail further in a later section to fully capture the nature of this effect.

For a given reference or base movie, films released afterward by the same director are called *sequences*. It's important to differentiate between a *sequence* and a *sequel*: while a sequel continues the story of the original film, a sequence merely shares the same director without necessarily continuing the storyline. When we examine the backward spillovers from sequels as depicted in Figure 1.3, they naturally demonstrate a significant impact on the sales of previous sequels, largely due to the continuity of the storyline. Consequently, we account for the sequel effect when investigating the director's backward spillovers on prior sequences.

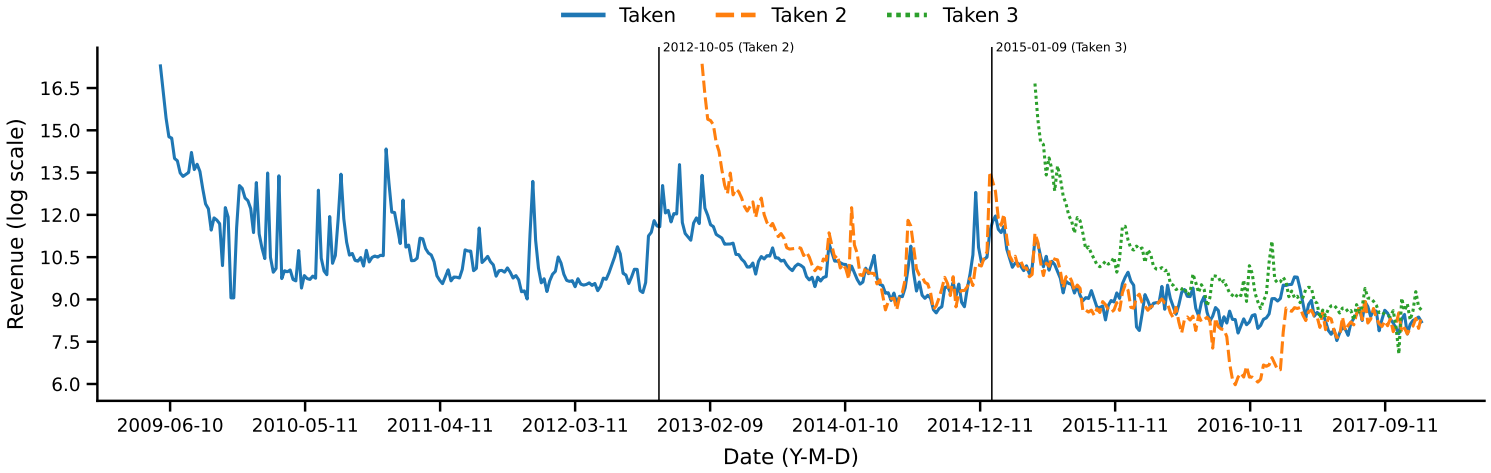

Figure 1.3: The Sequel Backward Spillovers in the "Taken" Series

Note: The time series includes video sales, both online and physical. "Taken" is considered the base film of the series. Indicating by the black vertical line, "Taken 2" was released in theaters on October 5, 2012, and "Taken 3" debuted on January 9, 2015. Typically, movie videos are released several months after their theatrical debut. There are significant backward spillovers from the release of the *sequels* to the base film.

In the following sections, we will focus on the *sequences*. We share three cases to provide a clearer understanding of the director's backward spillover effect on movie revenue. Although the unfiltered graphs appear noisy, the potential backward spillover effect remains discernible. Figure 1.4 showcases the films directed by Christopher Nolan. Notably, in his acclaimed documentary "Oppenheimer," released in IMAX theaters in July 2023, the director's name is prominently featured

on a splash screen in the official trailer weeks before the theatrical release. This could lead to a promotional effect for Nolan's previous non-sequel films if "Oppenheimer" proves to be a success.

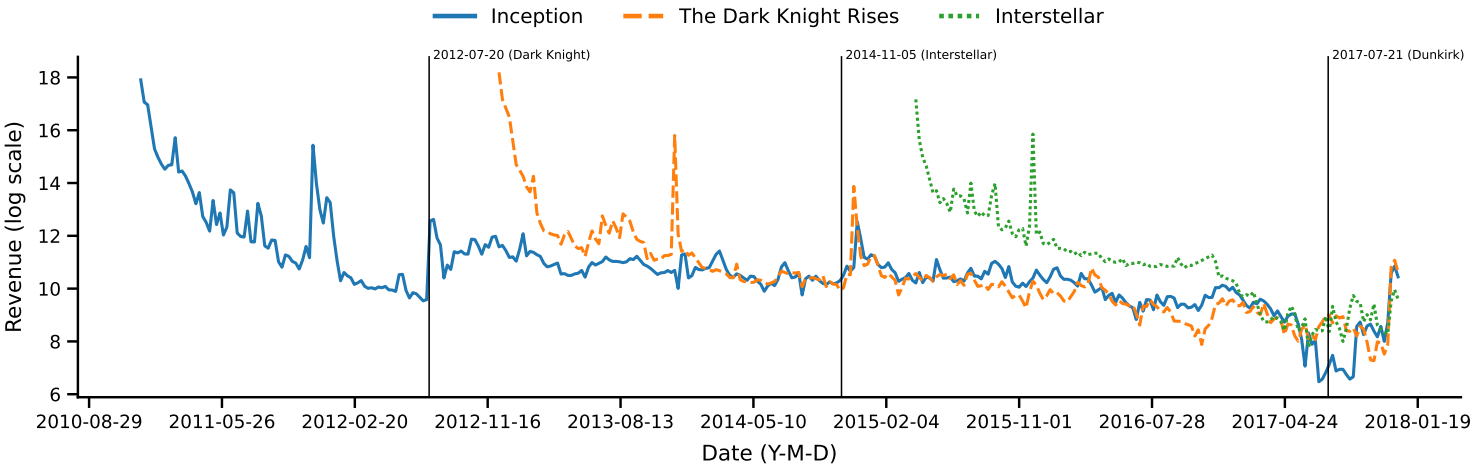

Figure 1.4: Films Directed by Christopher Nolan

Note: In North America, "The Dark Knight Rises" was released on July 20, 2012; "Interstellar" on November 5, 2014; and "Dunkirk" on July 21, 2017. The video of "Dunkirk" was released on December 24, 2017, but its sales are not depicted in the figure. Sales spikes due to the spillover effects within the 12-week period *backward spillover impact window* are significant for all sequences on their non-sequel base video.

In the figure above, the impact of backward spillovers from the sequences to the chosen base/reference film can be observed. The graph displays three log-scale revenue time series representing the weekly video sales of these films. Selecting "Inception" as the base film, the three black vertical lines represent the theatrical release dates of its three subsequent *sequences*: "The Dark Knight Rises," "Interstellar," and "Dunkirk." Sometimes, a sequence may also be a sequel to the previous video, but this effect has been accounted for in this study. It's important to note that there is typically a delay of a couple of months before the movie's video (represented in the time series) is released after its theater debut (black line). A noticeable jump in the sales of "Inception" (represented by the blue series) can be observed either directly on or shortly after each sequence's release date. Additionally, the spillover effect diminishes rapidly over the years.

In the same graph, when we select "The Dark Knight Rises" (represented by the orange series) as the reference film, we can see that its subsequent film, "Interstellar" (green series) which premiered

in theaters on November 5, 2014, led to a significant boost in sales for "The Dark Knight Rises" shortly after its debut.

In Figure 1.5, presenting films directed by David Fincher, "The Curious Case of Benjamin Button" is chosen as the reference film. However, the backward spillovers of "Gone Girl" on "The Social Network" and "The Girl with the Dragon Tattoo" are minimal. Nonetheless, there is a distinct backward spillover effect from "The Girl with the Dragon Tattoo" on the video sales of "The Social Network." Figure 1.6 shows the final spillover example from films directed by Martin Scorsese.

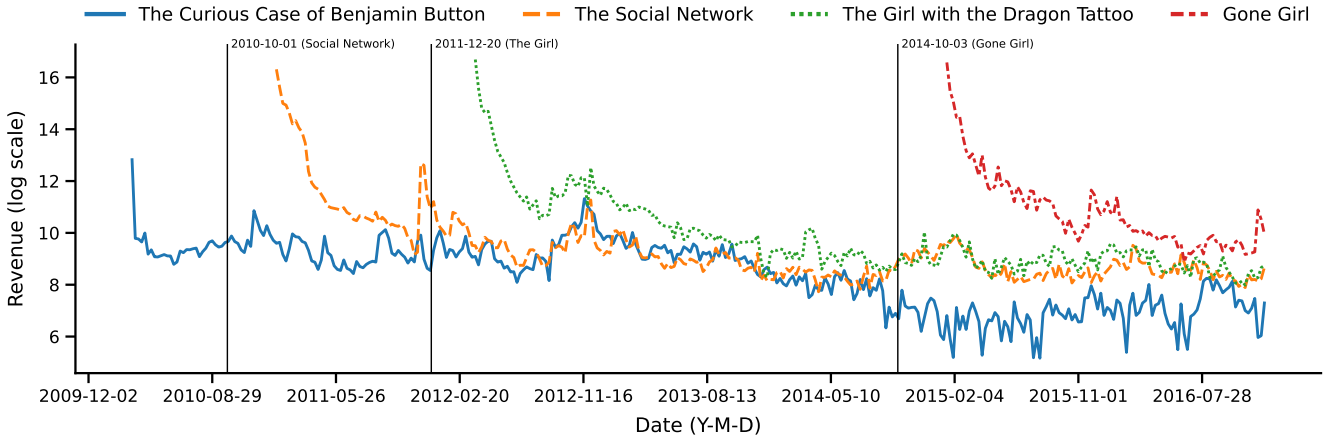

Figure 1.5: Films Directed by David Fincher

Note: Minor backward spillovers on the sales of "The Curious Case of Benjamin Button" were observed upon the release of its second sequence, "The Girl with the Dragon Tattoo," on December 20, 2011, and the third sequence, "Gone Girl," on October 3, 2014. The slight increase from the second sequence came after the Christmas holiday surge.

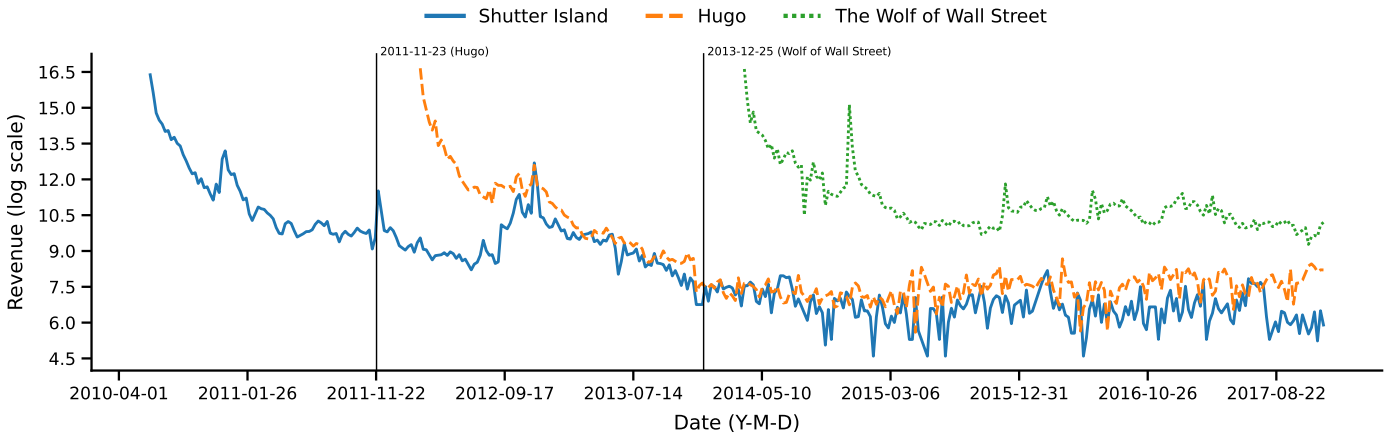

Figure 1.6: Films Directed by Martin Scorsese

Note: Only "Hugo," the first sequence to "Shutter Island," exhibits a spillover effect.

In contrast, the magnitude of backward spillovers is less evident for other directors. Figure 1.7 showcases films directed by Steven Spielberg. However, the backward spillovers from the subsequent films are minimal.

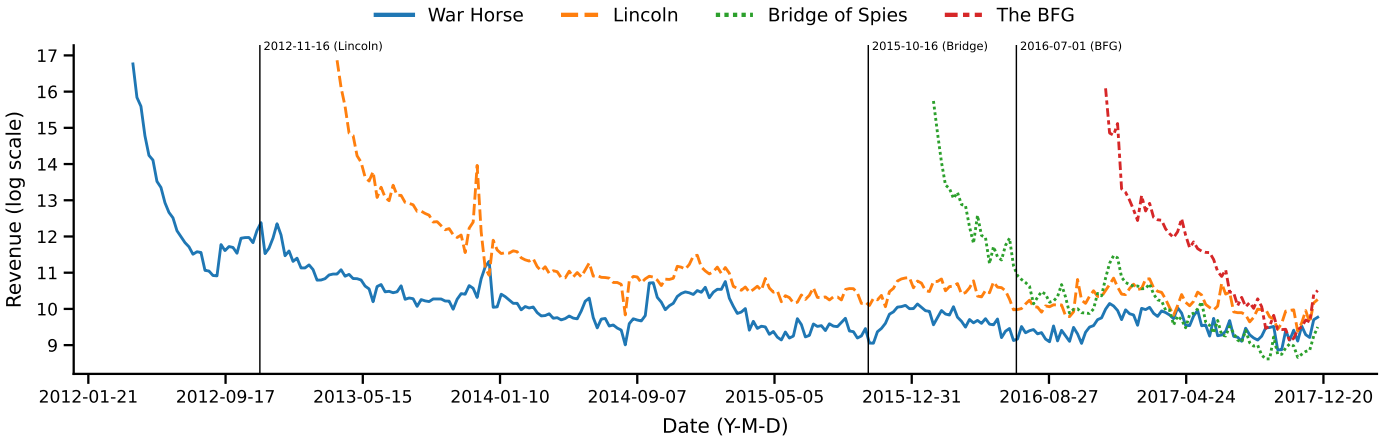

Figure 1.7: Films Directed by Steven Spielberg

Note: "War Horse" was released on December 25, 2011; "Lincoln" on November 16, 2012; "Bridge of Spies" on October 16, 2015; and "The BFG" on July 1, 2016. Only the first sequel saw a slight boost in the sales of "War Horse." The subsequent releases did not exhibit a clear spillover effect.

Potential concerns might arise from the belief that systematic factors, such as holiday sales, could obscure the true impact of the backward spillover effect such as in films directed by Quentin Tarantino in Figure 1.8 below, where subsequent films are released on Christmas.

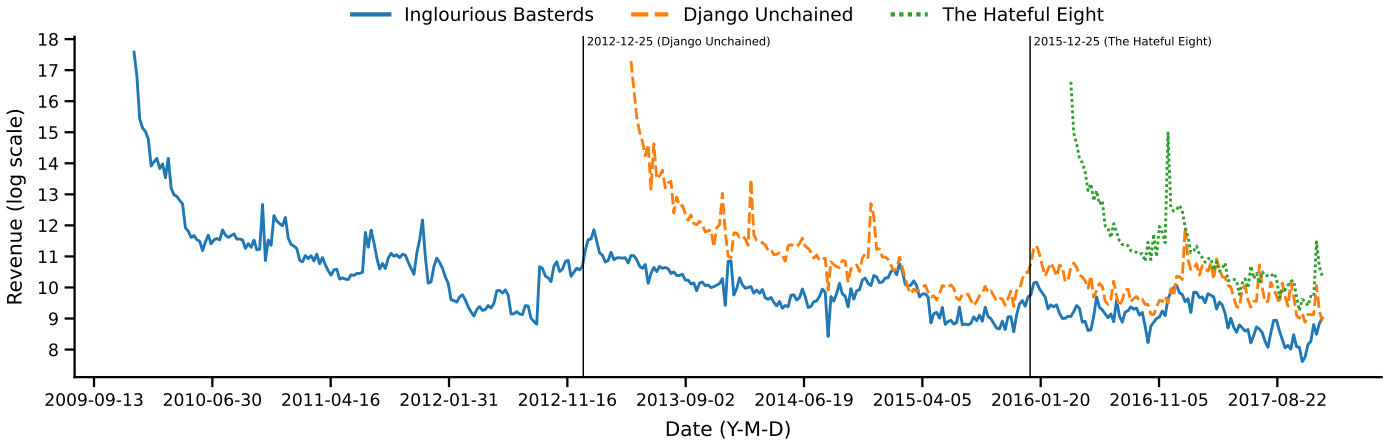

Figure 1.8: Films Directed by Quentin Tarantino

Note: Taking "Inglourious Basterds" as the base film, its sequels "Django Unchained" and "The Hateful Eight" were released in theaters on December 25, 2012, and December 25, 2015, respectively.

In certain cases, as illustrated in Figure 1.8, there is a pronounced holiday surge in the sales of all movies. This situation can be addressed by assuming that holidays affect all movies homogeneously. Consequently, rather than analyzing the spillover effect using a treatment effect model, we incorporate the backward spillovers into a demand model. This approach converts video sales numbers into market shares, effectively neutralizing systematic influences.

This paper is closely related to two strands of literature: one concerning the film industry and the other focusing on backward spillover effects in varied contexts. Most importantly, this is the first study to investigate the director's backward spillover effect in the film industry. Many studies in this field primarily focus on box office revenues. Einav (2007) investigates the seasonality of box-office revenue and uses it to measure weekly movie demand with a nested logit model. Similarly, Ferreira, Petrin and Waldfogel (2016) examines the influence of China on the preference composition of global movie consumers and producers using box office revenue data. The literature on the backward spillover effect is limited. Ebghaei (2016) delves into the backward spillovers of foreign direct investment on exports. Meanwhile, Hendricks and Sorensen (2009) examines the backward spillover effect of a new album release on older albums using a treatment effect model.

The remainder of this chapter is structured as follows: Section 1.2 provides background information on the film industry. Section 1.3 details the database construction. Section 1.4 defines the market and outlines the creation of the backward spillover indicators. Section 1.5 introduces the demand model. The findings, along with counterfactual analysis, are discussed in Section 1.6. Concluding remarks can be found in Section 1.7.

1.2 Industry Background

Before the era of high-definition content, movies were primarily distributed via DVDs. Apart from cinemas and broadcast channels, DVD players were the predominant means for watching movies. However, after 2006, the Blu-ray format was introduced, aiming to replace DVDs. Not long after,

online movie platforms such as *Netflix*, *Disney*, *Amazon Prime Video*, *Redbox*, *Hulu*, *Apple TV*, *Vudu*, and *AT&T* began to emerge. While some of these platforms offer movies for digital rental and purchase, others have set up a video-on-demand system to which users can subscribe. Nonetheless, even with the growing trend of online streaming, there remains a demand for disc-based content.

While DVD sales have significantly declined, there are still those who purchase DVDs[4], prioritizing affordability over top-tier image and sound quality. Additionally, DVD sales provided a reliable gauge of consumers' movie preferences for a lengthy period before the rise of streaming services. On the other hand, many continue to buy Blu-ray discs from wholesale clubs like *Target*, *Walmart*, and *Best Buy*. Others even opt to purchase Blu-rays from platforms such as *Amazon* and *eBay*. The predominant driving force behind the demand for Blu-rays in recent years is the joy of physical ownership. Collectors take pride in owning a physical disc and value a cabinet filled with Blu-rays. Occasionally, consumers also have the option to download high-definition videos using the digital code that comes with the Blu-ray discs. Furthermore, Blu-rays with Ultra HD content deliver superior sound and visual quality compared to streaming a 4K video[5].

For streaming services, the most pressing concern is the availability constraints that users encounter. Many theatrically released films have a short life span on these platforms; movies come and go unless you own them. For instance, according to *JustWatch*, a platform that monitors video service availability, the film *Air Force One*[6] is only available on *Amazon Prime Video* and not on *Netflix* or other streaming services. Another classic, *Avatar*, a 2009 science fiction film directed by James Cameron, despite being the highest-grossing movie[7] of all time[8], streams only on *Disney Plus* and *Direct TV* (AT&T). Additionally, it's noteworthy that *Titanic* is available only for rent

[4]Blu-ray and DVD (home movie video) sales peaked at $16.3 billion in 2005, then fell to approximately $3.29 billion in 2019 and further decreased to $1.97 billion in 2021. (Source: history-computer.com, *"DVD vs Blu-ray: How Do They Compare?"*)

[5]The 4K resolution refers to a horizontal display resolution of approximately 4,000 pixels, specifically 3840 x 2160, and is considered Ultra High Definition (UHD).

[6]Air Force One, released in 1997, is an American political action-thriller directed by Wolfgang Petersen. The film stars Harrison Ford as the U.S. President who must combat hijackers aboard the presidential aircraft.

[7]Source: wikipedia.org, "*List of highest-grossing films*."

[8]*Avengers: Endgame*, a 2019 superhero film based on *Marvel Comics*, is currently the second highest-grossing film worldwide. It once surpassed *Avatar*, but was overtaken when *Avatar* was re-released in Chinese theaters in March 2021. The film is exclusively available for streaming on *Disney Plus*.

or purchase and is not offered for streaming on any platform. Considering these limitations and even though online streaming has become the primary mode of movie consumption today, when compared to DVDs, Blu-rays, and digital formats, the offerings on streaming services remain limited and fluctuate over time.

In the digital era, powered by high-speed Internet connections, digital videos offer notable advantages in portability over physical discs. However, they require an Internet connection or a platform-specific device[9] for offline viewing. Using *Apple TV* as an example, digital movie purchase prices range from $9.99 to $14.99, and rentals are priced between $4.99 and $5.99. Although renting a digital movie is cheaper than purchasing, rentals come with time constraints. An online rented movie typically expires in 30 days, and once started, it's automatically removed from the library within 48 hours. Additionally, not every film offers a rental option, making the purchase selection somewhat broader than the rental pool.

Regardless of the ownership type, there are limits to the number of devices authorized for viewing. On the *Apple TV* platform, a movie can be played on up to five computers or mobile devices. For *Amazon Prime Video*, a purchased film can be downloaded to a maximum of four devices, while a Prime video rental can only be downloaded and streamed on a single device during the rental period.

1.3 Data Augmentation

The main market-level data[10] integrates three sources. Sales data and film production information are sourced from *OpusData*, a service provided by *Nash Information Service, LLC*. The front-end of the database is "The Numbers,"[11] which offers box office tracking information. The company's data collection processes are detailed in the first node of Figure 1.9.

[9] For example, movies downloaded from *Apple TV* can only be stored on *Apple TV* or Macs, while movies from Amazon Prime Video require Amazon's Fire tablet or a mobile device for offline viewing.

[10] Before model estimation, the weekly sales data from OpusData is aggregated to an annual level.

[11] *The Numbers* is a website dedicated to movie industry data: https://www.the-numbers.com.

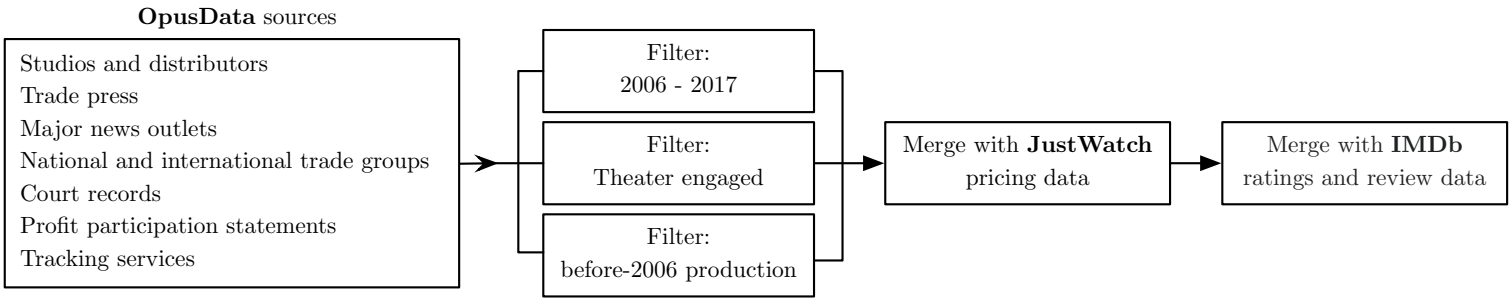

Figure 1.9: Data Processing Flow

Note: All movies produced before 2006 that have not debuted in theaters are removed from *OpusData*. Additional movie characteristics, pricing, and rental information are manually extracted from *JustWatch* and *IMDb*.

1.3.1 The OpusData

Drawing from multiple sources, *OpusData* offers weekly sales data and a plethora of film attributes. Information about the movie's high-frequency box office, coupled with lists of cast and crew as well as budget details, is sourced from both domestic and international theatrical distributors. Weekly sales data on physical discs (both DVD and Blu-ray formats) and digital copies (in standard, high, and ultra-high definitions), covering rental services and purchases from physical stores (point-of-sale) and online platforms such as *Amazon*, *Google Play*, *Apple TV*, *Vudu*, and *Xfinity*, are provided by retailers, rental outlets, and tracking services like *Rentrak Home Media Essentials* and *NPD VideoScan First Alert*[12].

For streaming platforms, estimated license fees[13] accrued by movies from subscriptions, such as those from *Netflix* and *Hulu*, are compiled. Additional financial data and movie characteristics are sourced from confidential profit participation statements from producers, court records, the press, news, and industry groups. However, supplementary sources are necessary to obtain subjective evaluations and pricing details for various movie distributions. We will look into these distinct facets of data aggregation subsequently.

[12] *Nielsen VideoScan* tracked sales of VHS cassettes, DVDs, HD DVDs, and Blu-ray Discs. The home video marketing research company was acquired by the NPD Group on January 6th, 2016.

[13] These fee estimates are based on the domestic box office and the rate cards employed by services for theatrically-released films.

1.3.2 Pricing Information

To prevent pricing interpolations, we scrap data from *JustWatch*[14] to determine the average price of each title across all available online digital rental and purchase options.

JustWatch serves as a comprehensive streaming guide. It presents digital offerings across multiple platforms, including paid subscriptions, streaming, rentals, and purchases, as depicted in Figure 1.10. For each title, the available offers are averaged to derive an estimated digital rental and purchase price. Given the limitations in tracking high-frequency digital pricing and considering that digital offer prices tend to be quite stable, with only infrequent promotions or discounts on individual platforms that don't significantly impact the average price, a consistent average price, varying by title, is used across all periods for online digital offers. Conversely, the weekly price for physical discs is directly sourced from *OpusData*.

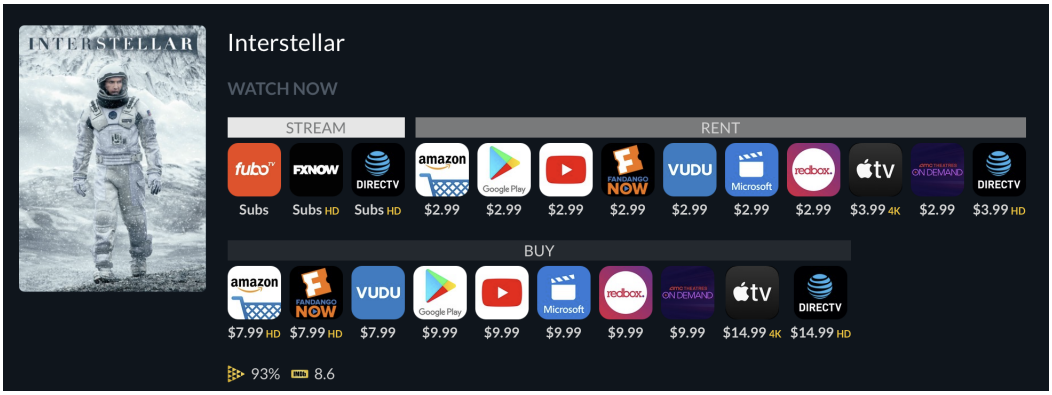

Figure 1.10: Purchase and Rental Options for Movie "Interstellar" on JustWatch

Note: JustWatch catalogs the movie purchase and rental options available on various online platforms.

For physical disc offerings, rental and purchase prices are established separately. For physical movie rentals, a consistent price is derived by averaging the costs of DVD and Blu-ray rentals from Redbox's Kiosk[15] over a 48-hour rental period. This rental duration is deliberately chosen to align with the standard online digital rental window. Weekly purchase prices for individual physical disc titles are deduced from *OpusData*'s weekly DVD and Blu-ray "revenue" and "units" data.

[14] *JustWatch*: www.justwatch.com.

[15] On average, the Kiosk 2-day rental price for DVDs is $1.75 and $2 for Blu-ray format.

Furthermore, to provide a more accurate measure of public choice behavior, subjective evaluations are integrated into the attributes.

1.3.3 Movie Ratings and Reviews

When selecting a movie to buy or rent, consumers encounter a wide array of choices, whether they are listed on a webpage or displayed on a shelf. While certain characteristics, such as plot design, execution, and acting performance, are revealed during a movie theater viewing, non-theatergoers can only access these experiences through ratings and reviews.

Among the three most prominent movie rating sites, namely *IMDb*, *Rotten Tomatoes*, and *Metacritic*[16], Rotten Tomatoes curates its reviews from trusted critics. While this allows access to professional opinions about a film, its scoring system lacks clear differentiation. A score marginally below 60% is ranked equally with a zero, making a film with a 59% rating seem as poorly received as one with no merits at all (Stegner, 2018). In contrast, *IMDb*[17] (Internet Movie Database) sources its ratings exclusively from users. Employing a 10-point scale, IMDb provides a nuanced subjective evaluation of films. Besides the ratings, the combined number of IMDb reviews from both users and critics is incorporated into our data. Given the critique that often only passionate fans or vehement critics leave a review on IMDb, rather than gauging the sentiment of a review, we utilize the total number of reviews as an indicator of a movie's popularity.

1.3.4 Sample Selection and Descriptive Statistics

The consolidated data comprises 40 movie attributes. These encompass domestic and international box office figures, domestic video sales, theater performance metrics, film and video production details, IMDb data, and metadata on films such as genre, whether it's a sequel, franchise affiliation, and information about the cast and crew. However, given the varied ways consumers can choose

[16] *IMDb*, *Rotten Tomatoes*, and *Metacritic* are the three most popular movie rating sites.
[17] *IMDb*: https://www.imdb.com.

to either buy or rent movies and the different formats of movie-watching incorporated in the data, our sample requires reevaluation based on the appropriate choice of consumers' consideration set. Indeed, for example, even after a consumer has shortlisted a specific title based on their preferences, they must decide whether to rent or purchase the movie, and this decision directly affects the modeling strategy.

Movie videos available for rent or purchase should not be treated as homogeneous products within the same choice set. Examining the issue from a pricing perspective, rental prices for movies typically fall below $4, while purchasing a movie averages at $14.99 and can rise to $17.99 or even higher[18]. Given the marked difference in pricing between rentals and purchases, their market shares should not be evaluated on an equal footing. However, there are exceptions, such as *The King's Speech*[19], where the price gap between rental and purchase is a mere dollar, making differentiation within the choice set challenging. Additionally, from a behavioral standpoint, the actions of renting and purchasing movies inherently differ.

Compared to renting a movie, people tend to purchase a film when they appreciate it and intend to watch it multiple times in the future. Before 2000, when renting physical VCDs or DVDs was prevalent, movie enthusiasts could repeatedly rent a film from outlets like Blockbuster. However, as digital streaming has largely replaced physical rentals, movie renters today mainly comprise those wanting to explore new releases and those keen on watching a film without visiting a theater. Given the constraints of digital movie rentals as discussed in Section 1.2, relying solely on sales data from movie purchases can yield insights driven by consistent consumer behavior, offering a more precise understanding of substitution patterns. Moreover, when purchases are influenced by backward spillovers, it suggests that a boost in sales results from consumers discovering a title because of the director, and importantly, from their desire to own that movie. To further streamline the choice set, we also exclude revenues generated from streaming license fees.

[18]The digital HD version of *Pirates of the Caribbean: On Stranger Tides* (2011) on Amazon Prime Video, directed by Rob Marshall and starring Johnny Depp, is priced at $17.99 for purchase and $3.99 for rental.

[19]*The King's Speech* (2010), directed by Tom Hooper and featuring Colin Firth, is priced at $3.99 for rental and $4.99 for purchase in HD digital format on Amazon.

Consumers tend to choose a streaming service[20] based on its subscription contents, rather than individual movies. The content libraries of platforms such as *Netflix*, *Hulu*, *Disney Plus*, and *HBO Max* differentiate them from standard online movie platforms. These limited choices on such platforms can obscure the observation of backward spillover effects. Furthermore, choices tend to be nested within these services, as many films are exclusive to particular platforms. For instance, popular movies[21] such as "The Wolf of Wall Street," "The Martian," and "Ready Player One" were unavailable on any of these four platforms in 2020. Additionally, both the highest-grossing domestic movie, "Star Wars: The Force Awakens," and the highest-grossing global film[22] in 2020, "Avengers: Endgame," which left Netflix in June 2020 due to contract expiration, are exclusively available on *Disney Plus*.

We utilize data spanning 12 years from 2006 to 2017, excluding movies produced before 2006 and those that never debuted in theaters. The number of movies produced each year is presented in Table 1.1, while the count of unique movies purchased by customers annually is summarized in Figure 1.11. We chose this specific timeframe to account for the uneven coverage of sales data from different movie purchase offerings in *OpusData*. Furthermore, movies released considerably before 2006 are more likely to become available for free to Amazon Prime members[23].

Table 1.1: Number of Movies Produced Per Year

Year	2006	2007	2008	2009	2010	2011
Number of Movies Produced	118	149	143	150	219	198
Year	2012	2013	2014	2015	2016	2017
Number of Movies Produced	152	330	329	317	243	40

Note: All produced movies premiere in theaters. However, not all movie videos are purchased by movie enthusiasts. Movies released in the fourth quarter are highly likely to have their videos released in the following year.

[20] Although users "stream" movies online or on a device after purchasing them on Amazon, Amazon is not classified as a "streaming service provider."

[21] The selected examples all feature well-known actors or directors: *The Wolf of Wall Street* stars Leonardo DiCaprio, *The Martian* features Matt Damon, and *Ready Player One* is directed by Steven Spielberg.

[22] For domestic and worldwide box office ranking information, refer to Table A.2 in the Appendices.

[23] While users don't need a Prime membership to buy or rent movies on Amazon Prime Video, viewing free videos on Amazon requires membership.

Furthermore, although the data extends to midway through 2020, the business model underwent a dramatic shift as movie theaters closed due to the COVID-19 pandemic. During this shutdown, movies such as *Mulan*[24] opted to have its premiere moved to the subscription service. The refined database includes 2,388 unique movies, 1,492 directors, 167 movie distributors, and 107 video distributors, and covers 12 genres with over 1.1 million weekly sales observations ranging from[25] April 30th, 2006, to December 31st, 2007.

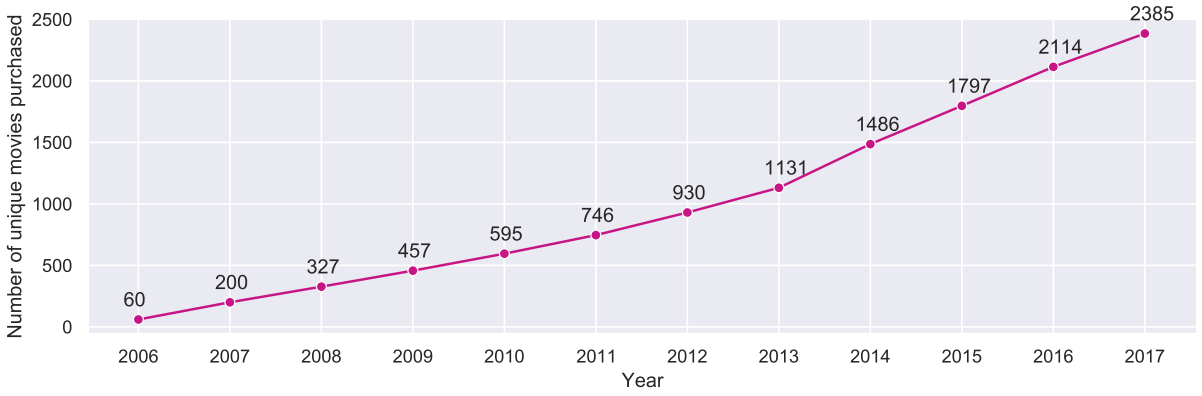

Figure 1.11: Number of Unique Videos Bought by at Least One Movie Enthusiast Per Year

Note: In 2006, 60 movies were produced, and their videos were purchased by at least one movie enthusiast. However, in 2017, there were 3 movies unpurchased by any enthusiast.

We have aggregated the genres into 5 major categories. Under *Comedy*, We have included "Comedy," "Romantic Comedy," and "Black Comedy." The *Horror* genre comprises both "Horror" and "Thriller and Suspense." The genres "Documentary," "Musical," "Western," and "Concert and Performance" are grouped under *Art*. The categorizations for movies are depicted in Figure 1.12. The *Drama* category includes films like "Life of Pi" and "Fifty Shades of Grey," which are crafted to evoke emotional responses from the audience. *Action-adventures* typically feature elements like fights, shootouts, stunts, and car chases. This genre can also present dangers in a more light-hearted manner, as seen in films such as the "Indiana Jones Series," the "Star Wars Series," and the "Transformers Series." The primary objective of *Comedy* films is to entertain and make the audience laugh; they often showcase characters in humorous situations. Films like "Pitch

[24] *Mulan* were directed by Niki Caro and stars Yifei Liu and Donnie Yen. It was exclusively available on Disney Plus.
[25] Movie videos are typically released a few months after their theatrical debut.

Perfect" and "The Devil Wears Prada" are examples of this category. *Horror* movies aim to elicit feelings of fear or disgust. Contrary to what might be expected, they can be characterized by rapid pacing and frequent action, utilizing plot-driven narratives to stir viewers' emotions more than action sequences. Films such as "Inception," "Angels & Demons," and "The Hunger Games Series" fall under this category. Lastly, *Art* encompasses films that depict real-life stories and those with musical components, examples being "Beauty and the Beast," "Les Misérables," and "La La Land."

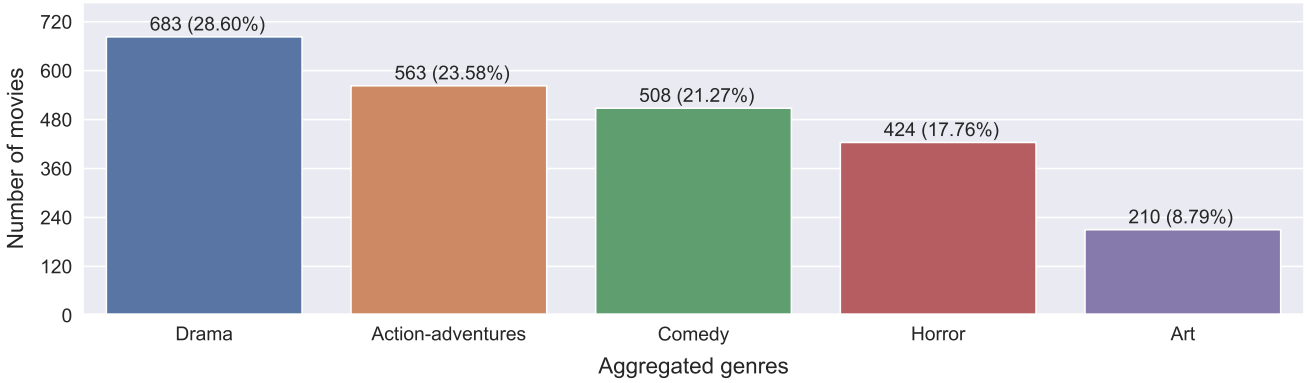

Figure 1.12: Summary of Movie Genres

Note: Compared to the IMDb genres, categories like "Fiction" are viewed as a type of "Creative Type." The genre aggregation combines smaller, less representative categories.

Creative type represents a form of class aggregation. *Dramatization and Factual* encompasses accounts of actual events as well as dramatizations of real-life incidents. The *Fiction* class has a wider scope, including fictitious works set in the real world, fiction incorporating future science or technology, and fiction targeted at children (kids fiction). The class of *Super Hero and Fantasy* includes movies featuring main characters endowed with superhuman abilities, as well as films that incorporate magic or supernatural elements. Other significant movie characteristics are detailed in Table 1.2. While genre and creative type shape the overall feel of a movie, another crucial factor influencing movie enthusiast choices is the makeup of the acting roles.

The number of stars and actors in a movie is determined by their acting roles. The most pivotal *leading* roles are identified by whether an actor or actress appears on the movie's theatrical poster. If more than four cast members meet this criterion, as seen in the "Harry Potter Series," they are

considered *lead ensemble members*. A cameo is a brief role, often appearing in just a single scene, but is played by a notable individual, such as Elon Musk's[26] appearance in "Iron Man 2" or Stan Lee's[27] in "Captain Marvel." Those listed in the credits who neither hold a leading nor a cameo role are categorized as supporting roles. To measure the impact of roles on movies, only leading roles are classified as "stars," while the rest are termed "actors."

Table 1.2: Movie Horizontal and Vertical Differentiation Attributes

Attributes	Definitions	Min	Max
Genre	Classes: *Comedy*, Action-adventure, Art, Drama, Horror.		
Creative type	Classes: *Dramatization/Factual*, Fiction, Super Hero/Fantasy.		
Production method	Classes: *Multiple Production Methods*, Animation, Live Action.		
Ratings	Classes: *Adult (R, NC-17)*, Unrestricted (G, PG, PG-13, not rated).		
IMAX	An indicator showing whether the movie has IMAX version or not.	0	1
Number of actors	Number of total roles in a movie.	1	166
Number of stars	Number of leading roles in a movie.	1	20
Running time	The length of the film (in minutes).	40 min	279 min
Sequel	An indicator showing whether the movie is a sequel or not.	0	1
IMDb ratings	A 10-point user-based rating reported on IMDb.	1.6	9.0
Total reviews	The combined numbers of reviews from users and critics on IMDb.	0	6762
Production budget	The movie's production budget.	$100,000	$379,000,000

Note: There are 2,388 unique movies spanning from 2006 to 2017, a 12-year period, with a total of 1,172,331 observations. The "genre" and "creative type" are re-categorized. The *first category* in each class is set as the base.

1.4 Market and Backward Spillover Measurements

In this analysis, we define the market at the year level. Targeting a shorter duration is impractical due to the limited number of unique titles purchased, which can result in skewed market share estimates. Additionally, measuring the market size for short periods is challenging, as irregular purchasing patterns can lead to significant fluctuations in market size over time.

[26] Elon Musk is the founder of SpaceX.
[27] Stan Lee was a primary writer for Marvel comics.

1.4.1 The Market Size

Following a similar construction as described in Berry, Carnall and Spiller (2006), we assume at the annual level that the market size, denoted as $\mathcal{M}$, is proportional to the population size:

$$\mathcal{M} = population \cdot \lambda, \quad \lambda = \frac{movie\ enthusiasts}{population} \cdot average\ purchase\ count \tag{1.1}$$

where λ is the proportional factor. Given that not everyone enjoys watching movies or purchasing them for their collection, we define the market based on the *ratio of movie "enthusiasts"* in the United States. This ratio is coupled with another choice variable that indicates the *average number of movies purchased by a movie enthusiast* annually. These two components are represented by λ.

Movie enthusiasts are individuals who enjoy watching movies and, crucially, are willing to purchase the videos[28]. Those who might watch movies but are often hesitant or disinclined to buy videos are not included in this market definition. To estimate the size of inside options, we use the total ticket sales of a top-selling movie. We specifically choose a blockbuster to best approximate the total number of movie enthusiasts, after adjusting for the corresponding annual average parameter. Given that *Netflix* and *Hulu* began offering stand-alone subscription-based services in November 2010 (Conlon, 2020), we aim to account for deviations caused by individuals transitioning to streaming alternatives instead of watching movies in theaters. Therefore, we select *Avatar*,[29] a top-grossing, non-sequel film released before 2010, as a reference. Using major franchises like "Star Wars" or "The Avengers" might introduce a significant "fan effect," resulting in a considerable upward bias. From there, we categorize *theatergoers* into three groups: targeted *movie enthusiasts*, *repeat viewers* who watch movies multiple times, and *non-enthusiasts*. The latter, although not considered part of the market, attend theaters due to herd behavior or sheer curiosity. As illustrated in Figure 1.13, selecting an exceptionally popular movie offers the advantage of capturing as many movie enthusiasts as possible. Under the allure of a popular film, the number of movie enthusiasts

[28] Recall that the term "videos" includes both physical and digital formats, which can be acquired in physical retail stores or online.

[29] Avatar (2009), directed by James Cameron, is still the highest-grossing movie worldwide in 2023.

who choose to wait for home video releases instead of going to the theater is minimized. We posit that this group will roughly counterbalance the overcount of tickets from the non-enthusiasts and the repeated viewers, especially since the number of these repeat viewers shouldn't be significantly larger[30] compared to other hit films. Using box office bestsellers to estimate market size offers both advantages and convenience, but it also comes with its own set of drawbacks.

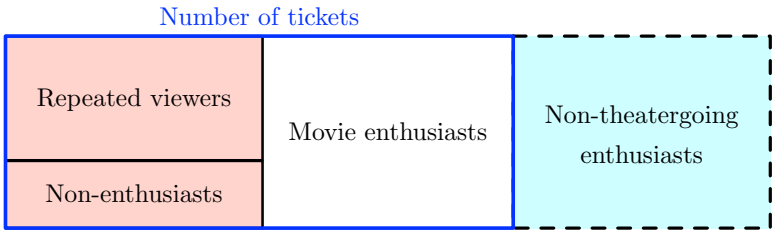

Figure 1.13: Tickets Composition and Potential Enthusiasts

Note: We assume that non-theatergoing enthusiasts are similar to both non-enthusiasts and repeat viewers. The non-enthusiasts contribute to "Avatar" ticket sales by herding effect and the novelty introduced by the 3D filming technology.

When determining the average number of movies purchased annually by a movie enthusiast, we also aim to account for potential pitfalls associated with using "Avatar" as a reference. Firstly, since the movie falls under the *Action-adventure* genre, films in this category often place a strong emphasis on special effects. The large IMAX screen and sound system of a theater provide a cinematic experience that home theaters or computer speakers/monitors simply cannot match.

Secondly, "Avatar" was heavily promoted due to its innovative use of "facial performance capturing technology" and the "3D fusion camera system." This likely drew more non-enthusiasts to theaters than usual, keen to witness the cutting-edge technology firsthand. Therefore, we opt for a more conservative annual movie purchase estimate for movie enthusiasts to offset any potential upward bias during estimation.

Choosing 2006 as the base year, we approximate that movie enthusiasts purchased four videos that year. By maintaining the expenditure consistent and adjusting for inflation, we can infer the

[30]The popularity of *Avatar* stemmed from its pioneering 3D photographic viewing experience depicting life on *Pandora*, rather than its plot. IMAX/3D tickets are more expensive than those for regular movies.

average yearly number of videos consumed in subsequent years from the average video price depicted in Figure 1.14. For example, a movie enthusiast would have bought 6.06 videos[31] in 2017.

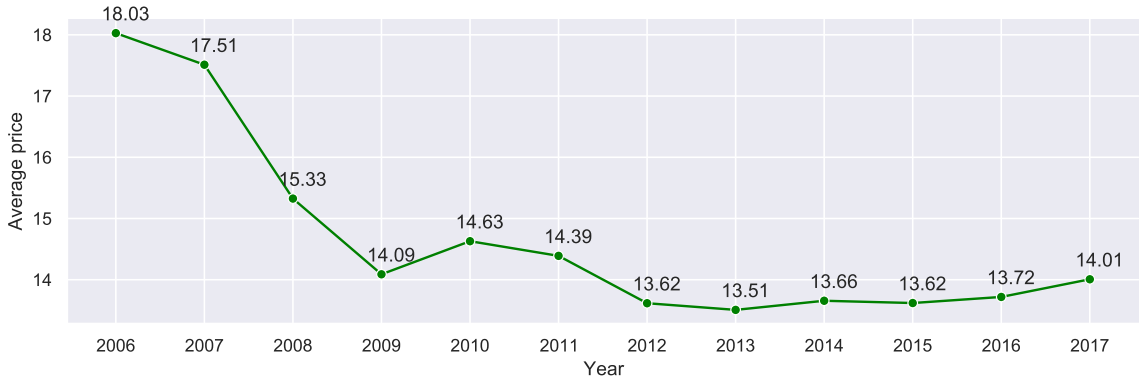

Figure 1.14: Average Video (Online Digital & Physical Disc) Price Per Year (Dollar)

Note: The average annual price is calculated from the mean of all sales observations for the respective year. For comparison, the average price for a DVD title was $22.29 in 2006 and $22.11 in 2007 (Raeford, 2020).

Building on our earlier discussion of market construction, we use *Avatar's* total ticket sales as a representation of movie enthusiasts in the 2009 population. This yields a 32.03 estimated ratio of movie enthusiasts to the total population, a ratio which we assume remains consistent across markets. To determine the market size for each year, we multiply the number of movie enthusiasts by the corresponding average yearly movie purchase count. The population and the relative "inside option" are depicted in Figure 1.15.

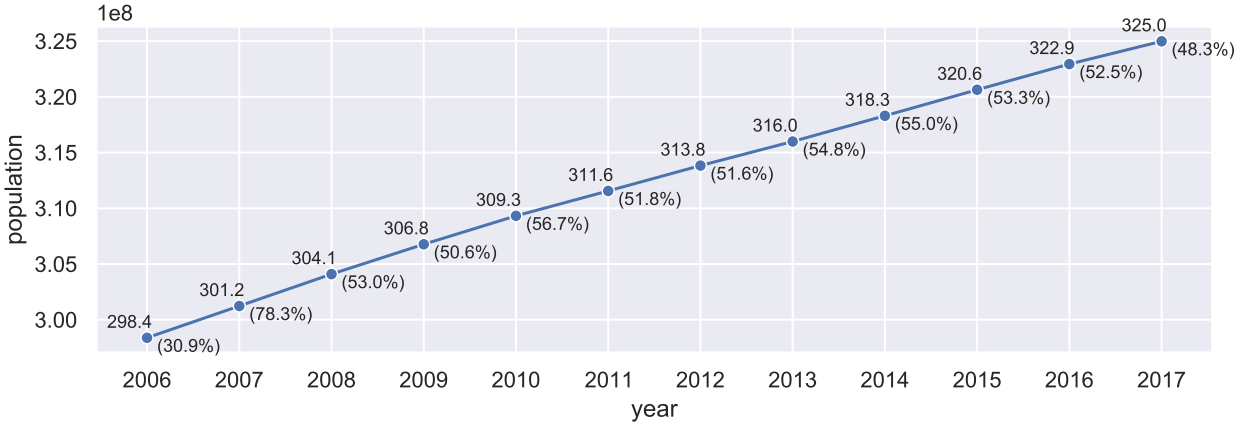

Figure 1.15: Population from 2006 to 2017 (Million)

Note: The population growth rate has remained steady over the years. The "inside options," represented by the total number of sales units in a year divided by the corresponding market size, are provided in parentheses.

[31] We allow for non-integer values since these numbers factor into the proportional ratio of the population.

1.4.2 Backward Spillover Indicators

The backward spillover effect is a phenomenon that emerges from the release of sequences under the same director. For directors who have directed multiple films, each movie experiences at least one backward spillover shock, except for the director's most recent work. The timeline for a new film typically begins with an early premium screening, accompanied by a promotional period.[32] The film then releases in theaters on a predetermined date. Depending on the theater's schedule, once the film ends its big-screen run, it will be distributed for home viewing within six months. The premium screening primarily targets critics and select invitees. If feedback from this screening is unsatisfactory, reshoots might be necessary. During the promotional phase, the main objectives are to ease the public's discovery of the film and to guarantee its quality. While the director's reputation may be highlighted in promotional materials, it is only after the audience watches and appreciates the film's post-theater release that true recognition is achieved. Consequently, a film experiences the most significant potential sales surge in the first week following the release of its sequences.

To illustrate the impact of a director's new releases, we incorporate a *sequence indicator* into the demand model. The indicator is assigned a value of one during the first week when the sequence-releasing shock occurs. We will explore the backward spillover effect up until the release of the third sequence[33]. As depicted in Figure 1.1, directors meeting the criteria for a fourth sequence comprise less than 2%. Instead of a persistent shock, the influence of the backward spillover diminishes over time and ultimately disappears. We refer to the duration from the sequel's release to the point where the effect mostly dissipates as the *backward spillover impact window*. The length of this window is dictated by the estimated *decay rate* of the backward spillover effect.

To model the decay function f, we start from the source of the spillover effects. The spillovers originate from the dissemination of the in-theater experience. During the initial week, an influx of reviews, social media shares, press reports, conversations, and occasionally news about record-

[32] Film promotion usually includes brief spoilers, several trailer versions, and press conferences featuring the director, cast, and crew.

[33] The third sequence refers to the fourth film released *after* the base film by the same director.

breaking box office performances enhances the likelihood of the public recognizing the film's director. This buzz wanes over the subsequent months, leading to diminished conversations about the film. We leverage the variations in box office revenues across different periods to track this decaying process. The right-skewed distribution of box office revenue is depicted in Figure 1.16.

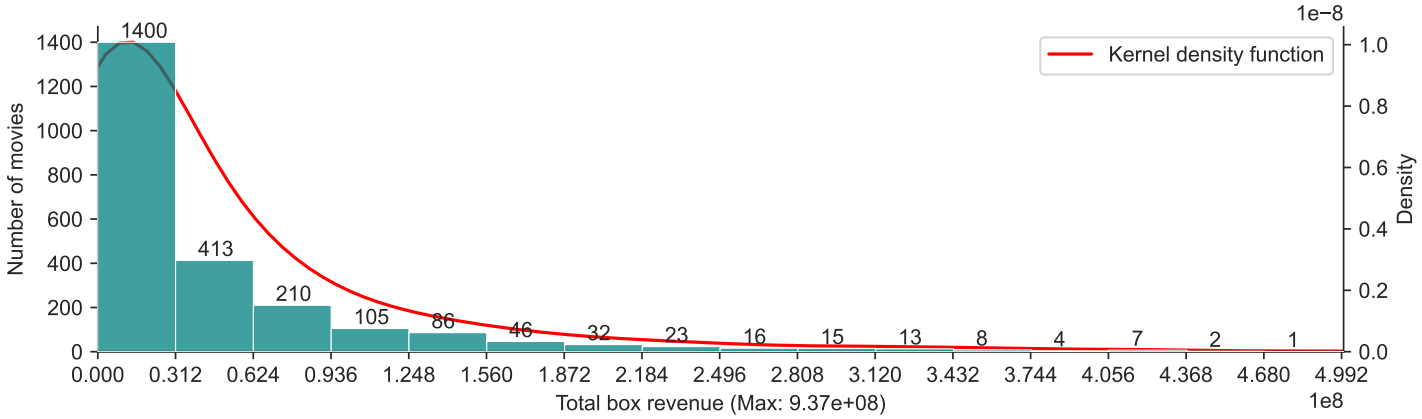

Figure 1.16: Box Revenue Distribution

Note: The average box office of the 2388 unique movies is 47.7 million. There are 702 movies with box revenue of over 50 million, 7 movies over 500 million, and 947 movies under 10 million.

We focus on films that have grossed a minimum of 50 million and have had a theatrical run of at least 12 weeks. By averaging the data for all selected films, we obtain the weekly box office revenues starting from the release week, as illustrated in Figure 1.17.

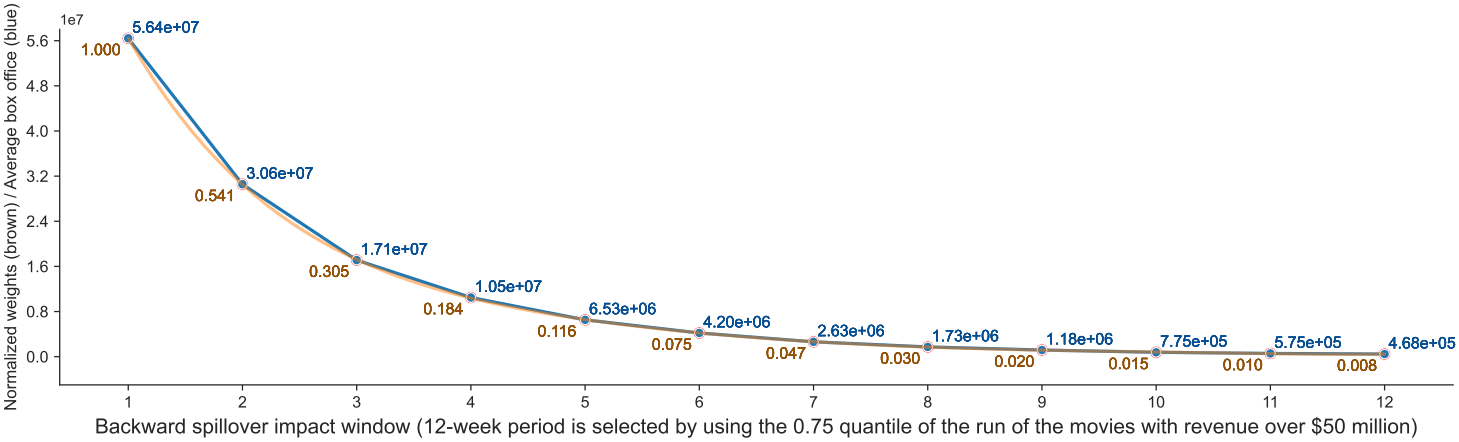

Figure 1.17: Average Weekly Box Office Revenues and the Decay Rate Function

Note: The 12-period weekly box office revenue levels are in blue. The orange line depicts the fitted polynomial function, with the normalized weights labeled. These function values correspond to the weekly box office levels.

We employed a seventh-degree polynomial to approximate the trend in weekly box office revenue. Additionally, $f(1)$, the impact factor for the first week of the treatment window, was normalized to one. If the decay function enters a domain (in terms of the number of weeks) that maps to negative values, it is set to zero. This polynomial spans 12 periods in the positive range, defining the backward spillover impact window. Beginning with the release week, the continuous treatment indicators adopt values from the decay rate function and continue to do so until they decline toward zero after 12 weeks. To allow flexible treatment variations, we will assign additional weight to the indicators throughout the impact window.

Despite the video receiving a full impact from the theoretical release of their sequences, the intensity of this treatment should vary among movies. For instance, the impact of a blockbuster differs from that of a mediocre movie. Although the decay rate is presumed to be the same, diminishing to zero in three months, the backward spillovers should be more pronounced for a video with a blockbuster sequence. Considering the heterogeneous popularity, we started with the continuously decaying treatment levels. From there, we constructed additional weights and integrated them into those initial treatment indicators. These additional weights were derived from each of the sequence's total box office earnings over the first 12 weeks, with the overall highest-grossing sequence movie normalized to 1. A single title might experience multiple waves of shocks due to sequence releases within a year. The cumulative yearly impact, combining these weekly impacts, is incorporated into the demand model.

By examining movie video sales, which include digital videos on various online platforms and physical discs in retail stores, we aim to gain a thorough understanding of the video market. Instead of focusing on theatrical consumption, our attention is on the home entertainment video sector. In our comprehensive view of the market, we're not only interested in the phenomenon of the director's backward spillover effect on video sales but also in understanding the market structure. This involves discerning movie enthusiasts' preferences regarding video selections and uncovering the cost structures and profit margins of videos, details often obscured from the consumer.

Given that the video availability is consistent across various cities and states, thanks to the expansion of online platforms and the presence of retail stores like Target, which can be found in almost every major city and many rural areas, we no longer view each city as a separate market. Instead, we treat each year as a market and encompass the entire nation. However, recall from the previous section that sales of DVDs peaked in 2005 with total sales surpassing 16 billion. This suggests that our data range captures a transition phase, seeing a decrease in physical disc market share while the digital video market share rises. As highlighted in Figure 1.14, there's a noticeable market shift starting in 2006, with the average video price dropping considerably for the subsequent years. Therefore, we will concentrate on the ten years from 2008 to 2017, during which the average video prices were more consistent, to estimate the demand model. The video prices observed during this period are depicted in Figure 1.18 below.

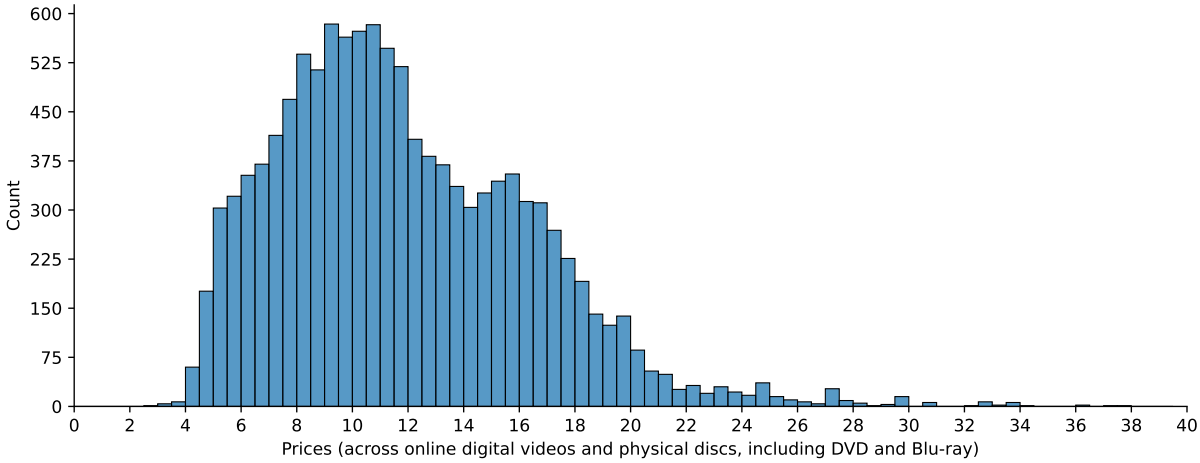

Figure 1.18: Video Price Distribution from 2008 to 2017

Note: The time range spans 10 years, from 2008 to 2017. The video prices encompass both online videos (averaged across various resolutions such as SD and HD) and physical discs, including DVDs and Blu-rays.

1.5 Model and Estimation

To describe movie enthusiast choices, we use a random coefficients (mixed) logit model to estimate the differentiated movie demand. Movie enthusiasts choose to purchase the movie that offers them the highest utility. Let us consider a series of markets from 2008 to 2017, denoted by

$t = 1, ..., 10$, where different movies that conditional on the choice set in each market[34] represented by $j = 1, ..., J_t$, are available for selection. The demand structure illustrates that the indirect utility, u_{ijt}, received by movie enthusiast i when purchasing movie j in year t is the following:

$$u_{ijt} = x_{jt}\beta_i + \tau_{jt}\gamma - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{1.2}$$

as a function of the movies' attributes, the impact from sequences shocks τ_{jt} and prices p_{jt} of the movies in market t. Here, x_{jt} is a vector of observed movie characteristics, including a constant.

These characteristics are divided into two categories. Under the category of non-linear attributes, which capture the product's horizontal differentiation, we have included five movie genres, three creative types, three production methods, and a movie rating that indicates whether the movie is unrestricted. These attributes reflect the heterogeneous preferences movie enthusiasts might have. For the linear attributes, which capture product vertical differentiation, we've incorporated an IMAX version indicator, the total number of actors, the total number of star actors, running time, sequel indicators, IMDb ratings, the total number of reviews from both critics and viewers on IMDb, and the production budget (scaled to millions).

α_i and β_i are individual-specific coefficients that represent the taste of movie enthusiasts i. γ indicates the conditional average taste from sequences' vertical differentiation. ξ_{jt} represents the market-level error term, which captures the common utility shocks that movie enthusiasts receive from the unobserved characteristics of movie j. By adding this constant term for each movie in each market, we remove the price endogeneity from the choice model. Finally, ϵ_{ijt} is an idiosyncratic shock distributed independently across movies, movie enthusiasts, and markets and follows a type-I extreme value distribution. To account for the heterogeneity among movie enthusiasts, let (α, β) represent the average preferences in the population that are common across movie enthusiasts. Across the population of households, enthusiasts' preferences for price and horizontally differentiated movie attribute (α_i, β_i) follow a multivariate normal distribution, where

[34] The video choice set varies by markets. See Figure 1.11.

the term ν_i is a vector of unobserved random tastes that influence purchasing decisions and Σ parameterizes the variances of the distribution of tastes

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Sigma\nu_i, \quad \nu_i \sim N(0, I_{n+1}). \tag{1.3}$$

The covariance matrix Σ allows for different variances between product non-linear characteristics. Following similar specification as Nevo (2000), equation 1.2 can be rewritten as $u_{ijt} = \delta_{jt} + \eta_{ijt} + \epsilon_{jt}$ where the mean utility

$$\delta_{jt} = x_{jt}\beta + \tau_{jt}\gamma - \alpha p_{jt} + \xi_{jt} \tag{1.4}$$

associated with movie j and varying across markets, is common across movie enthusiasts. The market specific idiosyncratic deviation from the mean utility η_{ijt} is specified as follows:

$$\eta_{ijt} = \big[p_{jt}, x_{jt}\big]\Sigma\nu_i.$$

If a movie enthusiast chooses not to purchase any of the movies, the indirect utility derived from this outside option is $u_{i0t} = \xi_{0t} + \epsilon_{i0t}$. Here, ξ_{0t} represents the unidentified mean utility from the outside good, which is normalized to 0. In market t, a decision to purchase movie j is made if and only if $\delta_{jt} + \eta_{ijt} + \epsilon_{ijt} \geq \delta_{kt} + \eta_{ikt} + \epsilon_{ikt}\ \forall j \neq k$. From the distributional assumption of ϵ_{ijt}, the probability of movie enthusiast i chooses to purchase movie j in market t is

$$\mathscr{s}_{ijt}(x_{jt}, \tau_{jt}, p_t, \xi_{jt}; \alpha_i, \beta_i, \gamma) = \frac{\exp(x_{jt}\beta_i + \tau_{jt}\gamma - \alpha_i p_{jt} + \xi_{jt})}{1 + \sum\limits_{k \in \{1,\ldots,J_t\}: k \neq j} \exp(x_{kt}\beta_i + \tau_{kt}\gamma - \alpha_i p_{kt} + \xi_{kt})}. \tag{1.5}$$

At the aggregate level, the model predicted market share of movie j in market t is the integration over the individual-level choice probability $\mathscr{s}_{ijt}$:

$$s_{jt}(\delta_t, \Sigma) = \int \frac{\exp(\delta_{jt} + \eta_{ijt})}{1 + \sum\limits_{k \neq j} \exp(\delta_{kt} + \eta_{ikt})} \mathrm{d}G_\nu(\nu_i|\Sigma) \tag{1.6}$$

where Σ is the scaling matrix to be estimated and G_ν is the movie enthusiasts preferences distribution. The integration in Equation 1.6 removes the independence of irrelevant alternatives (IIA) property. We follow Conlon and Gortmaker (2020) to estimate the nested fixed point. Let S_{jt} denote the observed market share of movie j in market t from the data. When attempting to match the model-predicted market shares with the observed shares from the data, Berry (1994) demonstrates the inversion such that, given Σ, δ_t has a unique solution[35] to the following equation:

$$S_{jt} = s_{jt}\big(\delta_t, \Sigma\big). \tag{1.7}$$

The converged mean utility δ_{jt} is used to estimate the linear parameters using an IV regression. As depicted in Equation 1.4, the unobserved characteristics ξ_{jt}, known only to the video distributors, may be correlated with the video price. Let z_{jt} be a vector of instruments. The IV moment conditions are

$$\mathbb{E}\Big[Z_t' \xi_t(\Sigma)\Big] = 0 \tag{1.8}$$

where $Z_t \equiv (z_{1t}', ..., z_{J_t t}')$, $\xi_t(\Sigma) \equiv \big(\xi_{1t}'(\Sigma), ..., \xi_{J_t t}'(\Sigma)\big)$[36]. To ensure that the moment condition is satisfied, Berry, Levinsohn and Pakes (1995) propose the GMM estimator that solves the following optimization problem

$$\widehat{\Sigma_{\text{GMM}}} = \underset{\Sigma \in \Theta}{\text{argmin}} \Big(\widehat{\xi}_t(\Sigma)' Z_t\Big)' \widehat{\Omega}^{-1} \Big(\widehat{\xi}_t(\Sigma)' Z_t\Big) \tag{1.9}$$

where $\widehat{\Omega}^{-1}$ is a positive semi-definite weight matrix. By matching the shares, the optimization ultimately searches a δ_t and its associated ξ_t that fulfill the moment condition. This optimization is addressed using two nested loops. The outer loop traverses the parameter space Θ using either grid search or gradient descent. Its goal is to minimize the GMM objective function, as defined in Equation 1.9. This function incorporates $\widehat{\xi}_t(\Sigma)$, which is derived from the fixed point δ_{jt} obtained in the inner loop, to construct the moments.

[35] $\delta_t \equiv (\delta_{1t}', ..., \delta_{J_t t}')$.

[36] For a fixed Σ, we derive a set of corresponding predicted shares by simulating with various sets of mean utility δ_t that solves the contraction mapping. Then $\widehat{\xi}_t(\Sigma)$ is obtained from the residuals of the IV regression.

For the inner loop, given Σ, the observed shares are matched with the predicted shares and are solved numerically using the following contraction mapping (Berry, Levinsohn and Pakes, 1995) to find the fixed point $\delta_{jt}(\Sigma) = s_{jt}^{-1}\big(S_{jt}; \Sigma\big)$ for the inversion in Equation 1.7:

$$\delta_t^{n+1}(\Sigma) = \delta_t^n(\Sigma) + \log S_t - \log s_t\Big(\delta_t^n(\Sigma), \Sigma\Big). \tag{1.10}$$

Here, S_t and s_t represent the stacked observed and predicted market shares for J_t movies, respectively. The objective function in Equation 1.9 is minimized when both the GMM gradient and the mean utility converge. The standard errors are estimated from the asymptotic covariance matrix of the GMM estimator.

The demand model is identified by the sufficient exogenous variation in prices from the instruments and the parametric assumption of the distribution of the random coefficients. The general model identifications are presented in Berry and Haile (2014). Excluding the *sequence* indicators, moments are generated using two sets of instruments for each video attribute listed in Table 1.2, with the inclusion of an additional variable: the movie's "maximum number of theaters." Both sets are derived from the video distributor ownership matrix. The first set aggregates the characteristics of competing movies from rival distributors. The second set focuses on the non-rival distributor, summing up the characteristics of the other movies from the same distributor. By design, these instruments are orthogonal to the unobserved product characteristic ξ_{jt}. The function of competitors' characteristics captures horizontal competition, which directly influences markup (profit margins), and thus is correlated with price. Furthermore, they act as approximations for the optimal instruments associated with the distribution parameters of random coefficients.

Of the 26 instruments constructed from the ownership matrix, we first correlated them with the price. Only those yielding more significant results (P-value < 0.01) are retained, leaving a total of 24 selected instruments. However, most of these instruments have small coefficients. Apart from these instruments, we exclude Waldfogel and Hausman instruments because the distribution of movie enthusiasts' demographics across the country has remained consistent over the years. Additionally,

by design, the prices of digital movies are rigid, showing no variation across different markets. For the outer loop optimization problem, we choose the L-BFGS-B optimization routine[37] from the Python `SciPy` library with gradient tolerance 10^{-6} for faster convergence. The tight fixed point converging tolerance for the inner loop is 10^{-14} as Dubé, Fox and Su (2012) find that the inner loop numerical error under high tolerance produces off estimates and can propagate preventing convergence of the outer loop.

For comparison, we estimate three demand models: linear regression, logistic IV regression (Berry, 1994) derived from the standard multinomial logit model,

$$\log(s_{jt}) - \log(s_{0t}) = x_{jt}\beta + \tau_{jt}\gamma - \alpha p_{jt} + \xi_{jt}, \tag{1.11}$$

where s_{0t} denotes the share of the outside option in market t, and the full model with random coefficients. In the full model, similar to Nevo (2000), we estimate the non-linear parameters with the video's fixed effects. The mean preferences for the horizontally differentiated characteristics are identified from the estimated mean utility $\widehat{\delta}_{jt}$ with the minimum-distance procedure.

1.6 Empirical Results and Conterfactual

From the demand model, we examine movie enthusiasts' video choice preferences, the director's spillover effects, and the market structure. Table 1.3 presents the results from the linear regression (LR) and logistic IV regression analyses. In the absence of random coefficients, the estimates represent the average preferences influencing movie video purchasing decisions. Without instruments, price endogeneity causes the linear regression model to fit the supply curve. From the logistic IV regression model, we observe that prices have a negative impact on movie enthusiasts'

[37] L-BFGS-B is a variant of BFGS that is more memory-efficient and supports constrained optimization. We observed that the BFGS routine takes significantly longer to converge.

utility, though the effect size is small. Also, only the impact of the director's first *sequence* exhibits backward spillover effects on the previous video.

Table 1.3: Linear Regression and Logistic IV Regression Results

Variable	Linear Regression	Logistic IV	**Variable**	Linear Regression	Logistic IV
Constant	-11.957^{***} (0.2032)	-11.065^{***} (0.2288)	Number of actors	0.0091^{***} (0.0009)	0.0079^{***} (0.0010)
Price	0.0807^{***} (0.0051)	-0.0261^{*} (0.0125)	Number of stars	0.0489^{***} (0.0078)	0.0552^{***} (0.0082)
Genre (Action-adventures)	-0.1475^{**} (0.0536)	-0.0863 (0.0585)	Running time	0.0055^{***} (0.0012)	0.0051^{***} (0.0012)
Genre (Art)	-1.2194^{***} (0.1046)	-0.8569^{***} (0.1096)	Sequel	0.0953 (0.0546)	0.1702^{**} (0.0598)
Genre (Drama)	-0.5491^{***} (0.0562)	-0.4455^{***} (0.0581)	IMDb rating	0.0703^{**} (0.0216)	0.1422^{***} (0.0242)
Genre (Horror)	-0.1491^{**} (0.0496)	-0.1365^{*} (0.0533)	Number of IMDb reviews	0.0007^{***} (0.0000)	0.0007^{***} (0.0000)
Creative type (Fiction)	0.0751 (0.0673)	0.0552 (0.0674)	Production budget	0.0067^{***} (0.0006)	0.0067^{***} (0.0006)
Creative type (Super Hero/Fantasy)	0.0226 (0.0812)	0.0569 (0.0835)	Sequence 1	0.0916^{***} (0.0089)	0.0905^{***} (0.0092)
Production method (Animation)	0.6890^{***} (0.1229)	0.6273^{***} (0.1316)	Sequence 2	0.0049 (0.0117)	-0.0213 (0.0119)
Production method (Live Action)	-0.2789^{**} (0.1030)	-0.3911^{***} (0.1116)	Sequence 3	-0.0460 (0.0205)	-0.0826^{***} (0.0205)
Rating (Unrestricted)	0.1302^{***} (0.0381)	0.1590^{***} (0.0392)	Sequence 4	-0.0967^{**} (0.0299)	-0.1412^{***} (0.0302)
IMAX	0.1280 (0.0750)	0.2233^{**} (0.0817)			

Note: Results based on 11,939 observations. Robust standard errors are given in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

On average, enthusiasts prefer *Comedy* (base) movies, such as "The Wolf of Wall Street," "Relatos Salvajes (Wild Tales)," and "The Grand Budapest Hotel," which bring joy and relaxation. Compared to films that employ *multiple* production methods, like "Inception" and "Avatar," people tend to enjoy more on animation videos such as "WALL-E," "Zootopia," and "How to Train Your Dragon." Conversely, pure action videos like "Django Unchained," "The Departed," and "The Da Vinci Code" are not as popular. Restricted films like "V for Vendetta" and "300" tend to be

less popular on average because their content may include strong language as well as bloody and violent scenes. However, the logit model produces counter-intuitive substitution patterns. Based on Equation 1.11, movie enthusiasts' taste heterogeneity arises solely from the movie-specific error term and is uncorrelated across similar videos. The own-price elasticity for video j and cross-price elasticity between videos j and k can be specified as $-\alpha p_j(1 - s_j)$ and $\alpha p_k s_k$. The cross-price substitution does not depend on the similarities between the videos (IIA property) but solely on market shares. To allow flexible substitution between videos, we add random coefficients. This enables consumers who favor videos with certain characteristics to prioritize selecting videos with similar bundles of features. The results of the full mixed-logit model are presented in Table 1.4.

Table 1.4: Random Coefficient (Mixed) Logit Model Results

Variable	Mean Utility (α, β)	Standard Deviation (σ)	**Variable**	Mean Utility (α, β)	Standard Deviation (σ)
Constant	-2.0330^{**} (0.5983)	0.1390	Number of actors	-0.0035 (0.0027)	-
Price	-1.0980^{***} (0.0642)	0.0860	Number of stars	0.1216^{***} (0.0218)	-
Genre (Action-adventures)	0.3264 (0.1734)	1.0351	Running time	0.0002 (0.0038)	-
Genre (Art)	-23.0497^{***} (0.2989)	14.8477	Sequel	0.9257^{***} (0.1736)	-
Genre (Drama)	0.5670^{***} (0.1557)	0.0000	IMDb rating	0.8638^{***} (0.0644)	-
Genre (Horror)	-1.4923^{***} (0.1549)	2.1396	Number of IMDb reviews	0.0004^{**} (0.0001)	-
Creative type (Fiction)	-0.2494 (0.1907)	0.0000	Production budget	0.0072^{***} (0.0017)	-
Creative type (Super Hero/Fantasy)	0.2728 (0.2420)	0.000	Sequence 1	0.0783^{**} (0.0285)	-
Production method (Animation)	-7.0651^{***} (0.3664)	7.3480	Sequence 2	-0.2804^{***} (0.0510)	-
Production method (Live Action)	-1.5582^{***} (0.2992)	0.2806	Sequence 3	-0.4367^{***} (0.0542)	-
Rating (Unrestricted)	0.1072 (0.1125)	2.4863	Sequence 4	-0.5817^{***} (0.0789)	-
IMAX	1.1686^{***} (0.2307)	-			

Note: Results based on 11,939 observations and 2,000 simulation draws. Robust standard errors are in the parentheses. σ is the estimated Cholesky root of the covariance matrix. The P-value for *Genre (Action-adventure)* is 0.0598.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Overall, the outcomes from the full model resemble those from the logistic regression for linear variables. However, the effect of video price on utility is much higher, which is more realistic. The director's backward spillover effects, though significant in the full model, are only apparent from the release of the first sequence. Beginning with the release of the second sequence, the increasing gap in years between film projects might make it more challenging for movie enthusiasts to discover the director's earlier works. Christopher Nolan directed films in 2007, 2008, 2010, 2012, 2015, and 2017, with an average interval of 2.17 years between releases. Taking "Dunkirk," his 2017 film, as an example, the second sequence was released 5 years ago. Figure 1.19 depicts the gap in years between a director's consecutive film releases.

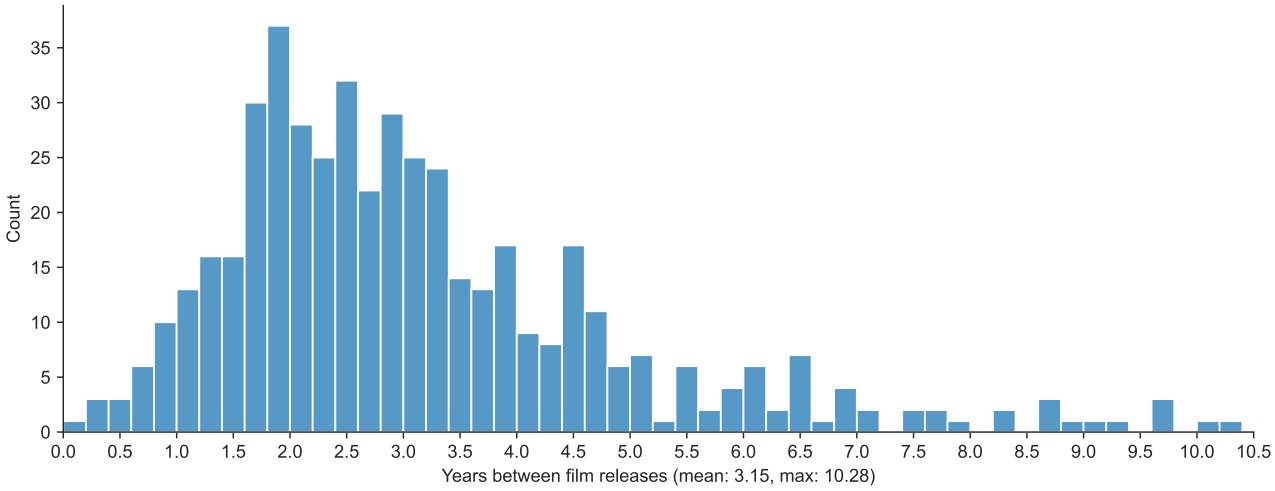

Figure 1.19: Gap in Years Between Director's Film Releases from 2008 to 2017

Note: Between the years 2008 and 2017, a total of 475 directors released more than one film.

On average, from 2008 to 2017, a director took 3.15 years to release a new film. Therefore, the decline in sales over time might outweigh the backward spillover effect. On the other hand, a director's backward spillovers primarily rely on their reputation and popularity. Compared to actors, whose performance and appearance can be directly observed, directors mainly work behind the scenes. They often require trailer promotions to not only leave a lasting impression on viewers but also to link the director to the film itself in the first place. Thus, while we do find evidence of the director's backward spillovers, the effect is not as pronounced as the sequel's backward spillover effect shown in Figure 1.3. Upon reviewing the preference estimates, the full model appears to have

captured intuitive choice behaviors, though it deviates somewhat from the logit model. Compared to *Comedy* movies, movie enthusiasts seem to have a preference for *Drama* films like "The King's Speech," "The Imitation Game," and "Whiplash," rather than horror movies such as "Shutter Island," "Gone Girl," "I Am Legend," and "Black Swan." Interestingly, while *art*-related movies like "Beauty and the Beast," "Alive Inside," "Into the Woods," and "Mamma Mia!" may appeal to a small group of people, they generally have a substantial negative impact on utility for the majority. Compared to movies that use a single production method, whether *live action* or *animation*, films that comprise a mixture of production techniques are preferred. This preference might stem from the unique viewing experiences that computer-generated images enhance when combined with *live action* scenes and performances.

The demand model also highlights the deviating preferences of movie enthusiasts when deciding between watching a film in a theater and buying a video for home entertainment. As outlined in Table A.1, most blockbuster movies fall under the category of *action-adventures*. These films emphasize special effects and scenes that benefit greatly from the immersive sound systems and large screens found in theaters. However, neither the logistic IV regression nor the random coefficient logit model provide sufficient evidence to confidently conclude that movie enthusiasts also prefer to select these types of movie videos for home viewing. These movies that do well in theaters are not also the case for video choices in home entertainment video markets.

In evaluating quality variables, movie enthusiasts tend to prefer movies with IMAX versions. When movies containing IMAX sequences are released on Blu-ray, they often maintain their original aspect ratio, offering an experience closer to full-screen viewing[38]. For instance, in the Blu-ray release of Christopher Nolan's "The Dark Knight," the aspect ratio adjusts to display more of the frame during the IMAX sequences (scenes). The results also align with our expectations that movie enthusiasts prefer films featuring more leading actors and actresses. While both *ratings* and the *number of reviews* have significant effects, the magnitude of their impact differs. Ratings, in

[38]Traditional movies use a cinemascope aspect ratio of 2.39:1, whereas IMAX films are shot with an aspect ratio closer to 1.43:1.

particular, have a significant impact on utility, as they serve as a screening criterion that people frequently focus on. As ratings can be seen as a broad yet subjective indicator of quality, they offer a quantitative measure. Our results confirm that ratings have a direct impact on choice decisions, even though there are concerns about bias in the numbers. As for the number of reviews, the combined count from both the public and critics on IMDb appears to be inconsequential to viewers. They may find the ratings alone sufficient for their screening needs. The production budget positively affects utility as well. Lastly, the pronounced effect of a sequel on the choice decision can be anticipated, these effects can be seen in series like "The Hunger Games," "Harry Porter," "Star Wars," and "The Pirates of the Caribbean." The continuity of a storyline often provides a compelling incentive for movie enthusiasts to purchase a video. Given the estimates indicating a preference for sequel movies, along with evidence of a director's backward spillover effect only in the first sequence, our findings suggest that producers don't need to retain the same director when scheduling the production of the entire story series. Next, we examine the responsiveness of the quantity demanded to changes in price and the substitution patterns across movie videos. The average own-price elasticity and the aggregate elasticity for the ten markets can be found in Table 1.5.

Table 1.5: Own-price and Aggregated Elasticities

Market	Average Own-price Elasticity	Aggregate Elasticity
2008	-17.0674	-2.2644
2009	-15.0735	-2.2034
2010	-14.1723	-2.2870
2011	-13.5428	-2.1411
2012	-11.8244	-2.0506
2013	-11.8443	-2.0464
2014	-12.3607	-1.9790
2015	-12.0794	-1.9665
2016	-12.3233	-1.9815
2017	-12.7727	-1.8135

Note: The demand for "purchasing movie videos" (as an entire product category) is generally less elastic than the demand for specific individual movie videos.

The aggregate elasticity of demand indicates the overall change in the market share of movie videos when there is a ten percent price increase. The elasticity of demand each year reveals that

movie video sales are highly sensitive to price fluctuations. However, movie enthusiasts are less likely to replace movie-watching with other leisure activities. Given that elasticity and substitution patterns are similar across markets, we use the own-price (η_{jt}) and cross-price elasticity (η_{jkt}) from 2,385 videos exclusively from the most recent market in our data, 2017, as illustrative example in the following Figure 1.20.

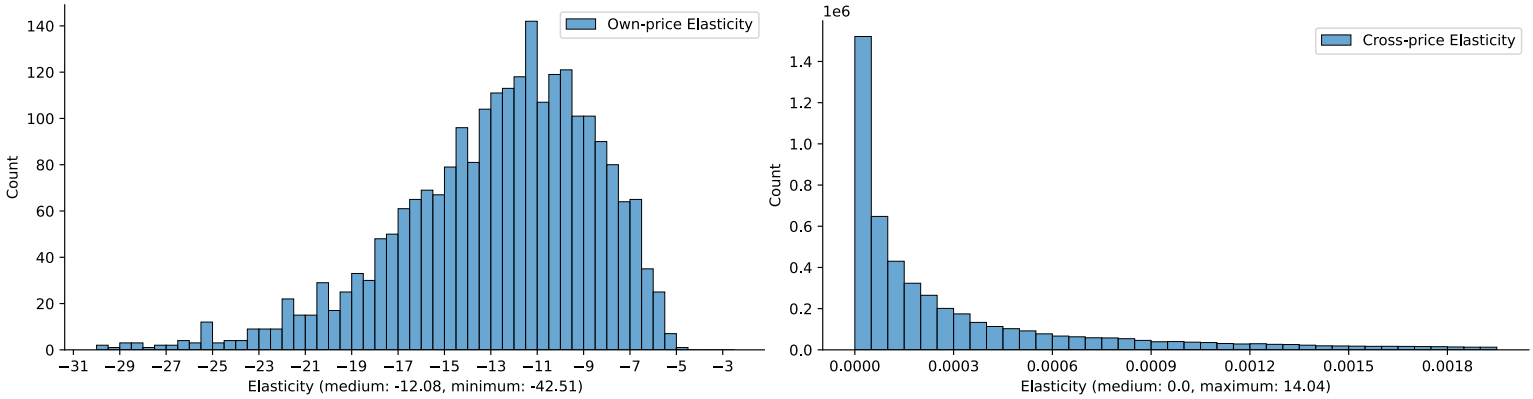

Figure 1.20: Video Own-price and Cross-price Elasticity in 2017

Note: Between the years 2008 and 2017, a total of 475 directors released more than one film.

We observe that the own-price elasticities of the video are highly elastic. From the cross-price elasticity depicted on the right panel, although all values are positive, only a few videos are strong substitutes. When the prices of similar videos change, and even for those within the same genre or of the same creative type, it generally has minimal impact on the purchasing choices of movie enthusiasts. Next, we'll examine the supply side and the market structure.

1.6.1 The Supply Side

Information from the supply side, which includes both marginal costs and markups, can be derived from the demand model. To derive the markup, we begin by representing the predicted market share for movie j in market t using the estimated mean utility:

$$\widehat{s}_{jt}(p_t) = \frac{\exp(x_{jt}\widehat{\beta} + \tau_{jt}\widehat{\gamma} - \widehat{\alpha}p_{jt})}{1 + \sum_{k \in \{1,\ldots,J_t\}: k \neq j} \exp(x_{kt}\widehat{\beta} + \tau_{kt}\widehat{\gamma} - \widehat{\alpha}p_{kt})} \tag{1.12}$$

where it is a function of the prices of all videos. Given that distributors influence the prices of videos sold across various platforms, and following Petrin (2002), let the set $\mathcal{V}$ contain J_t movie videos. Let $\mathcal{V}_\mathcal{D}$ denote the subset of videos released by distributor $\mathcal{D}$. The video distributors aim to maximize their profits over the set of movies they distribute:

$$\Pi_\mathcal{D}(j) = \mathcal{M} \sum_{j \in \mathcal{V}_\mathcal{D}} \big(p_{jt} - \text{mc}_{jt}\big) \cdot \widehat{s}_{jt}(p_t) - C_\mathcal{D} \tag{1.13}$$

where $C_\mathcal{D}$ represents the fixed cost distribution, and mc_{jt} is the constant marginal cost. Then the profit-maximizing decision is based on the following first-order condition:

$$\widehat{s}_{jt}(p_t) + \sum_{k \in \mathcal{V}_\mathcal{D}} \frac{\partial s_{kt}(p_t)}{\partial p_{jt}} \big(p_{kt} - \text{mc}_{kt}\big) = 0 \quad \forall j.$$

Let Ω represent the ownership matrix observed from the data. By stacking all the first-order conditions across all videos j in market t, we can rewrite the previous equation:

$$\widehat{s}_t(p_t) + \Omega \cdot \frac{\partial s_{kt}(p_t)}{\partial p_{jt}} \big(p_t - \text{mc}_t\big) = 0 \tag{1.14}$$

where the elasticities component is calculated by integrating movie enthusiast's choice probability from Equation 1.5:

$$\eta_{jt} = \frac{\partial s_{jt}(p_t)}{\partial p_{jt}} \cdot \frac{p_{jt}}{s_{jt}} = \frac{p_{jt}}{s_{jt}} \int \alpha_i \widehat{\mathscr{s}}_{ijt}(1 - \widehat{\mathscr{s}}_{ijt}) \text{d}G(\eta_{it}|\Sigma)$$

$$\eta_{jkt} = \frac{\partial s_{jt}(p_t)}{\partial p_{kt}} \cdot \frac{p_{kt}}{s_{jt}} = \frac{p_{kt}}{s_{jt}} \int \alpha_i \widehat{\mathscr{s}}_{ijt} \widehat{\mathscr{s}}_{ikt} \text{d}G(\eta_{it}|\Sigma).$$

The choice probabilities can be computed using the model's estimates, and the random coefficients included are obtained from Equation 1.3. After rearrangement, the marginal costs of videos in market t can be expressed as:

$$\text{mc}_t = p_t + \left(\Omega \cdot \frac{\partial s_{kt}(p_t)}{\partial p_{jt}}\right)^{-1} \widehat{s}_t(p_t). \tag{1.15}$$

Finally, the markup $\mathscr{M}$ of video j in market t can be inferred from the marginal costs:

$$\mathscr{M}_{jt} = \frac{p_{jt} - \widehat{\text{mc}}_{jt}}{p_{jt}}. \tag{1.16}$$

Focusing on the 2017 market year, which features the most recent market with the largest video choice set (2,385 movies), the recovered supply-side information is illustrated in Figure 1.21. Of the 2,385 films available in the 2017 market, the majority of movie videos have a markup ranging from 5% to 15%, with an average of 9.35%. The average markup for all years can be found in Figure A.4.

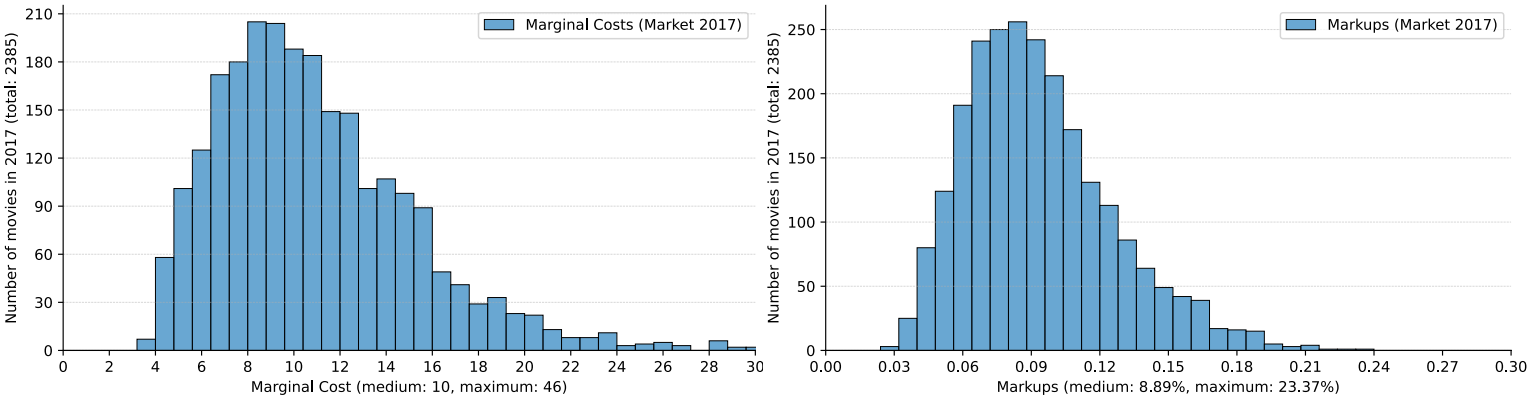

Figure 1.21: Video Marginal Cost and Markup Distributions for Market 2017

Note: The left panel presents the marginal costs for the 2,385 movies. The right panel shows the markups.

1.6.2 Market Structure and Merger Simulation

To look at the movie video market structure, we conduct a merger simulation to study the counterfactuals of how the market responds to a structural change. We focus on the 106 video distributors across all 10 markets from 2008 to 2017.

We have designed a hypothetical merger to analyze the structure of the movie video market. Numerous superheroes have been introduced via independent films. While each superhero has its unique storyline, they all share a common thread referred to as the "spindle task." Based on the

main storyline, there are two primary hero groups: *The Avengers*[39] and *The Justice League*[40]. *The Avengers* are owned by *Marvel Studios*, a subsidiary of *Walt Disney*, while *The Justice League* is produced by *DC Films*, under *Warner Bros*. Imagine a blockbuster film that unites all the superheroes from both groups. This would only be feasible if one parent company acquired the other. In this context, we aim to investigate the potential merger involving video distributors *Warner Home Video* and *Walt Disney Home Entertainment*. Using the 2017 market as an example, the descriptive statistics of the movies distributed by these entities are provided in Table 1.6. In 2017, neither of the entities involved in the merger were the first or second-largest movie distributors. Post-merger, they would only rank as the third-largest video distributor in that specific market. We will explore the changes in markup and market concentration resulting from the merger's impact across all market years.

Table 1.6: Number of Video Distributors in 2017

Video Distributors	Number of Movies
Sony Pictures Home Entertainment	319
Universal Home Entertainment	313
Fox Home Entertainment	289
Lionsgate Home Entertainment	245
Warner Home Video	**216**
Paramount Home Video	149
Walt Disney Home Entertainment	**93**
Anchor Bay Home Entertainment	86
Magnolia Home Entertainment	80
Other 97 distributors (less than 50 movies each)	595

Note: In the 2017 market, there are a total of 2,385 movies.

The post-merger structure results in a new price. Let Ω' represent the ownership matrix after the merger. Assuming the cost structure remains unchanged, by rearranging equation 1.14, we get:

$$p_t^* = \widehat{\text{mc}}_t - \left(\Omega' \cdot \frac{\partial s_{kt}(p_t^*)}{\partial p_{jt}}\right)^{-1} s_t(p_t^*) \tag{1.17}$$

where the equilibrium price p_t^* can be determined recursively. By substituting the new price into Equation 1.12, we obtain the post-merger market shares. With all the requisite information

[39] *The Avengers* includes characters such as Iron Man, Captain America, Doctor Strange, and Thor.
[40] *The Justice League* features heroes like Batman, Superman, and Wonder Woman.

concerning the post-merger scenario in hand, the new marginal cost can be computed using Equation 1.15, leading to the determination of the post-merger markup. Figure 1.22 displays the markups of the videos observed across all 10 market years. The right panel highlights the changes in markup following the merger. While the average markup for movie videos is 9.66% across all 10 markets, only a small fraction of videos experience a markup increase. In all 10 markets, the average markup sees a *percentage increase* of 1.56%. Around 16% of videos experience a markup percentage increase exceeding 1%, while a mere 5% see a percentage increase of more than 10%.

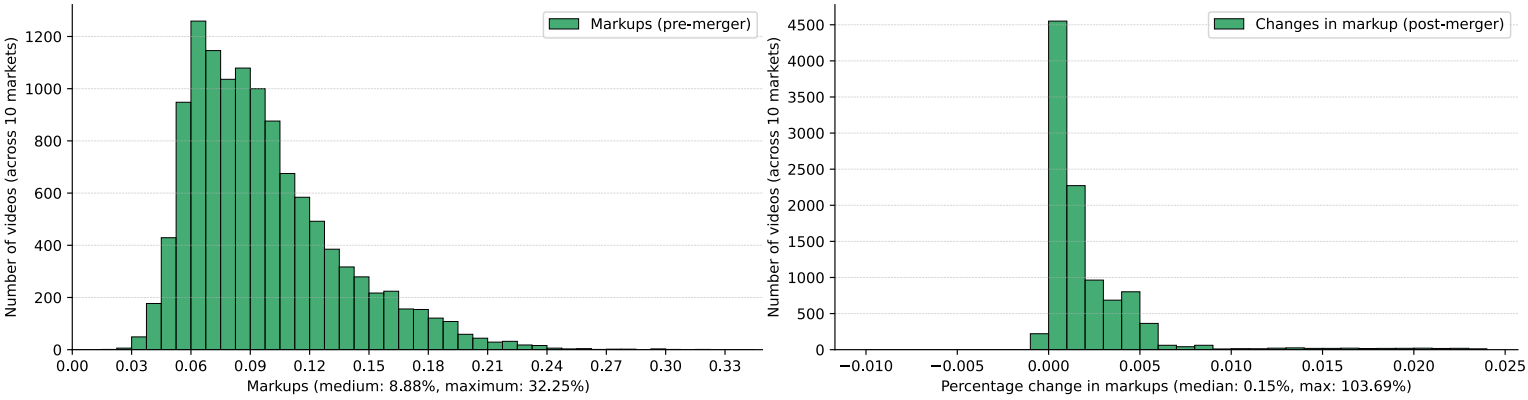

Figure 1.22: Markup Analysis Over All 10 Markets

Note: Both panels stack video observations from all 10 market years. The left panel displays the markup, while the right panel shows the difference in markup after the merger.

To provide a clearer perspective on the market structure changes caused by the merger, we compare the market concentration before and after the merger, as presented in Table 1.7.

Table 1.7: HHI of Movie Video Markets

Market	Before Merger	After Merger	Percentage Increase
2008	1154.09	1474.58	27.77%
2009	1136.79	1513.29	33.12%
2010	1241.11	1618.16	30.38%
2011	1300.68	1744.91	34.15%
2012	1182.39	1292.86	9.34%
2013	1208.63	1353.95	12.02%
2014	1060.61	1335.21	25.89%
2015	1296.76	1627.20	25.48%
2016	1211.33	1573.90	29.93%
2017	1351.95	1869.26	38.26%

Note: The Herfindahl-Hirschman Index (HHI) is obtained by summing the squared market shares of each distributor in the market and multiplying the result by 10,000. This indicates the pre-merger movie video market is monopolistic.

Before the merger, all 10 markets were competitive (concentrated)[41] with an average HHI of 1,214.43. However, when the two video distributors merge, even though their market share won't take the lead in the industry and the merger won't become the largest video distributor over the existing firms, the structure of the markets becomes moderately concentrated with less competition for 2009, 2010, 2011, 2015, 2016, and 2017 market years.

1.7 Conclusion

We have examined the movie video market, including both online digital videos and physical discs in retail stores. This study discloses consumer preferences and the various factors that influence their decisions when purchasing movie videos. With a focus on movie enthusiasts, we utilize a demand model to understand the structure of their buying behavior. Our investigation primarily targets consumers interested in buying movie videos, revealing preferences that differ from those selecting movies to watch in theaters.

We discovered evidence of backward spillover effects on video sales from the release of the directors' first sequence. An impact window was designed, starting from the theater release date of the sequence, to measure the shock from this spillover effect. The results indicate that although the impact of backward spillover on utility (and consequently on sales) isn't as pronounced as movie characteristics, it still offers a valid avenue for product discovery.

Movie enthusiasts tend to purchase popular movie videos, as evidenced by the number of reviews and high IMDb ratings. They show a preference for drama films, especially those with IMAX sequences. Both a higher production budget and a greater number of leading roles positively influence their buying decisions. The demand for movie videos remains highly elastic. Additionally, variations in prices of other movies have little influence on their purchasing choices. While the

[41] A market with an HHI under 1,500 points is classified as *competitive*. An HHI between 1,500 and 2,500 points means *moderate concentration*, while an HHI of 2,500 or more shows *high concentration*. The maximum possible HHI score is 10,000 points, which represents a monopoly.

"action-adventure" genre dominates as a popular choice for theater movies, there isn't sufficient evidence to support that movie enthusiasts have the same preference when purchasing movie videos.

By analyzing the demand model, we recover the cost structure of the movie video market. We conducted a merger simulation to gauge the effects of a structural change in the market. From the supply perspective, the average markup for movie videos is approximately 9.35%. Given the diversity among video distributors, the market remains competitive. The insights derived from this research complement existing literature centered on box office revenue, offering a comprehensive view of the movie video market.

Chapter 2

Understanding Household Choice of Leisure with Time Allocation and Expenditure Measurements

Everyone spends time in leisure but leisure is not costless. Leisure has often been overlooked in Economics studies due to the absence of direct measurements and available quantitative transformation. In this study, we combine data from the American Time Use Survey (ATUS) and mobile scanner data from retail markets to analyze household choices regarding leisure activities. Considering both time allocation and leisure costs, we investigate the geography of leisure using principal component analysis. Additionally, we construct leisure price indexes and examine time spent and expenditure variation in income across different leisure activities. By applying double machine learning and causal forest, we estimate leisure heterogeneous elasticity and identify leisure substitution patterns.

2.1 Introduction

Leisure has often been overlooked in economic research. Admittedly, the study of leisure is intrinsically complex and challenging owing to the absence of direct measurements and readily available quantitative transformations.

What is leisure? In classic Macroeconomics and Labor Economics, defined as the portion of the time not working. For instance, Atrostic (1982) separates total weekly hours into leisure and human capital. However, this dichotomy, originally intended to serve as a simplification, fails to capture the primary reasons households allocate time to leisure. Gronau (1977) introduces a trichotomy that divides total time into market activities, work at home, and leisure. He differentiates between the latter two based on the inability to enjoy or derive pleasure from the activities through surrogation. Bigoni et al. (2021) identify leisure as the time spent in activities that yield non-monetary rewards. We consider leisure as the time allocated to activities where the primary motivation is not measured by productivity. Through this lens, we can encompass a wide range of activities that fall under the umbrella of leisure pursuits.

Furthermore, it is important to consider the expenses/costs associated with leisure activities. Indeed, it is widely recognized that leisure is not costless, those activities come with a price tag and the cost of leisure should not be overlooked. For instance, going to the theater comes with a ticket price, attending a tennis match requires an entrance fee and reading might involve the cost of renting a book or purchasing a tablet/e-reader for accessing e-books. Likewise, to enjoy a brief nap in the afternoon sun, one might need to buy a chaise lounge. While these expenses are directly tied to leisure activities, they're frequently underestimated as households tend to concentrate on the primary activity itself rather than the associated expenditures. Other examples of leisure costs that are frequently disregarded include the expenses associated with purchasing goods necessary for activities such as *child care* and *socializing*.

To quantify leisure, we adopt a similar approach used in analyzing pricing and choice behaviors, focusing on the products. By aggregating the consumption of leisure-related products associated with leisure activities, we treat leisure as an indirect product measurable in terms of price. This approach offers a unique advantage in studying leisure choices based on their costs. By integrating data from the consumption of leisure-related products with time spent on various leisure activities, we offer comparisons that illuminate the amount of money necessary to allocate a desired amount of time to a specific leisure activity in a particular geographic location. This highlights how geographic factors can significantly affect costs. For example, an hour spent outdoors may incur different expenses in various states. Building a sand castle in a community lagoon or enjoying the waves at the surf ranch incurs additional costs compared to pursuing those activities in Maui.

To further examine the disparities in leisure expenses caused by demographics and location, we first need to develop a leisure price index. This index is standardized for comparisons across states and monitors the cost of leisure over time at a regional level. We then delve into the impact of income on leisure demand by estimating Engel curves and analyzing how various income groups allocate their time to leisure activities. Lastly, we examine own-price and cross-price elasticities to understand the sensitivities in heterogeneous household choices and the interplay between different leisure activities.

This article builds upon existing research on leisure, taking into account both the amount of time spent and dollar expenditure to offer a comprehensive view to study leisure as a whole and in a disaggregated perspective. We build on the work of Aguiar and Hurst (2007*a*), who provides a clear hierarchy of leisure activities, classifying them into 14 sub-categories. Most previous research on leisure-related issues has provided only partial analyses, largely focusing on the time individuals allocate to various leisure activities. Moreover, the definition and the scope of leisure in many studies tend to be narrow, encompassing only a limited range of activity categories. Aguiar, Hurst and Karabarbounis (2013) used data from the *American Time Use Survey* (ATUS) to explore shifts in how people divided their time between work and leisure during the Great Recession. Their

findings indicated that individuals primarily increased the time they spent sleeping and watching television. The study also highlighted notable increases in time dedicated to shopping, child care, education, and health. Meanwhile, Krueger and Mueller (2012) examined the variations in the time people devoted to leisure activities before and after embarking on a new job.

Luo, Ratchford and Yang (2013) used a dynamic panel data model to explore consumers' decisions regarding time allocation across a range of leisure activities. Their research revealed that consumer expertise, derived from past consumption experiences, is the principal factor influencing the dynamics of leisure activity consumption. Aguiar et al. (2021) studied the impact of quality variations within leisure activities on the marginal return to leisure and estimated leisure Engel curves to demonstrate how participation in leisure activities fluctuates based on one's total available leisure time.

Pawlowski and Breuer (2012) utilized the *Continuous Household Budget Survey* (CHBS) data from 2006 to examine expenditures on 18 distinct recreational leisure services, including entrance fees for swimming pools, music lessons, fitness centers, and admissions to theaters, museums, and circuses. They conducted a literature review on income and expenditure elasticities in tourism and recreational leisure, identifying two key shortcomings. Firstly, most research relied on data that was overly aggregated. Secondly, many analyses overlooked the issue of censoring samples from zero demand. They applied both type-I and type-II Tobit models to the data and discovered that the derived elasticities are highly model-sensitive. Our research addresses the issue of censoring through sample selection. Additionally, we estimate heterogeneous elasticities using a more flexible nonparametric approach.

This chapter is structured as follows: In the subsequent section, we begin by discussing the data source, pre-processing, and the re-categorization necessary to align the two datasets with the 14 leisure categories. We then merge time spent and expenditure measurements to present the geographical distribution of leisure. In Section 2.3, we calculate the leisure price index across various states and on a regional basis. Section 2.4 entails estimating the leisure Engel curves and

exploring the relationship between leisure time allocation and income. In Section 2.5, we present both the leisure own-price and cross-price elasticities. The final section, 2.6, offers our conclusion.

2.2 Data and Exploratory Analysis

The data integrates two sources: a time-allocation questionnaire and a user consumption tracking program. The flow of data is illustrated in Figure B.2. We utilize the time diary survey data from ATUS to measure the amount of time dedicated to leisure activities. Spanning from 2013 to 2018, this questionnaire engaged 540,000 respondents, prompting them to detail their previous day's allocation of time to various activities. It also records their demographic information. This breadth of data collection has established ATUS as a prominent tool for studying time allocations.

Burda, Hamermesh and Stewart (2013) employed this survey to investigate variations in weekly work hours, presenting an alternative productivity estimate. Mukoyama, Patterson and Şahin (2018) incorporated the data into a search model, discovering that the search effort did not amplify but rather dampened labor market fluctuations. Aguiar and Hurst (2007*b*) used the data to deduce the shopping technology that translates time-use and quantity-purchased into prices. The study also explored how households offset time with money through shopping and home production. Overall, the ATUS survey is an indispensable resource for researchers seeking to uncover insights into time-use patterns and their economic ramifications.

To represent the costs of leisure activities, we utilize consumption data that tracks purchasing behaviors in both online and physical retail stores in the US from August 2015 to February 2017, spanning 19 months. This data is gathered by a mobile-centric retail market research provider. Originating from a large user panel across two mobile apps, participants who have consented to share their purchase details photographs and upload their receipts post-visits to supermarkets and retail outlets. Additionally, these users participate in shopping trip-specific surveys. The company processes the receipt images and survey feedback, merging the purchasing details with specific

products and retailers. Such comprehensive purchase data allows retailers to delve into household buying decisions and acquire a deeper understanding of their customers' demographic nuances.

While both data collection processes are national in scope, only the consumption dataset extends to U.S. territories such as Puerto Rico, Guam, and the U.S. Virgin Islands. To ensure consistency in space and time, we limit our study to the fifty states and concentrate on data from the year 2016. Once this timeframe is established, it's crucial to ensure steady user participation. An ideal consumer representative should not only regularly photograph their receipts during the program's active period but also provide receipts spanning a broad spectrum of product categories. This is crucial because any lapses could skew demand estimation. For instance, if a participant solely shops from the secondary market, it might falsely suggest a lack of demand in retail outlets. Just because some user demand isn't immediately visible doesn't negate its existence. During the data preprocessing phase, we filter users based on their period of active usage. Later on, we will also take into account consumption variety.

Consistent participation is important when considering expenditure data. During the ATUS survey period, each respondent participates only once, ensuring no duplicate inputs. In contrast, the expenditure data collection spans 19 months, during which app users can be tracked multiple times. Evidence from the data indicates that some individuals may participate in the program out of curiosity rather than consistently recording their shopping lists. Several individuals exhibit gaps of months between scans. Such inactive behavior, leading to under-reporting, results in underestimated values when expenditures are aggregated. Consequently, we've filtered out buyers with a participation rate below 80% (or 15 months) from the pool to ensure only representative consumers are included. Within this vast set of receipt records, there are instances where the mobile app misinterprets price and quantity values. Such extreme values can significantly skew the average category expenditure. To address this, we've excluded 1% of data from both the upper and lower quantiles of the distribution. For ATUS, the time-spent data surveyed respondents about their previous day's schedule, including the time they spent on various activities. Some reported

spending over 21 hours in a single activity category. Such high values suggest potential errors or misinformation from the respondents. Given the relative size of the ATUS dataset (it's considerably smaller than the expenditure data), it's essential to preserve as many observations as possible. Thus, we've chosen to only remove the most extreme cases. Specifically, we've set a threshold of 1,300 minutes (approximately 21.7 hours) and excluded 13 out of 10,493 respondents from 2016 who exceeded this limit in a single activity.

To render the expenditure data and time-spent data comparable, we associate the two datasets by re-categorizing the product and time allocation categories into three groups: *leisure*, *market*, and *non-market* activities, following classifications found in the literature. This alignment allows us to produce results consistent with previous studies and deepens our understanding of household leisure choices by incorporating expenditure measurement, a critical component often overlooked in existing literature.

2.2.1 Leisure Activities

These datasets encompass the hierarchical structures of products and activities, as well as the demographic information of buyers and survey respondents. The ATUS data,[1] originates from 540,000 surveys conducted from January 2003 to December 2018. Respondents reported the number of minutes they spent on each third-tier activity, as depicted in Figure 2.1. This survey data is composed of three hierarchical structures, inclusive of 73 *second-tier* and 313 *third-tier* levels. To correlate the time spent with the expenditure on leisure activities, we primarily use Aguiar and Hurst (2007*a*) as a guide, categorizing leisure into 14 segments: *child care*[2], *eating*, *education*, *entertainment (not TV)*[3], *gardening/pet care*, *hobbies*, *own medical care*, *personal care*, *reading*, *religious/civic activities*, *sleeping*, *socializing*, *sports/exercise*, and *TV*.

[1]The ATUS data is sourced from the *U.S. Bureau of Labor Statistics*: https://www.bls.gov/tus/home.htm.

[2]Here, we group primary child care activities like breastfeeding and changing diapers, educational child care such as helping children with homework, and recreational child care like playing outdoors and attending a child's sports event under a singular "child care" category.

[3]The *entertainment* leisure category does not include television watching.

It may appear intuitive to integrate time spent into the cost of leisure, given that time availability inherently imposes constraints. For example, if someone spends six hours on leisure, the time consumed represents an opportunity cost. Yet, we consciously avoid conflating time spent with expenditure. Evaluating expenditure in isolation offers insights into the average dollar amounts people allocate to various leisure activities and enables us to investigate substitution patterns across leisure categories. Adding the cost of time complicates this, as it would necessitate a method to translate time into monetary value. Moreover, the valuation of time is subjective. For instance, a retiree might not value leisure time as highly as someone working 10 hours a day. As a result, our focus remains solely on monetary expenditure.

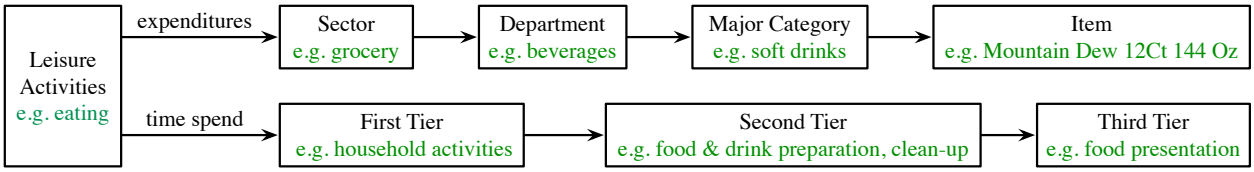

Figure 2.1: Data Hierarchical Structure

Note: The expenditure data comprises four levels, while the time-spent data has three levels. Both *Department* and *Second Tier* are employed to re-categorize items and time allocation categories into leisure activities. This figure showcases the various levels associated with *eating*.

There are some disputes in the literature regarding the categories of activities that should be considered leisure. At its core, leisure should be associated with enjoyable activities, bringing calm and excitement to the mind, or relaxing in nature. This makes categorizing activities such as *child care* and *own medical care* as leisure seem a bit far-fetched. However, when we step back and view the broader perspective, our definition of leisure, which considers activities not directly related to productivity, aligns with the categories recognized in the literature. In this context, our definition offers a more concise and inclusive understanding of leisure that accounts for a wider range of activities that people engage in for non-productive purposes.

From another perspective, many associate leisure solely with enjoyment. However, enjoyment is subjective and this limited view can be overly restrictive as it doesn't capture the diverse experiences and motivations of each person. The satisfaction gained from a leisure activity can differ depending

on an individual's specific circumstances, such as their goals and interests. Take *sports/exercise* as an example: jogging on a beach trail may be soothing and enjoyable for some, but for those primarily aiming for weight control, it might seem burdensome until they achieve their targets months later. By valuing motives beyond just productivity, we can acknowledge a wider range of activities that truly represent leisure. For example, even though *child care* can be challenging and exhausting, engaging with children and revisiting the joys of childhood can be deeply rewarding. Similarly, in the context of one's *own medical care*, individuals might derive a sense of accomplishment from participating in self-care activities or feel satisfaction believing their body is becoming healthier. Thus, by embracing a more inclusive definition of leisure, we can establish a consistent framework to categorize leisure activities and align our work more closely with existing literature.

Based on our categorizations, the 313 *third-tier* levels in the time allocation data can be mapped to one of 16 activities: 2 non-leisure and 14 leisure activities. In the Appendix, Table B.1 provides a summary aggregation of the 14 leisure categories, as delineated by Aguiar and Hurst. We further aggregate these categories to form four distinct leisure activity groups. These groups, ranging from level 1 to level 4, categorize leisure activities based on their level of detail, with level 4 including all the activities.

For the transformation of the expenditure data, purchased commodities are categorized by *sector*, *department*, and *major category*. The dataset tracks 19 months of both online and offline purchasing activities, spanning from August 1st, 2015 to February 15th, 2017. We applied fuzzy-matching to 17,000,000 product observations using product names to retrieve missing product *sector* information. After preprocessing, the data represents 303,909,188 items acquired by 301,890 distinct consumers across 70,796,084 receipts (purchasing trips). The 295 product *departments* from all purchased goods are manually matched to the previously mentioned 16 activities. The re-categorization process is depicted in Figure 2.2. Our choice to use *department* for expenditure data and *second-tier* for time-spent data is based on the need to relate these categories to leisure activities. This is because the matching process involves not just direct mapping, but also tier

reconstruction. In the matching process, the 295 product departments and the 108 second-tier categories are aligned with the 14 leisure categories in table B.4 based on their primary purpose[4]. For instance, the *appliances* category encompasses products like slow cookers, fryers, kettles, and pots, which are equipment primarily used to prepare food. Similarly, the time spent on *shopping for groceries* primarily involves purchasing items such as meat and fresh produce. As a result, both *appliances* and *shopping for groceries* are re-categorized under *eating*. Another case in point is condiments. While seasonings such as sauces and spices may not be the central items on a dining table, their primary purpose is to enhance the flavor of food. Consequently, they fall under the category of *eating*. An exception is smoking. Although people don't smoke for nourishment, we still classify smoking under *eating* because it is commonly associated with food and drinks.

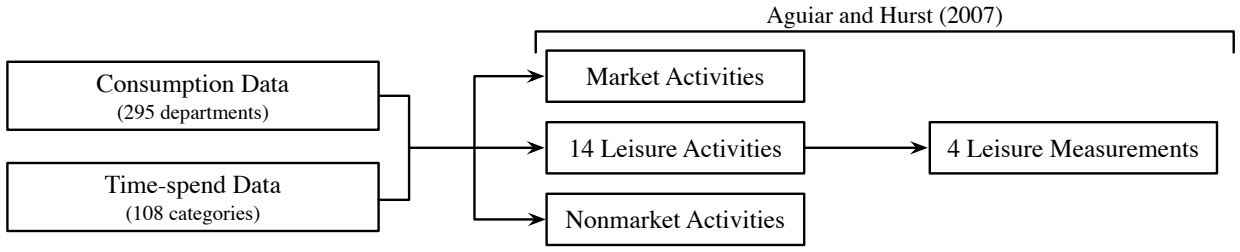

Figure 2.2: Products and Activities Recategorization

Note: Classifications from the two datasets correspond to the 14 leisure activities. Both market and non-market activities fall under the category of "non-leisure" activities.

The matching process is crucial for associating actual expenditures with leisure activities, allowing us to pinpoint the expenses linked to specific leisure pursuits. For example, the cost of *sleeping* can be connected to purchases of essential items like mattresses, bed frames, and pillows, essential equipment to facilitate and improve sleep quality. Nevertheless, the current matching process has its drawbacks. While it assigns existing categories to leisure activities, some categories have definitions that don't align well. While both datasets possess existing hierarchical structures, certain higher levels require reconstruction. In particular, some second-tier categories within the ATUS need regrouping. For instance, "relaxing and leisure" in the ATUS dataset is subdivided

[4]The primary purpose of a product refers to its main use.

and aligned to five leisure activities: *TV*, *entertainment*, *socializing*, *hobbies*, and *reading*, using third-tier descriptions. Conversely, instead of isolating travel activities, those related to "travel" in ATUS are distributed among the 16 activities[5]. Likewise, "services and activities" should be redefined and re-categorized into leisure classes, rather than being labeled as non-market work. This is because the purpose of such a service often correlates with a specific leisure activity. As an example, waiting times associated with government services are classified under *religious/civic activities*. Meanwhile, other non-leisure tasks still within "services and activities" are sorted into market and non-market work categories, including activities like working and cleaning respectively.

After recategorizing the two datasets, the 14 leisure activities were found to cover 189 expenditure *departments* and 227 *third-tier* time-spent categories. To provide a comprehensive overview of these departments and categories, including demographic heterogeneity, this study presents summary statistics in Table 2b through Table B.3 in the Appendix. Specifically, Table B.2 and B.3 offer detailed component lists of activity tiers and commodity departments for each of the 14 leisure categories, which are further illustrated by Table B.4. This combination of tables provides a clear and tangible understanding of the composition of each leisure category.

The unconditional average time an individual spends on different categories is measured in minutes and averaged across all respondents. The expenditure tables provide a summary of the annual conditional average amounts spent by individuals on different categories within a single receipt/purchasing trip. To clarify, we calculate the average expenditure for leisure activities based on the question: "On average, how much do you spend on a leisure activity when you shop?" However, it is important to note that these average expenditures are *conditional*, depending on whether the receipts include the relevant leisure categories. The "zeros" column in Table 2b and Table 2a sheds light on the extent of under-reporting by participants regarding their leisure activities. These tables provide a detailed summary of the preprocessed data and display the non-reporting ratio for each activity in the sample. Examining these ratios offers a deeper understanding of data coverage and helps us determine the steps to make the data more representative and informative.

[5]Time spent on "Travel" related activities is included in the *third-tier* classification.

Table 2a: Average Expenditure on Leisure Activities on a Receipt (in Dollars)

Category	Total Average	Shares	Female	Male	Asian	Black/AA	His./Latino	White/Cau.	Zeros
child care	19.39	0.08	19.32	20.12	23.14	18.78	20.24	19.00	0.07
eating	30.35	0.12	30.97	26.23	24.61	23.58	28.18	32.02	0.00
education	9.91	0.04	9.94	9.54	9.69	7.99	9.64	10.14	0.90
entertainment (not TV)	26.05	0.10	25.99	26.60	33.36	26.13	29.43	25.15	0.45
gardening/pet care	18.00	0.07	18.04	17.69	18.69	15.29	17.83	18.06	0.15
hobbies	7.73	0.03	7.72	7.74	7.78	6.78	7.34	7.82	0.26
own medical care	13.37	0.05	13.35	13.58	14.25	11.47	12.46	13.56	0.03
personal care	13.92	0.06	14.00	13.14	16.12	12.76	15.15	13.66	0.00
reading	6.47	0.03	6.49	6.32	7.39	6.65	7.68	6.30	0.49
religious/civic activities	12.99	0.05	13.01	12.86	20.99	12.86	12.39	12.06	0.97
sleeping	24.04	0.10	23.96	24.88	25.64	24.06	24.47	23.84	0.54
socializing	16.89	0.07	16.72	18.58	20.94	14.86	16.40	16.74	0.05
sports/exercise	20.44	0.08	20.38	21.00	24.99	20.48	20.71	20.07	0.53
TV	28.61	0.12	28.47	29.93	29.35	33.70	29.22	28.20	0.46

Note: Users are monitored multiple times throughout the year. Averages are calculated based on receipts containing relevant purchases, helping to mitigate under-reporting after data preprocessing. The column labeled "Zeros" denotes the percentage of responses that excluded the specific leisure activity. "Shares" and "Zeros" are ratios. The cost of *eating*, *TV*, and *entertainment* is higher. No major discrepancies exist between expenditures on genre and race, with Asians spending more, particularly on *religious/civic activities*.

Table 2b: Average Time Allocation on Leisure Activities in a Day (in Minutes)

Category	Total Average	Shares	Female	Male	Asian	Black	White	Others	Zeros
child care	39.36	0.03	48.85	27.50	62.97	20.90	41.51	43.48	0.71
eating	104.39	0.09	111.55	95.44	144.32	87.05	105.70	103.09	0.02
education	16.32	0.01	16.21	16.47	42.91	11.40	15.87	18.16	0.95
entertainment (not TV)	54.74	0.05	49.43	61.35	44.74	60.11	53.60	73.80	0.56
gardening/pet care	19.98	0.02	16.75	24.01	11.19	8.72	22.72	13.85	0.74
hobbies	4.53	0.00	6.02	2.67	3.36	2.81	4.93	3.92	0.97
own medical care	3.69	0.00	4.53	2.64	2.23	5.34	3.44	4.13	0.97
personal care	48.06	0.04	56.58	37.42	45.29	59.08	46.32	42.52	0.19
reading	20.68	0.02	23.35	17.35	18.58	12.23	22.55	14.82	0.78
religious/civic activities	25.63	0.02	29.68	20.58	19.86	37.17	23.73	26.51	0.82
sleeping	533.98	0.47	539.15	527.52	534.97	548.76	530.73	547.21	0.00
socializing	57.07	0.05	60.30	53.04	52.00	61.67	56.44	57.99	0.55
sports/exercise	22.81	0.02	17.37	29.60	26.62	14.44	24.28	20.06	0.80
TV	179.41	0.16	166.65	195.34	108.45	231.30	173.63	166.49	0.20

Note: In 2016, each user completed the questionnaire only once. The numbers presented are unconditional averages across all respondents. The reported average total time spent on leisure activities was 18.8 hours. "Shares" and "Zeros" are ratios. Households allocated more time to *sleeping*, *eating*, and *TV watching*. Women spent more time on *child care*, *eating*, and *personal care*, while men favored *entertainment* and *TV watching*. Black individuals allocated more time to *personal care*, *religious/civic activities*, and *TV watching*, and Asians invested more in *child care*, *eating*, and *education*.

A discrepancy is evident in the zeros of certain leisure categories between the two datasets, such as *child care*, suggesting issues with a censored sample. To counteract this, we introduced additional sample selection measures during expenditure analyses, encompassing control over participation rate and commodity variety. However, measures applicable to the time-spent data are somewhat restricted, given that the ATUS dataset's size is significantly smaller than that of the expenditure data. At the same time, it is possible that these zeros may also result from insufficient demand or time allocated to specific activities. To clarify the patterns of leisure across various states, we present the average dollar and minute amounts spent on 14 leisure activities in Table B.5 and B.6.

2.2.2 Two-way Comparisons

To gain deeper insights into leisure habits, we examine both the time spent and expenditures on various activities. By combining these measurements, we present a geographic distribution of leisure activities across different states. Figures 2.3 and 2.4 showcase the two-way graphs representing two selected activities.

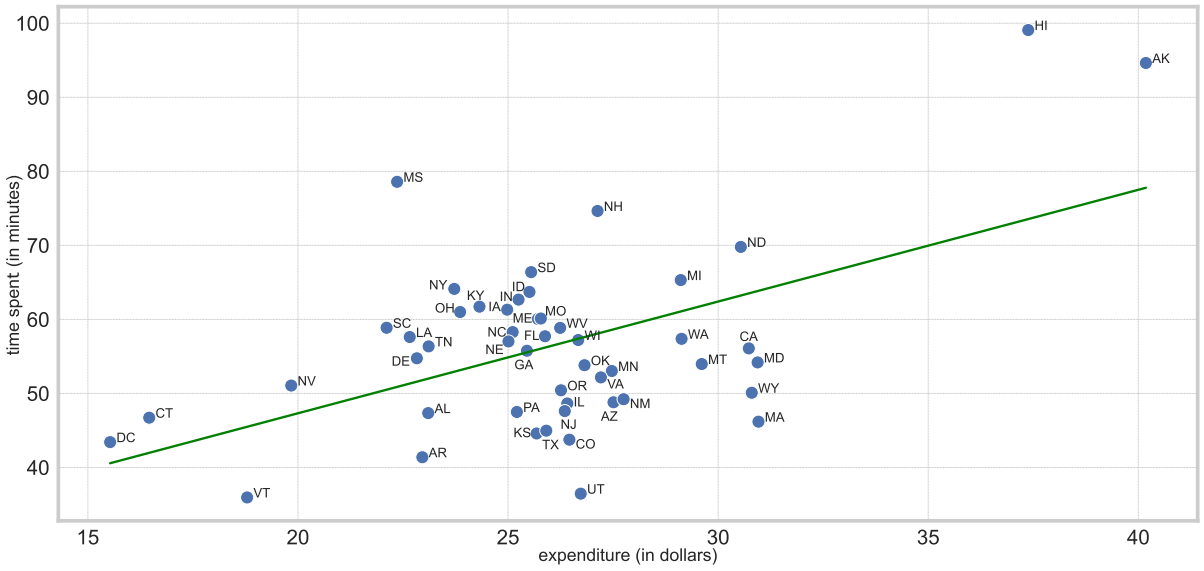

Figure 2.3: Two-way Comparison for *Entertainment (not TV)*

Note: Combining expenditures (conditional average on receipts involving relevant purchases) with time spent (unconditional average across all respondents), the scatter plot illustrates the household choice of engaging *entertainment (not TV)* in different states. The green line depicts the conditional mean of time allocation for this leisure activity.

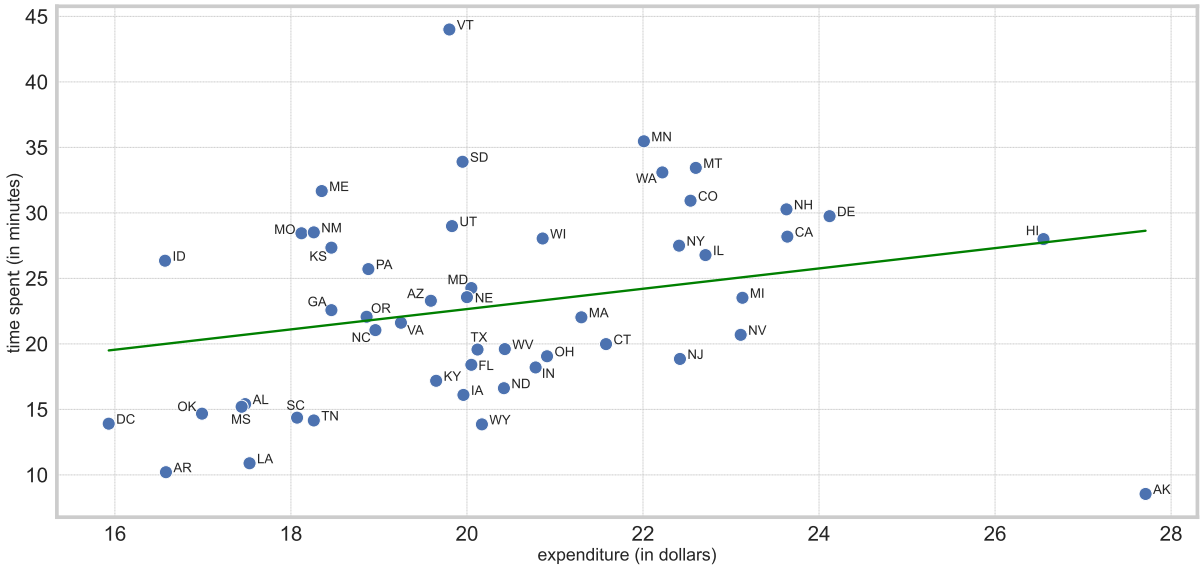

Figure 2.4: Two-way Comparison for *Sports/Exercise*

Note: The scatter plot combines the measurements of time allocation and cost of doing *sports/exercise* across different states. There is a positive relationship between time spent and expenditures.

As an illustrative example, we delve into two specific leisure activities: *entertainment (not TV)* and *sports/exercise*. Several factors, including geographical location and associated costs, influence the selection of leisure activities. We employ a two-way plot to study households' leisure preferences across states, along with the expenses tied to them. The green line on this plot depicts the conditional mean of time allocations. Through the integration of both time and expenditure data, we explore how individuals distribute their leisure time and the financial implications across different states.

The choice of leisure varies based on geographic areas and the cost of associated commodities. For these two leisure activities, the majority of states exhibit similar tendencies. Notably, given Hawaii (HI) and Alaska (AK)'s appeal as tourist destinations, residents spend significant time on *entertainment (not TV)*. Additionally, households in these states incur higher costs for goods related to this activity than do other states. For *sports/exercise*, households in both HI and AK spend more on associated commodities compared to other regions. However, Alaskans dedicate significantly less time to this activity, suggesting that *sports/exercise*-related activities are more expensive in Alaska. Following, we will examine the patterns across states using separate measurements.

2.2.3 Geographic Clustering

We explore more thoroughly the geography of leisure, clustering based on the principal components of time spent and expenditures. To uncover the underlying patterns, we standardized and normalized the data, and utilized the first and second principal components to change the basis and perform clustering with the EM algorithm. The clustering results are showcased in Figures 2.5 and 2.6. Utilizing the principal components condenses the influences of the 14 leisure categories into two dimensions. Furthermore, the majority of the variance is captured by these first two components.

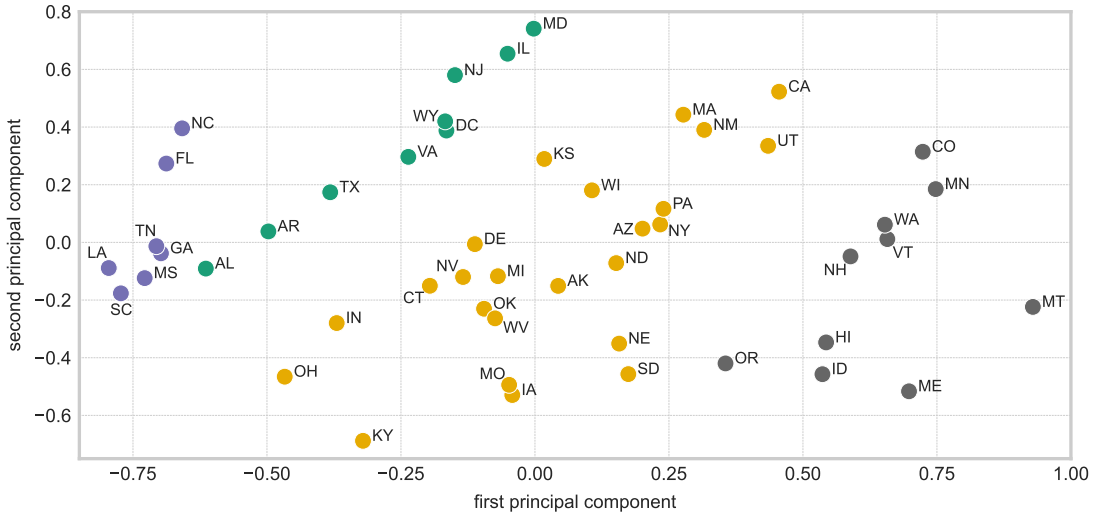

Figure 2.5: Principal Component Analysis of Time Spent on 14 Leisure Activities

Note: The explained variance ratios for the first five principal components are [0.208, 0.115, 0.112, 0.096, 0.081]. Four clusters were formed using the EM algorithm. The silhouette scores for cluster assignments of [3, 4, 5] are [0.249, 0.281, 0.250]. The colors solely represent the clusters.

Compared to k-means clustering, the expectation-maximization approach provides the benefit of more flexible cluster assignments. It overcomes the constraints of spherical clusters inherent to k-means. This flexibility is granted by permitting each Gaussian mixture component to possess its distinct covariance matrix. To ensure robust clustering outcomes, we established a minimal convergence threshold and significantly increased the maximum iteration count in the algorithm. The optimal number of clusters is determined using silhouette scores[6], which vary between -1 and

[6]The silhouette value measures how similar an object is to its cluster relative to other clusters. A score of 1 suggests that the clusters are densely packed and distinctly separated.

1. These scores offer insights into cluster density and separation. For our analysis, we opted for a cluster count that aligns closely with the number of census regions, ultimately dividing both datasets into four groups.

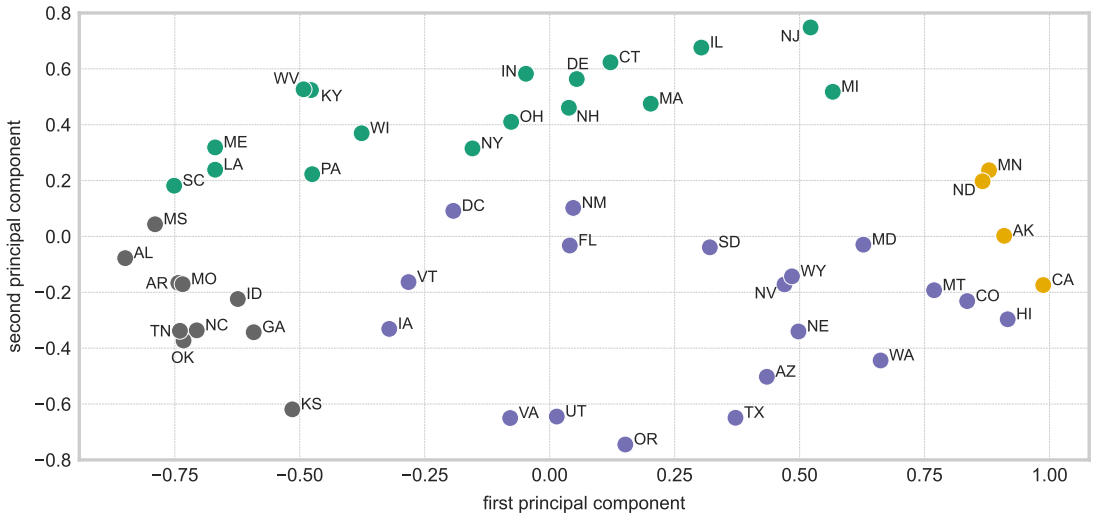

Figure 2.6: Principal Component Analysis of Expenditures on 14 Leisure Activities

Note: The explained variance ratio of the first five principal components are [0.404, 0.125, 0.097, 0.075, 0.074]. The silhouette scores for [3, 4, 5] clusters are [0.397, 0.271, 0.266].

The clustering of the expenditure data yielded a silhouette score of 0.271, whereas the time spent data received a score of 0.281. The clusters reveal discernible geographical patterns. Figure 2.7 displays the map projections of these clusters with expenditure data clustered on the right.

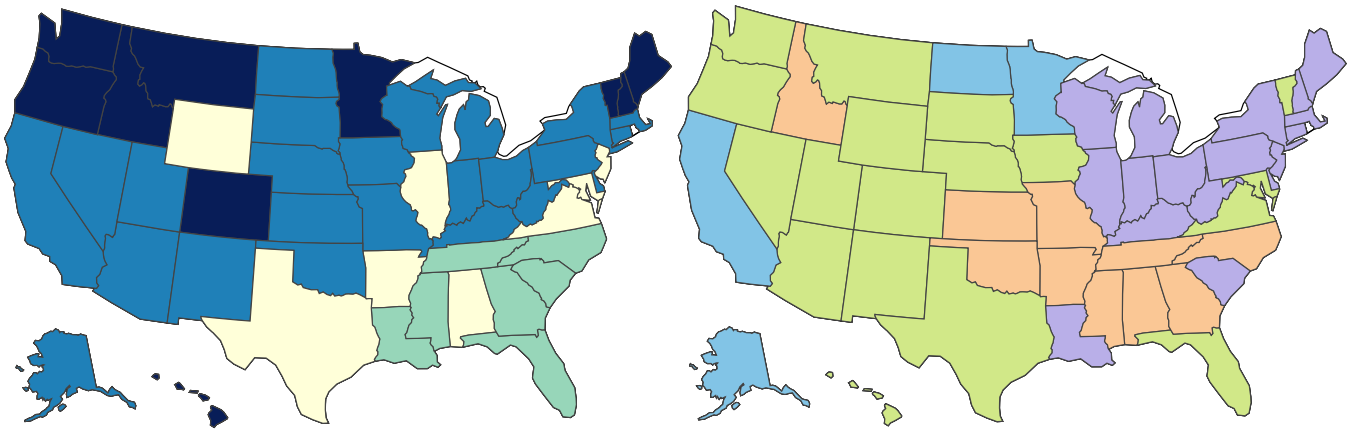

Figure 2.7: Geographic Visualization of Time Spent and Cost in Leisure Activities

Note: The clustering for time spent is shown on the left, while the cluster assignments for the expenditure data appear on the right. The colors between the two panels are independent and solely indicate the clusters.

The clusters reveal geographical patterns. From the time spent data, states with similar latitudes exhibit analogous time allocations for leisure activities. For the expenditure data, similarities are observed along longitudes. Moreover, expenditure patterns align more closely with the divisions of the census regions: Midwest, Northeast, South, and West.

2.3 Cost of Leisure

Although often viewed as "costless," pursuing leisure actually has a cost. To measure the cost of leisure, we follow Hill (2004) and Zhen et al. (2019) to construct leisure price indexes, as detailed in Appendix B.2. Unlike the U.S. Bureau of Labor Statistics, which uses the Consumer Price Index (CPI) to measure the cost of living over time, we construct cross-sectional price indexes by state to assess the cost of leisure. However, because product availability and variety differ significantly across states, we cannot create a uniform "basket of goods" for all. As a result, for each state, we aggregate all 14 leisure categories to represent a single commodity in the basket. This modification causes the four types of price indexes to converge into a price ratio p^t/p^0, representing the proportion of the leisure price in period t relative to the national average in the base period. We summarize the cost of leisure across states in Table 2.2.

The calculation of leisure prices is influenced by the variation in quantities demanded across various products. As such, normalization becomes essential when aggregating leisure activities. For instance, within the *TV* leisure activity, the annual consumption of an "OLED smart TV" diverges considerably from that of "cable accessories." As a result, there is a considerable risk that prices deduced for the *TV* activity based on total expenditure might be underestimated, especially when compared to *eating*. This discrepancy arises because the *major categories*[7] within *eating* display less variation. Furthermore, smart TV prices are markedly higher than those in other *major categories*. For example, when compared to the price of a bible in the *religious/civic activities* category, leisure

[7]The hierarchy of leisure, starting from the broadest level and narrowing down, is: *leisure category*, *department*, and *major category*.

categories containing pricier *major categories* can intensify the bias. To address this issue, we demeaned both the price and quantity of each commodity according to its associated *major category* before aggregation.

Table 2.2: Leisure Price Index (Base Period: National Average)

Ranking	State	Normalized Index	**Ranking**	State	Normalized Index	**Ranking**	State	Normalized Index
1	HI	1.3169	18	CT	1.0221	35	WV	0.9695
2	AK	1.2831	19	PA	1.0172	36	IL	0.9693
3	MT	1.1043	20	LA	1.0109	37	NH	0.9655
4	ND	1.1029	21	NM	1.0093	38	ME	0.9599
5	CA	1.0735	22	NE	1.0088	39	IN	0.9592
6	NJ	1.0657	23	VT	1.0033	40	IA	0.9588
7	FL	1.0488	24	MI	1.0027	41	WI	0.9549
8	WA	1.0457	25	MA	0.9990	42	MO	0.9517
9	MD	1.0449	26	OR	0.9981	43	MS	0.9494
10	MN	1.0420	27	VA	0.9900	44	GA	0.9484
11	WY	1.0420	28	AZ	0.9831	45	TX	0.9435
12	SD	1.0349	29	OH	0.9830	46	AL	0.9403
13	CO	1.0315	30	SC	0.9820	47	OK	0.9403
14	DE	1.0300	31	ID	0.9784	48	KY	0.9402
15	UT	1.0291	32	NC	0.9776	49	AR	0.9330
16	NY	1.0240	33	KS	0.9729	50	TN	0.9324
17	NV	1.0230	34	DC	0.9702			

Note: To ensure comparability across categories, individual items within each leisure activity are normalized according to their respective major category, as illustrated in Figure 2.1. Following aggregation, the resulting price indexes devolve into price ratios, using the national average as the base period.

From the table, while some discrepancies are noted between the state rankings for our computed "cost of leisure" and the "cost of living" from MERIC[8], the majority of the entries appear consistent. The price indexes for each leisure activity across all states are presented in Table B.7. However, as demonstrated in Table 2a, the limited number of observations for *education* and *religious/civic activities* implies that most of the price indexes for these two leisure activities are biased at the state level. For instance, both Alaska (AK) and Minnesota (MN) possess few observations, or even lack them entirely, leading to unrepresentative results. Following this, we further aggregate the price indexes to explore the disparities in locations (regions) across various periods.

[8]Cost of living states ranking in the third quarter of 2021 by the Missouri Economics Research and Information Center: https://meric.mo.gov/data/cost-living-data-series.

2.3.1 Price Indexes at the Regional Level

The regional price level similarities are, to some extent, comparable to the EM clusters found in the state expenditure PCA plots, as illustrated in Figure 2.6. Our analysis focuses on the four census regions[9]: Midwest, Northeast, South, and West. Instead of previously selecting a single representative leisure product, we now choose the top five states that best represent each region, as listed in Table 2.3. These states serve as the "products" in the basket of leisure activities used to compute the price indexes for specific leisure across different months in each region.

Table 2.3: Five Representative States in the Four Census Regions

Regions (in Alphabetical Order)	Representative States (in Alphabetical Order)
Midwest	IL, IN, MI, MN, OH
Northeast	CT, MA, NJ, NY, PA
South	FL, GA, NC, TX, VA
West	AZ, CA, OR, UT, WA

Note: The five states in each census region with the most consumer observations are considered representative.

The leisure price indexes at the regional level capture price fluctuations across periods for various leisure activities. In this section, we present comparable leisure Laspeyres price indexes[10] spanning 12 months and covering the four census regions. We set the Midwest price level in January as the base period. Our focus will be on leisure activities that exhibit more pronounced fluctuations across periods[11]. As depicted in Figure 2.8, the West region generally has higher price levels than the other regions. The cost of *child care* remains similar for most periods but sees peaks in January and October. Regarding *eating*, the coastal regions exhibit higher prices. In general, the cost of eating escalates during the holiday season. Nevertheless, the holiday price surge is even more pronounced for *hobbies* and *TV*. Price levels for *hobbies* dip in the summer and then rise

[9]The U.S. Census Bureau's map displays the four census regions. Before June 1984, the Midwest Region was referred to as the North Central Region. https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.

[10]There are only subtle differences between the calculated numbers of the four price indexes.

[11]The regional price indexes over 12 months for *gardening & pet care*, *personal care*, *socializing*, and *sleeping* demonstrate relatively flat curves. Only regional price level differences are observed.

steadily through to December. In contrast, *TV* prices, despite experiencing larger fluctuations in all regions, decline continuously through the first three quarters. They hit their lowest point around Thanksgiving, only to spike again in December.

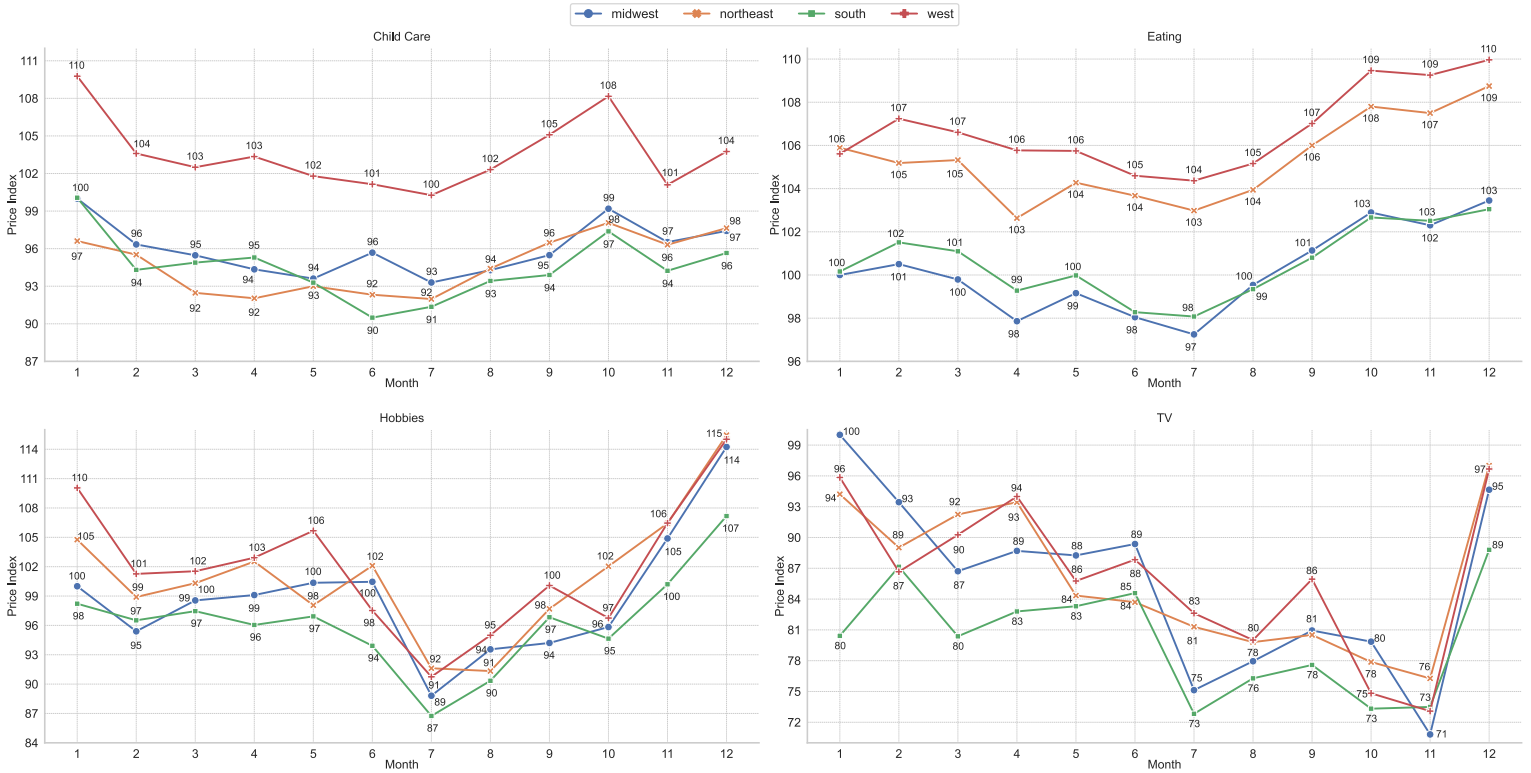

Figure 2.8: Regional Leisure Laspeyres Price Indexes (Base Period: January in the Midwest)

Note: The price index for the four leisure activities is compared at the regional level. Further details on the Laspeyres Price Indexes can be found in Appendix B.2.

2.4 Demand and Income Variation

In this section, we explore demographic disparities in conjunction with income variations employing a nonparametric regression smoother. In the context of location dissimilarity, how do income changes influence the demand for leisure? With limitless desires yet limited resources, households are compelled to make judicious decisions regarding their expenditure. This reveals a direct correlation between both the quantity demanded and the time allocated to wage (w). To gain insights into household spending behavior for different income groups, we direct readers to Figure

B.1 in the appendix which depicts the total number of receipts held by households across different income brackets. Our findings indicate that, on average, a household makes 166 purchases annually. The graph shows that households across different income levels have similar purchasing patterns.

Unlike our previous analysis of price indexes, we now focus on the total expenditure of each household instead of the average expenditure for a fixed set of goods. When comparing the use of aggregated numbers to the average expenditure measurement, we observe greater variation across different income groups when fitting the model. Even though we have already set a minimum household participation rate of 80% to account for household engagement, we still focus on the median to ensure that the model is not swayed by individuals with extreme consumption patterns.

In choosing an estimator, ordinal least squares assumes that the relationship between wage and expenditure, represented by $g(w)$, is linear. However, this assumption may not be valid in many real-world scenarios. As a result, this estimator might yield unrealistic results since the true relationship is likely non-linear. To overcome this limitation, we use a more flexible approach using basis splines (b-splines). This sieve regression leverages a sequence of increasing complexity functions to provide enhanced flexibility and offer a more accurate approximation of the underlying relationship. We select b-splines[12] of order 4, which are suited for capturing non-linear relationships in the data. This piece-wise cubic polynomial ensures the model remains smooth up to the second-order derivatives. This balance between model fit and model complexity results in a more reliable representation of the relationship. To prevent overfitting and avoid generating curves with large oscillations, we constrain the model complexity by setting the degree of freedom for the cubic b-spline to 3. By doing so and eliminating any internal knot, the spline adopts the following degenerate form:

$$g(w) = \sum_{i=0}^{3} \beta_i w^i + \epsilon \tag{2.1}$$

where w^i represents the (truncated) power basis. Without any internal knots, $g(w)$ is also commonly referred to as the Bézier curve. We fit the curve to the household expenditure and time spent on

[12]The spline is of degree 3, resulting in a cubic b-spline.

various leisure activities across a spectrum of wage levels. In Figure 2.9, we showcase our estimated Engel curves for selected leisure activities. The 0.5 quantile estimate is depicted by a solid blue line, which is encompassed by a 95% confidence band, represented by dotted red lines.

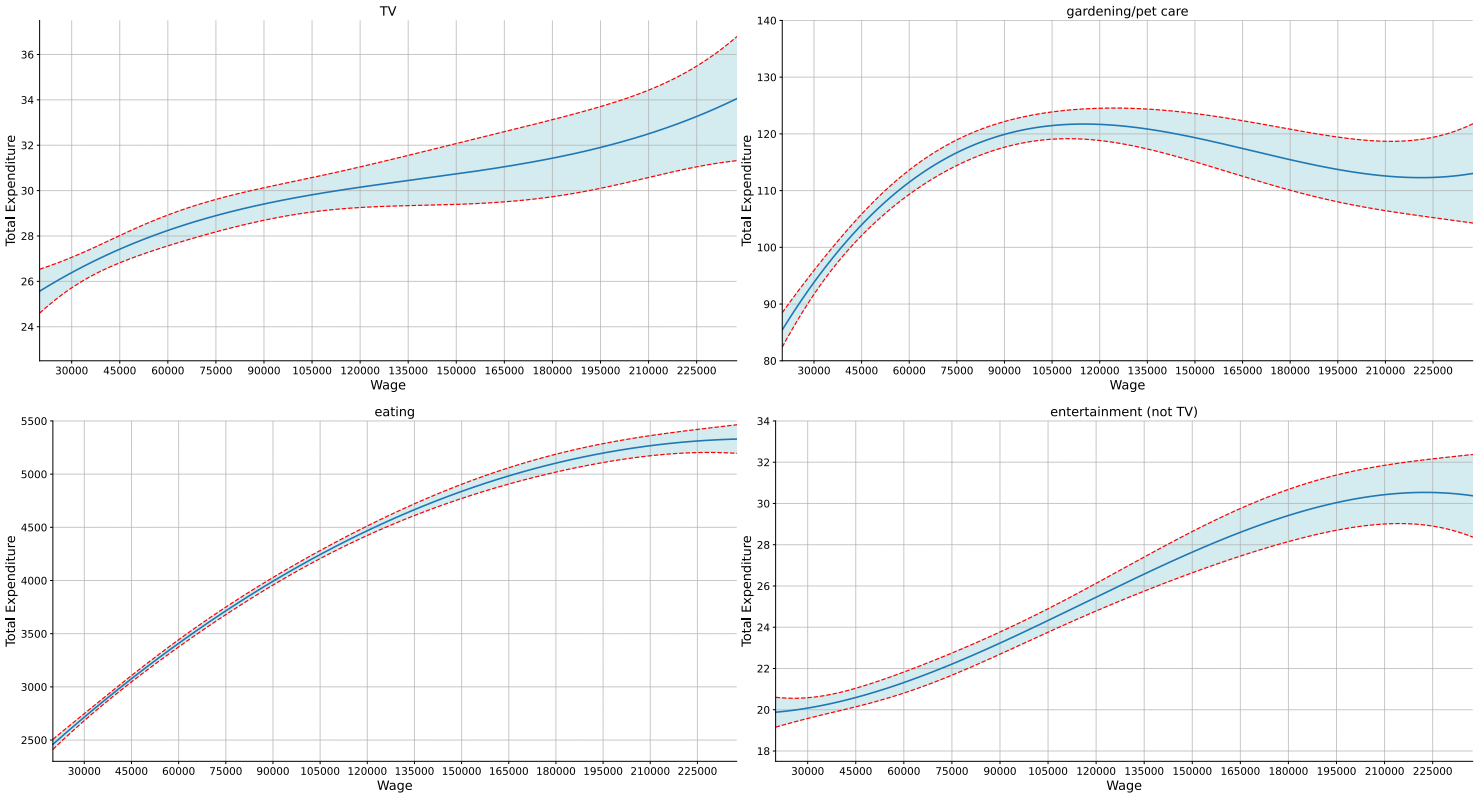

Figure 2.9: Engel Curves

Note: The Engel curves for the leisure categories resemble the shape observed in these instances. The shading represents the 95% confidence interval.

The Engel curves for most leisure activities are concave and resemble the shape of the curve for *eating*, except for *gardening/pet care*. The latter shows only a slight variation in expenditure for households with wages above $60,000. These activities demonstrate a clear positive correlation between increased expenditure and rising wages. It's important to note that the median yearly total expenditure varies substantially across different leisure categories. For example, expenditure on *eating* is markedly higher than for other leisure activities, corroborating the expectation that food consumption remains a top priority for household spending. Furthermore, a significant disparity exists between the "average" and "median" expenditure for certain leisure activities. As an example,

the average total expenditure on *TV* for households on the lower income spectrum is roughly $70, yet the range of its Engel curve (0.5 quantile) lies between $25 and $35. This disparity suggests that even after winsorization during data pre-processing, a step that mitigates potential data misrecordings or number misinterpretations by the app, there remain households making substantial purchases. In parallel, we also analyze households' leisure time allocation across different wage levels, with the findings presented in Figure 2.10.

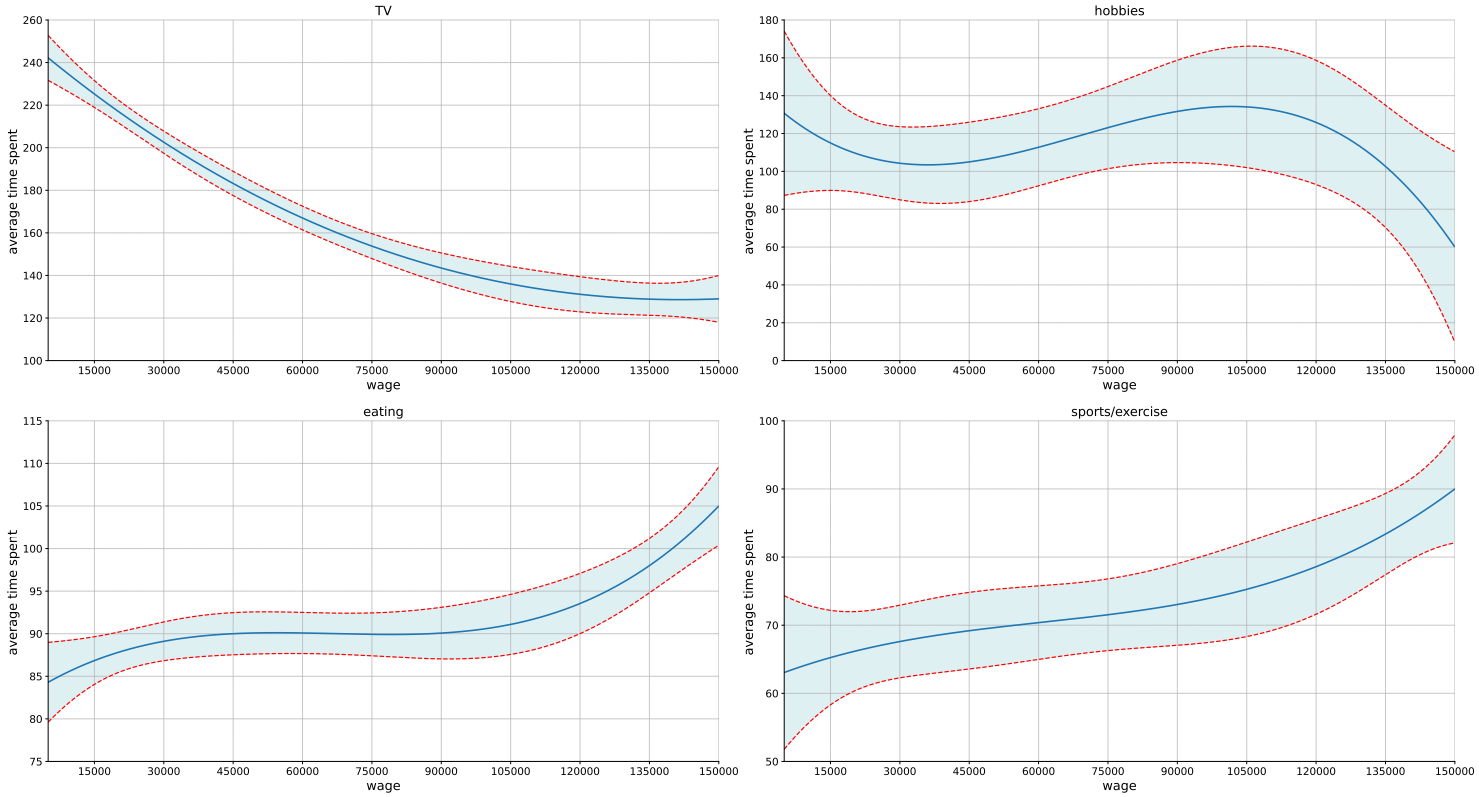

Figure 2.10: Time Spent Variations in Income

Note: Most of the curves except for eating, are flat or portray a negative relationship between time spent and income. The shade represents the 95% confidence interval.

To avoid redundancy, only four representative figures are shown to illustrate the results. However, we will discuss findings from the leisure activities that aren't included. Similar to *TV*, the amount of time spent on *entertainment (not TV)* and *sleeping* decreases as income increases. Households with mid-range incomes spend approximately 80 minutes on *entertainment*. For *sleeping*, there's a noticeable division at the 75K wage mark: the upper half sleeps around 8.5 hours per day, while

those in the lower-income bracket sleep for 9.5 hours. The time dedicated to activities such as *child care*, *gardening/pet care*, *own medical care*, *personal care*, *reading*, and *socializing* remains relatively consistent across all households. However, time allocation for *eating* and *sports/exercise* rises for higher-income households but remains relatively stable for those with mid-range incomes.

2.5 Choice Sensitivity and Substitution Patterns

Leisure consumption demands not just time but frequently entails monetary expenditures. In this section, we examine consumers' choice sensitivity to price fluctuations across different leisure activities by analyzing heterogeneous treatment effects. The elasticity of demand is identified by evaluating the correlation between price variations and corresponding shifts in observed quantity demanded. Our goal is to calculate the conditional average treatment effect (CATE) to determine both the leisure heterogeneous price elasticity and cross-price elasticity, thereby gaining a deeper understanding of consumer behavior.

Using a continuous demand model, we first address the issue of "zero demand." This contrasts with the discrete choice setting, where purchases are probabilistic and choices result from maximizing the likelihood of selecting a particular bundle of characteristics. Here, consumers determine the number of "units" to purchase. In contrast to "no demand" situation, the absence of information on leisure categories that aren't part of the households' consideration sets can lead to potential corner solutions (cases that are on the price axis at zero quantity demanded). In making their purchases, consumers often consider only a subset of available goods, leaving their demand for the remaining goods to zero (Phaneuf, 1997). For example, when choosing recreation sites during a season, some individuals repeatedly visit only a subset of available places. To address this concern and mitigate bias in our estimates, we aggregate the data by the consumer on an annual level. We also regroup activities with related products consumed by fewer than 35% of buyers in a year. Activities including *education*, *religious/civic activities*, and *reading* are merged into the *other* leisure activity.

Compared to the other 11 leisure alternatives, these 3 activities have a significantly higher incidence of zero demand. In addition, we further refine our analysis by controlling for participation by filtering with consumption variety, a detail to be elaborated upon later. The aggregation provides a summary of individuals' average annual consumption of leisure-related goods. For each consumer, we have two data points: the total units consumed and the average price of those products for the year, broken down by leisure categories. This data captures both the quantity of leisure goods demanded and their prices. To account for price variations across products, both the observed price and quantity were demeaned at the *major category* level, adding the product fixed effect to the model before aggregating these product prices to determine an overall price for the leisure category.

Finally, to ensure accurate results, we take into account not only household participation rates but also abnormal consumption behaviors. At the yearly level, we exclude individuals whose average leisure prices and quantity demanded simultaneously fall within either the higher or lower quantiles. This approach also helps minimize the number of buyers who typically purchase just a single or very few products. As the app tracks retail stores and supermarkets, we've noticed that many buyers purchase only a few items during most of their visits[13]. Insufficient variability in purchase quantities obscures the effect of price changes, making it challenging to identify choice sensitivity among these consumers. Regarding app-based reporting participation, recall that the data is collected when consumers scan their receipts using their phones. People are deemed representative if they consistently report their purchases. A low level of participation, like infrequent receipt scanning, can lead to inaccurate reporting and zero demand. Merely setting an 80% participation threshold during data cleaning is no longer sufficient, as households with a limited variety of observed consumed products can cause ambiguity between under-reporting and actual no demand. Therefore, in this sample selection, we have narrowed down our sample to 45,847 consumers who have purchased leisure categories no less than the floor of the mean. By concentrating on more popular leisure categories and engaged participants, the average treatment effect on the treated offers a reliable approximation of the overall average treatment effect.

[13]This will lead to a region densely populated with observations in the bottom-left corner of the price-quantity graph.

In our effort to understand leisure elasticities, we will estimate three effects: the average treatment effect, the heterogeneous treatment effect, and the multiple treatment effect. Utilizing the log-linear demand model, let $\widetilde{Q}$ denote the log quantity, and $\widetilde{P}$ indicates the log price. Let X be a vector of individual-level consumption behavior and demographic attributes, which includes state, gender, education level, marital status, presence of children, employment status, ethnicity, age, the total number of receipts[14], average expenditure per receipt, household size, and average prices of top eight leisure substitute categories[15]. The latter are unique to each household. Let I represent income, while g and m are the nuisance functions. U and V stand for the error terms. The general partial-linear model for estimating heterogeneous treatment effect can be described as follows:

$$\begin{aligned} \widetilde{Q} &= \widetilde{P} \cdot \theta(I) + g(I, X) + U \\ \widetilde{P} &= m(I, X) + V \end{aligned} \tag{2.2}$$

which satisfies the following moment conditions: $\mathbb{E}\big[U|I, X\big] = 0$ and $\mathbb{E}\big[V|I, X\big] = 0$. The nuisance functions are the following:

$$\begin{aligned} g(I, X) &= \mathbb{E}\big[\widetilde{Q}|I, X\big] \\ m(I, X) &= \mathbb{E}\big[\widetilde{P}|I, X\big]. \end{aligned} \tag{2.3}$$

To isolate confounders and obtain the debiased average treatment effect (represented as own-price elasticity, θ, with no interaction with income I), we employ a two-stage cross-fitting[16] approach. This process utilizes K-fold partitioning to reduce overfitting bias. During a single iteration, all observations except for fold k are used to estimate the nuisance functions using standard machine learning (ML) techniques. In the second stage, residuals are computed using the left-out k^{th} fold:

$$\begin{aligned} \widehat{U}_k &= Q_k - \widehat{g}_{-k}(I_k, X_k) \\ \widehat{V}_k &= P_k - \widehat{m}_{-k}(I_k, X_k). \end{aligned} \tag{2.4}$$

[14]The number of receipts indicates the total purchases a household made within a year.
[15]On average, a household consumes products from 9 leisure categories annually.
[16]Cross-fitting guarantees convergence without the necessity for stringent assumptions (Chernozhukov et al., 2018).

In the first stage, the nuisance functions in Equation 2.2 are estimated by ML methods. In the second stage, we derive the leisure own-price elasticity estimate from the residuals in Equation 2.4:

$$\widehat{\theta}^{\,k}_{\text{elasticity}} = \left(\frac{1}{n} \sum_{\{i=1:i\in k\}}^{n} \widehat{V}_i^k \widehat{V}_i^k \right)^{-1} \left(\frac{1}{n} \sum_{\{i=1:i\in k\}}^{n} \widehat{V}_i^k \widehat{U}_i^k \right) \tag{2.5}$$

where n is the number of transactions in the k^{th} fold. Finally, we calculate the cross-fitted elasticity estimates by averaging across all K iterations. For this analysis, we set $K = 2$. The nuisance functions g and m are estimated using 1000 boosted trees, each with a maximum depth of 20 and a learning rate of 0.5.

We chose the gradient boosting method over the random forest based on the fact that the boosting method prunes the tree by comparing the similarity score of the original node with its children. The splitting process stops if the gain is minimal, which helps prevent overfitting. Additionally, the initialization of a boosted tree requires fewer hyperparameters than a random forest, and these exogenous values only affect one tree at the beginning. We present the estimated price elasticity results in Table 2.4.

Table 2.4: Leisure Own-price Elasticity

Leisure Activities	Elasticity	Leisure Activities	Elasticity
child care	-0.211^{***} (0.013)	personal care	-0.117^{***} (0.012)
eating	-0.613^{***} (0.017)	sleeping	-0.039^{***} (0.006)
entertainment (not TV)	-0.064^{***} (0.004)	socializing	-0.204^{***} (0.009)
gardening/pet care	-0.281^{***} (0.011)	sports/exercise	-0.026^{***} (0.005)
hobbies	-0.186^{***} (0.008)	TV	-0.173^{***} (0.006)
own medical care	-0.068^{***} (0.010)	others	-0.058^{***} (0.006)

Note: The price elasticity of demand for leisure is inelastic, meaning the quantity demanded for leisure-related products isn't highly sensitive to price changes. Standard errors are in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The leisure own-price elasticities are inelastic. Especially, the demand for *entertainment (not TV)*, *own medical care*, *sleeping*, *sports/exercise*, and *other* (*education*, *religious/civic activities*, and *reading*) are insensitive to price. The price elasticity of *eating* is 0.61 (absolute value). As the primary component of *eating* is food (as seen in Table B.3), this number can be cross-referenced with other studies. Andreyeva, Long and Brownell (2010) reported that the price elasticities for food and nonalcoholic beverages range between 0.27 and 0.81. In Table 5 of the USDA Economics Research Report (Okrent and Alston, 2012), the price elasticity for cereals and bakery is 0.58, for meat and eggs it's 0.31, dairy is at 0.05, fruits and vegetables at 0.79, and alcohol at 0.71. These elasticities represent the average treatment effect of price on the quantity demanded.

Next, we will examine the elasticity among different income groups. As described in Equation 2.2, we estimate the heterogeneous treatment effect, denoted as $\theta(I)$, at different income levels in two steps. We use double machine learning to estimate the nuisance functions then we fit a causal forest, a generalized random forest method (Athey, Tibshirani and Wager, 2019), to estimate the heterogeneous elasticity with the orthogonalized price and quantity demanded. We specifically focus on the elasticity at three income quantiles: 0.25, 0.50, and 0.75.

Following Wager and Athey (2018) and Athey and Imbens (2019), we construct the forest using a weighted average of 1,000 causal trees. Each tree has a maximum depth of 20 levels. In our setting, the choice of hyper-parameters does not have a significant effect on the elasticity estimates. To prevent overfitting, causal trees are grown with honesty; the data used to construct the tree structure, i.e., to create splits, are separate from the data used to estimate the treatment effect within the leaves. For tree construction, each tree is grown using a randomly selected fraction of the data without replacement. This introduces variation among the regression trees, thereby reducing the risk of overfitting. Furthermore, trees that serve as imperfect predictors are likely to offer better generalizability for the estimated causal effect. By subsampling, the trees aim to partition neighborhoods with similar CATE in the covariate space[17]. They achieve this by recursively splitting

[17] In some regions of the feature space defined by the tree splits, a small price change might have a larger effect on the quantity demanded, while in other regions, the effect might be smaller.

along the three income levels, which results in smaller subgroups. During each split, the chosen income level by the tree aims to maximize the heterogeneity in treatment effects between the leaves. The ultimate goal is to form leaf nodes comprised of observations where the differences in outcomes between treated and untreated units are roughly consistent. The tree ceases to grow when there's insufficient variation in treatment effects[18] within the partitions (Gulen, Jens and Page, 2020).

Considering price as a continuous treatment variable, this ensemble method uses each trained regression tree to predict the change in quantity demanded corresponding to a relative change in price, conditional on different income levels. During the prediction phase, the forest leverages its learned non-linear structure to make predictions. As a test data[19] passes through the forest, each tree provides an estimate of the elasticity for that observation. Note that this prediction arises not from explicitly modeling the log-log relationships but from the localized[20] average treatment effects associated with a specific region of the feature space.

From the grown tree, we then calculate the weights for each tree to form the forest as a weighting function. The final elasticity is then a weighted average of all these estimates, based on the weights assigned to each tree. The weights $\alpha_i(x)$ is a data-adaptive kernel that measures the frequency with which training data falls into the same leaf as the test data (Athey and Wager, 2019). The elasticity for a target income quantile $x \in I$ is determined by solving the following moment condition, which is represented by the score function ψ:

$$\sum_{i=1}^{n} \alpha_i(x)\psi(\widetilde{Q}, \widetilde{P}; \theta, \eta) = \sum_{i=1}^{n} \alpha_i(x)\Big[\widetilde{Q}_i - g_x\left(I_i, X_i\right) - \theta(\widetilde{P}_i - m_x(I_i, X_i))\Big]\Big[\widetilde{P}_i - m_x(I_i, X_i)\Big] = 0$$

which is equivalent to minimizing the following squared loss:

$$\widehat{\theta}(x) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \alpha_i(x)\Big[\widetilde{Q}_i - g_x\left(I_i, X_i\right) - \theta\big(\widetilde{P}_i - m_x(I_i, X_i)\big)\Big]^2. \tag{2.6}$$

[18] For the units within the subgroup, both $\text{Var}(Y|T=1)$ and $\text{Var}(Y|T=0)$ are small, meaning individual treatment effects closely align with the average treatment effect.

[19] The observation is a price and quantity demanded pair.

[20] The model, during training, internalizes the structure of that region, drawing from similar observations in the training data.

The $\eta = (g, m)$ is the nuisance parameter and similar to the double machine learning setup, the local estimates (g_x, m_x) are defined as $g_x = \mathbb{E}\left[\widetilde{Q}|I, X\right]$ and $m_x = \mathbb{E}\left[\widetilde{P}|I, X\right]$. The score function satisfies the Neyman orthogonality condition (Neyman, 1959). The moment condition is insensitive to the value of the nuisance parameters. These parameters are estimated similarly, using 1,000 boosted trees, each with a maximum depth of 20, and a learning rate of 0.5. The heterogeneous elasticities are shown in Figure 2.5.

Table 2.5: Heterogeneous Income Elasticity

Leisure Activities	0.25 Quantile ($52,500)	0.50 Quantile ($90,000)	0.75 Quantile ($168,750)
child care	-0.151^{***}	-0.236^{***}	-0.216^{**}
	(0.039)	(0.056)	(0.080)
eating	-0.568^{***}	-0.646^{***}	-0.778^{***}
	(0.038)	(0.043)	(0.055)
entertainment (not TV)	-0.054^{***}	-0.056^{***}	-0.042^{**}
	(0.008)	(0.008)	(0.018)
gardening/pet care	-0.283^{***}	-0.245^{***}	-0.174^{***}
	(0.017)	(0.016)	(0.035)
hobbies	-0.155^{***}	-0.145^{***}	-0.236^{***}
	(0.022)	(0.025)	(0.042)
own medical care	-0.060^{***}	-0.047^{**}	-0.023
	(0.011)	(0.014)	(0.035)
personal care	-0.113^{***}	-0.132^{***}	-0.192^{***}
	(0.019)	(0.026)	(0.029)
sleeping	-0.037^{***}	-0.039^{***}	-0.069^{**}
	(0.010)	(0.008)	(0.022)
socializing	-0.208^{***}	-0.196^{***}	-0.129^{**}
	(0.012)	(0.012)	(0.046)
sports/exercise	-0.018	-0.033^{***}	-0.050^{**}
	(0.009)	(0.008)	(0.016)
TV	-0.159^{***}	-0.157^{***}	-0.202^{***}
	(0.013)	(0.010)	(0.030)
others	-0.046^{***}	-0.054^{***}	-0.134^{***}
	(0.007)	(0.011)	(0.037)

Note: Income range: $20,000 to $250,000. Standard errors in parentheses. The point estimate for *sports/exercise*, corresponding to an income of $52,500, has a P-value of 0.051. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The heterogeneous income elasticities are inelastic. Though there are some fluctuations between the second and third quartiles, we observe an increasing price sensitivity with the level of income

for half of the leisure activities. However, the order is reversed for *gardening/pet care*, *own medical care*, and *socializing*.

To explore this in more detail, we excluded the *others* category, concentrating on the substitution patterns households exhibit when purchasing goods from the remaining 11 leisure categories. By assessing how the demand for one leisure-relative product is influenced by the price changes of other leisure products, the results illustrate the sensitivity of the quantity demanded for leisure activities represented on the vertical axis to percentage changes in the prices of other leisure activities depicted on the horizontal axis, as shown in Figure 2.11. These cross-price elasticities were derived by extending the double machine learning method to accommodate multiple treatment effects, incorporating various leisure activities from the horizontal axis. By substituting $\widetilde{Q}$ and $\widetilde{P}$ with vectors representing multiple quantities and prices of leisure activities, we can fit the substitution patterns while considering other alternative activities. The cross-price substitution matrix indicates that while many of the leisure activities are unrelated, some exhibit a wide range of effects.

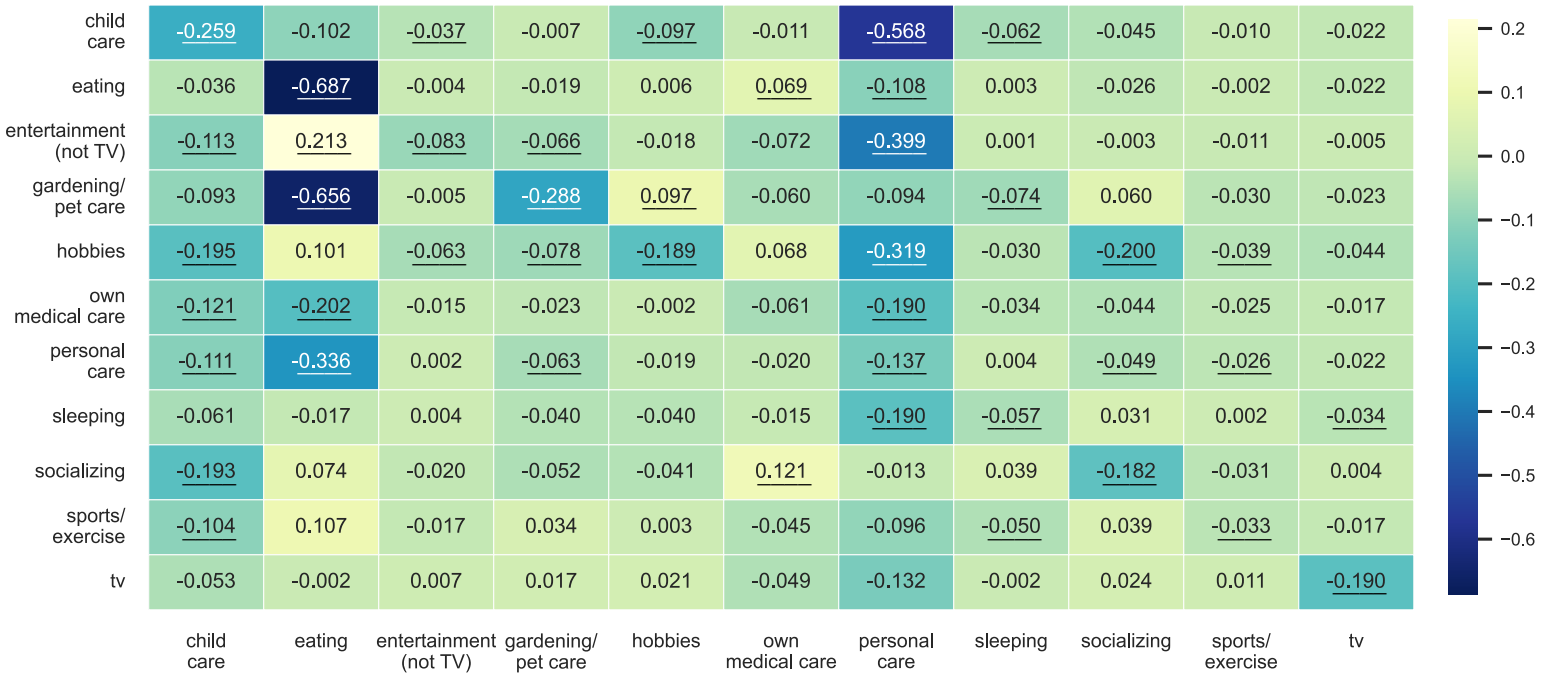

Figure 2.11: Cross-price Elasticity of Demand

Note: The prices of the leisure activities represented on the horizontal axis serve as the treatments. Only households whose consumption encompasses all 11 leisure categories are included in the estimation. Statistically significant estimates are underlined, and the intensity of the fill color corresponds to the magnitude of the cross-price elasticities.

Price changes in both *child care* and *personal care* impact nearly all leisure categories. As complementary goods, the prices of products related to *child care*, compared to *personal care*, exert a relatively minor effect on products from other categories. A price shift in *personal care* significantly affects the quantity demanded for *child care*, *entertainment (not TV)*, and *hobbies* related commodities. The most notable cross-price effect is the complementary relationship between *eating* and *gardening/pet care*. *eating* also acts as a complement for products related to *own medical care* and *personal care*. Conversely, *eating* serves as a substitute for *entertainment (not TV)*.

2.6 Conclusion

This study carries out a comprehensive empirical investigation into households' leisure choices, taking into account both the time invested and the associated costs. Despite the moderate attention that leisure has received in past research, most studies have incorporated only partial measurements. Our work aims to bridge this gap in the field by offering an in-depth comprehensive analysis that takes into account both time spent and the cost of leisure. The classifications employed in this study facilitate the incorporation of a wide range of non-productivity-oriented activities and yield results that are comparable to those found in the literature.

Focusing on the disparities in demographics and geographical locations, we conduct both integrated and disaggregated analyses. In examining leisure holistically, we showcase geographic clusters of households' choice of leisure across different states and visualize the similarities in time spent and expenditure expenditures on leisure activities. Subsequently, we develop a leisure price index to elaborate on the cost of leisure and investigate seasonality across the four census regions. Diving into demographic discrepancies, we estimate nonparametric "Engel curves" to examine the relationship between income and our two key leisure measurements. Lastly, we investigate household choice sensitivity by estimating the fully heterogeneous own-price and cross-price elasticities, capturing the substitution patterns across different leisure categories.

Chapter 3

Refugee Migration During the 2022 Russia-Ukraine War: Evidence from Queer Social Network Users

This chapter examines the refugee migration during the 2022 Russia-Ukraine War. Using queer social network data, we develop an interactive visualization map and investigate the city-level factors influencing user choice of migration destinations. Moreover, we utilize a count model to evaluate how a city's cost of living and geographical factors impact the number of users residing in cities and the movements between them. Lastly, we depict the migration patterns of refugee users.

3.1 Introduction

On February 24th, 2022, the ongoing tension between Russia and Ukraine developed into a full-scale war. The conflict between the two countries has roots dating back to 2014 when Moscow-backed separatists launched a rebellion in the Donbas[1] region. A summary of events can be found in

[1]The Dobas region is most commonly defined as the Donetsk and Luhansk regions.

Korovkin and Makarin (2023). The conflict between the two countries has had a profound impact on the global economy and the residents of the affected regions. The primary objective of this study is to gain insights into refugee migration. We utilize novel queer social network data to create an interactive visualization map and examine the factors that influence migration decisions and uncover migration patterns. We offer a brief overview of the war during the time frame that is the focus of our study. Figure 3.1 is a regional map identifying certain cities discussed in this research.

Figure 3.1: Regional Map

Note: The map indicates the neighboring countries of Ukraine and several cities referenced in this study.

In January 2022, Russian troops arrived in Belarus and began a 10-day military exercise on February 10th. Starting from February 17th, conflicts arose in the separatist regions of eastern Ukraine. Following the outbreak of war, Russian military forces attacked the Ukrainian capital of Kyiv, as well as Mariupol and Kharkiv, the country's second-largest city.

In the first week of the military assault, refugees, mostly women, children, and the elderly, poured into neighboring countries, with long queues of cars at the borders. Men of fighting age were

largely prohibited from leaving the country. In early March, both sides agreed to open humanitarian corridors for the evacuation of civilians. On March 5th, the Russian armed forces announced a ceasefire to allow approximately 200,000 civilians to evacuate Mariupol. On March 29th, Moscow announced the withdrawal of forces from Kyiv and other areas, turning to focus on the South and East of Ukraine. In the following months, Russia launched attacks in Kramatorsk[2], Kremenchuk[3], and Lysychansk. In August, Ukrainian forces launched a counteroffensive in Kherson and retook the northeast of the Kharkiv region in late September. On November 9th, Russia announced a pullback from the city of Kherson.

Numerous studies have been conducted to examine the impact of the war on the global economy, but only a few have focused on the refugees. The majority of information regarding refugee migration is organized and presented through the United Nations High Commissioner for Refugees (UNHCR) data portal. Kalogiannidis et al. (2022) demonstrate how Russia's aggressive actions and the subsequent international sanctions have impacted the European economy, particularly the petroleum and gas market. They find that through the immediate effects on the energy supply, the sanctions have not only impacted Russia's economy but have also affected European economies equally. Khudaykulova, Yuanqiong and Khudaykulov (2022) examines the economic consequences of war and explores the possible effects that the conflict between Russia and Ukraine could have on both the local and global economies. Boungou and Yatié (2022) provide the first empirical evidence of the effect of the impact of the Ukraine-Russia war on global stock market returns. They use daily stock market returns from a pool of 94 nations to investigate how global stock market indexes reacted to the ongoing conflict between Ukraine and Russia. Their findings suggest that the response of worldwide stock markets was notably influenced in the initial two weeks, but gradually weakened in the subsequent weeks.

Morariu (2022) compares the migration phenomenon within the European Union before and during the war and analyzes the reasons for the migration of Ukrainians to Poland, Russia, and

[2]On April 9th, a Russian missile strike on a train station in Kramatorsk.
[3]On June 27th, a shopping mall in the city of Kremenchuk (southeast of Kyiv) was struck by missle.

Romania through a qualitative approach. Moreover, in 2022, Poland emerged as the primary destination for war refugees from Ukraine due to geographic factors and history of migration (Duszczyk and Kaczmarczyk, 2022). They also argue that the arrival of refugees in Poland has presented challenges for public services and institutions. This influx of refugees has transformed Poland from an emigration country to an immigration one without an intermediate phase. Furthermore, the authors provide forecasts of future immigrant populations and highlight the obstacles faced by both Poland and the refugees themselves. Lloyd and Sirkeci (2022) adopt the perspective of migration conflict to investigate the departure of individuals from Ukraine during the war. They contend that the welcoming attitude of several European nations has played a crucial role, while Ukraine has historically experienced insecurity that has already fueled large waves of significant emigration from the country. Dumont and Lauren (2022) noted that adult refugees of working age who migrate will seek to work during their stay in the new country, but they will encounter particular challenges in integrating into the labor market, compared to other migrants. According to their calculations, the refugee influx is expected to have roughly twice the impact on the labor force as the number of refugees who entered the European Union between 2014 and 2017.

Assessing migration is a complex task that poses significant challenges. International organizations rely on country border statistics or information from refugee stations to gain insight into the number of refugees. Unfortunately, these types of data do not provide detailed information about migration movements, as they lack the necessary level of granularity. Our research employs novel social network data from Hornet to gain an understanding of the overall refugee migration by examining the migration of individuals within the network.

Founded in 2011, Hornet is a feed-first social media platform primarily used by members of the LGBTQIA+ community. The company is based in Los Angeles, US, and is popular worldwide, with core markets representing 85% of all users located in Thailand, Turkey, Russia, Brazil, Indonesia, and Taiwan. Hornet handles its users' activities within the app by assigning a unique, non-personally identifiable user code to each user. This allows for tracking of user activity and location without

compromising user privacy. This research was conducted while ensuring the highest privacy standards for Hornet's users.

Although our study is limited to a specific group of users from a social network app, we believe our findings provide a reasonable approximation. Hornet is a well-established brand in Russia and Ukraine and the only queer social network available in those markets. Over the past decade, the app has had an average of 240,000 active users per month. Based on Boyon (2021) assuming that 1% of the 20.4 million male population in Ukraine and 66.8 million in Russia[4] identifies as queer, Hornet's market penetration would be around 28%. However, this estimate does not take into account individuals who lack internet access or do not use social media, nor does it include those outside our target demographic, such as individuals under the age of 18, which could further increase Hornet's penetration rate.

The organization of this chapter is as follows: the next section will cover the process of data exploration and wrangling, as well as sample selection and an interactive visualization map. In Section 3.3, we introduce a choice model and a count model designed to assess the impact of city characteristics on migration destination preferences and the number of users migrating, considering both static city-level counts and average user flows between cities. After that, we estimate the user migration pattern. Finally, Section 3.4 will provide the conclusion.

3.2 Data and Sample Selection

The information obtained from the LGBTQIA+ social network presents a valuable chance to gain an understanding of the migration patterns of refugees. For privacy concerns, the original user log files from Hornet are aggregated to the city level after sampling, and all personal identifying information is removed from the data. Additionally, to mitigate security issues, we put in place supplementary filters that eliminate small cities, which prohibits the identification of specific refugee

[4]The population estimates are obtained from The World Bank: "*Population, male - Russian Federation*" (2022), https://data.worldbank.org/indicator/SP.POP.TOTL.MA.IN?locations=RU.

movements through our interactive map visualization. The user log files are processed by the data funnel depicted below in Figure 3.2. Compared to the overall refugee population, the data we have is selective. Nevertheless, we concentrate on the average migration of the LGBTQIA+ community and aim to offer some insights into the broader migration patterns.

Figure 3.2: Data Processing Funnel

Note: The user log files from the Hornet app undergo six stages of data aggregation and sample selection.

The sampled aggregated data spans from June 1st, 2021 to November 11th, 2022. The data undergoes six stages of processing before proceeding with the analysis. The participants in the study had to have been active in Russia or Ukraine at least once during the time frame between when they created their accounts and the end of our sampling period. The data covers more than 0.7 million unique users from over 200 countries and 18,000 cities. Next, it is necessary to identify the nationalities of the observed individuals.

3.2.1 User Groups and User Log Imputation

The app users in the sample are categorized into three classes: Ukrainians, Russians, and Foreigners. However, in the absence of a declared nationality by the user, we attempt to infer their group membership through their recorded activities. We choose the classification period from June 1th 2021 to November 30th 2021, purposely selected to exclude the holiday season to avoid any abnormal user behavior. As a consequence, users who joined the network in December 2021 or later will not be classified. If a user is new to the sample and their residency has not been established, they will be categorized as unclassified and later removed from the sample.

The class assignment is established based on the users' residency ratio. The residency is determined by the region where the majority (over 50%) of the user's activities occurred during the classification period. Following the classification, our data comprises 111,090 (14.53%) Ukrainians, 552,016 (72.20%) Russians, and 101,459 (13.27%) Foreigners with a total of almost 70 million logs. However, it is common for users to use the app repeatedly in a single day. To account for this, we eliminate individual duplicated daily log entries and only consider the most latest instance to establish the user's city location for that day. The migration pattern is determined by monitoring alterations in the number of users in each city, however, these counts are easily influenced by app usage since users may not open the app every day. This leads to a significant discrepancy between the recorded number of users and the actual number of residents in a particular city for a specific day. To overcome this challenge, a 30-day buffer period is introduced before considering a user as churned. As we monitor the count of users in various cities over time to sketch the migration pattern, we have noticed irregular fluctuations in the numbers due to the inherent limitations of the data collection process. This phenomenon is a result of missing city information in the log files caused by factors such as poor GPS signal or intentional blocking by users, as well as instances where users may not utilize the app daily or while they are in the process of migrating.

Nevertheless, even though the users may not show up in the app data for a certain day, they are still residing in the city. Therefore, in the case of a missing log, we presume that the user continues

to reside in the same city for the subsequent 30 days unless we obtain a new log entry. If there are still no updates on the user after that period, we consider them to have churned until they eventually restart the app at a later date. The log file imputation not only rationalizes the data, resulting in smoother changes in the count of users in each city, but it also backfills the missing city information for some users. Eventually, those users whose city and country locations cannot be recovered are dropped from the data. The cities are geocoded for data visualization.

3.2.2 Geocoding and Migration Visualization

Explore refugee migration patterns through our interactive map, showcasing the movement of refugees in three categories, available at haochehsu.com/migration/map.

By augmenting the Hornet data, we obtain coordinates for each unique city-country pair by leveraging both OpenStreetMap (OSM), a geographic database, and the Google Maps API, as illustrated in Figure C.1. Our visualization emphasizes cities within 62 European countries and their neighboring countries. We further highlight the alterations in logarithmic user counts across these cities before and during the war, as presented from Figure 3.3 to 3.5. The number of users is represented by the size of the spikes, with each group's spike[5] size determined independently.

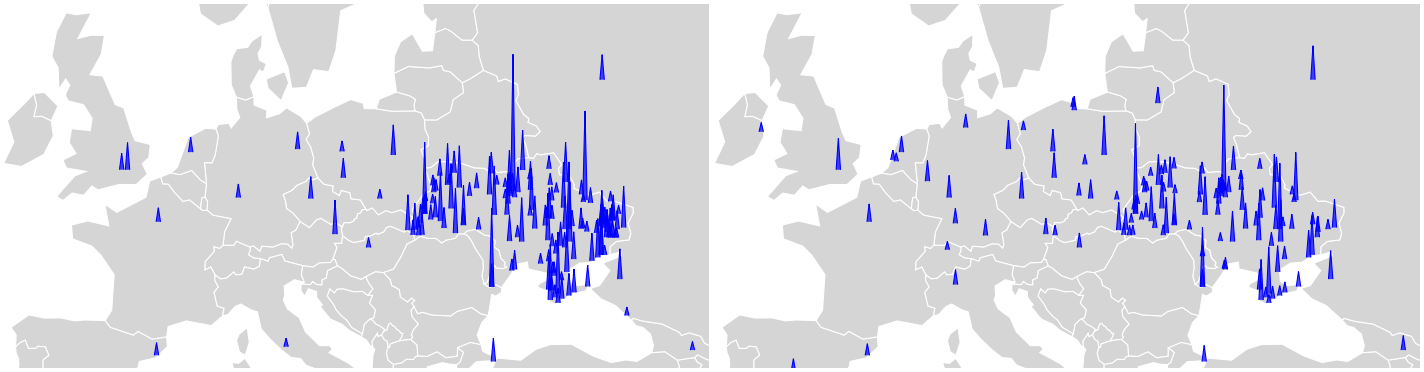

Figure 3.3: Distribution of Ukrainian Refugee Users

Note: The panels depict the distribution in October 2021 (left) and May 2022 (right). A major shift towards the left region can be seen with the decline in the number of users in the capital and major cities located at the right border of Ukraine. Notable migration influxes are present in Poland, Russia, Hungary, Moldova, Slovakia, Czechia, and Germany.

[5]The size of the spikes in the map is determined independently within each group.

To offer context for the magnitude of the spikes, the spike in Paris depicted in the right panel of Figure 3.3 corresponds to 30 Ukrainians, while Kyiv had 20,203 users in October 2021 and 5,811 users in May 2022. Refugees in Ukraine are leaving the capital, Kyiv, and cities in the East including Kharkiv, Luhansk, Donetsk, and Mariupol, and moving to Western cities such as Lviv, Mukachevo, and Uzhhorod, and cities in Western European countries. In Figure 3.3, we can roughly identify the refugees' migration destination cities from the spikes including Moscow (Russia), Rzeszów, Gdańsk, Katowice, Wrocław, Kraków, and Warsaw (Poland), Hamburg, Berlin, Munich, Stuttgart, Frankfurt, and Düsseldorf (Germany), Prague (Czechia), London (United Kingdom), Milan (Italy), (Paris) France, Bratislava (Slovakia), Vienna (Austria), Vilnius (Lithuania), Zürich (Switzerland), and Budapest (Hungary). Additionally, a significant number of refugees opt for more distant destinations, such as the United States, beyond just neighboring countries (UNHCR, 2022).

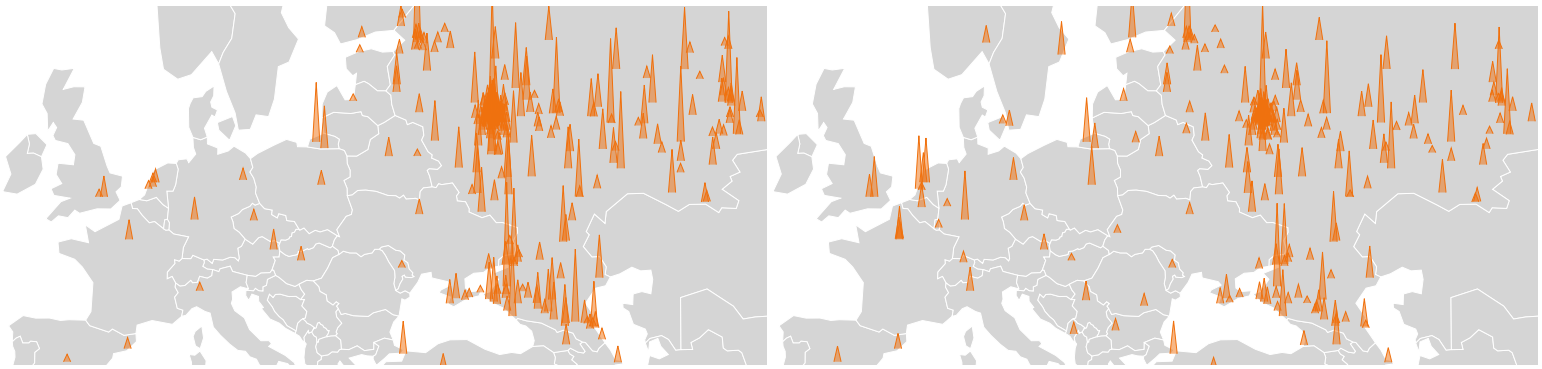

Figure 3.4: Distribution of Russian Refugee Users

Note: Noticeable migrations are captured from Russia to Western European countries. The panels depict the distribution in October 2021 (left) and May 2022 (right). The spike in Berlin depicted in the right panel represents 79 Russians.

Likewise, Russian refugees are relocating to major cities in Western Europe such as Helsinki (Finland), Stockholm (Sweden), Oslo (Norway), Warsaw (Poland), Bucharest (Romania), Belgrade (Serbia), Brussels (Belgium), Amsterdam, Rotterdam (Netherlands), London, Berlin, Frankfurt, Milan, and Paris. On the contrary, as demonstrated in Figure 3.5, a portion of the foreign users in Kyiv and the eastern cities of Ukraine depart, while the numbers remain relatively unchanged in the other cities. Nevertheless, the data on city users still needs further adjustments before we can draw any conclusions about the factors that influence migration and the selection of a destination.

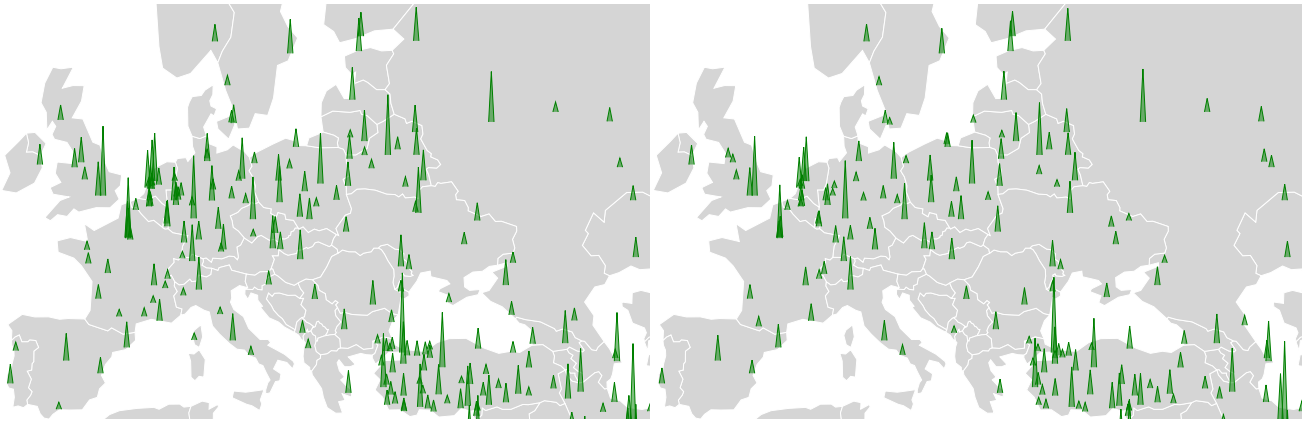

Figure 3.5: Distribution of Foreign Refugee Users

Note: The panels depict the distribution in October 2021 (left) and May 2022 (right). There are no significant movements of foreigners in Ukraine and Russia. The Kyiv spike shown in the right panel of the figure represents 166 foreigners.

3.2.3 Churn Control

The social network comprises heterogeneous users who form connections and friendships with each other. Some users are highly engaged with the app and heavily rely on their network connections. In contrast, others may only join the network temporarily out of interest but quickly lose interest and leave the app due to boredom. With our implementation of residency classification, we can address the issue of continuously adding new people and the sudden influx of a large number of new users due to certain activities or approaching holiday periods. As a result, we also reduced the duration of our sample period to the start of 2022. This adjustment mitigates the influence of substantial holiday travel that occurred prior to the war in our analysis, allowing for a more stable city user count by tracking the movement of a consistent group of individuals over time.

To ensure that the counts are more representative, it is important to avoid having significant fluctuations in the numbers due to unreliable user participation. To address this, based on the user active ratio presented in Figure C.2, we introduce a user participation rate filter that removes users with an app usage rate of less than 90%. There may be worries about the impact of selecting a specific sample, especially under the participation rate filter, which could further limit an already selective group and result in the movements of the remaining users not being a referable reflection

of the overall migration patterns of refugees. Nevertheless, with the help of previous user log imputation, we can still maintain a sizable and consistent group of app users in our sample after implementing churn control, without overly restricting their natural app usage behavior. Figure 3.6 below illustrates the composition of the final user base for each of the three refugee groups.

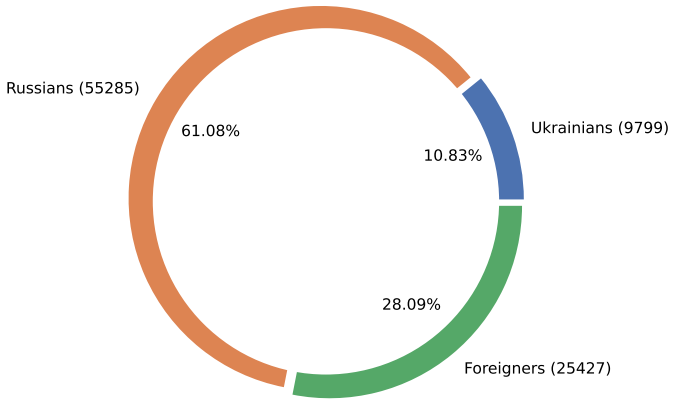

Figure 3.6: Group Composition in the Final User Base

Note: The majority of the users are from Russia. These figures represent the number of consistent app users for each of the three groups, with the actual total number of users shown in parentheses.

Our primary objective is to ensure that our sample consists of only consistent users and to exclude users who only use the app for a short time, as their presence can compromise the accuracy of our analysis, which relies on the changes in the number of app users in each city to track migration. Users who typically only open the app every few days or weeks are included in our sample based on our assumption that the user log can be extended up to 30 days. Consequently, following the imputation process and based on our data observations, we can consider the majority of users as daily users, and the participation filter will only exclude users who have churned before our sample period ends. Table 3.1 summarizes an overview of the changes in the number of users at each stage of the selection process.

While our analysis relies on aggregated city-level data, implementing churn control is advantageous for filtering out irregular users; however, it does result in a significant loss of data. However, this process yields a more accurate characterization of migration, based on actual user counts. The daily active users of the three groups are shown in Figure C.3 to C.5. Our next step is to

apply choice-based and count-based models to study migration patterns and destinations, and to understand the factors that play a role in refugees' decisions on migration locations.

Table 3.1: Summary Statistics of Social Network Users at Different Stages

Stages	Number of Users	Number of Countries	Number of Cities
Preprocessing	770625	222	18628
Cleaning	470394	220	16629
Imputation	464847	192	16629
Geocoding	464828	176	12130
Churn control	90511	170	9927

Note: The table illustrates the number of selected users at each stage. To safeguard privacy and security, we rely on aggregated city-level user counts to infer user migration. Although a significant number of users are excluded in the final selection stage, it allowed us to create a more precise depiction of changes in user counts over time.

3.3 Refugee Migration

Migration patterns are represented by the number of daily app users in each city. We also monitor user relocation from one city to another to better understand migration patterns. Following the outbreak of war, several neighboring countries established refugee camps to offer assistance and provide humanitarian aid. Refugees can either locate the nearest shelter or choose a destination based on where acquaintances or family members reside.

This section aims to examine the economic factors of destination cities that may affect refugees. Although the war officially began on February 24th, we use February 22nd as the starting date in our study to account for the pre-war preparations. Additionally, conflicts had already erupted in the eastern region of Ukraine before the war, and people had started to migrate as the conflict unfolded. First, we provide an overview of the migration flow, followed by an estimation of the average migration patterns among three distinct user groups. Then we analyze the contributing factors from a choice probability perspective and through the lens of an user event-based approach.

We employ a large transition matrix to capture the combined movements of individuals across various cities. To gain insight into the changes in the migration flow induced by the war, we examine

the average of 50 days before and following the war. The migration flow is presented in Figure 3.2 and 3.3. In particular, we show the average daily outflow of Ukrainians and Russians from their capital and their top 15 migration destination cities.

Table 3.2: Average Daily Migration Flow of Ukrainian Refugee Users from Kyiv

50 Days Before War		50 Days After War	
Destination City	Average Flow	Destination City	Average Flow
Staying at Kyiv	5820	*Staying at Kyiv*	3256
Lviv	Ukraine	119	Lviv	Ukraine	232
Kharkiv	Ukraine	73	Dnipro	Ukraine	70
Dnipro	Ukraine	70	Odesa	Ukraine	21
Odesa	Ukraine	45	Vinnytsia	Ukraine	16
Zaporizhzhia	Ukraine	23	Kharkiv	Ukraine	12
Vinnytsia	Ukraine	19	London	United Kingdom	5
Kryvyi Rih	Ukraine	14	Kryvyi Rih	Ukraine	5
Mykolaiv	Ukraine	11	Zaporizhzhia	Ukraine	5
London	United Kingdom	7	Frankfurt	Germany	3
Frankfurt	Germany	4	Mykolaiv	Ukraine	3
Warsaw	Poland	3	Warsaw	Poland	2
Istanbul	Turkey	2	Chicago	United States	1
Barcelona	Spain	1	Prague	Czechia	1
Paris	France	1	New York	United States	1

Note: The numbers indicate the average flow from Kyiv to various cities during 50 days before and after the war.

Table 3.3: Average Daily Migration Flow of Russian Refugee Users from Moscow

50 Days Before War		50 Days After War	
Destination City	Average Flow	Destination City	Average Flow
Staying at Moscow	19868	*Staying at Moscow*	19795
Kazan	Russia	208	Saint Petersburg	Russia	156
Krasnodar	Russia	169	Frankfurt	Germany	144
Saint Petersburg	Russia	117	Krasnodar	Russia	81
Nizhny Novgorod	Russia	55	Nizhny Novgorod	Russia	69
Ufa	Russia	34	Warsaw	Poland	57
Frankfurt	Germany	31	Amsterdam	Netherlands	46
Yaroslavl	Russia	25	Stockholm	Sweden	39
Voronezh	Russia	20	Yekaterinburg	Russia	35
Yekaterinburg	Russia	18	Paris	France	35
Volgograd	Russia	18	London	United Kingdom	34
Samara	Russia	16	Helsinki	Finland	31
Warsaw	Poland	15	Kazan	Russia	28
Irkutsk	Russia	12	Yaroslavl	Russia	24
London	United Kingdom	12	Volgograd	Russia	23

Note: The data illustrates the average migration flow of Russians from Moscow (Russia) to different locations.

We monitor the day-to-day movements of refugee users from Ukraine and Russia. All selected users were residing in Kyiv or Moscow the previous day. Then we compute the average number of users present at each destination for 50 days before and after the outbreak of war. The migration flow provides a glimpse of where people are seeking refuge. The ongoing conflict in Ukraine is causing residents to evacuate from urban areas that are near the attacks including eastern cities and places near the border.

After the war started, many individuals fled from Kyiv to Lviv and refrained from traveling to major cities in the eastern region, like Kryvyi Rih and Mykolaiv. Nevertheless, many people have decided to remain in Ukraine. This is further supported by our interactive map, which demonstrates a sharp decline in the number of users in Kyiv during March, followed by a gradual return in May, and a rise in user numbers in other western cities throughout Ukraine.

The migration flow numbers indicate that Russian users also tend to migrate from Russia toward Western Europe, with notably fewer movements from Moscow. Before the war, people were traveling within Russia, but following the outbreak of the war, they started leaving Russia for countries such as Germany, Poland, Turkey, Netherlands, Sweden, France, the United Kingdom, and Finland. Next, we will concentrate on factor analysis.

3.3.1 City Preferences and Factor Analysis of Refugee User Counts

When seeking refuge, the ability to ensure a safe and secure environment and to maintain an adequate standard of living are crucial factors influencing the choice of migration destination. Beyond countries with established refugee stations, a city's cost of living and other economic factors significantly influence how individuals select their destination. To address these concerns, we incorporate worldwide city-level attributes obtained from a third-party database.

The Euromonitor data comprises information regarding the demographics, price indexes, infrastructure-related, and geographical characteristics of the cities. Due to the war taking place in early 2022, we have chosen to utilize the 2021 Euromonitor data in our analysis. It was necessary to

manually match the cities to resolve the inconsistency in the city names present in both Euromonitor and Hornet data. Ultimately, we successfully matched 953 cities worldwide in the final dataset. Additionally, to narrow our focus to cities where users have traveled, we computed the total count of unique visitors for each city during the sampling period, and we have limited the scope of the cities to those with more than 100 accumulated visitors. This also helps ensure greater data privacy and reduces the risk of identifying the exact locations of individual users. The study will examine the three groups separately. After the data goes through the processing funnel illustrated in Figure 3.2, the numbers of distinct cities[6] belonging to each group are 44 for Ukrainians, 126 for Russians, and 210 for Foreigners. This is the number of averaging cities for the count model, but in the choice model, the inclusion of an "outside option" (staying in cities in Ukraine) will result in a slight reduction in the number of cities selected.

Before the analysis, the added attributes are refined. We compute the correlations between the city attributes and eliminate redundant characteristics, as well as those that have exceptionally high correlations. Fifteen variables have been chosen, and they are displayed in Figure C.6. We then specify the models to examine the perspectives of user count from the choice and actual number standpoints. By examining the user count from both angles, we gain a comprehensive understanding of the factors that influence the choice of migration destination and the actual migration patterns.

3.3.1.1 Multinomial Logit Model

Throughout the selected sample range for analysis, spanning from January 1st, 2021 to November 5th, 2021, the choice model prioritized the period following the outbreak of the war on February 22nd, 2021 (256 days). The model that analyze actual user counts, discussed in the following section, utilizes the entire time range. Each day during the selected period[7] is considered a separate market. The aggregated city-level counts within each market consist of distinct individuals[8] with a

[6]The consideration set consists of the cities that the users have visited during our sampling period.

[7]Our analysis focuses on the city choices made by users during the war, with the selected time frame for the choice model spanning from February 22nd to November 5th.

[8]By our construction, there are no duplicate user logins on the same day and almost every user is present every day.

post-imputation participation rate of at least 90%. During the war, the utility obtained by a user i traveling to cities located outside of Ukraine $j = 1, ..., J$ on day t can be expressed as follows:

$$u_{ijt} = x_j\beta + \xi_{jt} + \epsilon_{ijt} \tag{3.1}$$

where the vector $x_j \in \mathbb{R}^K$ denotes the time-invariant 2021 city j attributes and a constant. Our primary focus is on the price indexes and city attributes that are closely related to migration decisions. We are also controlling for GDP and inflation. The unobserved factor ξ_{jt} refers to a demand shock that captures the latent preference variation and latent city characteristics common to the users on day t. The term ϵ_{ijt} is an i.i.d. error drawn from a standard type-I extreme value (Gumbel) distribution. The share of counts to each city on day t is a nonlinear function of the mean utility $\delta_{jt} = x_j\beta + \xi_{jt}$ and expressed as follows:

$$s_{jt} = \frac{e^{\delta_{jt}}}{1 + \sum_{m=1}^{J} e^{\delta_{mt}}} \tag{3.2}$$

where migrating to or staying in cities in Ukraine ($m = 0$) is normalized as an outside option. The number of unique cities covered for the three groups is 35 for Ukrainians, 124 for Russians, and 205 for Foreigners. We present the X-standardized[9] factor analysis for the three groups in Table 3.4. The coefficients reveal the change in the mean utility of each group for every standard deviation change in the factors. We explain some of the attributes that have been taken into account. The annual average inflation rate calculates the average percentage increase in the price of goods and services. It does so by comparing each month of the year with the corresponding month of the previous year. The housing price is determined through a weighted average of the various price indexes, such as rental rates, costs of maintenance and repairs, and utility expenses. The transport prices include the cost of purchasing cars, motorcycles, and transport services. Similarly, communication price consists of the cost of postal services and the price of telecommunication equipment and services. When it comes to obtaining access to broadband internet, it includes access via computers,

[9]The standardization process excludes categorical and endogenous variables.

smartphones, and tablets. The population numbers are reported as of mid-year for the majority of countries, except European and some Asian countries. Lastly, the net migration rate measures the yearly variation between the number of people who enter and leave the city, for every 1,000 individuals in the population.

Table 3.4: Coefficients for Factors that Influence User Preferences

City Attribute	Ukrainians	Russians	Foreigners
Constant	-6.2476^{***}	0.4885^{***}	-1.6622^{***}
	(0.0083)	(0.0096)	(0.0055)
GDP	0.2458^{***}	-0.4687^{***}	0.1577^{***}
	(0.0262)	(0.0142)	(0.0071)
Unemployment Rate	-0.3778^{***}	0.0357^{***}	-0.1274^{***}
	(0.0126)	(0.0107)	(0.0060)
Inflation	-1.0902^{***}	1.3990^{***}	-0.0257^{***}
	(0.0691)	(0.0334)	(0.0068)
Index of Food and Non-Alcoholic Beverage Prices	-2.0818^{***}	-2.2624^{***}	0.8746^{***}
	(0.1435)	(0.0730)	(0.0686)
Index of Alcoholic Beverage and Tobacco Prices	2.0422^{***}	1.8193^{***}	-0.0218
	(0.0650)	(0.0446)	(0.0222)
Index of Clothing and Footwear Prices	-0.4745^{***}	-0.7470^{***}	-0.1055^{***}
	(0.0490)	(0.0359)	(0.0263)
Index of Housing Prices	-0.5088^{***}	1.1614^{***}	-0.2994^{***}
	(0.0362)	(0.0407)	(0.0236)
Index of Health Goods and Medical Services Prices	1.6316^{***}	1.1255^{***}	-0.0480
	(0.0895)	(0.0421)	(0.0297)
Index of Transport Prices	-0.5331^{***}	1.7211^{***}	0.0352
	(0.0713)	(0.0966)	(0.0533)
Index of Communication Prices	2.1609^{***}	-0.8608^{***}	0.4585^{***}
	(0.0775)	(0.0262)	(0.0102)
Index of Hotel and Catering Prices	-0.4932^{***}	-3.5156^{***}	-0.2005^{***}
	(0.0438)	(0.0977)	(0.0396)
Total Population	-0.0719^{**}	1.0382^{***}	0.3854^{***}
	(0.0267)	(0.0172)	(0.0095)
Net Migration Rate	-0.0820^{***}	0.2579^{***}	-0.1415^{***}
	(0.0106)	(0.0105)	(0.0073)
Percentage of Households with Access to Broadband Internet	0.0040	0.0660^{***}	0.1474^{***}
	(0.0226)	(0.0124)	(0.0072)
Mean Temperature	-0.1503^{***}	-1.1673^{***}	0.0953^{***}
	(0.0130)	(0.0124)	(0.0069)

Note: The coefficients have been X-standardized to provide easier interpretation and within-group comparisons. The groups should not be cross-compared. Standard errors are given in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Different factors significantly affect the preferences of refugee users across various groups. The cost of food and communication constitute two fundamental factors that significantly impact the

user preferences of all groups. Ukrainian users' preferences are further influenced by the prices of alcoholic beverages/tobacco and medical services. In addition to these factors, the choices of Russian users are significantly influenced by transportation costs and, predominantly, hotel prices. However, aside from food and communication costs, other factors do not largely impact foreigners' preferences. The findings provide insight into the factors that influence general user preferences. However, these choices can also directly reflect the characteristics of the respective cities. Additionally, due to unobserved individual heterogeneity, deviations may exist in the cities to which people are migrating. In the following sections, we will illustrate the user groups' migration patterns and examine the factors that affect the actual user numbers in the cities as well as the user flows between the cities during the 309 days of observations.

3.3.1.2 Regularized Generalized Linear Model

The number of users in city $j = 1, ..., J$ on day $t = 8, ..., T$ conditional on the average count μ_{jt} and the overdispersion parameter θ follows a negative binomial distribution:

$$P(Y = y_{jt}|\mu_{jt}, \theta) = \frac{\Gamma\left(\theta^{-1} + y_{jt}\right)}{\Gamma\left(y_{jt} + 1\right)\Gamma\left(\theta^{-1}\right)} \left(\frac{1}{1 + \theta\mu_{jt}}\right)^{\theta^{-1}} \left(\frac{\theta\mu_{jt}}{1 + \theta\mu_{jt}}\right)^{y_{jt}} \tag{3.3}$$

which allows the modeling of Poisson heterogeneity with Gamma distribution. This formulation is also known as the NB2 model in Cameron and Trivedi (1986). The variance of this mixture distribution is adjusted from the Poisson model by θ on the quadratic term:

$$\text{Var}\left(y_{jt}|x_{j1}, \ldots, x_{jk}, y_{jt-1}, \ldots, y_{jt-\ell}\right) = \mu_{jt} + \theta\mu_{jt}^2 \text{ where } \ell = 7 \tag{3.4}$$

which relaxes the equidispersion restriction, and is more suitable for our situation. The log link function links the mean and the linear predictor:

$$\mu_{jt} = \exp\left\{\alpha_j + \lambda_t + D_t\gamma + \sum_{m=1}^{k} \beta_m x_{jm} + \sum_{n=1}^{\ell} \phi_n y_{jt-n}\right\} \tag{3.5}$$

where α_j is city j's fixed effect, λ_t is the time-fixed effect with the first day (January 8$^{\text{th}}$) serving as the baseline and θ is pre-determined from the Poisson model.

Additionally, the indicator D_t specifies the period during the war, while x_{jm} denotes the vector of k time-invariant city j attributes in 2021 and y_{jt-n} are a week of lags in city j. To generate the history counts of cities on a daily basis, we also impute the city-level user counts across the entire sampling period to fill in the gaps in the city daily records. The two-stage model is estimated by first maximizing the penalized log-likelihood (based on Equation 3.3) on the standardized data to select the variables:

$$\underset{\gamma,\alpha,\beta,\phi}{\text{argmin}} - L(\gamma, \alpha, \beta, \phi) + \psi \left| \gamma + \sum_{j=1}^{J} \alpha_j + \sum_{m=1}^{k} \beta_m + \sum_{n=1}^{\ell} \phi_n \right| \tag{3.6}$$

where the regularization excludes λ_t which captures the migration pattern. In the second stage, the coefficients of the post-LASSO selected variables of the negative binomial model are estimated. The penalty factor ψ is determined by cross-validation to minimize the following deviance[10]

$$D = 2 \sum_{j=1}^{J} \sum_{t=1}^{T} \left\{ y_{jt} \ln \left(\frac{y_{jt}}{\mu_{jt}} \right) - \left(y_{jt} + \theta^{-1} \right) \ln \left(\frac{1 + \theta y_{jt}}{1 + \theta \mu_{jt}} \right) \right\}. \tag{3.7}$$

In contrast to the Poisson distribution, which models user count, the negative binomial distribution emerges as the marginal distribution when unobserved heterogeneity is integrated out from the conditional Poisson distribution. The heterogeneity in the counts is presumed to follow a $Gamma(\theta, \theta)$ distribution. Such latent heterogeneity maintains the conditional mean of the Poisson model but induces overdispersion, making the negative binomial model preferable as it accurately accommodates the increased variability in the data. Following Cameron and Trivedi (1986), the estimated overdispersion parameter $\widehat{\theta}$ in the NB2 model[11] can be determined[12] by estimating the

[10] Deviance is the distance between the log-likelihood of the model and the log-likelihood of the saturated model (model with a free parameter for each observation).

[11] Instead of choosing the NB1 model, Greene (2008) suggests that data with large positive values would favor NB2. Also, NB1 and NB2 models produce similar results.

[12] The auxiliary regression is based on the overdispersion test of whether $\theta = 0$ given the alternative hypothesis $\text{Var}(y) = \mu + \theta g(\mu)$ where $\mathbb{E}(y) = \mu$.

following auxiliary regression where under the NB2 setting $g(\mu) = \mu^2$ and $\widehat{\mu}_{jt}$ is the fitted value from the Poisson model:

$$\frac{(y_{jt} - \widehat{\mu}_{jt})^2 - y_{jt}}{\widehat{\mu}_{jt}} = \theta \frac{g(\widehat{\mu}_{jt})}{\widehat{\mu}_{jt}} + \epsilon_{jt}. \tag{3.8}$$

The results of the count model are presented in Table 3.5. The coefficients indicate the %-change in the average count of city users for every unit change in the factors. This model includes count lags, a binary war indicator, and city time-invariant characteristics as well as both city- and time-fixed effects. The choice of variables is established by the penalty term in Equation 3.6.

Table 3.5: Factors Affecting the Percentage Change in the Number of Users

City Attributes	Ukrainians	Russians	Foreigners
Count_{t-1}	0.0500^{***}	0.0800^{***}	0.1301^{***}
Count_{t-6}	–	–	-0.0200^{***}
Count_{t-7}	-0.0071^{***}	–	-0.0200^{***}
GDP	–	–	0.0001^{***}
Unemployment Rate	-5.1999^{***}	3.2001^{***}	-2.7028^{***}
Inflation	–	14.0082^{***}	-0.4888^{***}
Index of Food and Non-Alcoholic Beverage Prices	–	-1.2225^{***}	–
Index of Alcoholic Beverage and Tobacco Prices	0.9142^{***}	0.2904^{***}	-0.1199^{***}
Index of Housing Prices	0.3305^{***}	0.0085^{*}	-0.1499^{***}
Index of Health Goods and Medical Services Prices	0.5817^{***}	0.5515^{***}	0.5013^{***}
Index of Transport Prices	–	0.7427^{***}	–
Index of Communication Prices	–	-0.1000^{***}	0.1301^{***}
Index of Hotel and Catering Prices	-1.6856^{***}	-0.7770^{***}	0.0500^{***}
Total Population	0.0016^{***}	0.0063^{***}	0.0042^{***}
Net Migration Rate	-2.7028^{***}	4.4564^{***}	-6.1526^{***}
Percentage of Households with Access to Broadband Internet	3.2311^{***}	-0.9851^{***}	0.8032^{***}
Mean Temperature	-3.6613^{***}	-8.5977^{***}	1.5215^{***}

Note: The counts represent the mean daily user count in different cities. Regularization is used to eliminate certain variables from the models. In this case, the "Index of Communication Prices" was removed from all three groups. Additionally, all of the estimates have been converted into effects of percentage change on the city average user counts. The intercept has been omitted from the report. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Different groups have different sets of characteristics selected, except the "Index of Clothing and Footwear Prices," which is excluded for all groups. Overall, the average number of Ukrainian and Russian refugee users increases in the prices of "alcoholic beverages/tobacco" as well as "medical services and health goods" but decreases in "hotel and catering" prices. The factors that influence user count largely mirror those affecting user preferences. Ukrainian users, including those in domestic cities, tend to reside in more developed cities characterized by lower migration and unemployment rates, better access to broadband internet, and higher housing prices. Conversely, more Russian users are observed in cities that experience higher inflation, greater population flux, and elevated unemployment rates. Foreign refugees, who are less impacted by the war, tend to reside in more developed cities with higher costs for health-related goods and hotel accommodations.

One of the important findings of this research is the identification of refugee migration patterns. The model leaves the time-fixed effects unpenalized to sketch the daily trend, which is represented by the average percentage change in the number of users. These trends are presented in Figure 3.7 to 3.9, showcasing the migration pattern of the three refugee user groups.

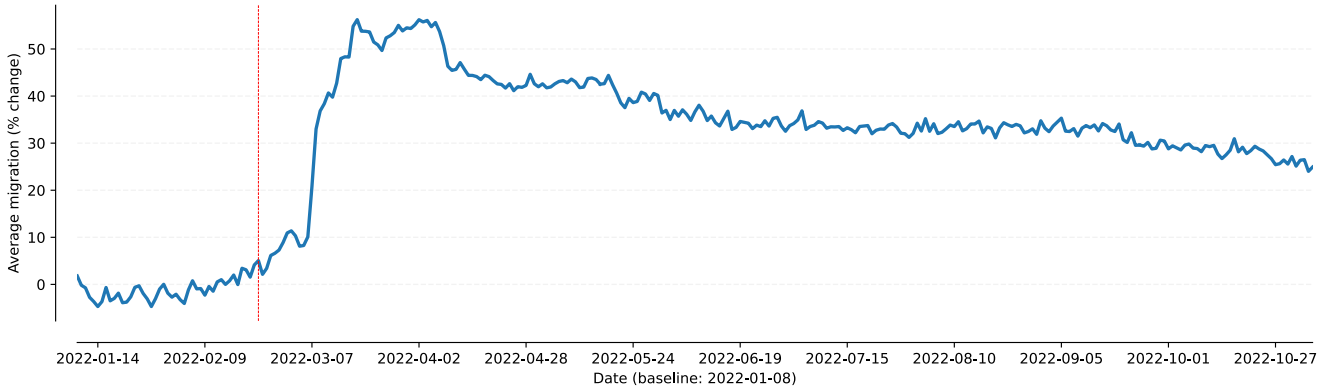

Figure 3.7: Migration Patterns of Ukrainian Refugees

Note: The onset of war results in a significant increase in migration, which peaks in mid-March and subsequently decreases gradually, but remains relatively high.

Immediately after the war began, the migration of Ukrainian users increased significantly, reaching over 50% increments within a month. In April, there was a slight reduction in the movements, which were maintained at a nearly consistent level thereafter. The migration of Russian

users also increased, but only about half as much as that of Ukrainians. The average migration remained approximately stable until September. On September 21st, 2022, the announcement of mobilization by the Russian Federation resulted in a doubling of the migration of Russian users, which exceeded 50% within two weeks.

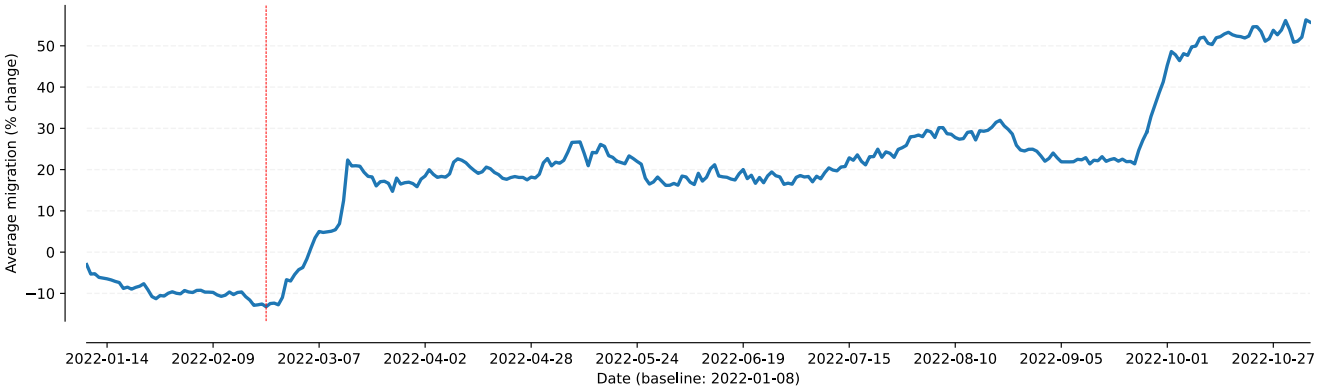

Figure 3.8: Migration Patterns of Russian Refugees

Note: The war's impact on migration is only half as pronounced as observed with Ukrainian migration. Migration levels stayed consistent for about six months before experiencing a surge in October.

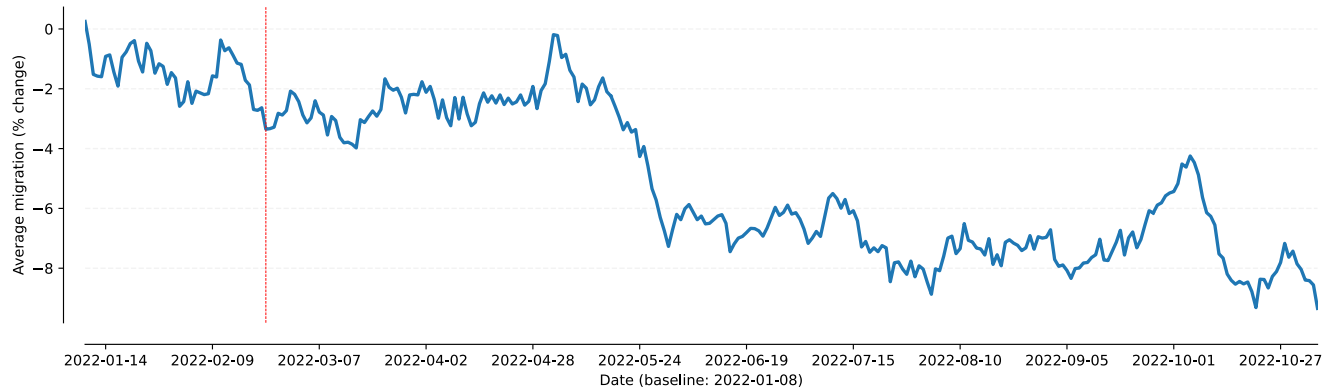

Figure 3.9: Migration Patterns of Foreign Refugees

Note: Foreigners' migration has slightly decreased during the observation period and is largely unaffected by the war.

The impact of the war on foreign users was less substantial. Following the holiday season, there was a decrease in movements compared to the baseline day. After April, the migration saw a modest decrease. In general, the migration of foreigners appeared to be within the range of normal fluctuations. These migration patterns are recovered from the change in the number of users in each city. Although the UNHCR data portal offers current refugee statistics and tracks the total number

of refugees in each country, the aggregated data without the micro-level information cannot provide a comprehensive understanding of migration patterns. Such patterns not only demonstrate how a particular group of refugees responds to the effects of war but also help us comprehend the scale of migration during times of conflict, providing insights into general refugee migration patterns.

Up to this point, the primary emphasis has been on analyzing the daily user counts in various locations. However, moving forward, we will shift our attention toward studying the average movements by analyzing the factors that affect migration flows.

3.3.1.3 Factor Analysis of the Migration Flows

Migration flows refer to the total number of users who were present in City A on the previous day (day $t-1$) and are currently located in City B (day t). These flows are monitored via a sizeable transition matrix. The user movements exclude people who remain in the same city throughout the next day. To highlight the representing migration flows, we eliminate the less significant routes, i.e., the origin-destination city pairs taken by fewer than 100 cumulative unique users throughout our sampling period of 309 days. After the selection process, there are now 117 distinct routes for Ukrainian users, 867 for Russian users, and 1075 for foreign users. The analysis employs a slight variation of the two-stage model specified in Equations 3.3 to 3.8 that changes the city user counts to the user flows. The variables are selected by the regularized maximum likelihood and the results are shown in Table 3.6.

The flow y_{abt} follows a negative binomial distribution, given the average flow μ_{abt} and the overdispersion parameter θ. These represent the number of users who travel from city $a \in \{1, ..., J\}$ to city $b \neq a$ on day $t = 9, ...T$, where the destination city b is conditional on the original city a in the transition matrix. The mean μ_{abt} is a function of the route$_{ab}$ fixed effects, time-fixed effects λ_t with the first day (January 9$^{\text{th}}$) serving as the baseline, the war indicator D_t, pairs of time-invariant city attribute x_{abm} with $m = 1, ..., k$ including the characteristics of the origin and the destination city, and $y_{ab(t-n)}$ seven days of lagged data on the flow of route$_{ab}$ with $n = 1, ..., 7$.

Table 3.6: Factors Affecting the Percentage Change in the Flow of Users

City Attribute/Lag Count/Travel Points		Ukrainians	Russians	Foreigners
	Flow_{t-1}	0.1601^{***}	0.7125^{***}	0.4410^{***}
	Flow_{t-2}	0.2102^{***}	0.3908^{***}	0.8133^{***}
	Flow_{t-3}	0.2002^{***}	0.1802^{***}	0.8637^{***}
	Flow_{t-4}	0.1301^{*}	0.1101^{***}	0.8234^{***}
	Flow_{t-5}	0.0800^{*}	0.0700^{**}	0.7830^{***}
	Flow_{t-6}	0.0400	0.0700^{**}	0.7125^{***}
	Flow_{t-7}	0.1401^{***}	0.1201^{***}	0.7730^{***}
GDP	origin	–	0.0000^{***}	0.0000^{***}
	destination	0.0000	0.0000^{***}	0.0000^{***}
Unemployment Rate	origin	-4.3907^{***}	-2.4495^{***}	-0.0300
	destination	-4.6866^{***}	-2.5957^{***}	-0.2896^{***}
Inflation	origin	-4.8771	3.9251^{***}	-0.2098^{***}
	destination	–	5.1166^{***}	-0.1199^{***}
Index of Food and Non-Alcoholic Beverage Prices	origin	–	-0.4788^{***}	0.0300^{**}
	destination	0.2603	-0.4291^{***}	-0.0048
Index of Alcoholic Beverage and Tobacco Prices	origin	0.1701^{***}	–	-0.0100^{***}
	destination	0.0800^{*}	–	-0.0066^{***}
Index of Clothing and Footwear Prices	origin	–	-0.0800^{***}	0.0032
	destination	-0.0200	-0.0800^{***}	0.0043
Index of Housing Prices	origin	–	-0.1000^{***}	0.0200^{***}
	destination	–	-0.0800^{***}	–
Index of Health Goods and Medical Services Prices	origin	0.1201^{*}	–	-0.0084
	destination	0.0700	-0.0800^{***}	-0.0300^{***}
Index of Transport Prices	origin	0.0700	0.6924^{***}	-0.0100
	destination	-0.1798	0.5917^{***}	0.0500^{***}
Index of Communication Prices	origin	0.0900	-0.0700^{***}	0.0054
	destination	-0.0200	-0.0700^{***}	0.0080^{**}
Index of Hotel and Catering Prices	origin	–	-0.1399^{***}	–
	destination	–	-0.0700^{***}	–
Total Population	origin	-0.0016^{***}	0.0011^{***}	0.0001^{*}
	destination	-0.0010	0.0011^{***}	0.0001^{**}
Net Migration Rate	origin	-5.1525^{***}	0.2403^{*}	-0.1998^{**}
	destination	-4.6676^{***}	0.1802	-0.1199
Percentage of Households with Access to Broadband Internet	origin	-0.5783^{***}	0.2303^{***}	0.0500^{*}
	destination	-0.6876^{***}	0.1802^{***}	0.0048
Mean Temperature	origin	-1.3015^{**}	-0.5187^{***}	0.1802^{***}
	destination	-1.8526^{***}	-0.6777^{***}	0.1201^{**}

Note: The counts represent the mean daily user count in different cities. The intercept has been omitted from the report.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Based on the findings, it is apparent that there is no supporting force that creates a push-pull effect from both ends. Specifically, the migration flow of Russian users is slightly more influenced by the price index factors in their cities of origin. The Ukrainian users are similar, particularly in the pricing of "Alcoholic Beverages and Tobacco" and "Health Goods and Medical Services." Conversely, the migration flow of foreign users is marginally more impacted by the price index factors in the destination cities. Upon examining overall inflation in both origin and destination cities, we observe a positive correlation with the flow of Russian users, indicating that refugees from Russia are mainly moving between cities with higher inflation levels. In contrast, inflation appears to adversely affect the flow of refugees from other foreign countries.

By examining user counts from two perspectives, we aim to gain insights into migration patterns. We analyze three user groups individually, considering their respective cities and routes. The choice model focuses on city selections outside of Ukraine, the city-level model reviews changes in each city's resident numbers, and the flow model studies movements between cities. These models provide unique viewpoints on how city attributes influence refugee migration.

3.4 Conclusion

In summary, this research investigates refugee migration patterns observed during the Russia-Ukraine war in 2022, leveraging data obtained from a queer social network. We have created an interactive map tool to visualize the movements of refugee users. Then we compare migration outflows from the capital cities of both countries before and after the war. Moreover, from the perspective of choice probability, we explored how factors such as city price indexes, demographics, and geography characteristics influence users' migration decisions using a choice model. By observing user counts and migration flows between cities, we conducted a factor analysis to explain the relationship between the frequency of occurrence of two user events and city attributes. Finally, we estimated the migration patterns of three distinct user groups.

Bibliography

Aguiar, Mark, and Erik Hurst. 2007*a*. "Life-Cycle Prices and Production." *American Economic Review*, 97(5): 1533–1559.

Aguiar, Mark, and Erik Hurst. 2007*b*. "Measuring Trends in Leisure: The Allocation of Time Over Five Decades." *Quarterly Journal of Economics*, 122(3): 969–1006.

Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis. 2013. "Time Use During the Great Recession." *American Economic Review*, 103(5): 1664–1696.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst. 2021. "Leisure Luxuries and the Labor Supply of Young Men." *Journal of Political Economy*, 129(2): 337–382.

Andreyeva, Tatiana, Michael W. Long, and Kelly D. Brownell. 2010. "The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food." *Am J Public Health*, 100(2): 216–222.

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics*, 11: 685–725.

Athey, Susan, and Stefan Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application." *Observational Studies*, 5(2): 37–51.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics*, 47(2): 1148–1178.

Atrostic, Barbara K. 1982. "The Demand for Leisure and Nonpecuniary Job Characteristics." *American Economic Review*, 72(3): 428–440.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2): 242–262.

Berry, Steven T., and Philip A. Haile. 2014. "Identification in Differentiated Products Markets Using Market Level Data." *Econometrica*, 82(5): 1749–1797.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven T., Michael Carnall, and Pablo T. Spiller. 2006. "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity." *Competition policy and antitrust*, 183–213.

Bigoni, Maria, Davide Dragone, Stéphane Luchini, and Alberto Prati. 2021. "Estimating Time Preferences for Leisure." *CEPR Discussion Papers*.

Boungou, Whelsy, and Alhonita Yatié. 2022. "The Impact of the Ukraine-Russia War on World Stock Market Returns." *Economics Letters*, 215: 110516.

Boyon, Nicolas. 2021. "LGBT+ Pride 2021 Global Survey Points to a Generation Gap Around Gender Identity and Sexual Attraction." *IPSOS*.

Burda, Michael C., Daniel S. Hamermesh, and Jay Stewart. 2013. "Cyclical Variation in Labor Hours and Productivity Using the ATUS." *American Economics Review*, 103(3): 99–104.

Cameron, A. Colin, and Pravin K. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics*, 1(1): 29–53.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal*, 21(1): C1–C68.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best Practices for Differentiated Products Demand Estimation with PyBLP." *RAND Journal of Economics*, 51(4): 395–421.

Conlon, Kimberly. 2020. "Quantifying the Benefits of Streaming Video Services." *Working paper*.

Diewert, W. E. 1978. "Superlative Index Numbers and Consistency in Aggregation." *Econometrica*, 46(4): 883–900.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation." *Econometrica*, 80(5): 2231–2267.

Dumagan, Jesus C. 2002. "Comparing the Superlative Törnqvist and Fisher Ideal Indexes." *Economics Letters*, 76: 251–258.

Dumont, Jean-Christophe, and Ave Lauren. 2022. "The Potential Contribution of Ukrainian Refugees to the Labour Force in European Host Countries." *OECD*.

Duszczyk, Maciej, and Paweł Kaczmarczyk. 2022. "The War in Ukraine and Migration to Poland: Outlook and Challenges." *Intereconomics: Review of European Economic Policy*, 57(3): 164–170.

Ebghaei, Felor. 2016. "Effect of Firm's Export-Orientation on Backward Spillovers of Foreign Direct Investment in Turkish Manufacturing Industry." *Journal of Business Economics and Finance*, 5(4): 351–359.

Einav, Liran. 2007. "Seasonality in the U.S. Motion Picture Industry." *RAND Journal of Economics*, 38(1): 127–145.

Ferreira, Fernando, Amil Petrin, and Joel Waldfogel. 2016. "Preference Externalities and the Rise of China: Measuring their Impact on Consumers and Producers in Global Film Markets." *Working paper*.

Greene, William. 2008. "Functional Forms for the Negative Binomial Model for Count Data." *Economics Letters*, 99(3): 585–590.

Gronau, Reuben. 1977. "Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited." *Journal of Political Economy*, 85(6): 1099–1123.

Gulen, Huseyin, Candace Jens, and T. B. Page. 2020. "An Application of Causal Forest in Corporate Finance: How Does Financing Affect Investment?" *Microeconomics: Intertemporal Firm Choice & Growth*.

Hendricks, Ken, and Alan Sorensen. 2009. "Information and the Skewness of Music Sales." *Journal of Political Economy*, 117(2): 324–369.

Hill, Robert J. 2004. "Constructing Price Indexes across Space and Time: The Case of the European Union." *American Economic Review*, 94(5): 1379–1410.

Kalogiannidis, Stavros, Fotios Chatzitheodoridis, Dimitrios Kalfas, Stamatis Kontsas, and Ermelinda Toska. 2022. "The Economic Impact of Russia's Ukraine Conflict on the EU Fuel Markets." *International Journal of Energy Economics and Policy*, 12(6): 37–49.

Khudaykulova, Madina, He Yuanqiong, and Akmal Khudaykulov. 2022. "Economic Consequences and Implications of the Ukraine-Russia War." *International Journal of Management Science and Business Administration*, 8(4): 44–52.

Korovkin, Vasily, and Alexey Makarin. 2023. "Conflict and Intergroup Trade: Evidence from the 2014 Russia-Ukraine Crisis." *American Economic Review*, 113(1): 34–70.

Krueger, Alan B., and Andreas I. Mueller. 2012. "Time Use, Emotional Well-Being, and Unemployment: Evidence from Longitudinal Data." *American Economic Review*, 102(3): 594–599.

Lloyd, Armağan Teke, and Ibrahim Sirkeci. 2022. "A Long-Term View of Refugee Flows from Ukraine: War, Insecurities, and Migration." *Migration Letters*, 19(4): 523–535.

Luo, Lan, Brian T. Ratchford, and Botao Yang. 2013. "Why We Do What We Do: A Model of Activity Consumption." *Journal of Marketing Research*, 50(1): 24–43.

Morariu, Alunica. 2022. "The Impact of the Russian-Ukrainian Conflict on the Current Migration Phenomenon." *Ovidius University Annals, Economic Sciences Series*, 22(2): 99–108.

Mukoyama, Toshihiko, Christina Patterson, and Ayşegül Şahin. 2018. "Job Search Behavior Over the Business Cycle." *American Economic Review*, 10(1): 190–215.

Nevo, Aviv. 2000. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, 31(3): 395–421.

Neyman, Jerzy. 1959. "Optimal Asymptotic Tests of Composite Statistical Hypotheses." *Probability and Statistics: The Harald Cramer Volume*, 213– 234.

Okrent, Abigail M., and Julian M. Alston. 2012. "The Demand for Disaggregated Food- Away-From-Home and Food-at-Home Products in the United States." *Economic Research Report*, , (139).

Pawlowski, Tim, and Christoph Breuer. 2012. "Expenditure Elasticities of the Demand for Leisure Services." *Applied Economics*, 44(26): 3461–3477.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705–729.

Phaneuf, Daniel James. 1997. "Generalized Corner Solution Models in Recreation Demand."

Raeford, Michael. 2020. "Average Price Per DVD Title (Movie Industry Stats)." *Entertainment Industry Market Statistics Report, Motion Picture Association of America*.

Stegner, Ben. 2018. "IMDb vs. Rotten Tomatoes vs. Metacritic: Which Movie Ratings Site Is Best?" *Make Use Of*.

UNHCR. 2022. "Refugees Fleeing Ukraine (Since 24 February 2022)."

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

Zhen, Chen, Eric A. Finkelstein, Shawn A. Karns, Ephraim S. Leibtag, and Chenhua Zhang. 2019. "Scanner Data-Based Panel Price Indexes." *American Journal of Agricultural Economics*, 101(1): 311–329.

Appendix A

Supplementary material for Chapter 1

A.1 OpusData Data Components

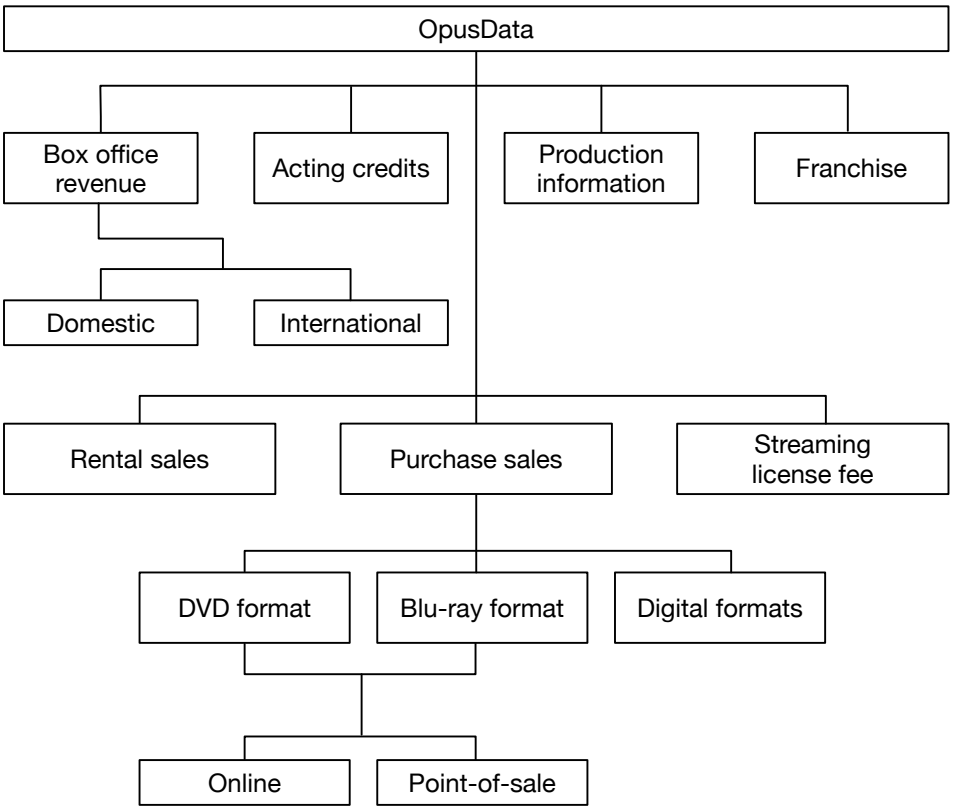

Figure A.1: *OpusData* Structure

Note: The data includes high-frequency box office revenue, various video formats sales, and production information.

A.2 Unfiltered Annual Movie Production Count

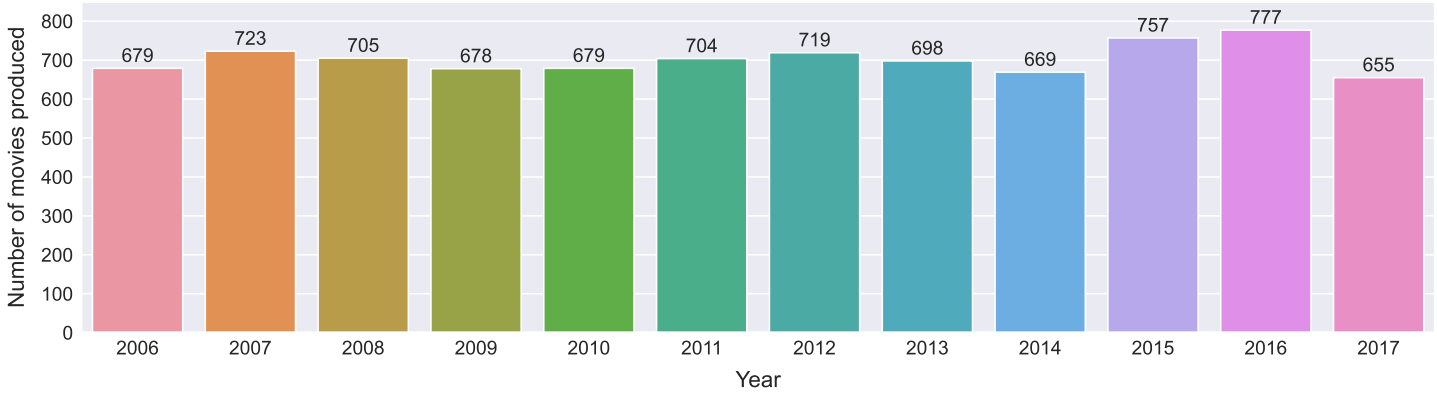

Figure A.2: Number of Movies Produced in Each Year (Unfiltered)

Note: The number of movies produced annually. To be included in this research, movies must have debuted in theaters and demonstrated positive sales across all video formats, both physical and digital. The curated dataset comprises 2,388 unique movies spanning 12 years.

A.3 Revenue from Box Office and Video Sales

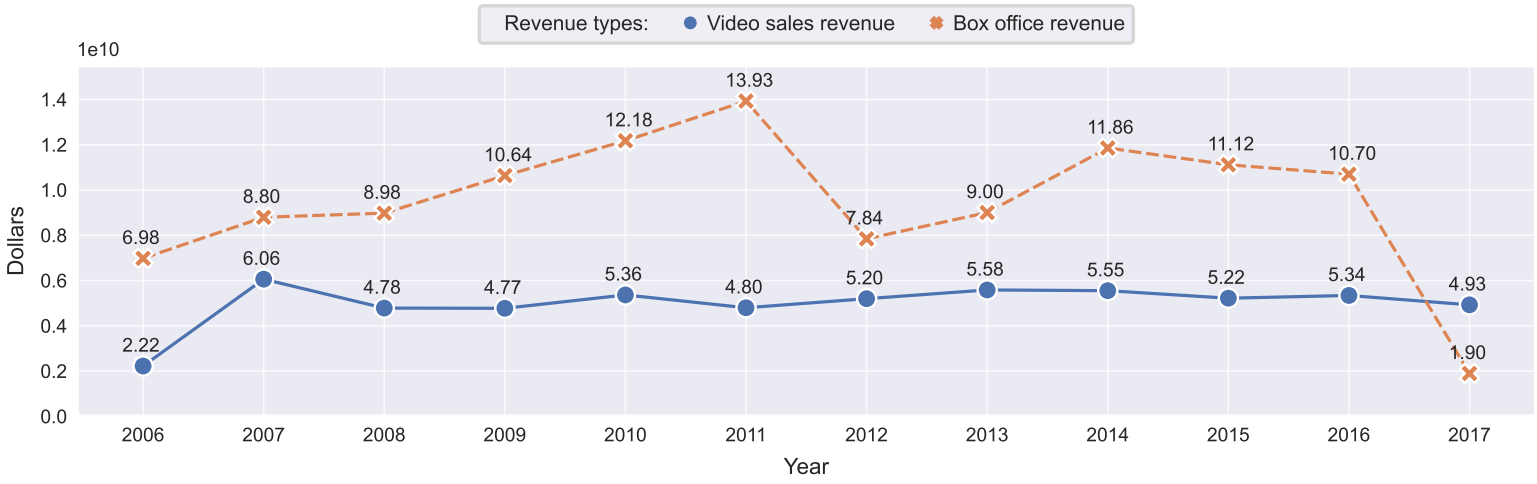

Figure A.3: Total Revenue from Box Office and Video Sales (10 Billion)

Note: The overall average theater ticket price is $8.03. AMC ticket price for an adult is $13.69. The video's average price for each year is presented in Figure 1.14.

A.4 Average Video Markup by Markets

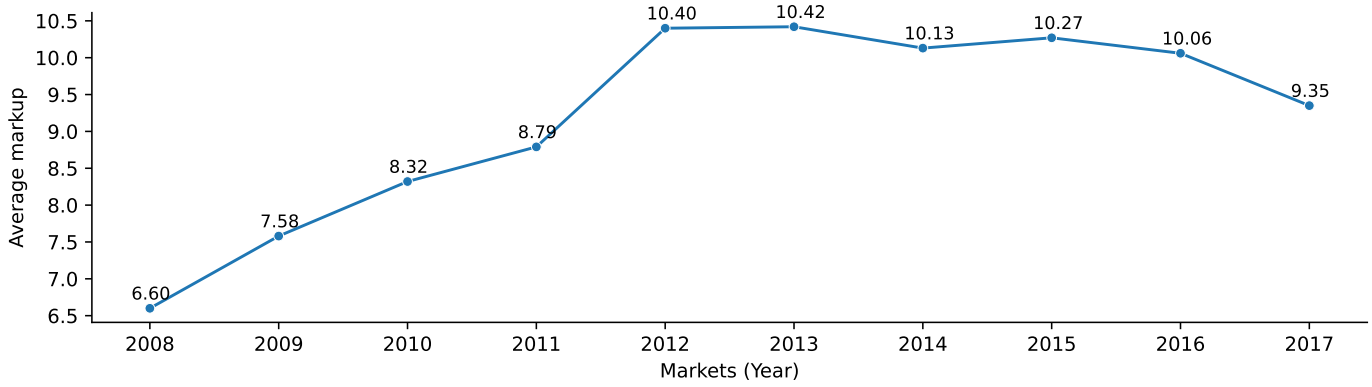

Figure A.4: Video Markup Averages Across Ten Markets

Note: The average markups increased in the first few years and then fluctuated around 10.11%.

A.5 Movie Rankings

Table A.1: Blockbusters Ranking from 2006 to 2017

Rank	Domestic Box Office	Worldwide Box Office
1	Star Wars Ep. VII: The Force Awakens	Avatar
2	Avatar	Star Wars Ep. VII: The Force Awakens
3	Jurassic World	Jurassic World
4	The Avengers	Furious 7
5	The Dark Knight	The Avengers
6	Rogue One: A Star Wars Story	Avengers: Age of Ultron
7	Beauty and the Beast	Harry Potter and the Deathly Hallows: Part II
8	Finding Dory	Beauty and the Beast
9	Avengers: Age of Ultron	Frozen
10	The Dark Knight Rises	The Fate of the Furious

Note: There are 60 movies produced and have been bought by at least one customer in 2006. The domestic box office refers to the US box office. Of the 14 movies listed in this table, according to our data source from the-numbers.com, only "Beauty and the Beast" and "Frozen" belong to the *Art* genre (originally it is the *Musical* genre). The remaining 12 movies are categorized as *Action-adventures*.

Table A.2: All-Time Ranking (Until September 2020)

Rank	Domestic Box Office	Worldwide Box Office
1	Star Wars Ep. VII: The Force Awakens	Avengers: Endgame
2	Avengers: Endgame	Avatar
3	Avatar	Titanic
4	Black Panther	Star Wars Ep. VII: The Force Awakens
5	Avengers: Infinity War	Avengers: Infinity War
6	Titanic	Jurassic World
7	Jurassic World	The Lion King
8	The Avengers	Furious 7
9	Star Wars Ep. VIII: The Last Jedi	The Avengers
10	Incredibles 2	Frozen II

Note: As of August 2023, the top five films in the all-time domestic box office rankings are "Star Wars Ep. VII: The Force Awakens," "Avengers: Endgame," "Spider-Man: No Way Home," "Avatar," and "Top Gun: Maverick." For the all-time worldwide box office, the top five are "Avatar," "Avengers: Endgame," "Avatar: The Way of Water," "Titanic," and "Star Wars Ep. VII: The Force Awakens." Every movie listed in this table is categorized as *Action-adventures*.

Appendix B

Supplementary material for Chapter 2

B.1 Number of Receipts in Income Groups

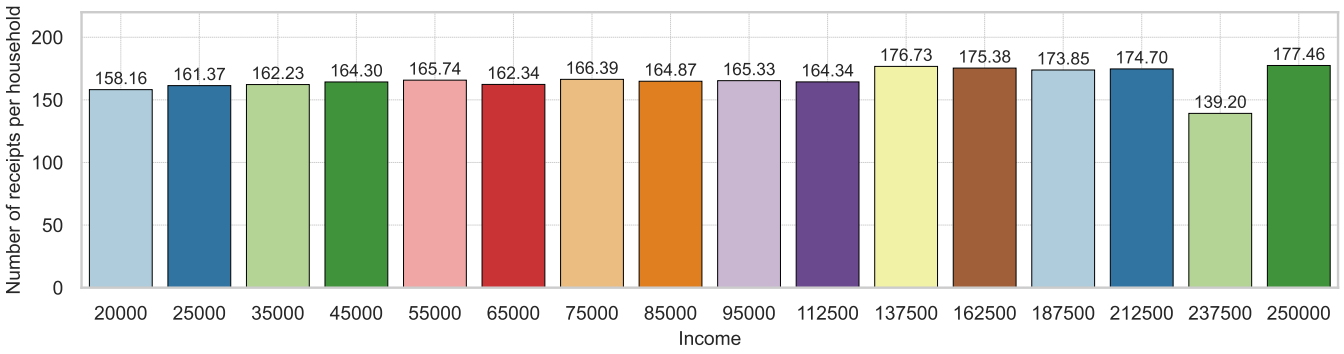

Figure B.1: Number of Total Receipts per Household (Different Income Groups)

Note: The total number of receipts uploaded by households in each income group is similar, except for those at the 237,500 level.

B.2 Four Price Indexes

The bilateral price comparison between countries j and k for periods s and t is a function determined by the prices and quantities of a fixed basket of commodities:

$$P = f(p_i^{kt}, p_i^{js}, q_i^{kt}, q_i^{js}) \tag{B.1}$$

where i denotes the commodities in the basket. In our single-country context, bilateral price comparisons vary only across periods. Next, we'll expand the comparison to encompass several states within the census regions over time. In a fixed-country scenario, we compare current expenditures relative to the spending in the base period. For measuring price fluctuations, the base period is defined as the first period of the time series. For cross-comparisons, we designate a specific region as the common base period. We use four bilateral comparisons to measure both equivalent and compensating variations, as well as to adjust for inflation bias.

The Paasche index, also known as the quantity deflator, measures the price difference between consuming with today's dollars and consuming with the base period dollar. It is defined as:

$$P_P^{0t} = \frac{\sum\limits_{i \in \Theta} p_i^t q_i^t}{\sum\limits_{i \in \Theta} p_i^0 q_i^t} \tag{B.2}$$

where Θ is the common basket in both period 0 and t. The equivalent variation indicates the income adjustments made prior to the price change, bringing the consumer's utility to the level it would have reached if the price change had occurred. Conversely, the Laspeyres index, represented by the following equation, measures the compensating variation:

$$P_L^{0t} = \frac{\sum\limits_{i \in \Theta} p_i^t q_i^0}{\sum\limits_{i \in \Theta} p_i^0 q_i^0}. \tag{B.3}$$

This index indicates the current price level if the consumer were to purchase the same basket as in the base period. These adjustments represent the necessary income changes after a price shift to bring the consumer back to the utility level of the base period. However, the Laspeyres index, without accounting for the current price-adjusted quantity, is also prone to substitution bias. While the Paasche and Laspeyres indexes typically understate and overstate inflation respectively, the Fisher ideal index mitigates this bias by taking the geometric average of the two indexes:

$$P_F^{0t} = \sqrt{P_P^{0t} \times P_L^{0t}}. \tag{B.4}$$

The Fisher index, also known as the price deflator, corrects for both price and substitution biases. By measuring the unbiased basket value in constant dollars, the Fisher index produces a value that falls between the Laspeyres and Paasche indexes. However, in the absence of information from the base year, the Fisher index can be biased when determining the amount of price change associate to inflation or alterations in basket quality. Lastly, the Törnqvist index, sometimes referred to as the Törnqvist-Theil index, represents the weighted price ratio determined by the average of the expenditure shares from the two periods:

$$P_T^{0t} = \prod_{i \in \Theta} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}} \quad \text{where} \quad s_j^n = \frac{p_j^n q_j^n}{\sum\limits_{i \in \Theta} p_i^n q_i^n} \ \ \forall n \text{ periods.} \tag{B.5}$$

As demonstrated in Diewert (1978), the Törnqvist and Fisher indexes are numerically close to each other. Furthermore, Dumagan (2002) later showed that a component capturing the growth rates in both indexes is approximately the same.

B.3 Leisure Time Spent and Expenditure Data Flow

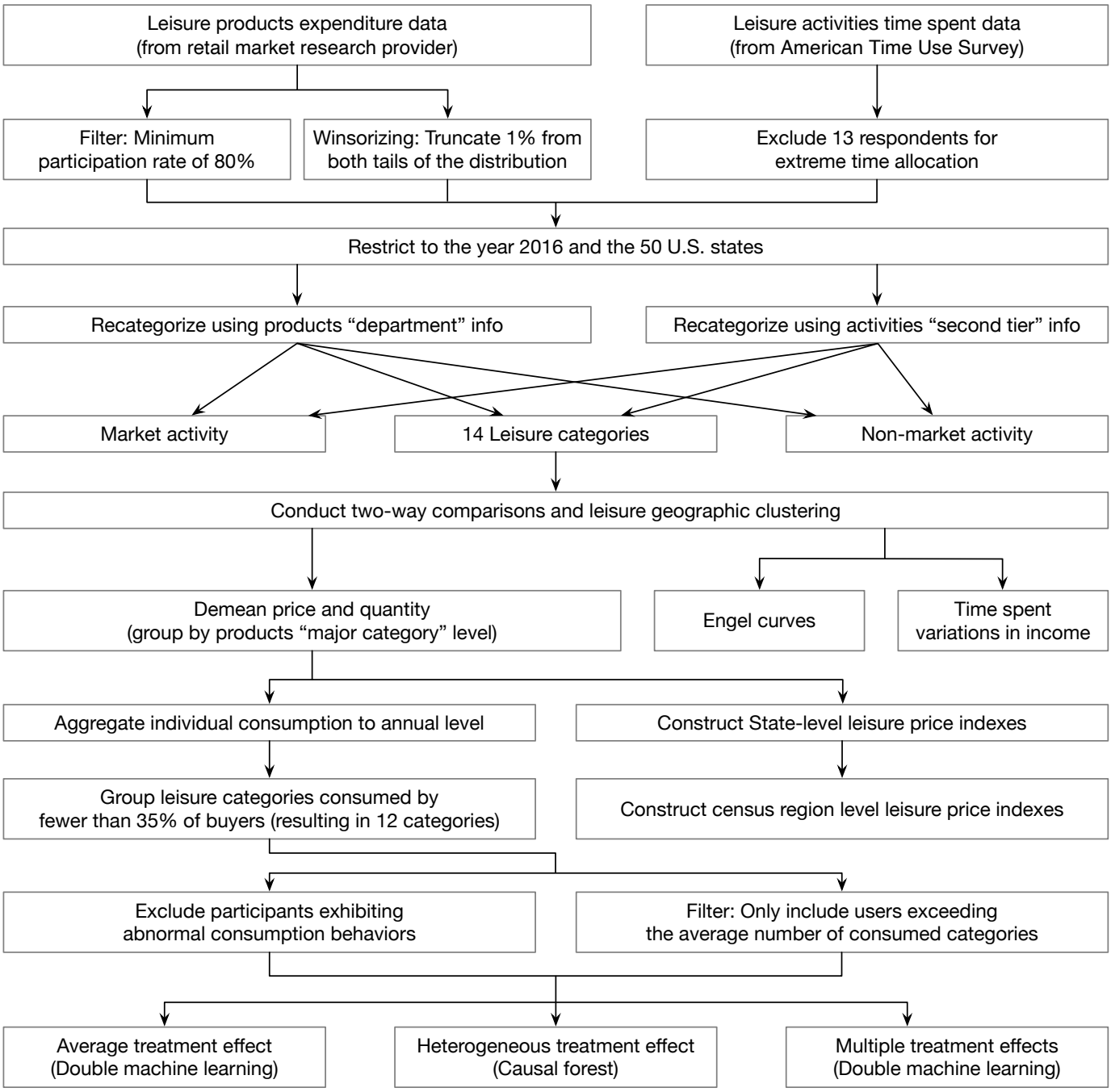

Figure B.2: The Flow of Time Spent and Expenditure Data

Note: The data from both leisure measurements are processed to eliminate outliers, such as abnormal consumption behaviors and recording errors, aiming to ensure they accurately represent the consumption patterns of typical households. This filtering also seeks to preserve sufficient variation in consumption observations.

B.4 Supplementary Tables

Table B.1: Four Leisure Levels

Leisure Levels	Classifications
Level 1	sports or exercise TV entertainment (not TV) socializing reading gardening or pet care hobbies religious or civic activities
Level 2	**Everything in Level 1** sleeping personal care eating
Level 3	**Everything in Level 2** child care
Level 4	**Everything in Level 3** education own medical care

Note: The *leisure measurements* aggregate the *leisure activities* into four levels.

Table B.2: Time Allocation

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
child care	Waiting for/with non-household(nonhh) children	0.17	0.22	0.11	0.08	0.20	0.17	0.16
	Waiting associated with household(hh) children's health	0.09	0.12	0.04	0.65	0.03	0.06	0.31
	Physical care for nonhh children	0.78	1.13	0.35	1.03	0.35	0.86	0.47
	Reading to/with nonhh children	0.05	0.06	0.03	0.00	0.01	0.06	0.00
	Playing with nonhh children, not sports	1.19	1.37	0.97	0.37	0.86	1.31	0.61
	Arts and crafts with nonhh children	0.03	0.05	0.01	0.00	0.00	0.04	0.00
	Playing sports with nonhh children	0.07	0.08	0.06	0.00	0.00	0.09	0.00
	Talking with/listening to nonhh children	0.12	0.05	0.19	0.01	0.06	0.03	3.31
	Organization & planning for nonhh children	0.81	1.15	0.38	0.15	0.58	0.90	0.33
	Looking after nonhh children (as primary activity)	0.42	0.39	0.44	0.15	0.52	0.42	0.11
	Attending nonhh children's events	0.08	0.13	0.02	0.05	0.03	0.07	0.79
	Dropping off/picking up nonhh children	0.19	0.24	0.13	0.33	0.04	0.22	0.15
	Homework (nonhh children)	0.06	0.07	0.04	0.00	0.05	0.06	0.11
	Meetings and school conferences (nonhh children)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Homeschooling of nonhh children	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Waiting associated with nonhh children's education	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Providing medical care to nonhh children	0.01	0.01	0.00	0.00	0.00	0.01	0.11
	Obtaining medical care for nonhh children	0.00	0.01	0.00	0.00	0.00	0.01	0.00
	Waiting associated with nonhh children's health	0.01	0.02	0.00	0.00	0.00	0.01	0.00
	Travel related to using childcare services	0.02	0.03	0.00	0.04	0.02	0.01	0.00
	Waiting associated w/purchasing childcare svcs	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Using paid childcare services	0.04	0.06	0.01	0.01	0.03	0.04	0.00
	Providing medical care to hh children	0.22	0.35	0.06	0.26	0.15	0.23	0.01
	Waiting associated with hh children's education	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	Obtaining medical care for hh children	0.25	0.33	0.13	0.18	0.13	0.26	0.44
	Meetings and school conferences (hh children)	0.13	0.14	0.11	0.51	0.01	0.13	0.11
	Telephone calls to/from paid child or adult care providers	0.01	0.01	0.01	0.00	0.00	0.01	0.00
	Travel related to caring for and helping hh children	4.13	5.07	2.96	5.77	2.41	4.36	4.33
	Travel related to caring for and helping nonhh children	1.00	1.17	0.79	0.80	1.06	0.98	1.33
	Homeschooling of hh children	0.25	0.42	0.04	0.00	0.00	0.32	0.00
	Reading to/with hh children	1.03	1.32	0.66	2.12	0.52	1.02	2.35
	Playing with hh children, not sports	7.19	7.25	7.11	14.79	1.64	7.88	6.32
	Arts and crafts with hh children	0.09	0.11	0.06	0.11	0.00	0.11	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
	Playing sports with hh children	0.38	0.32	0.45	0.07	0.06	0.46	0.20
	Physical care for hh children	10.87	14.64	6.16	20.91	6.47	11.21	10.39
	Organization & planning for hh children	2.46	3.26	1.46	3.96	1.24	2.59	3.26
	Looking after hh children (as a primary activity)	1.81	2.14	1.41	1.12	0.64	2.06	1.97
	Attending hh children's events	0.50	0.60	0.37	1.07	0.38	0.47	1.16
	Waiting for/with hh children	1.09	1.42	0.68	1.35	0.74	1.13	1.46
	Homework (hh children)	1.94	2.45	1.29	4.00	1.48	1.93	1.67
	Picking up/dropping off hh children	1.12	1.54	0.59	1.59	0.63	1.18	1.29
	Talking with/listening to hh children	0.31	0.51	0.07	0.54	0.27	0.31	0.31
eating	Purchasing food (not groceries)	1.53	1.55	1.51	1.11	1.76	1.52	1.34
	Waiting associated w/eating & drinking	0.30	0.36	0.23	0.28	0.30	0.30	0.48
	Travel related to eating and drinking	7.22	6.98	7.51	9.00	4.75	7.69	4.17
	Food presentation	0.45	0.61	0.25	1.16	0.19	0.47	0.40
	Using meal preparation services	0.01	0.01	0.00	0.00	0.00	0.01	0.00
	Tobacco and drug use	0.29	0.25	0.34	0.05	0.44	0.26	0.72
	Eating and drinking	64.73	63.81	65.88	84.91	48.58	67.00	57.35
	Food and drink preparation	29.86	37.99	19.72	47.81	31.04	28.47	38.53
education	Travel related to education (except taking class)	0.13	0.13	0.14	0.56	0.12	0.12	0.00
	Travel related to taking class	0.77	0.78	0.75	1.39	0.60	0.77	0.62
	Administrative activities: class for personal interest	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Waiting associated w/admin. activities (education)	0.00	0.00	0.00	0.04	0.00	0.00	0.00
	Administrative activities: class for degree, certification, or licensure	0.03	0.04	0.02	0.00	0.00	0.03	0.00
	Taking class for personal interest	0.55	0.63	0.45	0.59	0.39	0.58	0.44
	Research/homework for class for pers. interest	0.27	0.14	0.43	3.75	0.23	0.10	0.44
	Research/homework for class for degree, certification, or licensure	6.66	6.76	6.53	21.92	5.48	6.04	9.30
	Extracurricular student government activities	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Extracurricular music & performance activities	0.05	0.06	0.05	0.00	0.00	0.07	0.00
	Extracurricular club activities	0.03	0.05	0.00	0.00	0.00	0.04	0.00
	Security procedures rel. to taking classes	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Waiting associated with taking classes	0.07	0.10	0.03	0.19	0.01	0.08	0.00
	Telephone calls to/from education services providers	0.01	0.02	0.00	0.00	0.00	0.01	0.02
	Taking class for degree, certification, or licensure	7.41	7.22	7.66	12.17	4.44	7.74	7.34
	Waiting associated with research/homework	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
entertainment (not TV)	Attending movies/film	1.54	1.44	1.66	1.62	1.16	1.56	2.92
	Travel related to arts and entertainment	3.07	2.63	3.62	2.27	2.89	3.10	4.34
	Security procedures rel. to arts & entertainment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Attending gambling establishments	0.53	0.55	0.50	0.29	0.63	0.53	0.11
	Relaxing, thinking	19.67	18.70	20.88	14.18	31.00	17.55	27.80
	Playing games	13.21	11.37	15.50	9.08	11.11	13.42	24.95
	Listening to/playing music (not radio)	2.05	1.53	2.69	1.17	2.41	2.02	2.04
	Listening to the radio	1.47	0.98	2.08	0.11	2.85	1.29	1.07
	Computer use for leisure (e.g. Games)	9.50	8.66	10.56	13.87	6.40	10.00	5.47
	Waiting associated with relaxing/leisure	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Attending museums	0.45	0.45	0.45	0.44	0.02	0.54	0.00
	Attending performing arts	1.05	0.96	1.15	0.65	0.74	1.13	0.99
	Waiting associated with arts & entertainment	0.07	0.08	0.05	0.00	0.08	0.07	0.12
gardening/ pet care	Lawn, garden, and houseplant care	12.24	8.51	16.90	9.22	5.88	13.75	7.08
	Ponds, pools, and hot tubs	0.28	0.16	0.43	0.00	0.02	0.35	0.29
	Care for animals and pets (not veterinary care)	7.18	7.83	6.37	1.97	2.76	8.30	6.20
	Travel related to using veterinary services	0.06	0.02	0.10	0.00	0.03	0.07	0.00
	Travel related to using lawn and garden services	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Waiting associated with veterinary services	0.03	0.01	0.06	0.00	0.00	0.04	0.00
	Using pet services	0.05	0.06	0.02	0.00	0.01	0.05	0.13
	Waiting associated with pet services	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Travel related to using pet services (not vet)	0.04	0.05	0.02	0.00	0.00	0.04	0.15
	Waiting associated with using lawn & garden services	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Using veterinary services	0.07	0.05	0.09	0.00	0.01	0.08	0.00
	Using lawn and garden services	0.02	0.02	0.02	0.00	0.00	0.02	0.00
hobbies	Arts and crafts as a hobby	2.13	2.32	1.88	0.18	1.84	2.30	1.51
	Collecting as a hobby	0.04	0.00	0.09	0.00	0.00	0.05	0.00
	Hobbies, except arts & crafts and collecting	0.22	0.21	0.24	1.34	0.17	0.18	0.00
	Sewing, repairing, & maintaining textiles	1.89	3.40	0.00	1.10	0.59	2.17	1.95
	Writing for personal interest	0.26	0.09	0.46	0.74	0.21	0.23	0.46
own medical care	Travel related to using medical services	1.06	1.33	0.72	0.53	1.44	1.00	1.60
	Using health and care services outside the home	2.13	2.55	1.62	1.70	3.11	1.98	2.09
	Waiting associated with medical services	0.41	0.55	0.24	0.00	0.52	0.41	0.44
	Personal emergencies	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Using in-home health and care services	0.08	0.10	0.06	0.00	0.27	0.05	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
personal care	Telephone calls to/from professional or personal care svcs providers	0.16	0.18	0.13	0.09	0.30	0.14	0.09
	Washing, dressing and grooming oneself	40.96	47.65	32.59	40.34	48.79	39.59	38.68
	Travel related to personal care	1.31	1.38	1.23	0.83	0.97	1.40	1.38
	Health-related self care	3.98	4.83	2.92	2.76	5.34	3.86	1.68
	Waiting associated w/personal care services	0.04	0.04	0.04	0.07	0.09	0.03	0.00
	Travel related to using personal care services	0.29	0.41	0.15	0.35	0.47	0.26	0.26
	Using personal care services	1.06	1.67	0.31	0.85	2.18	0.89	0.42
reading	Reading for personal interest	20.68	23.35	17.35	18.58	12.23	22.55	14.82
religious/ civic activities	Indoor & outdoor maintenance, repair, & clean-up	0.56	0.38	0.77	0.66	1.05	0.42	1.56
	Building houses, wildlife sites, & other structures	0.02	0.00	0.04	0.00	0.00	0.02	0.00
	Waiting associated w/religious & spiritual activities	0.10	0.11	0.08	0.04	0.15	0.09	0.00
	Teaching, leading, counseling, mentoring	0.91	0.93	0.88	0.65	0.93	0.92	0.83
	Performing	0.39	0.34	0.45	0.73	0.22	0.39	0.62
	Writing	0.04	0.05	0.03	0.00	0.00	0.05	0.00
	Security procedures rel. to religious & spiritual activities	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Religious education activities	1.01	1.30	0.64	0.71	1.61	0.93	0.24
	Collecting & delivering clothing & other goods	0.10	0.12	0.09	0.00	0.04	0.13	0.00
	Food preparation, presentation, clean-up	0.62	0.92	0.24	0.00	0.90	0.60	0.66
	Computer use	0.74	0.79	0.69	0.11	0.54	0.82	0.46
	Organizing and preparing	0.67	0.80	0.51	0.29	0.51	0.73	0.57
	Reading	0.10	0.08	0.13	0.00	0.12	0.11	0.00
	Fundraising	0.33	0.51	0.11	0.78	0.01	0.38	0.00
	Telephone calls (except hotline counseling)	0.16	0.10	0.23	0.00	0.20	0.15	0.17
	Providing care	0.42	0.63	0.16	0.44	0.00	0.49	0.68
	Serving at volunteer events & cultural activities	0.44	0.45	0.44	0.00	0.22	0.50	0.58
	Attending religious services	9.13	10.83	7.02	6.98	16.09	7.93	9.44
	Public health activities	0.02	0.03	0.02	0.00	0.04	0.02	0.00
	Participation in religious practices	3.12	3.78	2.29	3.93	6.79	2.38	3.42
	Civic obligations & participation	0.17	0.10	0.24	0.76	0.10	0.15	0.00
	Waiting associated with using government services	0.03	0.06	0.00	0.15	0.06	0.02	0.00
	Waiting associated w/civic obligations & participation	0.03	0.00	0.07	0.10	0.01	0.03	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
	Travel related to religious/spiritual practices	1.97	2.33	1.52	2.17	3.41	1.72	1.03
	Attending meetings, conferences, & training	0.56	0.65	0.46	0.00	0.26	0.65	0.55
	Television (religious)	0.25	0.42	0.02	0.00	0.66	0.18	0.22
	Travel related to using government services	0.06	0.04	0.09	0.28	0.05	0.05	0.00
	Travel related to volunteering	0.97	1.03	0.89	0.57	0.90	0.94	2.77
	Telephone calls to/from government officials	0.01	0.00	0.02	0.00	0.03	0.01	0.00
	Public safety activities	0.04	0.00	0.09	0.00	0.00	0.05	0.00
	Travel related to civic obligations & participation	0.09	0.08	0.11	0.19	0.13	0.07	0.22
sleeping	Sleeplessness	4.48	5.03	3.78	1.30	6.07	4.28	6.10
	Sleeping	529.50	534.12	523.73	533.67	542.69	526.45	541.11
socializing	Travel related to socializing and communicating	5.76	5.63	5.91	3.41	6.09	5.82	5.55
	Travel related to attending or hosting social events	0.99	1.10	0.85	1.48	0.85	0.96	1.72
	Telephone calls to/from friends, neighbors, or acquaintances	2.07	2.55	1.45	1.74	2.98	1.92	1.90
	Telephone calls to/from family members	2.67	3.79	1.28	3.96	3.69	2.43	2.48
	Travel related to phone calls	0.24	0.13	0.37	0.05	0.43	0.20	0.49
	Socializing and communicating with others	38.11	38.61	37.49	32.35	38.65	38.36	36.07
	Attending or hosting parties/receptions/ceremonies	4.87	5.50	4.07	7.49	3.92	4.90	5.28
	Attending meetings for personal interest (not volunteering)	0.53	0.65	0.38	0.15	0.79	0.46	1.70
	Waiting assoc. w/socializing & communicating	0.01	0.01	0.01	0.04	0.02	0.01	0.00
	Waiting assoc. w/attending/hosting social events	0.01	0.01	0.01	0.00	0.04	0.01	0.00
sports/ exercise	Watching climbing, spelunking, caving	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Hunting	0.61	0.00	1.37	0.00	0.00	0.77	0.00
	Participating in martial arts	0.05	0.05	0.04	0.00	0.03	0.06	0.00
	Playing racquet sports	0.19	0.18	0.21	0.26	0.00	0.21	0.66
	Participating in rodeo competitions	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	Rollerblading	0.07	0.02	0.13	0.13	0.00	0.08	0.00
	Playing rugby	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Playing soccer	0.19	0.11	0.29	0.44	0.00	0.20	0.44
	Skiing, ice skating, snowboarding	0.18	0.10	0.28	0.00	0.00	0.22	0.44
	Playing hockey	0.04	0.00	0.08	0.00	0.00	0.05	0.00
	Softball	0.08	0.07	0.08	0.00	0.00	0.10	0.00
	Using cardiovascular equipment	0.44	0.42	0.46	0.75	0.40	0.43	0.40
	Vehicle touring/racing	0.07	0.00	0.17	0.29	0.00	0.08	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
	Playing volleyball	0.08	0.07	0.08	0.00	0.00	0.10	0.00
	Running	0.86	0.82	0.92	2.23	0.47	0.87	0.85
	Hiking	0.61	0.55	0.69	2.07	0.00	0.68	0.00
	Playing football	0.14	0.05	0.26	0.00	0.19	0.14	0.00
	Golfing	1.12	0.23	2.24	0.00	0.56	1.32	0.00
	Travel related to attending sporting/recreational events	0.39	0.36	0.43	0.22	0.02	0.47	0.17
	Travel related to participating in sports/exercise/recreation	2.28	1.67	3.04	2.66	1.41	2.41	2.58
	Doing aerobics	0.12	0.15	0.08	0.00	0.10	0.12	0.55
	Playing baseball	0.11	0.02	0.22	0.00	0.00	0.14	0.00
	Playing basketball	0.47	0.16	0.87	0.15	1.39	0.30	1.10
	Biking	0.67	0.33	1.10	0.15	0.26	0.77	0.88
	Doing gymnastics	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Playing billiards	0.19	0.03	0.38	0.37	0.50	0.10	0.88
	Bowling	0.19	0.18	0.20	0.00	0.00	0.23	0.33
	Climbing, spelunking, caving	0.02	0.00	0.05	0.00	0.00	0.03	0.00
	Participating in equestrian sports	0.17	0.16	0.18	0.15	0.00	0.21	0.00
	Fencing	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Fishing	1.14	0.10	2.44	0.00	1.46	1.14	1.10
	Walking	3.93	4.16	3.65	6.46	3.28	3.94	3.56
	Boating	0.36	0.41	0.29	0.15	0.00	0.44	0.00
	Watching bowling	0.03	0.03	0.03	0.00	0.00	0.03	0.00
	Participating in water sports	1.88	2.00	1.72	2.29	0.40	2.14	1.69
	Working out, unspecified	1.99	1.60	2.47	2.58	1.85	2.00	1.48
	Watching soccer	0.16	0.13	0.21	0.00	0.00	0.19	0.44
	Watching skiing, ice skating, snowboarding	0.00	0.01	0.00	0.00	0.00	0.01	0.00
	Watching running	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching rugby	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching rollerblading	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching rodeo competitions	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching racquet sports	0.01	0.01	0.00	0.00	0.00	0.01	0.00
	Watching softball	0.08	0.10	0.06	0.00	0.00	0.10	0.00
	Watching martial arts	0.01	0.00	0.03	0.00	0.00	0.01	0.00
	Watching gymnastics	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching golfing	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching football	0.32	0.21	0.46	0.71	0.14	0.35	0.00

Table B.2 continued from previous page

Category	Sub-category	Total	Female	Male	Asian	Black	White	Other
	Watching fishing	0.02	0.00	0.04	0.00	0.00	0.02	0.00
	Watching fencing	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching equestrian sports	0.04	0.06	0.00	0.00	0.00	0.04	0.00
	Watching dancing	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching hockey	0.11	0.10	0.13	0.00	0.00	0.14	0.00
	Weightlifting/strength training	0.84	0.43	1.35	1.04	0.77	0.83	1.08
	Watching vehicle touring/racing	0.11	0.03	0.21	0.00	0.00	0.13	0.00
	Watching walking	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Wrestling	0.01	0.00	0.03	0.00	0.00	0.01	0.00
	Doing yoga	0.33	0.47	0.16	1.34	0.09	0.34	0.00
	Watching aerobics	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching baseball	0.32	0.33	0.31	0.00	0.07	0.39	0.00
	Watching basketball	0.20	0.13	0.27	0.40	0.14	0.18	0.66
	Watching biking	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching billiards	0.02	0.03	0.01	0.00	0.00	0.03	0.00
	Watching volleyball	0.09	0.17	0.00	1.09	0.00	0.06	0.00
	Watching boating	0.02	0.00	0.04	0.00	0.00	0.02	0.00
	Security related to playing sports or exercising	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Waiting related to attending sporting events	0.06	0.05	0.06	0.00	0.03	0.07	0.00
	Waiting related to playing sports or exercising	0.06	0.08	0.04	0.00	0.09	0.05	0.18
	Watching wrestling	0.03	0.03	0.03	0.00	0.00	0.04	0.00
	Watching people working out, unspecified	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Watching weightlifting/strength training	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Watching water sports	0.04	0.00	0.10	0.00	0.00	0.06	0.00
	Security related to attending sporting events	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Dancing	0.37	0.37	0.37	0.22	0.29	0.41	0.07
TV	Television and movies (not religious)	179.41	166.65	195.34	108.45	231.30	173.63	166.49

Note: The table presents the re-categorization of ATUS sub-categories into fourteen leisure categories.

Table B.3: Expenditure

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
child care	Action Figures	15.14	15.04	16.10	16.71	15.27	14.85	15.01
	Baby & Child Care	6.48	6.43	7.13	8.84	6.65	7.01	6.09
	Bathing & Skin Care (Baby)	7.69	7.70	7.58	10.61	7.15	7.52	7.44
	Bedding & Decor (Baby)	18.91	18.77	20.44	20.35	17.74	20.25	18.65
	Boys Apparel	13.83	13.81	14.17	14.35	14.58	13.92	13.71
	Building (Toys)	15.79	15.72	16.66	16.95	14.65	15.17	15.84
	Celebrate Children's Books	3.97	3.98	3.48	N/A	N/A	4.49	3.95
	Character Corner	15.13	14.18	24.35	17.69	16.68	15.14	14.79
	Children	8.47	8.47	8.45	9.13	7.46	8.60	8.46
	Clothing	8.51	8.44	9.27	9.42	8.47	8.50	8.43
	Daycare & Learning	5.66	5.66	N/A	N/A	N/A	N/A	5.99
	Development	16.38	16.12	19.09	17.53	19.19	16.24	16.02
	Diapering	20.00	19.94	20.75	28.27	20.13	21.63	19.0
	Dolls	17.75	17.72	18.09	18.35	18.29	17.74	17.67
	Dress Up	11.72	11.73	11.62	13.11	12.39	12.29	11.45
	Electronics (Toys)	20.79	20.84	20.28	24.88	20.63	18.88	20.69
	Equipment (Baby)	23.25	23.12	24.46	24.83	24.62	27.47	22.45
	Girls Apparel	14.28	14.24	14.73	14.87	14.13	14.50	14.18
	Health (Baby)	7.93	7.89	8.35	8.75	6.39	7.36	7.99
	Infant Toddler Nutrition	20.51	20.36	21.84	24.32	22.06	21.69	19.84
	Kids & Teens Rooms	9.69	9.70	9.64	9.38	8.97	8.95	9.87
	Learning	7.45	7.39	8.34	9.39	7.03	6.79	7.42
	Lego	25.68	25.64	26.15	26.47	22.96	23.34	25.95
	Novelty (Toys)	6.88	6.93	6.09	7.46	5.09	6.28	7.03
	Novelty Apparel	14.19	14.31	12.81	17.80	12.19	14.3	13.99
	Nursery Furniture	30.91	30.58	33.98	35.15	28.37	35.62	30.42
	Potty	15.40	15.36	15.68	14.04	17.21	15.97	15.32
	Pregnancy & Maternity	20.68	20.73	20.20	18.20	25.90	15.32	20.39
	Pretend	10.96	10.88	11.92	13.30	10.00	11.16	10.86
	Riding Toys	36.02	36.07	35.49	40.47	43.58	37.19	34.93
	Safety (Baby)	14.20	14.23	13.63	14.09	14.88	14.52	14.04
	Shoes (Baby)	12.01	11.99	12.23	12.25	11.20	11.70	12.10
	Sports & Outdoor Play (Toys)	11.93	11.91	12.17	11.91	9.79	10.79	12.13
	Stuffed	8.14	8.15	8.12	9.14	8.14	9.86	7.93
	Toddler Furniture	30.49	29.74	37.61	22.26	35.29	36.26	30.59
	Toys	8.92	8.90	9.15	9.93	8.14	9.49	8.84

Table B.3 continued from previous page

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
	Toys (Baby)	13.00	13.01	12.84	15.74	11.61	12.42	12.87
	Vehicles	12.39	12.30	13.27	13.92	13.55	11.95	12.26
eating	Alcohol Beverages	16.57	16.66	16.04	16.58	13.97	16.09	16.85
	Appliances	7.93	8.18	4.67	6.35	9.05	4.86	7.95
	Bakery & Bread	5.87	5.92	5.52	5.73	5.29	5.87	5.92
	Baking & Cooking	6.93	6.97	6.65	6.93	6.57	6.82	6.97
	Balanced Nutrition and Snacks	9.53	9.56	9.34	11.20	9.04	9.64	9.45
	Beans & Grains	5.16	5.07	5.79	10.04	4.84	5.16	4.54
	Beverages	9.85	9.99	8.86	9.91	8.24	9.56	10.05
	Breakfast	7.30	7.33	6.97	7.28	6.98	7.23	7.33
	Canned	7.84	7.85	7.70	8.02	7.67	7.58	7.86
	Cigarettes and Loose Tobacco	20.59	20.26	22.39	30.58	14.14	23.18	20.28
	Cigars	13.57	13.27	14.78	20.16	6.73	32.26	12.23
	Condiments	6.08	6.12	5.80	5.95	5.29	5.63	6.18
	Dairy	10.15	10.25	9.40	9.00	8.10	9.58	10.48
	Deli & Prepared Foods	8.37	8.48	7.62	8.46	7.75	8.31	8.43
	Electronic Cigarettes	24.67	26.43	20.83	51.33	13.44	14.80	24.37
	Food	7.49	7.63	6.49	6.52	6.96	6.70	7.73
	Frozen Foods	11.53	11.61	10.91	10.71	11.24	11.03	11.68
	Gourmet	7.38	7.42	6.97	8.12	7.52	7.00	7.33
	Herbs & Spices	4.32	4.34	4.12	4.39	4.17	3.99	4.35
	Ice	5.28	5.26	5.37	6.18	4.75	5.39	5.23
	Kitchen & Dining	7.99	7.98	8.12	8.54	6.88	7.84	8.06
	Meal Solutions	6.35	6.36	6.26	7.17	6.08	6.34	6.31
	Meat	14.27	14.41	13.21	14.00	13.58	14.48	14.33
	Pasta & Noodles	4.57	4.55	4.65	5.44	4.08	4.41	4.53
	Performance Nutrition	12.36	12.57	11.13	13.92	10.84	13.41	12.25
	Produce	10.34	10.37	10.19	10.64	8.74	10.47	10.42
	Resto Beverages	15.82	15.86	15.99	9.94	18.77	16.00	15.66
	Seafood & Fish	12.68	12.72	12.45	13.29	13.48	13.37	12.30
	Smokeless Tobacco	19.83	20.36	17.84	50.11	11.99	18.61	19.00
	Snack	9.41	9.52	8.53	9.20	7.35	8.22	9.76
education	Office & School Supplies	9.92	9.95	9.63	9.58	7.94	9.47	10.21
	Textbooks	9.74	9.84	8.79	10.48	8.48	11.28	9.54
entertainment (not TV)	Audio	18.43	18.26	19.94	24.04	16.49	19.60	18.19
	Event Tickets	34.16	34.2	34.05	49.83	21.09	36.87	31.59
	Grown-Up Toys	6.39	6.38	6.56	6.54	5.39	5.29	6.49

Table B.3 continued from previous page

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
	Mobile App Downloads	59.77	59.16	63.77	70.54	52.90	45.34	59.56
	Music (Entertainment)	9.40	9.34	9.84	11.24	8.55	8.55	9.40
	Music (Toys)	14.83	15.00	13.74	21.87	9.94	16.67	14.73
	Tablets & eReaders	134.36	135.42	122.93	180.69	109.62	149.12	128.06
	Travel (Party & Occasions)	14.59	13.92	16.59	N/A	N/A	N/A	14.59
	Video Games, Consoles, & Accessories	36.97	37.02	36.53	41.94	41.45	42.71	35.70
	Virtual Reality	26.23	26.84	21.58	48.61	22.98	24.72	25.20
	iPods & MP3 Players	26.27	26.25	26.58	28.49	26.55	27.38	26.00
gardening/ pet care	Gardening & Lawn Care	13.78	13.65	14.93	14.05	13.65	12.50	13.89
	Pet Food & Treats	17.43	17.48	17.01	18.59	14.94	17.74	17.41
	Pet Supplies	13.24	13.20	13.55	14.60	11.39	12.28	13.31
hobbies	Adult Coloring Books	10.69	10.71	10.48	13.31	9.52	9.64	10.73
	Arts & Crafts	7.36	7.37	7.26	7.41	6.43	6.87	7.48
	Camcorders & Accessories	36.22	36.60	32.95	39.86	26.95	64.18	33.84
	Cameras & Camera Supplies	16.62	16.57	17.07	13.62	17.49	17.18	16.87
	Drones & Accessories	41.13	41.31	38.67	44.98	46.33	42.03	40.63
	Fan Shop	10.20	10.3	9.25	11.02	8.96	10.66	10.17
	Hobbies	7.62	7.65	7.40	9.20	6.65	9.69	7.47
	Musical Instruments	36.49	39.73	25.4	36.27	98.65	32.75	32.99
	Posters	5.48	5.48	N/A	N/A	4.98	N/A	5.99
	Sewing & Mending	4.39	4.37	4.69	4.04	3.63	4.07	4.45
own medical care	Health (Health & Beauty)	11.73	11.18	15.87	22.68	2.94	10.70	11.01
	Medical Products	9.38	9.35	9.66	11.15	8.25	9.10	9.35
	Personal Health Care	12.04	12.05	11.95	13.53	10.01	11.16	12.18
	Prescription (RX)	18.69	18.67	18.8	17.54	16.80	18.11	18.97
personal care	Apparel Accessories	10.18	10.15	10.53	10.99	9.62	9.98	10.19
	Bath & Body	7.25	7.29	6.89	9.53	8.56	8.07	6.76
	Deodorants & Antiperspirants	7.34	7.40	6.86	8.70	7.52	8.13	7.15
	Ear	7.08	7.06	7.24	7.62	6.44	6.84	7.07
	Eye	11.13	11.16	10.86	13.61	9.45	10.5	10.99
	Feminine Care	8.49	8.49	8.46	10.26	8.17	8.84	8.28
	Foot	9.85	9.82	10.11	12.96	8.50	9.40	9.75
	Fragrance	11.35	11.42	10.73	12.60	10.20	12.88	11.12
	Hair	10.42	10.48	9.78	13.05	10.05	12.62	9.89
	Hand	4.95	4.98	4.66	5.76	4.70	4.63	4.94
	Health (Health & Beauty)	5.62	5.61	5.77	6.22	3.99	5.07	5.98
	Makeup	10.55	10.57	10.16	11.95	9.15	11.26	10.43

Table B.3 continued from previous page

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
	Massage, Therapies & Relaxation	11.00	10.98	11.25	12.44	7.95	10.55	11.47
	Oral Hygiene	8.28	8.30	8.09	11.03	8.18	8.80	7.88
	Sexual Wellness Products	9.44	9.41	9.57	9.98	8.82	9.23	9.46
	Shaving & Hair Removal	11.72	11.76	11.40	14.13	11.54	13.55	11.35
	Skin	9.31	9.32	9.13	11.57	8.35	9.37	9.15
	Tools (Health & Beauty)	5.90	5.93	5.59	6.61	4.99	5.71	5.95
	Vision Care & Supplies	10.87	10.86	10.92	11.69	10.11	10.43	10.89
	Vitamins & Supplements	13.65	13.55	14.41	19.12	12.09	13.93	13.19
	Weight Management (Diet)	12.66	12.68	12.55	13.68	12.41	12.92	12.61
reading	Archive	7.85	7.80	8.41	6.06	6.49	10.86	7.95
	Art (Books)	8.10	8.07	8.46	8.11	10.37	6.70	8.00
	Biography & Memoirs	11.93	11.87	12.4	14.58	15.15	10.41	11.77
	Business & Investing	10.76	10.85	10.12	13.82	10.42	10.71	10.67
	Computing & Internet (Books)	6.98	6.81	9.23	6.36	7.84	6.13	7.16
	Cooking, Food & Beverages	8.26	8.26	8.24	8.81	9.58	9.30	7.94
	Health, Mind & Body (Books)	7.12	7.17	6.60	8.45	8.32	7.04	6.80
	History	11.82	11.82	11.76	10.82	9.90	15.72	11.84
	Home, Hobbies & Garden	7.39	7.39	7.40	7.95	7.37	6.51	7.43
	Journals	5.22	5.28	4.71	5.27	5.86	6.79	5.06
	Literature & Fiction	9.37	9.32	9.78	9.73	8.62	9.77	9.38
	Magazines	10.62	10.52	11.46	11.78	8.15	9.18	10.75
	Miscellaneous (Books)	11.70	11.77	10.82	11.47	10.86	11.65	11.51
	Music Books	1.46	1.50	0.80	3.12	0.88	0.50	1.53
	Mystery & Suspense	8.32	8.36	7.77	11.57	6.95	9.61	7.99
	Parenting & Families	9.49	9.69	3.95	11.56	N/A	27.40	8.61
	Performing Arts	4.32	4.32	N/A	N/A	N/A	N/A	4.32
	Photography	11.39	11.39	N/A	10.68	N/A	N/A	11.02
	Political & Social Sciences	9.05	8.04	14.00	14.18	10.62	5.92	8.53
	Reference	4.62	4.56	5.16	7.20	4.98	3.94	4.29
	Romance	13.87	13.14	19.22	12.26	10.06	13.33	13.83
	Sports & Recreation (Books)	7.35	7.32	7.67	8.55	5.31	6.19	7.62
	Travel & Nature (Books)	5.41	5.35	6.00	5.37	3.95	5.56	5.48
religious/ civic activities	Bibles	12.94	12.94	13.02	21.57	12.90	12.15	11.88
	Religion	13.53	13.78	11.53	10.26	12.37	15.12	13.88
sleeping	Bedding (Home & Garden)	24.04	23.96	24.88	25.64	24.06	24.47	23.84
socializing	Balloons	9.06	9.07	9.00	9.02	8.74	8.90	9.18
	Banners, Streamers & Confetti	3.43	3.52	2.45	2.67	2.42	3.71	3.54

Table B.3 continued from previous page

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
	Birthday	17.76	17.69	18.73	20.37	14.29	16.16	17.94
	Cake Supplies	3.82	3.82	N/A	N/A	0.43	1.04	5.10
	Cell Phones	50.91	51.09	49.31	44.96	47.94	45.44	51.61
	Christmas	12.14	12.07	12.98	14.11	11.63	13.18	11.89
	Easter	5.99	5.98	6.20	5.58	7.57	6.57	5.86
	Fathers Day	4.16	4.16	N/A	N/A	12.95	N/A	3.53
	Floral	13.87	13.77	14.52	15.53	15.29	14.28	13.54
	Funeral	10.41	10.41	N/A	22.00	N/A	N/A	8.48
	Games	17.14	16.88	19.53	18.37	17.08	17.30	17.01
	Gift Bags & Wrapping Paper	7.61	7.61	7.59	8.62	6.58	6.46	7.73
	Gift Cards	34.17	33.74	38.30	32.95	28.27	29.18	35.58
	Gift Registry	15.61	15.46	16.82	23.50	10.85	19.11	15.02
	Gift Sets (Health & Beauty)	12.58	12.80	10.3	12.35	13.41	16.80	12.16
	Gifts (Baby)	14.51	14.52	14.27	15.50	12.96	12.71	15.01
	Gifts (Party & Occasions)	6.83	6.82	7.00	7.50	5.81	5.79	6.93
	Halloween	6.63	6.56	7.55	4.07	8.15	5.86	6.76
	Holiday Guest Headquarters	6.43	6.25	9.00	9.00	9.00	12.00	4.95
	Invitations & Cards	6.08	6.08	6.07	6.05	5.58	5.37	6.15
	Mothers Day	7.67	7.72	7.33	8.60	6.50	7.65	7.64
	Noisemakers	2.33	2.36	2.14	N/A	2.00	N/A	2.38
	Party Supplies	7.89	7.87	8.31	9.64	7.16	8.42	7.76
	Phones & Two-way Radios	18.13	18.40	16.56	28.78	19.28	21.02	17.33
	Puzzles	7.55	7.56	7.54	8.95	6.19	6.42	7.68
	Stationery (Baby)	11.10	11.12	10.9	10.56	9.87	7.94	11.40
	Table covers, Tableware & Centerpieces	6.55	6.48	7.45	6.75	6.52	6.69	6.58
	Thanksgiving	3.59	3.94	0.50	0.50	N/A	5.82	3.70
	Tickets	53.00	54.65	46.66	54.10	44.8	69.08	47.38
	Wedding	4.84	4.91	4.00	4.38	3.74	4.33	5.02
sports/ exercise	Accessories (Sports)	6.07	6.05	6.37	9.99	8.09	6.62	5.75
	Action & Extreme Sports	27.7	27.98	25.56	28.91	31.99	27.06	27.46
	Cargo Storage & Racks	13.66	13.66	N/A	12.99	N/A	N/A	12.99
	Dance & Gymnastics	10.49	10.49	N/A	N/A	N/A	N/A	10.49
	Exercise & Fitness	13.17	13.01	14.74	15.69	12.44	12.64	13.11
	Home - Indoor / Game Room	5.85	6.03	4.71	4.46	6.13	4.43	6.02
	Home - Outdoor	15.45	15.35	16.52	15.77	13.41	17.07	15.39
	Leisure Sports & Games	16.35	16.50	14.91	17.18	11.84	15.54	16.53
	Outdoors	15.17	15.11	15.63	18.75	14.12	15.11	14.95

Table B.3 continued from previous page

Category	Department	Total	Female	Male	Asian	Black/AA	His./Latino	White/Cau.
	Racquet Sports	8.12	8.11	8.20	8.15	7.60	9.06	8.09
	Team Sports	15.86	15.78	16.74	17.36	18.59	15.66	15.60
	Wearable Technology	105.49	104.61	113.53	132.18	98.00	116.79	102.26
TV	Home Audio & Theater	27.72	27.50	29.47	30.90	21.68	30.71	27.94
	Movies & TV	17.19	17.07	18.32	18.16	16.04	16.17	17.30
	TV & Video	134.49	135.28	128.17	119.13	137.03	141.03	134.84

Note: The table presents the product Departments that form the fourteen leisure categories.

Table B.4: Activity Examples

Classification	Time-spend Data 3rd-tier Activity Examples	Consumption Data Product Examples
Child care	Activities related to household/non-household children's education, caring for & helping household/non-household children	Baby or children's daily necessities, books, equipment, and furniture
Eating	Eating and drinking, waiting associated w/eating & drinking, travel related to eating and drinking	Food, drinks, condiments, herbs and spices, and kitchen supplies
Education	Taking classes, extracurricular school activities (except sports), research/homework, registration/administrative activities, travel related to education	School supplies and textbooks
Entertainment	Relaxing, thinking, using tobacco, listening music, playing games, attending arts events, and travel related to arts and entertainment	Audio, tobacco, events tickets, grown-up toys, video games tablets and eReaders
Gardening/Pet Care	Household activities related to garden, gardening, and pet/veterinary services, and travel related to these activities and services	Gardening and lawn appliances, pet food and supplies
Hobbies	Writing for personal interest, arts and crafts as a hobby, collecting as a hobby, hobbies, except arts & crafts and collecting	Action figures, arts and crafts, camera supplies, sewing and mending, and musical instruments
Own Medical Care	Personal emergencies and waiting for and using health care services	Prescriptions, and medical products
Personal Care	Washing, dressing, grooming, waiting for and using health-related self-care, and telephone calls and travel related to personal care	Apparel accessories, deodorants, makeup, personal-care products, and massage
Reading	Reading for personal interest	Books, magazines, and archives
Religious/Civic Activities	Religious and spiritual activities, volunteer activities, civic obligations and participation telephone calls and travel related to government and volunteer	Apparel accessories, deodorants, makeup, Bibles or religious related books
Sleeping	Sleeping and sleeplessness	Bedding
Socializing	Attending/hosting meetings/parties/receptions/ceremonies, telephone calls to/from families and friends, and travel related to social events	Gifts, festival/occasions/party decorations and preparations
Sports/Exercise	Attending/participating sports events/exercises, travel/waiting related to sports events/exercises	Outdoor/indoor sports equipment, sporting goods, and wearable technology
TV	Television and movies (not religious)	TV, movies, and home audio/theater

Note: The leisure activities classification adopted in Aguiar and Hurst (2007*a*).

Table B.5: Average Time Spend on Leisure Activities Per Person in a Day by State (in Minutes)

State	Child Care	Eating	Education	Entertainment (Not TV)	Gardening/ Pet Care	Hobbies	Own Medical Care	Personal Care	Reading	Religious/ Civic Activities	Sleeping	Socializing	Sports/ Exercise	TV	Respondents
AK	33.36	133.32	13.41	94.64	10.05	4.41	0.00	45.91	12.05	23.68	550.45	68.86	8.55	118.50	22
AL	27.26	100.60	12.53	47.35	30.41	1.87	2.32	54.86	11.16	29.87	527.80	51.69	15.40	225.92	193
AR	30.27	100.13	29.84	41.38	31.50	3.42	0.87	57.12	19.06	26.44	565.64	64.87	10.21	209.64	117
AZ	55.49	102.81	15.25	48.81	20.03	6.14	1.02	43.69	21.20	29.55	534.01	46.25	23.29	181.76	236
CA	46.57	110.76	24.43	56.09	21.40	5.23	3.05	51.15	20.20	20.98	529.03	57.76	28.19	157.87	994
CO	43.47	114.77	24.69	43.76	22.63	2.97	0.96	41.28	24.82	14.15	525.18	54.56	30.93	165.41	172
CT	38.77	111.75	3.36	46.73	18.88	2.64	3.76	46.74	18.88	26.66	534.44	60.26	19.98	183.02	125
DC	42.48	91.09	31.06	43.42	10.52	0.00	6.82	60.52	33.30	8.03	549.33	46.82	13.91	136.48	33
DE	18.55	101.30	0.00	54.75	8.75	0.00	1.50	60.20	28.00	29.25	579.80	56.10	29.75	124.45	20
FL	31.46	104.02	14.57	57.72	16.60	3.44	3.39	55.08	24.61	31.37	538.25	51.13	18.40	187.78	619
GA	35.98	102.50	9.12	55.77	16.66	3.19	5.15	52.45	15.70	30.29	549.35	62.33	22.58	198.01	340
HI	61.11	116.91	0.00	99.09	35.03	6.57	0.00	48.43	50.09	34.14	497.40	46.69	28.00	136.80	35
IA	46.92	106.31	8.08	61.32	18.00	5.82	3.30	44.57	20.31	21.06	530.75	64.77	16.11	173.08	142
ID	59.25	97.78	18.27	63.71	24.33	1.75	0.71	34.52	17.59	25.86	540.67	91.71	26.35	122.32	63
IL	41.49	105.90	15.72	48.64	16.02	4.98	5.02	48.82	17.94	27.09	526.34	56.83	26.78	179.39	442
IN	33.57	94.00	19.02	62.68	24.18	6.07	2.51	44.65	18.07	25.43	541.57	48.09	18.20	182.31	242
KS	37.22	95.90	14.12	44.59	19.29	13.68	5.68	47.98	20.11	35.64	542.86	47.46	27.35	165.71	133
KY	50.98	92.54	8.83	61.72	27.18	2.56	1.11	52.59	21.47	21.78	534.38	65.31	17.18	191.01	180
LA	34.00	105.71	14.50	57.62	21.24	2.90	3.43	53.88	17.74	30.79	557.22	57.10	10.89	194.24	169
MA	38.95	103.28	27.71	46.19	19.00	3.19	4.01	44.85	29.84	25.77	528.07	66.12	22.03	175.00	203
MD	30.46	107.85	27.24	54.19	11.74	8.30	5.53	49.82	24.70	31.07	513.36	58.34	24.26	187.64	188
ME	58.40	109.33	1.56	60.06	32.62	11.25	0.62	35.88	28.65	24.08	534.83	92.85	31.67	132.75	48
MI	43.63	102.74	16.23	65.31	17.19	4.75	5.16	46.97	22.65	23.05	533.86	64.14	23.52	180.37	329
MN	38.80	114.04	16.84	53.03	21.72	11.42	3.39	38.97	23.54	19.48	532.16	55.12	35.47	159.61	227
MO	27.74	92.24	12.99	60.12	27.12	1.42	3.52	44.61	24.18	24.45	547.94	61.92	28.44	185.31	198
MS	25.22	85.02	16.64	78.59	12.84	1.94	5.90	53.32	14.25	31.09	567.13	47.64	15.20	202.87	134
MT	56.64	119.54	16.15	53.97	31.36	4.36	0.51	29.79	41.69	25.62	523.46	75.23	33.44	145.69	39
NC	38.50	98.33	20.43	58.29	16.68	6.82	5.14	53.59	16.48	26.95	538.99	49.65	21.05	193.00	311
ND	15.88	78.47	21.47	69.79	15.88	15.74	0.44	43.82	42.76	16.24	490.00	55.47	16.62	180.38	34
NE	49.01	89.92	5.07	57.03	16.26	5.07	2.57	40.19	16.53	14.15	509.28	47.96	23.57	174.00	74
NH	23.51	109.27	21.89	74.65	25.35	4.05	1.49	37.19	21.22	30.68	521.92	49.27	30.27	138.00	37
NJ	36.82	114.99	17.64	47.62	10.66	3.89	2.72	47.31	18.59	25.12	533.02	50.53	18.85	174.41	311
NM	33.22	110.97	20.24	49.22	29.52	5.24	4.03	50.49	18.91	35.08	530.34	52.90	28.51	149.03	103
NV	35.49	99.62	16.81	51.07	17.49	7.23	2.31	43.71	24.10	12.03	531.87	64.84	20.69	214.18	91
NY	38.43	109.95	10.95	64.12	11.73	3.03	5.64	45.97	24.36	19.46	523.94	65.85	27.50	173.90	526
OH	38.62	97.31	12.07	61.00	21.47	5.42	3.87	47.22	21.31	16.80	533.29	55.39	19.06	196.04	356
OK	44.39	103.59	15.84	53.82	32.71	1.94	6.21	40.94	17.63	34.33	542.48	48.84	14.67	171.22	147
OR	40.04	102.36	12.68	50.43	28.04	0.53	5.02	37.04	24.29	21.95	532.66	61.34	22.07	171.33	169
PA	35.12	105.46	15.24	47.50	18.06	2.94	3.63	42.08	22.37	24.04	530.54	64.25	25.71	183.95	424
SC	43.13	100.62	10.13	58.87	22.58	6.21	4.51	56.33	14.36	31.77	544.17	60.37	14.37	199.13	203
SD	28.55	83.94	5.32	66.39	22.65	8.55	0.00	46.45	21.77	36.06	534.61	68.26	33.90	161.45	31
TN	36.39	97.70	12.84	56.36	13.87	6.38	2.99	51.42	17.53	30.64	527.64	62.68	14.16	198.51	245
TX	39.96	106.40	15.94	44.97	24.72	2.64	3.52	49.31	11.60	32.59	534.72	54.10	19.57	180.09	801
UT	62.58	113.48	17.97	36.46	25.93	3.36	1.64	47.49	30.95	62.16	513.59	40.24	28.99	142.61	110
VA	32.94	104.93	18.82	52.17	20.80	2.93	3.81	48.08	25.98	19.93	536.00	51.57	21.61	188.74	297
VT	92.19	111.43	2.14	35.95	15.24	0.00	0.00	35.00	25.62	5.24	503.05	28.43	44.00	136.48	21
WA	36.04	104.84	16.34	57.38	23.31	7.00	4.48	44.94	31.84	13.16	539.12	52.05	33.09	158.59	232
WI	47.07	103.65	18.18	57.22	17.35	9.08	4.18	44.34	17.63	20.85	537.03	55.19	28.05	192.22	195
WV	31.79	112.68	8.10	58.86	16.41	2.62	2.03	38.54	17.98	12.14	538.13	50.9	19.60	199.70	63
WY	27.52	88.71	43.19	50.10	11.05	0.00	18.81	55.67	17.67	69.14	489.29	87.67	13.86	154.67	21

Note: The per person time allocation of a day across leisure activities is measured in minutes. The last column presents the number of respondents in a state in the year 2016.

Table B.6: Average Expenditure on Leisure Activities Per Receipt/Trip by State (in Dollars)

State	Child Care	Eating	Education	Entertainment (Not TV)	Gardening/ Pet Care	Hobbies	Own Medical Care	Personal Care	Reading	Religious/ Civic Activities	Sleeping	Socializing	Sports/ Exercise	TV	Participants
AK	22.12	36.89	8.48	40.18	20.56	8.84	14.87	15.41	6.73	9.69	28.4	19.71	27.71	23.82	352
AL	17.19	28.26	8.50	23.10	16.48	7.20	12.80	12.59	6.09	11.20	21.9	15.52	17.48	30.65	2987
AR	16.06	29.70	9.59	22.96	15.84	7.60	12.03	12.76	6.37	9.83	21.3	18.39	16.58	27.79	1941
AZ	19.75	28.63	9.85	27.51	18.79	7.44	13.89	13.98	6.90	15.23	24.98	18.29	19.59	27.99	2452
CA	21.95	28.87	9.80	30.73	20.05	8.17	13.87	16.06	7.70	16.76	26.58	19.11	23.64	29.26	15827
CO	19.68	31.60	9.77	26.46	19.62	8.14	14.28	14.52	7.45	22.90	25.88	21.07	22.54	26.56	1866
CT	20.37	31.32	11.70	16.46	18.73	8.06	13.54	14.05	6.05	15.99	22.91	15.37	21.58	31.73	1547
DC	20.25	22.83	8.59	15.53	15.75	9.32	14.74	11.96	5.27	10.90	28.11	11.49	15.93	34.45	135
DE	18.83	31.15	9.05	22.83	18.57	7.61	13.61	13.86	5.17	9.23	26.50	15.00	24.12	32.14	529
FL	18.66	32.17	9.21	25.88	18.61	7.44	13.46	13.65	6.74	12.39	24.39	17.42	20.05	29.49	10785
GA	18.19	28.25	9.14	25.45	17.56	7.14	13.41	13.00	6.35	13.48	24.26	16.97	18.46	28.63	4887
HI	23.52	28.34	6.41	37.38	23.81	9.32	13.69	16.65	8.80	12.38	27.55	22.45	26.55	29.78	1120
IA	19.44	27.99	6.83	24.98	17.17	8.15	12.65	13.91	6.80	12.72	21.63	15.82	19.96	29.51	1442
ID	16.69	31.41	9.94	25.51	17.64	7.15	12.64	13.14	6.60	7.08	23.11	15.14	16.57	23.97	624
IL	19.14	30.35	11.99	26.41	18.05	8.09	13.61	14.35	6.09	12.00	23.29	16.13	22.71	29.50	7100
IN	20.07	32.82	9.64	25.25	18.06	7.67	13.48	13.82	5.99	9.58	22.38	16.18	20.78	29.62	3957
KS	18.71	28.63	8.55	25.68	17.21	7.71	13.18	13.36	6.98	7.55	22.39	16.77	18.46	25.72	1809
KY	18.22	31.34	11.15	24.32	16.95	7.61	13.17	12.60	5.61	10.28	26.07	15.69	19.65	30.42	3123
LA	18.11	30.84	10.99	22.66	17.01	7.05	12.77	13.17	5.85	17.13	22.53	15.01	17.53	28.57	2525
MA	20.68	31.41	12.42	30.96	16.95	7.30	13.16	14.45	5.98	13.84	24.25	15.95	21.30	28.35	2267
MD	20.64	30.89	11.02	30.94	19.55	7.87	14.12	14.62	7.44	7.49	23.92	16.46	20.05	29.58	2671
ME	16.88	30.54	11.61	25.71	17.25	7.57	12.44	12.99	4.98	11.21	19.57	17.02	18.35	28.85	565
MI	20.04	36.30	10.49	29.11	18.54	7.86	13.69	14.33	6.40	13.17	23.73	15.61	23.13	27.40	6008
MN	21.76	31.88	10.26	27.47	18.87	8.85	13.75	15.52	7.19	13.75	25.77	17.56	22.01	32.55	2359
MO	17.59	28.99	9.53	25.78	16.96	7.75	12.55	13.59	6.54	7.96	21.28	15.50	18.12	27.34	4079
MS	17.08	29.28	7.46	22.36	15.75	7.61	12.46	12.57	5.79	15.00	21.41	14.64	17.44	31.50	1619
MT	22.04	33.16	5.99	29.61	19.25	7.67	14.50	14.78	7.77	28.33	26.34	18.20	22.60	30.35	272
NC	18.05	28.65	8.05	25.11	17.20	7.56	13.38	12.97	6.53	9.07	23.42	15.28	18.96	28.17	5626
ND	21.94	34.74	14.07	30.54	20.99	8.90	15.18	16.00	6.79	22.21	24.70	17.59	20.42	25.65	529
NE	20.87	31.73	11.32	25.01	18.52	8.09	14.17	14.46	10.62	10.15	22.28	16.58	20.00	27.40	1033
NH	19.61	34.16	10.37	27.13	16.78	8.21	12.14	13.07	5.65	17.49	23.94	16.35	23.63	25.73	492
NJ	19.94	33.87	13.41	26.35	18.97	7.95	14.28	14.26	5.86	12.08	26.05	15.76	22.42	31.18	5285
NM	17.82	32.17	11.69	27.75	17.70	7.71	12.10	13.78	7.24	31.25	24.79	16.28	18.26	29.53	560
NV	19.64	30.73	7.58	19.84	19.48	8.15	13.23	15.10	6.78	9.87	24.09	19.20	23.11	28.07	1282
NY	19.31	27.51	10.02	23.72	17.71	7.67	13.38	13.62	5.48	14.25	25.06	15.94	22.41	29.18	8881
OH	19.74	32.67	9.52	23.86	17.65	7.61	13.13	13.48	6.23	12.45	23.25	18.02	20.91	29.57	6417
OK	17.44	28.17	7.29	26.82	17.68	7.57	12.29	13.04	5.77	8.85	22.26	16.16	16.99	26.44	2094
OR	18.53	28.29	9.17	26.26	19.52	7.74	13.91	13.87	7.24	14.20	23.66	17.13	18.86	23.96	1389
PA	19.48	30.44	10.56	25.21	17.95	7.33	13.27	13.67	5.76	9.93	22.58	17.01	18.88	28.40	9397
SC	18.11	29.96	10.63	22.11	17.45	7.23	13.07	13.01	5.61	12.56	23.15	16.54	18.07	28.47	2477
SD	19.44	32.08	7.85	25.55	18.05	8.20	12.20	13.99	6.32	19.09	32.38	17.94	19.95	28.96	380
TN	17.01	28.84	7.84	23.11	16.28	7.45	12.39	12.82	5.97	12.79	22.82	17.90	18.26	26.06	3446
TX	19.87	29.20	9.66	25.91	17.92	7.74	13.52	13.95	7.32	13.39	24.60	17.02	20.12	27.78	11529
UT	20.27	29.83	8.07	26.73	17.81	7.35	13.53	13.30	7.03	11.42	25.46	16.41	19.83	25.34	1066
VA	18.92	29.81	7.94	27.21	18.16	7.76	13.71	13.42	7.42	12.14	23.50	16.22	19.25	29.01	4825
VT	20.95	33.74	5.99	18.79	14.58	8.61	12.31	11.11	5.68	N/A	20.16	17.16	19.80	17.77	138
WA	20.08	29.80	9.18	29.13	18.92	7.67	13.33	14.14	6.84	16.47	25.55	18.28	22.22	24.97	2482
WI	19.45	29.74	10.33	26.67	16.17	8.04	12.78	13.69	6.20	13.99	22.86	15.30	20.86	28.76	3496
WV	18.02	32.05	9.65	26.24	18.10	7.73	13.39	12.78	5.30	12.56	23.25	15.89	20.43	29.12	1436
WY	20.11	36.36	7.07	30.80	19.34	8.10	13.77	14.43	6.72	8.35	21.72	19.79	20.17	27.40	210

Note: The table displays the expenditure per person across products spending on different leisure categories. The last columns list the participants (buyers) for each state.

Table B.7: Leisure Activities Price Indexes for Each State (Base Period: National Average)

State	Child Care	Eating	Education*	Entertainment (Not TV)	Gardening/ Pet Care	Hobbies	Own Medical Care	Personal Care	Reading	Religious/ Civic Activities*	Sleeping	Socializing	Sports/ Exercise	TV
AK	1.16	1.30	0.15	1.42	1.27	1.11	1.13	1.15	0.85	N/A	1.26	1.34	1.20	0.87
AL	0.93	0.94	0.59	0.92	0.92	0.95	0.89	0.91	0.90	0.92	0.92	0.96	0.97	0.93
AR	0.86	0.94	0.17	1.01	0.93	0.97	0.90	0.91	0.91	1.23	0.87	0.90	0.91	0.92
AZ	1.06	0.97	0.77	1.13	1.07	0.97	1.03	1.02	0.97	0.83	1.11	1.06	0.89	0.98
CA	1.12	1.06	0.92	1.19	1.13	1.06	1.10	1.14	1.02	0.90	1.21	1.13	1.12	1.06
CO	1.06	1.03	0.86	0.96	1.19	1.04	1.04	1.03	1.14	0.66	1.12	1.07	1.03	1.00
CT	0.97	1.03	1.59	0.61	0.86	1.08	1.05	1.03	0.92	0.57	0.93	1.04	1.18	1.08
DC	1.12	0.96	N/A	0.86	0.80	1.11	1.20	1.02	0.86	N/A	1.20	1.08	0.81	1.08
DE	0.92	1.04	0.63	0.95	0.96	1.01	1.09	0.99	0.87	N/A	1.09	0.98	1.16	1.03
FL	1.00	1.06	0.70	0.99	0.99	0.98	1.00	1.00	0.94	0.58	0.96	1.02	1.03	0.96
GA	1.00	0.95	0.59	1.05	0.95	0.96	0.93	0.94	0.97	1.22	0.96	0.98	0.98	0.91
HI	1.17	1.33	0.69	1.21	1.44	1.10	1.09	1.22	1.03	N/A	1.25	1.27	1.19	1.11
IA	0.96	0.95	0.77	0.98	1.02	1.00	0.98	0.98	0.91	1.40	0.91	1.05	0.94	1.00
ID	0.88	0.98	0.31	0.97	1.15	0.92	0.92	0.94	1.23	N/A	1.16	1.05	0.83	0.98
IL	0.96	0.96	1.15	1.04	0.99	1.01	1.04	1.02	1.00	0.74	0.98	0.98	1.07	1.03
IN	1.02	0.96	0.69	1.01	1.01	0.99	0.99	0.97	0.96	0.66	0.93	0.94	0.96	0.91
KS	0.99	0.97	0.69	0.95	1.12	0.96	0.99	0.98	0.89	0.92	0.91	1.01	0.98	0.96
KY	0.99	0.94	1.28	0.93	0.91	0.97	0.92	0.91	0.97	0.82	0.97	0.92	0.96	0.96
LA	0.97	1.02	1.14	0.94	0.95	0.95	0.96	0.95	0.89	0.87	0.93	0.94	0.94	0.95
MA	1.03	1.00	1.15	1.10	0.85	1.01	1.07	1.05	0.95	0.93	0.98	1.02	1.02	0.97
MD	1.01	1.05	1.20	1.06	0.97	1.03	1.06	1.05	1.02	0.80	1.02	1.04	0.81	1.02
ME	0.85	0.97	0.74	1.00	0.89	0.97	0.92	0.91	1.18	0.77	0.86	0.95	0.92	0.96
MI	1.04	1.00	0.94	0.99	1.03	1.03	1.03	1.01	1.09	0.99	1.03	0.96	1.07	1.00
MN	1.04	1.04	0.88	1.03	1.10	1.06	1.06	1.07	0.88	1.07	1.03	1.04	1.08	1.20
MO	0.90	0.95	1.00	1.01	1.05	0.98	0.98	0.97	0.98	0.83	0.87	0.95	0.95	0.98
MS	0.93	0.96	0.45	0.87	0.97	0.94	0.88	0.91	1.04	1.04	0.86	0.90	0.91	0.90
MT	1.12	1.11	1.04	1.06	1.32	1.00	1.11	1.03	0.82	1.16	1.15	1.04	0.99	1.20
NC	0.94	0.98	1.11	0.99	0.97	0.99	0.95	0.95	1.00	0.94	0.95	0.94	0.97	0.94
ND	1.06	1.11	0.98	1.23	1.22	1.00	1.07	1.06	1.10	N/A	0.92	1.11	0.97	1.09
NE	0.98	1.01	1.13	1.02	1.09	0.98	1.06	1.02	1.09	0.57	0.91	1.01	0.99	0.98
NH	0.96	0.97	1.40	1.07	0.84	1.04	0.98	0.97	0.92	1.10	0.91	1.04	1.02	0.97
NJ	0.99	1.07	1.27	1.07	0.88	1.07	1.09	1.04	1.16	0.85	1.05	1.05	1.08	1.10
NM	0.91	1.01	2.70	1.03	1.13	0.99	0.96	0.96	1.33	N/A	0.97	1.03	1.02	1.05
NV	1.05	1.02	0.53	1.02	1.04	1.02	1.01	1.05	1.10	N/A	1.02	1.13	1.02	1.01
NY	1.02	1.03	1.07	0.89	0.94	1.03	1.05	1.03	0.88	1.00	1.03	0.99	1.06	1.05
OH	1.02	0.98	0.77	0.94	1.00	0.97	0.97	0.96	1.12	0.94	0.94	0.92	0.98	1.03
OK	0.89	0.94	1.21	1.03	1.02	0.99	0.93	0.93	0.96	0.82	0.90	0.98	0.91	0.93
OR	0.99	0.99	1.77	0.88	1.17	1.00	1.02	1.01	1.21	N/A	1.14	1.03	0.99	0.97
PA	0.98	1.03	1.10	0.95	0.96	1.00	0.99	0.98	1.09	0.94	0.93	0.97	0.89	1.01
SC	0.93	0.99	1.37	0.94	0.94	0.97	0.93	0.95	0.96	0.85	0.99	1.00	1.03	0.95
SD	0.92	1.04	0.96	0.96	1.21	1.01	0.95	1.00	0.83	N/A	0.84	1.11	0.95	1.02
TN	0.96	0.93	0.69	0.93	0.91	0.99	0.91	0.91	0.99	1.89	0.91	0.93	0.99	0.91
TX	1.01	0.93	0.82	0.98	1.08	0.96	0.99	0.99	0.99	1.13	0.99	1.00	0.96	0.99
UT	1.01	1.03	0.70	1.03	1.15	0.96	0.98	0.97	1.00	1.22	1.14	1.02	0.98	1.02
VA	0.99	0.99	0.28	0.97	0.99	0.99	1.01	0.98	0.94	1.40	0.98	0.99	0.99	1.02
VT	1.27	1.01	N/A	0.74	0.79	1.11	1.03	0.91	0.90	N/A	0.87	1.02	0.88	1.00
WA	1.02	1.04	0.46	1.12	1.12	1.00	1.03	1.05	1.21	0.26	1.22	1.11	1.06	1.00
WI	0.95	0.95	0.85	1.03	1.02	1.04	1.01	0.98	0.89	0.91	0.96	0.99	1.03	1.05
WV	1.00	0.98	1.08	0.97	0.97	1.01	0.92	0.91	0.90	1.53	0.91	0.93	1.04	1.02
WY	0.97	1.05	1.19	0.99	1.10	1.00	1.00	0.99	0.88	N/A	0.93	1.12	0.91	1.13

Note: Price indices are calculated using price ratios. Leisure activities marked with an asterisk (∗) have very few observations, as shown in Table 2a. Both the price and quantity of commodities are demeaned within their respective *major categories*. The national average serves as the base period for each leisure activity. The state price indices for *education* and *religious/civic activities* are skewed due to limited observations in states like Alaska and Minnesota.

Appendix C

Supplementary material for Chapter 3

C.1 World Cities with Hornet App Users

Taken from the database, the app city coverage is indicated in Figure C.1 below.

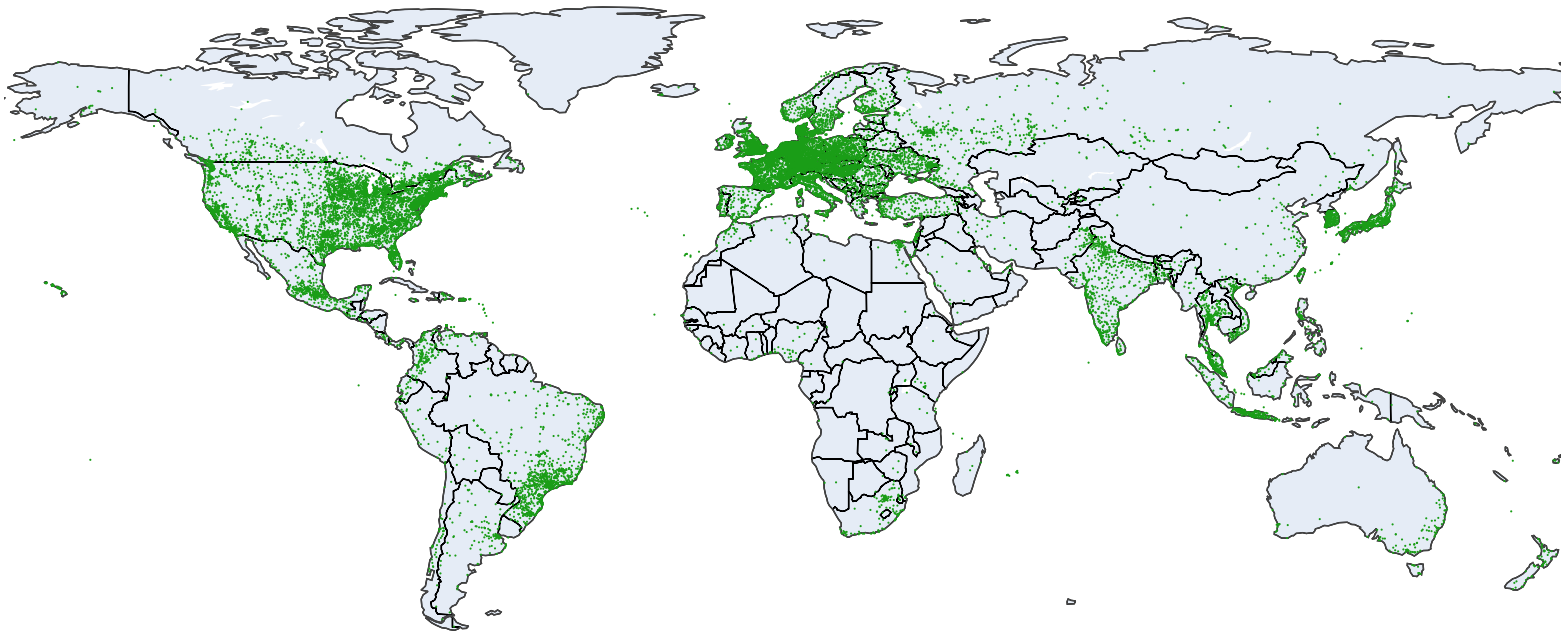

Figure C.1: Hornet Database World City Coverage

Note: Each green dot represents a city where Hornet users are located.

C.2 Active Rate of Refugee Users

The active rate is calculated after imputing the 30-day extension missing data in the user log files and limiting the dates starting from January 1st, 2022.

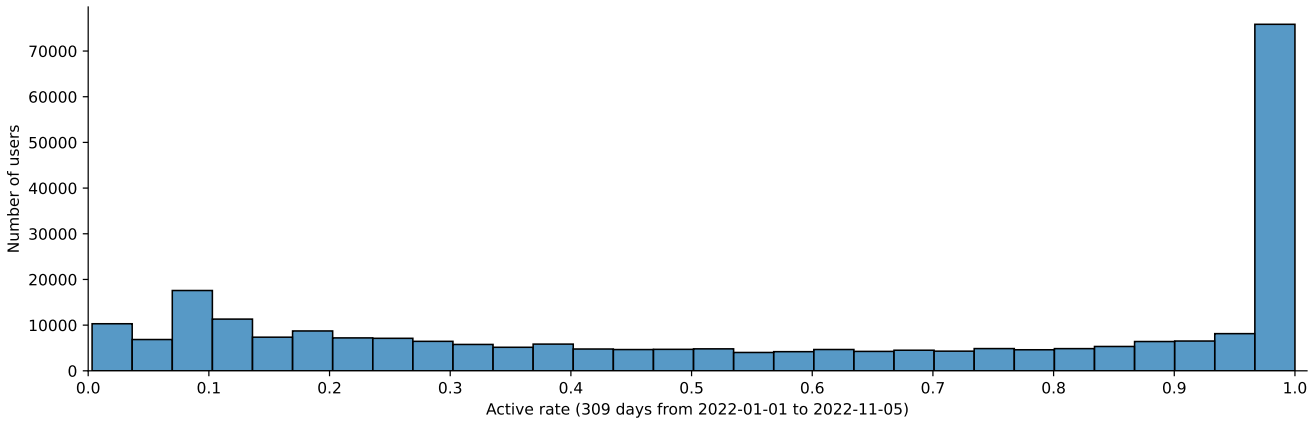

Figure C.2: User Active Rate Distribution

Note: The active rate is calculated by examining the frequency of app usage over 309 days.

C.3 Active Users After Sample Selection

After the 30-day user log extension and churn control, the number of daily active users in all three groups remains consistent, with only minor fluctuations of a few hundred users.

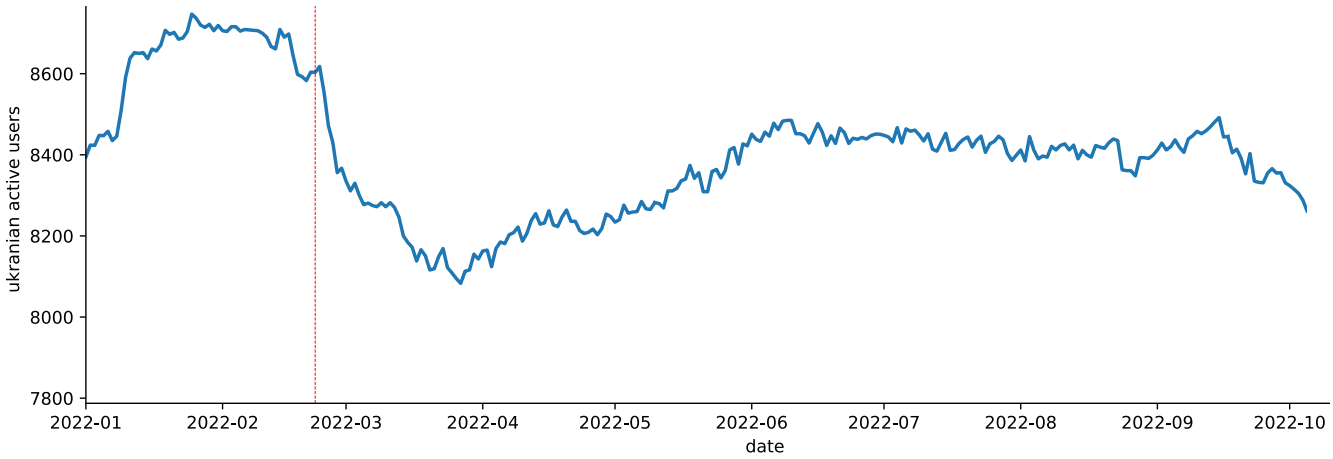

Figure C.3: Daily Active Ukrainian Refugee Users

Note: The number of active users experiences a decline immediately following the start of the war but eventually recovers over time. The vertical red line on the graph represents the date when the war began (02/22/2022).

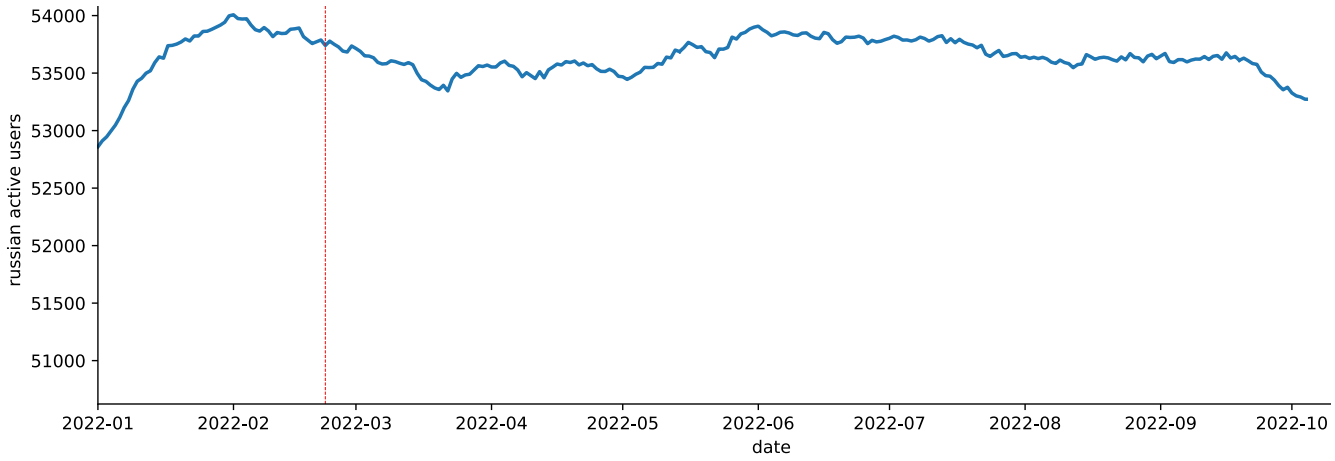

Figure C.4: Daily Active Russian Refugee Users

Note: Over time, Russian users remain engaged and active.

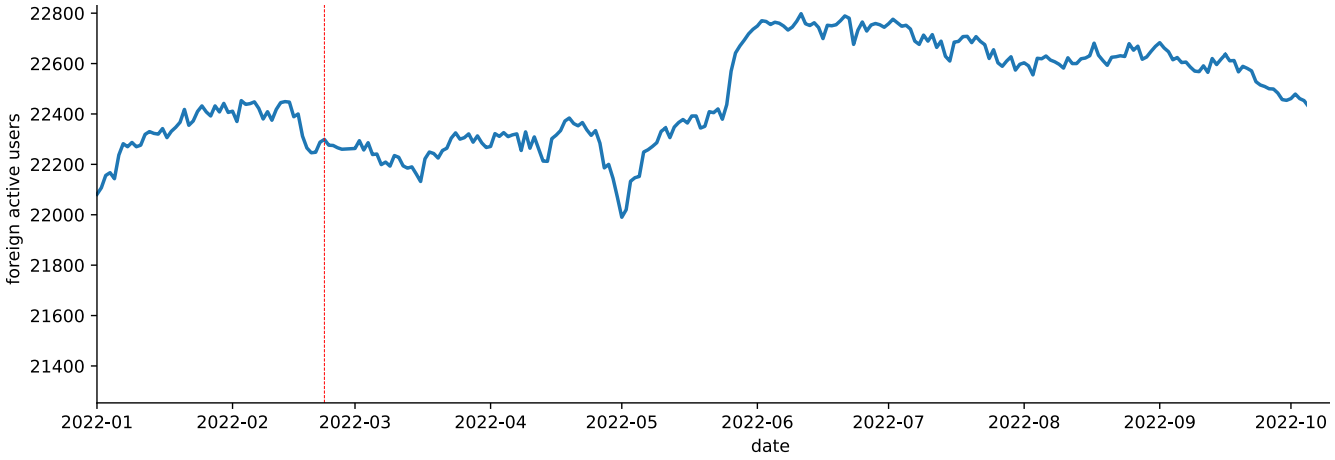

Figure C.5: Daily Active Foreign Refugee Users

Note: Foreign users are less affected by the war and experience increased activity, particularly starting in the summer.

C.4 Correlation of Selected City Attributes

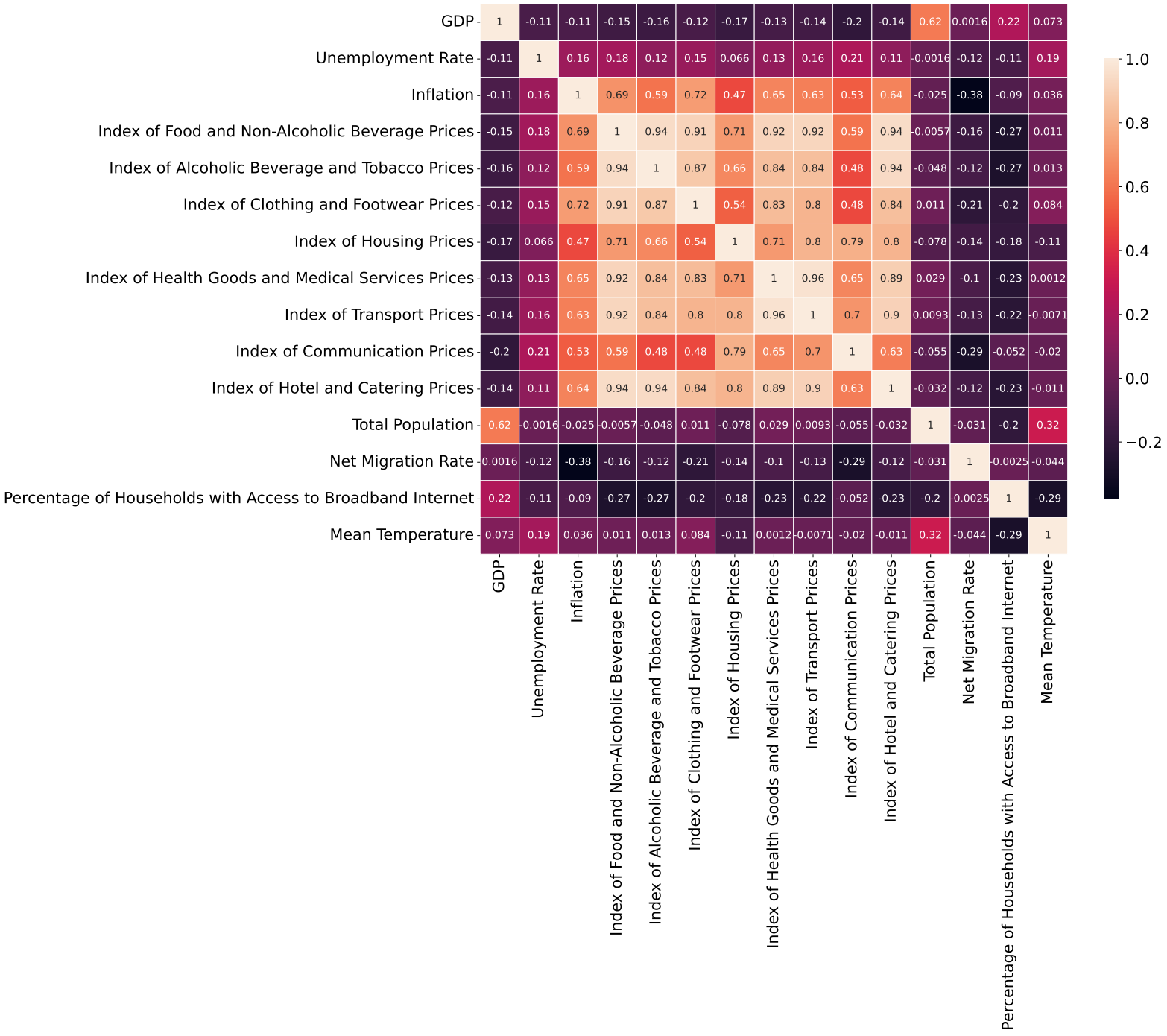

Figure C.6: City Attributes Covariance Matrix

Note: To eliminate highly correlated variables, we utilize the covariance matrix. The correlations among the fifteen chosen variables are displayed.