Food for Thought - Consumer Mobility and Nutritional Choices

Frédéric Kluser

Inauguraldissertation zur Erlangung der Würde eines

DOCTOR RERUM OECONOMICARUM

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Bern

August 2024

The faculty accepted this thesis on September 19, 2024, at the request of the reviewers Prof. Maximilian von Ehrlich and Prof. Dr. Jessie Handbury as a dissertation, without wishing to comment on the views expressed therein.

Die Fakultät hat diese Arbeit am 19. September 2024 auf Antrag der Gutachter Prof. Maximilian von Ehrlich und Prof. Dr. Jessie Handbury als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

Out of the night that covers me,
Black as the pit from pole to pole,
I thank whatever gods may be
For my unconquerable soul.

In the fell clutch of circumstance
I have not winced nor cried aloud.
Under the bludgeonings of chance
My head is bloody, but unbowed.

Beyond this place of wrath and tears
Looms but the Horror of the shade,
And yet the menace of the years
Finds and shall find me unafraid.

It matters not how strait the gate,
How charged with punishments the scroll,
I am the master of my fate,
I am the captain of my soul.

William Ernest Henley, Invictus

Here's to the crazy ones. The misfits. The rebels. The troublemakers. The round pegs in the square holes. The ones who see things differently. They're not fond of rules. And they have no respect for the status quo. You can quote them, disagree with them, glorify or vilify them. About the only thing you can't do is ignore them. Because they change things. They push the human race forward. And while some may see them as the crazy ones, we see genius. Because the people who are crazy enough to think they can change the world, are the ones who do.

Steve Jobs

Preface

Books are mirrors: you only see in them what you already have inside you.

Carlos Ruiz Zafón, The Shadow of the Wind

I always wanted to become a writer. So far, I did not. But writing this dissertation came surprisingly close to it. Novelists turn chaotic thoughts into stories, always hunting for the perfect words. And, in some sense, so do economists. This dissertation tells the story of an intense four-year-long roller coaster ride during which I was fortunate to enjoy the company, help, and guidance of amazing mentors, colleagues, friends, and my family. Nothing of this would have been possible without them.

Among them, I thank Max von Ehrlich for his never-ending support and understanding. His personality always calmed me when things got rough. Blaise Melly, for his sharpness and directness. Jessie Handbury for her enthusiasm and honesty. Gilles Duranton for inviting me to Wharton, allowing me to discover an entirely unknown world. I thank my Ph.D. friends, especially Elisabeth Preyer, for her emotional support, Lukas Hauck and Dino Collalti for stimulating debates and coffee breaks, and Moritz Weik for challenging me on the climbing wall and Catan board. But foremost, I thank Martina Pons. For your patience, empathy, and passion. You never stopped believing in me and always challenged me to grow.

Bern, August 2024 *Frédéric Kluser*

Contents

Preface ix

Introduction xix

1 The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits 1

1.1 Introduction 2
1.2 Data 6
1.2.1 Data Sources 6
1.2.2 Sample Construction 8
1.2.3 Summary Statistics 9
1.3 Measuring Mobility 12
1.3.1 Rank-Rank Slope 14
1.3.2 Intergenerational Elasticity 15
1.3.3 Transition Matrix 15
1.3.4 Conditional Expected Rank 16
1.4 Main Results 16
1.4.1 Dietary Mobility 17
1.4.2 Comparison to Income Mobility 20
1.5 Heterogeneities 22
1.6 Mechanisms 26
1.6.1 Socioeconomic Status 27
1.6.2 Current Location 31

1.6.3 Genetic Family Background 33
1.6.4 Habits 36
1.7 Model Setup 37
1.7.1 Identification and Estimation 40
1.8 Conclusion 43

2 Cross-Border Shopping: Evidence from Household Transaction Records 47
2.1 Introduction 48
2.2 Data 51
2.3 Empirical Strategy 56
2.4 Results and Discussion 59
2.4.1 Dynamic Treatment Effects 59
2.4.2 The Distance Decay Function 60
2.4.3 Variation Across Socioeconomic Characteristics 62
2.4.4 Culture 66
2.4.5 Commuting and Trip Chaining 68
2.4.6 Variation Across Cross-Border Locations 69
2.4.7 Robustness 70
2.5 Conclusion 71

3 Spatial Frictions in Retail Consumption 73
3.1 Introduction 74
3.2 Data 78
3.3 Conceptual framework 83
3.4 Empirical Strategy 86
3.4.1 Staggered DiD: Average Effect of Entry 87
3.4.2 Staggered DiD: A Robust Approach (Wooldridge, 2022) 88
3.4.3 Dynamic DiD: An Event-Study Style Approach 89
3.5 Empirical Results 89
3.5.1 Conventional Gravity Estimates 90

3.5.2 Store-Opening Effects 90
3.6 Robustness and Sensitivity 97
3.7 Consumption Areas 98
3.7.1 Recovering Store Valuations 98
3.7.2 Spatial Consumption Access 100
3.7.3 Type-Specific Spatial Frictions 104
3.8 Conclusions 106

A The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits 109
A.A Data: Matching Procedure 109
A.B Data: Additional Figures and Tables 111
A.C Model: Derivations 115

B Cross-Border Shopping: Evidence from Household Transaction Records 117

C Spatial Frictions in Retail Consumption 125
C.A Data: Matching Procedure 125
C.B Summary Statistics 127
C.C Additional Results 128
C.C.1 Staggered DiD: OLS With log(expenditures) Instead of Poisson . . 128
C.C.2 Robust DiD Estimator and DiD With Kinks in the Log-Distance . . 134
C.D Robustness 137

List of Figures

1.1 Distribution of Fruit and Vegetable Consumption 9
1.2 Life Cycle in Income and Diet of Children 13
1.3 Intergenerational Diet: RRS and IGE 18
1.4 Rank-Rank Slope: Lifecycle 19
1.5 Intergenerational Diet 21
1.6 Intergenerational Diet vs. Income: RRS 22
1.7 Intergenerational Diet: Heterogeneous RRS 25
1.8 Intergenerational Diet: the Role of Income and Education 30
1.9 Intergenerational Diet: the Role of Location 32
1.10 Lifestyle-Related Death of a Parent 36
1.11 Habit Persistence Parameters 42
1.12 Income Heterogeneities in the Model 44

2.1 Distance to the Closest Cross-Border Shopping Location 57
2.2 Dynamic Treatment Effects 60
2.3 Decay of the Treatment Effect 62
2.4 German-French Language Border 66

3.1 Spatial Distribution of Grocery Shop Openings 80
3.2 Distance Gradients 94
3.3 Log-Kink Distance Gradients 95
3.4 Estimated vs. Predicted Store Fixed Effects 99
3.5 Spatial Distribution of Grid-Level Retail Market Access 101

3.6 Relating to the Spatial Equilibrium 102

A.B1 Match Rate 111
A.B2 Municipality Averages: Sample vs. Population 113

B.1 Distribution of Travel Times 117
B.2 Robustness of the Dynamic Treatment Effects: Different Control Distance . . 119
B.3 Robustness of the Dynamic Treatment Effects: the Full Grocery Transaction Data 120
B.4 Robustness of the Dynamic Treatment Effects: Different Definitions of Cross-Border Locations 124

C.B1 Ranking of Favorite Stores 127
C.B2 Descriptives of the Treatments 131
C.C3 Distance Gradients (Robust Estimator) 134
C.C4 Non-Parametric and Log-Kink Specification of the Incumbent Shift 135
C.C5 Non-Parametric and Log-Kink Specification of the Competitor Shift 136
C.D6 Dynamic Treatment Effects 139
C.D7 Spatial Distribution: City of Zurich 142

List of Tables

1.1 Summary Statistics for Children's Expenditures 10
1.2 Comparison of Mobility Measures 17
1.3 Heterogeneities 24
1.4 Divorced Parents 34

2.1 Household Summary Statistics 54
2.2 Transactions Summary Statistics 55
2.3 Average Treatment Effects 61
2.4 Treatment Effects by Socioeconomic Subgroups 63
2.5 Treatment Effects by Cultural and Spatial Subgroups 65
2.6 Cultural Differences: Effect at Language Border 67
2.7 Treatment Effect for Different Commuting Behaviors 69
2.8 Prices in Neighboring Countries, 2015–2020 70

3.1 Transactions Summary Statistics 81
3.2 Summary Statistics for Households 82
3.3 Conventional Gravity (Poisson Model) 91
3.4 Incumbent Expenditure Shift (Poisson Model) 92
3.5 Competitor Expenditure Shift (Poisson Model) 93
3.6 Grocery Market Access Across Household Characteristics 103
3.7 Heterogeneous Distance Costs 105

A.B1 Summary Statistics for Children 112
A.B2 Comparison of Mobility Measures 114

B.1 Cross-Border Locations 118
B.2 Average Treatment Effects (With a 100 min. Control Group) 120
B.3 Treatment Effects by Socioeconomic Subgroups (With a 100 min. Control Group) 121
B.4 Treatment Effects by Cultural and Spatial Subgroups (With a 100 min. Control Group) 122
B.5 Cultural Differences: Effect at Language Border (With a 100 min. Control Group) 123
B.6 Treatment Effect for Different Commuting Behaviors (With a 100 min. Control Group) 123

C.C1 Incumbent Expenditure Shift – Intensive Margin (Log Model) 129
C.C2 Incumbent Expenditure Shift (Log+1 Model) 130
C.C3 Competitor Expenditure Shift – Intensive Margin (Log Model) 132
C.C4 Competitor Expenditure Shift (Log+1 Model) 133
C.C5 Log Specifications With Kinks (Incumbent Shift) 134
C.C6 Log Specifications With Kinks (Competitor Shift) 135
C.D7 Incumbent Expenditure Shift Pre-COVID-19 (Poisson Model) 137
C.D8 Competitor Expenditure Shift Pre-COVID-19 (Poisson Model) 138
C.D9 Incumbent Expenditure Shift, Only Once-Treated Households (Poisson Model) 140
C.D10 Competitor Expenditure Shift, Only Once-Treated Households (Poisson Model) 141
C.D11 Correlations for Figure C.D7 141
C.D12 Heterogeneous Distance Costs (Rural Areas) 143
C.D13 Heterogeneous Distance Costs (Urban Areas) 144

Introduction

Individual shopping and consumption decisions of households living across the globe have far-reaching consequences for the national and global economy, public health, biodiversity, and climate change. Consequentially, the UN's Sustainable Development Goals officially aim to "*end hunger, achieve food security and improved nutrition, and promote sustainable agriculture*" by 2030. Undoubtedly, significant progress was made – bringing, for example, the global prevalence of undernourishment from 34% to 12% in the last 50 years – but considerable challenges remain. One in ten people in the world still do not get enough to eat. 21% of the global population faces high levels of acute food insecurity and requires urgent assistance, while 148 million children under the age of five suffer from stunting and 45 million from wasting (Ritchie, Rosado, and Roser, 2023). Hence, significant challenges remain for policymakers, researchers, and activists until we are finally able to eradicate hunger.

Paradoxically, a very different type of malnutrition has materialized among developed countries in recent decades, with dire outlooks for the foreseeable future. The reason is that not only hunger and a general lack of food harm our health and well-being but also the (over)consumption of inadequate and monotonous diets poor in essential micronutrients – although the effects are very different. Today, 65% of Americans and 54% of Swiss residents are overweight, and 37% and 21% are considered clinically obese, respectively (Ritchie and Roser, 2024). Yet, contradicting common beliefs, healthy diets do not necessarily cost more than a standard diet (see, for the UK, Law et al., 2024) such that 98.5% of people living in high-income countries can theoretically afford a healthy diet (Ritchie, 2021). However, sufficient access to retail stores is far from certain – let alone to stores selling healthy products – especially for people living in disadvantaged neighborhoods and cities. For example, nearly 40 million people in the United States (12.8% of the U.S. population) live in low-income areas with limited access to a supermarket or grocery store (Rabbitt et al., 2023). Therefore, understanding the (spatial) economic, cultural, and behavioral factors that shape consumer choices and mobility – including product and store choices, residential location decisions, and the planning of shopping trips – is crucial for informing public policy and fostering healthier societies.

This thesis includes three papers investigating different dimensions of consumer behavior in Switzerland within the fields of urban and health economics: eating patterns within families across generations, consumer mobility and grocery market access within cities, and shopping trips across national borders. Chapter One, titled *The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits*, studies the intergenerational persistence of healthy eating patterns. Chapter Two, titled *Cross-Border Shopping: Evidence from Household Transaction Records*, analyzes the consumers' response to the COVID-19-induced national border closure in Switzerland. Chapter Three, titled *Spatial Frictions in Retail Consumption*, exploits supermarket openings to estimate distance decay functions and incorporates them into a simple framework of spatial shopping. Addressing these topics contributes to (i) the design of effective public health interventions and (ii) land-use restrictions and urban planning that account for the complexities of spatial consumer behavior.

All chapters of this thesis include empirical investigations based on two main ingredients. First, I use the universe of grocery transaction records recorded by the loyalty program of the largest Swiss retailer. The second ingredient is a set of restricted administrative data sets provided by the Swiss Federal Statistical Office (BFS) and the Central Compensation Office (ZAS). Supplementary data includes public health surveys, economic reports, web-scraped travel times and locations of amenities and stores, as well as various spatial open-source data sets.

In the first chapter of my thesis, co-authored with Martina Pons, we address the determinants of unhealthy eating. We provide novel evidence on the persistence of dietary habits between parents and children, finding that this link is strong and exceeds income transmission across all measures we consider. At the same time, substantial heterogeneities in the persistence of diet indicate that the socioeconomic background and location of children may be crucial to fostering beneficial eating habits and breaking unhealthy ones. We discuss potential mechanisms and show in a counterfactual analysis that only 10% of the intergenerational link in diet can be explained by the transmission of income and education. In line with these results, we introduce a habit formation model and argue that the formation of dietary habits during childhood and their slow alteration are key drivers of our findings.

In the second chapter of my thesis, I investigate cross-border shopping, a phenomenon allowing consumers from high-price countries to obtain comparable goods at lower prices in foreign markets. Yet, cross-border shopping can also reduce domestic consumption, sales, or tax collection. During the COVID-19 pandemic, many countries restricted cross-border movements to mitigate the virus's spread, thereby also prohibiting cross-border shopping. This chapter exploits the random timing of the Swiss border closure to study heterogeneities in the willingness to travel for lower prices. I find that domestic grocery expenditures temporarily increase by 10.9% in border regions. My results show that the effect increases in household size and decreases in age, income, education, and the cross-border locations' price index. Furthermore, I find novel evidence that citizens working close to the border combine their commuting trips with cross-border shopping, providing evidence for strategic trip chaining.

In the third chapter of my thesis, co-authored with Maximilian von Ehrlich and Tobias Seidel, we analyze spatial consumption frictions by estimating the causal effect of store openings on individual shopping behavior. Our findings reveal that spatial frictions significantly influence shopping behavior, with the distance elasticity of expenditures and the number of visits being approximately 0.15. Our estimates suggest that consumption areas extend to about 10-20 minutes of travel time, depending on household type, and

we show that traditional gravity estimates are considerably biased due to the endogenous nature of store locations. By combining distance elasticities with a simple model of shopping behavior, we derive store-specific attraction parameters and compute a measure of local grocery market access. Market access varies significantly across different locations, and, consistent with spatial equilibrium theory, this variation is reflected in local rents. Consumption frictions are more pronounced for older and smaller households and vary with income, primarily in non-urban areas. Overall, spatial variations in market access are more significant than the dispersion in income. Combined with the positive correlation between income and market access, this suggests an important role for real income disparities.

Chapter 1

The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits

joint with Martina Pons

How wild it was, to let it be.

Cheryl Strayed, Wild

There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.

Albert Einstein

Acknowledgments: I am grateful to Blaise Melly, Maximilian von Ehrlich, and Jessie Handbury for their invaluable support. I thank David Atkin, Jean-Michel Benkert, Simon Büchler, Gilles Duranton, Marcel Henkel, Benjamin Lockwood, and various seminar and conference participants for their helpful comments and suggestions. I gratefully acknowledge funding from the UniBE Doc.Mobility grant to visit the University of Pennsylvania, the Swiss National Science Foundation with ref. 187443 (NRP 77), and the Berne University Research Foundation. I thank the Migros Genossenschafts Bund, the Swiss Federal Statistical Office, and the Central Cooperation Office for sharing their data.

1.1 Introduction

Unhealthy eating habits not only impact our personal health and well-being but also put a substantial economic burden on our healthcare systems. A variety of health conditions, including obesity, cardiovascular diseases, and diabetes, has been linked to inadequate diet, accounting for 18% of all North American deaths (Afshin et al., 2019). Additionally, these lifestyle-related diseases generate high medical costs. For example, according to the American Diabetes Association, every fourth healthcare dollar in the United States is spent on people with diabetes, and patients with diabetes generate more than twice as many medical costs as those without the disease. The detrimental consequences of poor dietary choices highlight the need to investigate the origins of unhealthy eating, opening the way for targeted interventions and policy recommendations. A growing literature has taken on the challenge of understanding determinants of dietary choices, and the general consensus is that eating patterns are highly persistent (see Bronnenberg, Dubé, and Gentzkow, 2012, Atkin, 2013, 2016, Hut, 2020, Hut and Oster, 2022) and withstand major personal shocks and interventions (see Oster, 2018, Allcott et al., 2019, Hut and Oster, 2022).

This paper studies the role of the family in determining dietary patterns by analyzing how parents transmit their nutritional choices to their children. To this end, we exploit unique grocery transaction records matched with Swiss administrative data to analyze the intergenerational persistence of diet. Switzerland is an insightful case to study dietary patterns, as almost everyone has sufficient access to healthy food.[1] Our data contains customer-linked spending by product categories from 1.7 billion shop visits between 2019Q1 and 2021Q2 at the largest Swiss retailer.[2] We enrich this consumption data

[1]Switzerland has a high density of grocery stores, households travel on average 600 meters to the nearest one, and 80% have a store within 2 kilometers (Swiss Federal Statistical Office). In comparison, the median distance to the nearest food store in the United States is 1,450 meters, and only 40% of the population lives less than a mile from the closest store (USDA). In addition, healthy eating is also relatively affordable in Switzerland. According to the World Bank, less than 0.1% do not have the financial means to follow a healthy diet in Switzerland. In comparison, this is the case for 1.5% of households in the United States, 12% in China, and 97% in Madagascar. The World Bank considers a healthy diet as unaffordable if the lowest-cost basket fulfilling national guidelines for a healthy diet costs more than 52% of a household's income.

[2]Our findings are robust if we concentrate our analyses on the pre-COVID-19 period.

with family linkages and individual socio-demographic information from the Federal Statistical Office, allowing us to observe the shopping behavior of 270,000 individuals (12% of the population of interest) and their parents. The main variable of interest and our measure of the healthiness of a household's diet is the expenditure share of fresh fruits and vegetables relative to total food expenditures.

Our findings show that family is a crucial determinant of dietary choices. We document an extensive intergenerational persistence in fruit and vegetable shares, indicating a strong transmission of eating choices from parents to children. We estimate a rank-rank slope of 0.250, and children whose parents spend one percentage point more on fruits and vegetables have a 0.252 percentage point higher spending themselves at the median of parental consumption. Further, the children's probability of reaching the top quintile when parents are in the bottom quintile is 11.5%. This is substantially smaller than the probability that children with parents at the top quintile remain at the top of the distribution (31.9%). A comparison of our findings to income mobility suggests that the intergenerational persistence of diet exceeds income transmission across all measures we consider, indicating that the development of dietary habits during childhood might be a persistent channel through which parents impact their children's future. Yet, the children's socioeconomic background may be crucial to fostering beneficial habits and breaking unhealthy ones. Therefore, we look at different sub-samples and observe that the parents' influence is stronger in rural areas and among children with lower education and income, while the transmission mechanism weakens as the geographical distance between parents and children increases. Hence, high socioeconomic status and exposure to new environments seem to foster healthy eating.

Additional factors beyond the direct transmission of dietary habits influence children and their diet in many interconnected ways, and these could partly explain our findings. Such mechanisms include the transmission of socioeconomic status across generations, location and network effects, and unobserved family backgrounds, such as genetic variations in taste, genetic predispositions to diseases, or unobserved family shocks. For example, if highly educated and high-income individuals eat healthier, the transmission of these socioeconomic variables could (at least partially) drive our results. To understand the importance of these mechanisms, we apply the counterfactual analysis proposed in Chernozhukov, Fernández-Val, and Melly (2013) and find that the transmission of income and education can only explain 10% of the persistence in diet, while the transmission of location preferences accounts for 6%. In addition, we analyze the impact of the lifestyle-related death of a parent to assess whether information on genetic predisposition impacts dietary choices, and we find no significant response.

These results indicate that parents impact their children's nutrition directly – for example, through the transfer of nutritional knowledge and dietary habits – rather than indirectly through socioeconomic variables. To this end, we introduce a model of dietary habit formation in which agents inherit a habit stock from their parents and childhood environment. These habits influence the agents' diet by creating a trade-off. On the one hand, agents want to eat healthily while, on the other hand, deviating from one's habit causes disutility. The solution of our model suggests that fruit and vegetable consumption is a weighted average of current habits and a known optimal diet. The most important determinants of these weights are the strength of habit formation and adaptation costs. The results from our model estimation suggest that sticky habits are an important determinant of dietary persistence. Further, we find that better-earning households are more efficient producers of healthy eating habits.

The existing literature on intergenerational mobility predominantly focuses on income. For example, Chetty et al. (2014) document strong transmissions of income from parents to their children in the United States. Related papers show substantial spatial variation in mobility and disproportional disadvantages for non-white groups and Chetty et al. (2022a,b) document the importance of social networks in fostering upward income mobility for low-income people.[3] In recent years, various papers conducted comparable analyses for other countries (Bratberg et al., 2017, Corak, 2020, Deutscher and Mazumder,

[3]See also Chetty, Hendren, and Katz (2016); Chetty et al. (2020), and Chetty and Hendren (2018). Rothstein (2019) tries to disentangle the channels behind income persistence and concludes that job networks, as well as the local labor and marriage markets, drive income mobility rather than the transmission of education or human capital.

2020, Acciari, Polo, and Violante, 2022, Asher, Novosad, and Rafkin, 2024), including Switzerland (Chuard and Grassi, 2020).[4]

Yet, a much scarcer literature analyzes mobility in non-pecuniary dimensions like education, jobs, health, and consumption, which may partially be due to the limited data availability. For example, Halliday, Mazumder, and Wong (2020) analyze health mobility and find striking gaps by race, region, and parent education, while Black, Devereux, and Salvanes (2005) show that sons of better-educated mothers also attain higher education levels.[5] Nonetheless, the literature analyzing the behavior of consumers is surprisingly scarce. Exceptions rely on self-reported survey data for small samples (less than 3,000 observations), including Waldkirch, Ng, and Cox (2004) and Charles et al. (2014) who use total food expenditures and imputed consumption based on the PSID and find an intergenerational correlation in food expenditures from 0.14 to 0.20. Similarly, Bruze (2018), using the Danish Expenditure Survey, calculates an intergenerational elasticity of 0.41 for consumption. While informative, these studies do not address the composition of consumers' shopping baskets. In comparison, our study is the first to analyze the intergenerational transmission of specific dietary choices rather than aggregate expenditures, offering novel insights into dietary behaviors and their persistence across generations.

We further contribute to the literature on dietary choices. This strain of the literature primarily focuses on evaluating the impact of policies promoting healthier eating behavior, but most papers find results with limited economic or statistical significance. These policies include food subsidies (Hastings, Kessler, and Shapiro, 2021, Goldin, Homonoff, and Meckel, 2022, Bailey et al., 2024), food labels (Cook, Ostermann, and Sloan, 2005, Araya et al., 2022, Barahona, Otero, and Otero, 2023), sin taxes (Allcott, Lockwood, and Taubinsky, 2019, Dubois, Griffith, and O'Connell, 2020, Aguilar, Gutierrez, and Seira,

[4]Some studies show that wealth is also persistent within families, sometimes even after four to five generations (Charles and Hurst, 2003, Clark and Cummins, 2015, Adermon, Lindahl, and Waldenström, 2018, or Belloc et al., 2024).

[5]In addition, Halliday, Mazumder, and Wong (2020) find a rank-rank slope of 0.11-0.15 for health in the United States, while Andersen (2021), using Danish register data, estimates a higher rank-rank slope of 0.28. Furthermore, intergenerational persistence has been documented for longevity (Black et al., 2024), labor force participation (Fernandez, Fogli, and Olivetti, 2004), and tax evasion (Frimmel, Halla, and Paetzold, 2019).

2021, Dickson, Gehrsitz, and Kemp, 2023), carbon pricing of food (Springmann et al., 2018), or school-food programs (Berry et al., 2021, Handbury and Moshary, 2021). In contrast, this paper contributes to the understanding of the origins of eating behaviors in the first place.

The paper is structured as follows. Section 1.2 introduces the data and presents summary statistics while Section 1.3 discusses our measures of intergenerational mobility. Section 1.4 documents the intergenerational patterns in diet and compares them to income mobility. Section 1.5 dives into heterogeneities and we discuss potential mechanisms in Section 1.6. Emphasizing the importance of dietary habits, Section 1.7 introduces and estimates a model framework on habit formation. Section 1.8 concludes.

1.2 Data

We analyze the intergenerational transmission of diet by combining (i) individual transaction data from the largest Swiss retailer with (ii) administrative data from the Federal Statistical Office. Throughout this paper, we refer to *children* as adult residents for which we observe at least one parent in the administrative data. They are our population of interest, and we treat their parents' characteristics as observable covariates. To introduce the data, we refer to individuals in the grocery data as *customers* and those in the administrative data as *residents*.

1.2.1 Data Sources

Grocery Transaction Data – The consumption data stems from the loyalty program of the largest Swiss grocery retailer. We observe expenditures on 41 product groups for 1.7 billion customer-linked purchases between 2019Q1 and 2021Q2, and customer characteristics include their residence location, age, and household type. Locations are coded on a grid of 350,000 100×100–meter cells with a mean population of 25 residents.[6] In

[6]The retailer holds a market share of 32.7% in 2020. The major product groups include, among others, *fruits and vegetables*, *meat and fish*, *milk products and eggs*, and *bakery and convenience*. The household types include the categories *small households*, *young families*, *established families*, *golden agers*, and *pensioners*. To be a

this program, participants identify themselves at the checkout with their loyalty cards in exchange for exclusive offers and discounts. The program has substantive coverage, tracking expenditures of 2.1 million active users (32% of all Swiss residents above legal age), spending on average at least 50 Swiss francs monthly (USD 56 on July 29, 2024), and capturing 79% of the retailer's total sales. Notably, the retailer charges the same prices throughout the country, independent of local purchasing power, wages, and costs, and stores of comparable size generally offer similar goods, except for local products.

Our analysis focuses on a child's share of fresh fruits and vegetables relative to total food expenditures. This is a suitable measure for a healthy diet because (i) fruits and vegetables are highly correlated with the healthy eating index in Allcott et al. (2019) (0.57 and 0.41, respectively), (ii) a diet low in fruits or vegetables is among the most frequent reasons for nutrition-related mortality in Afshin et al. (2019), and (iii) our measure correlates strongly with the intake of important micronutrients across age groups.[7] Furthermore, this measure provides a transparent and objective approximation of dietary quality as it requires no weighting of different nutrients or products.

Administrative Data – We enrich this unique consumption data with administrative records for the entire Swiss population (8.7 million inhabitants in 2020). Pseudo social security numbers allow linking residents across three different administrative data sets. The *Population and Households Statistics* provides socio-demographic characteristics for each resident for the years 2016–2021. This includes, among others, information on gender, age, marital status, residence location, household identifiers, and the pseudo-identifiers of spouses and children.[8] The residence locations are coded on the same

family, consumers have to register their children. This registration gives access to additional benefits related to family products.

[7]We compare our data's fruit and vegetable shares to the micro-nutrient intake reported in the *National Nutrition Survey* (by age group). This survey inquired in 2014 and 2015 2,000 participants between the ages of 18 and 75 about their previous day's diet. We find that the expenditure share of fruits and vegetables has a correlation of 0.4 with the intake of fibers, 0.38 with phosphorus, 0.33 with zinc, 0.22 with Vitamin A, and 0.29 with magnesium.

[8]Family linkages, including pseudo-identifiers for mothers and fathers, have been collected since 2005. This information is available for all individuals unless their parents never lived in Switzerland, died before 2005, or if there was no civil status change either for them or their parents since the 1990s (for example, wedding, divorce, or birth). Consequently, the *Population and Households Statistics* includes information on the parents of 84% of the Swiss residents under age 60, and of 22% above age 60. The coverage for foreigners

100×100–meter grid as in the grocery transaction data. The *The Old-Age and Survivors Insurance* dataset contains annual gross labor market income for every resident for the years 2016 to 2021.[9] We average annual household income for the years 2016–2021 to reduce biases in permanent income from transitory shocks and adjust, in most cases, average household income by the square root of household size.[10] Finally, the *Structural Survey* gives information on the highest completed education in a household for the years 2010–2021.[11]

1.2.2 Sample Construction

We restrict our analysis to customers that we can uniquely match to a resident based on the common variables of age and location. Appendix A.A describes the individual steps of the matching procedure. The matching links 337,000 children to at least one of their parents. We focus on children and parents with average monthly grocery expenditures between CHF 50 and 1,000 per capita. This is because too-small monthly baskets might not accurately capture the overall consumption, while too-large monthly baskets are unlikely to suit personal use but are from business customers. We keep households with at most ten members to exclude large cohabiting arrangements and retirement homes. Ultimately, we focus on *children* between the ages of 21 and 70 and parents between the

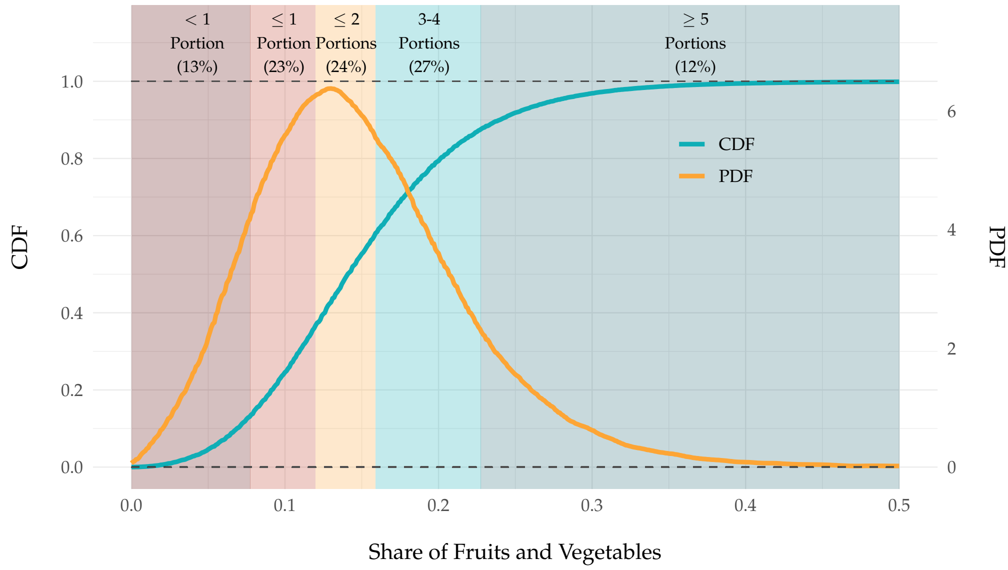
is lower because many of their parents live abroad. Yet, we include foreigners with known parents in our analysis.

[9] Contribution to this insurance is mandatory for everyone except for individuals younger than 25 with an annual income below 750 Swiss Francs. The contributions amount to a fixed share of the gross labor market income, including official awards, gifts, and bonuses, and are also mandatory for self-employed individuals.

[10] The calculation is $income\_adjusted = \frac{income\_total}{\sqrt{\#household\_members}}$, where we consider all household members, including small children. The adjustment follows one of the equivalence scales suggested by the OECD. We compute *income_total* as the household's annual income by summing the income of all household members but excluding grown-up children who still live with their parents, as they likely do not contribute to the household's budget.

[11] The survey questions a representative sample of 200,000 people above age 15 every year on housing, employment, mobility, and education. Participation is mandatory. Education is categorized as either primary, secondary, or tertiary education. Primary (or compulsory) education ends at the latest after eleven mandatory school years (including kindergarten). Individuals who completed high school or an upper-secondary specialized school have a secondary education. The completion of any degree at a university, university of applied sciences, or university of teacher education results in a tertiary degree. As education stabilizes for most individuals after a certain age, we use educational variables only for individuals above age 25 at the time of the survey.

FIGURE 1.1: Distribution of Fruit and Vegetable Consumption



Notes: The figure shows the cumulative distribution function and the probability density function of the fruit and vegetable share in our final data. The colored bars show additional data on the fruit and vegetable portion intake in Switzerland from the *National Nutrition Survey*.

ages of 48 and 97 to avoid too small age groups in our estimation.[12] Further, we generate parents' variables as the average value of the father and mother weighted by their respective food expenditures.[13] This results in a final sample of 271,000 children.

1.2.3 Summary Statistics

Table 1.1 displays summary statistics for the consumers' monthly food expenditures and the share allocated to fruits and vegetables. The average household spends 399 Swiss francs per month (450 USD on July 29, 2024) and allocates 15% of this money to fresh fruits and vegetables. To put the latter observation into perspective, we plot in Figure 1.1 the distribution of fruit and vegetable expenditures and overlap it with data on portion

[12]Because we detect minor life cycles in diet, we provide all our results conditional on age groups and want to ensure that groups are large enough (see Section 1.3, for details).

[13]If parents live together, their household characteristics and consumption behavior are identical, while individual variables vary. If parents have separate living arrangements, household characteristics and consumption behavior differ, and we average all characteristics in the same way we average the shares of fruit and vegetables.

TABLE 1.1: Summary Statistics for Children's Expenditures

	Total Spending			% Fruit & Vegetable			Budget Survey	
	Mean	p50	SD	Mean	p50	SD	Spending	Share
Overall	399	323	284	0.15	0.14	0.07	616	0.65
By Age								
< 34	298	239	207	0.15	0.14	0.07	459	0.65
35–44	425	357	287	0.15	0.14	0.07	654	0.65
45–54	459	382	316	0.14	0.14	0.07	728	0.63
55–64	393	325	274	0.16	0.15	0.08	663	0.59
65+	345	286	237	0.17	0.16	0.08	616	0.56
By Household Income								
< 4,530	269	212	191	0.14	0.13	0.08	409	0.66
4,530–6,717	294	230	215	0.14	0.13	0.08	485	0.61
6,718–9,288	374	305	259	0.14	0.13	0.07	604	0.62
9,289–12,855	422	354	283	0.15	0.14	0.06	713	0.59
12,856+	458	384	312	0.16	0.16	0.07	869	0.53
By Highest Education								
Primary	275	222	190	0.13	0.12	0.07		
Secondary	376	304	264	0.14	0.13	0.07		
Tertiary	442	368	303	0.16	0.16	0.07		
By Pop. Density								
Rural	386	317	266	0.14	0.13	0.06		
Suburban	407	332	288	0.15	0.14	0.07		
Urban	389	303	289	0.17	0.16	0.08		

Notes: This table shows summary statistics for the transaction records of food expenditures in our final data. The columns titled *Budget Survey* show the average grocery expenditures for food and beverages from the administrative Household Budget Surveys (2015–2017) and the average expenditures in our data relative to the survey. *Household Income* is a household's average gross labor market income 2016-2020 in 1,000 CHF. *Highest Education* is the highest education completed by anyone within the household, and *Pop. Density* is the municipality's population density.

intake from a representative administrative nutrition survey.[14] Only 12% of Swiss households fulfill the recommended fruit and vegetable intake of five daily portions, while the mass of households in our data consume only between one and two portions of produce a day. The last two columns of Table 1.1 compare expenditures in our data to the administrative *Household Budget Survey*, showing that our transaction data covers 65% of the

[14]The Federal Food Safety and Veterinary Office conducted the *National Nutrition Survey* between 2014 and 2015 to document the diet of 2,000 Swiss adults.

average household grocery expenditures on food and beverages.[15]

Looking at different household characteristics, we observe that households increase their grocery expenditures throughout their life from a young age (298 Swiss francs) until age 45-54 (459 Swiss francs) before decreasing them again towards retirement (345 Swiss francs). Meanwhile, the share of these expenditures allocated to fruits and vegetables increases with age from 15% to 17%. This gives a first indication of a potential lifecycle in diet. Food expenditures also grow with income and education, such that, for example, the top income quintile spends 458 Swiss francs per month compared to 269 Swiss francs for the bottom quintile. Wealthier and better-educated households also consume relatively more fruits and vegetables, providing evidence of nutritional inequality across different socioeconomic status as previously observed in Allcott et al. (2019). Finally, we observe a larger fruit and vegetable share in urban than suburban or rural areas. One explanation could be that households in sparsely populated areas are more likely to buy fresh products from a farmer or own their own garden. Yet, households in rural areas spend with 386 Swiss francs only marginally less on grocery products than households in urban areas (389 Swiss francs), and we do not expect this to affect our results.

To assess the representativeness of our data, Table A.B1 shows summary statistics for the 271,000 matched children and compares them to the 2.3 million children in the population fulfilling the same selection criteria. Figure A.B2 plots municipality-level sample averages against the population values. The average *child* in the final dataset is 43.7 years old with an adjusted household income of 83,000 Swiss francs. 54% of them are female and 62% married. Further, 53% hold a tertiary degree, and 90% live in multi-person households. Regarding geographical characteristics, 76% of the children in our sample live in the German-speaking part of Switzerland, 19% in the French- and 4% in the Italian-speaking region. Our sample resembles the population of children well, with some differences in marital status and the degree of urbanization. The latter discrepancy is because we are less likely to identify unique combinations of customers and residents

[15]This survey continuously selects 2,500 households each year, and participants take notes on their income and expenditure for an entire month. Note that as we do not observe beverage expenditures, our actual coverage of food products is even higher.

when more people live in a raster cell.[16] Our findings remain qualitatively unchanged if we re-weight undersampled locations. In summary, our sample represents the target population well, and our expenditures cover a large share of grocery expenditures.

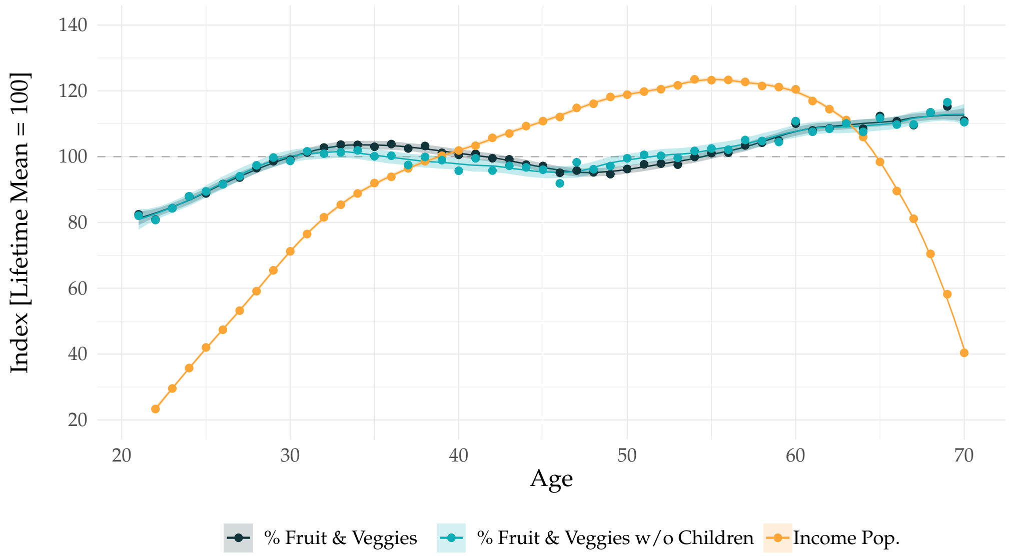
1.3 Measuring Mobility

Different statistics capture different aspects of mobility, which are not necessarily positively correlated (see Deutscher and Mazumder (2023) for an extensive discussion and classification of different mobility measures). For this paper, we need to consider that the focus is on diet and not income, and the two outcomes exhibit important differences. First, our measure of diet is bounded from below and above, while income is not. Second, we usually assume a positive marginal utility of income so that more real income leads to better living standards and higher welfare. Hence, having a higher real wage than your parents is a good thing in most cases. Differently, with diet, there is an optimal level or interval for fruit and vegetable shares, and an increase beyond a certain threshold might not be beneficial. Yet, Figure 1.1 shows that most of the population seems to be on the left of this threshold.

Papers analyzing intergenerational mobility face two challenges: (i) how to approximate the lifetime outcome well enough to handle transitory fluctuations and (ii) how to deal with lifecycle issues. The general approach in the recent literature is to average the outcome of children and parents over longer periods and to restrict the analysis to certain age bins of children and parents, ensuring that children, in the case of income, are old enough to be a regular part of the labor market and that parents are not yet retired to

[16]We illustrate this in Figure A.B1a by plotting the share of residents in a municipality linked to a child against the number of children living within this municipality. The final data set includes more than 10% of residents in smaller municipalities; this share declines as the population grows and lies around 5% for the largest cities. This result is not driven by the difference in penetration rates of the loyalty program across municipalities, as shown in Figure A.B1b. Further, Figure A.B2 shows that the representativeness of the matched customers is not different for larger cities. See further discussion of the data in Kluser and Pons (2024) and Kluser, Seidel, and von Ehrlich (2024), studying spatial consumption mobility from quasi-experimental shocks.

FIGURE 1.2: Life Cycle in Income and Diet of Children



Notes: The figure shows the average of three household variables for each age group between 21 and 70: (i) annual household income in the target population adjusted by the square root of household members (3.1 million households), (ii) the households' expenditure share of fruits and vegetables in the sample (270,000), and (iii) the households' expenditure share of fruits and vegetables for households in the sample who currently do not live together with their children (105,000). All values are normalized to 100 for the lifetime average of each variable. The regression lines and uniform confidence bounds are estimated by a local regression weighted by the size of the age groups.

avoid lifecycle and attenuation biases.[17]

Figure 1.2 compares the lifecycle variation of diet and income, displaying the average income and the share of fruit and vegetable consumption as a function of age. Both variables are normalized to the respective lifetime mean to make the results comparable.

[17] There is a large variety of specific approaches. For example, Chetty et al. (2014) rank children's income at ages 29 and 30 within birth cohorts and compare it to their parents' five–year average family income when the children were 15 to 19 years old. Chetty and Hendren (2018) use children's income at the household level at age 26. Parents' income is measured as the five–year average household income from 1996 to 2000 (independent of their children's age), and ranks are conditional on birth cohorts. Corak (2020) measures children's individual income at age 38–45, arguing this age approximates average lifetime income very well. He compares this to parents' income measured by a five–year average when the child was 15–19 years old. He addresses lifecycle concerns with robustness using children at ages 31 and 32. Acciari, Polo, and Violante (2022) restrict their analysis for Italian children's income at age 34–38 in 2016. The parents' and children's income is the average from 2016 to 2018. They compare the children's income to parents jointly and fathers and mothers separately. Acciari, Polo, and Violante (2022) address lifecycle issues with an error component model, simulating lifetime income. Similar strategies are also used in papers that do not concern income. For example, Andersen (2021) documents mobility in health, measuring parental health at ages 60–70 and the children's health at ages 36–50.

While income and diet exhibit both some variation over the lifecycle, the variation in diet is substantially smaller than for income. Income more than doubles from age 21 to 60 before declining again towards retirement age. Diet exhibits an s-shaped pattern. Young people tend to have a relatively poor diet, which improves by 30 percent until age 35.[18] After that, there is a decline of 10 percent until age 50, when diet ameliorates again.[19] If we exclude instead households with children, the curve flattens, providing interesting insights. At the age where many households have small children, their diet improves above the lifetime mean. At the same time, they eat unhealthier around the age where they live together with older children.[20] Given the visible, albeit small, lifecycle in diet, and since we observe children and parents at the same point in time, we will estimate ranks conditional on age as in Chetty et al. (2014) for the positional measures, and we control for age if the measure directly relies on the share of fruits and vegetables. If not indicated otherwise, we always compare a child's household diet to the weighted average of their parent's household diet, where the weights are proportional to the expenditure.

1.3.1 Rank-Rank Slope

Our first measure of intergenerational mobility is the rank-rank slope (RRS), where the percentile ranks of parents and children are computed within each age category. Let r_{ci} denote child i's percentile rank (from 1 to 100) among children conditional on their age. Similarly, let r_{pi} be the percentile rank of their parents within their parents' age group. The rank-rank regression is estimated by regressing the children's rank on the parents' rank:

$$r_{ci} = \alpha + \beta r_{pi} + \epsilon_i, \tag{1.1}$$

where β is the rank-rank slope, which provides a measure of transmission of the parents' position in their generation. The intercept α is the average rank for the lowest percentile ($r_{ci} = 1$). Without any correlation between r_{ci} and r_{pi}, the slope coefficient would be zero,

[18]Note that both age and cohort effects could drive these differences.

[19]This effect toward the end of life could also be driven by higher survival rates of individuals following a healthy diet.

[20]For both variables, the graph shows the values of the variable at a point in time. Thus, the changes could also be due to differences in diet across cohorts and not age effects.

and the intercept corresponds to the median rank. A value of $\beta = 0.3$ tells that if you compare two sets of parents one decile apart, their children are expected to be three percentiles apart. A steeper slope reflects a less mobile society (meaning more persistence). For instance, if each child were in the same percentile as their parents, the slope would be one, and the line would correspond to the 45-degree line.

1.3.2 Intergenerational Elasticity

As a second measure, we directly examine the relationship between children's diet and their parents. This measure is similar to the well-established intergenerational elasticity computed by regressing the logarithm of children's income on the logarithm of parents' income.[21] For our measure of diet, we do not take the logarithm, but we use a quadratic model since it better fits the data. Further, we control for the lifecycle in diet by including parent and child age as well as their squares in the following regression:

$$s_{ci} = \delta_1 s_{pi} + \delta_2 s_{pi}^2 + x_i' \gamma + \nu_i, \tag{1.2}$$

where s_{ci} and s_{pi} are, respectively, the child's and parents' fruit and vegetable share, and x_i contains the age control variables. Since we fit a polynomial regression, the slope changes over s_{pi}, and we will report the slope at the $\{25, 50, 75\}$ percentiles of s_{pi}.

1.3.3 Transition Matrix

Transition matrices break down the children's and parents' distribution into groups of equal size. We group children and parents into quintiles and compute the conditional probability that a child is in bin p_j given her parents are in bin p_k:[22]

$$TP_{j,k} = Pr(s_{ci} \in p_j | s_{pi} \in p_k). \tag{1.3}$$

[21] With a slight abuse of terminology, we refer to this measure as the *intergenerational elasticity*.
[22] We omit here the dependence of p_j and p_k on age to simplify notation.

This transition matrix answers questions like, *"What is the probability that an individual whose parents are in the bottom quintile of the distribution is in the top quintile?"* or *"What is the probability that this individual stays at the bottom of the distribution?"*. Hence, transition probabilities compare children to their parents at a fixed part of the parents' distribution. As for the previous measures, we compute quintiles again for each generation and age group separately. This implies that the bins p_j and p_k are age-dependent.

1.3.4 Conditional Expected Rank

The *Conditional Expected Rank* (CER) is the expected rank of children having parents at population percentile p:

$$CER(p) = \mathbb{E}(r_{ci}|r_{pi} = p). \tag{1.4}$$

We focus on the CER at the 25th and 75th percentiles, denoted CER25 and CER75. The CER can be estimated parametrically (using directly the information from the rank-rank regression) or nonparametrically. Both have different advantages and disadvantages. On the one hand, the parametric CER for children with parents at the 25th percentile also depends on the observations with parents at the top of the distribution as these observations influence both the intercept and slope of the regression. Hence, the parametric CER may be misspecified. On the other hand, with a large enough data set, one can calculate the CER directly from the sub-sample of parents at the percentile of interest, which is a fully nonparametric model. This measure is resilient against misspecification, but susceptible to larger variance. We opt for a middle ground and use a nonparametric local linear regression evaluated at percentile p.

1.4 Main Results

This section presents results on the overall persistence of dietary habits across generations. Table 1.2 reports coefficients and standard errors for all our results. Across all

TABLE 1.2: Comparison of Mobility Measures

	(a) Rank-Rank Reg.		(b) IGE			(c) CER		(d) Transition Prob.		
	Intercept	Slope	25	50	75	25	75	Q1Q1	Q1Q5	Q5Q5
Diet	37.75	0.250	0.274	0.252	0.226	44.97	56.07	31.26	11.47	31.89
	(0.1)	(0.002)	(0.003)	(0.002)	(0.002)	(0.67)	(0.68)	(0.17)	(0.13)	(0.17)
Income	43.22	0.143	0.117	0.120	0.122	47.84	54.22	26.48	14.12	28.45
	(0.12)	(0.002)	(0.003)	(0.003)	(0.004)	(0.81)	(0.80)	(0.22)	(0.18)	(0.22)

Notes: The diet results are estimated using 270,957 observations. The income results are estimated using 161,504 observations and we restrict the sample to children between 32 and 38 and fathers between 50 and 62. The IGE uses the log of the father's income as an explanatory variable and the log of the children's income as a dependent variable. Standard errors are computed using 1,000 bootstrap replications.

the reported mobility measures, we compute standard errors using 1,000 nonparametric bootstrap replications. Further, to assess the magnitude of the persistence of dietary choices, we compare the findings to intergenerational mobility in income.

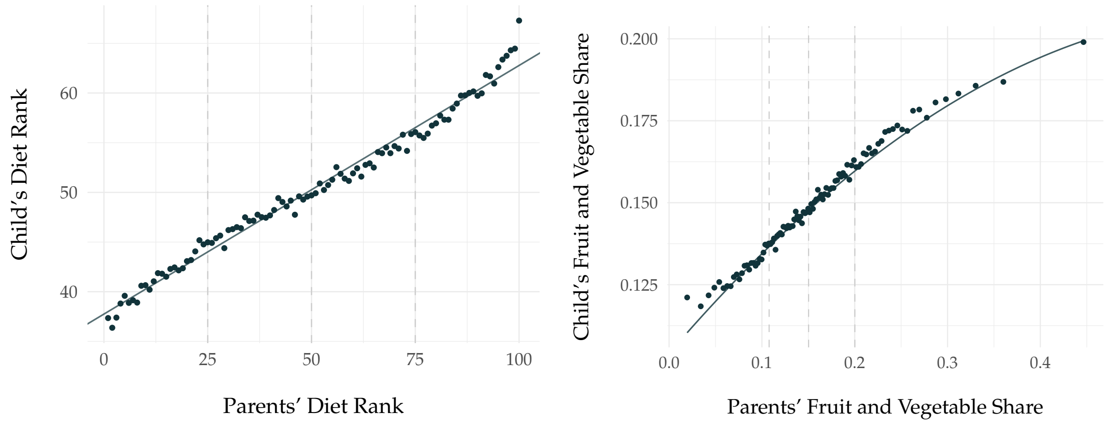
1.4.1 Dietary Mobility

Rank-Rank Regression – The estimated rank-rank slope in Panel (a) of Table 1.2 is 0.250, which shows that an increase in the parental percentile rank by one decile corresponds to an increase of 2.5 percentile ranks for the child. To put these results into perspective, it takes 3.16 generations to close the gap between two families at the first and the ninth decile.[23]

Figure 1.3a graphically illustrates the positional relationship between parents and children, plotting the estimated RRS regression line. The dots represent the average child percentile rank for each parental rank. The linear model approximates dietary patterns particularly well, which aligns with previous findings on income mobility. To show that conditioning the percentile ranks on age solves the lifecycle issues, we compare the results using conditional and unconditional ranks where we allow the intercept and the

[23] The number of generations N to close the gap of $\Delta_{10,90} = 80$ percentile ranks between the first and ninth decile solves $\beta^N \Delta_{10,90} = 1$, such that $N = \frac{log(1/\Delta_{10,90})}{log(\beta)}$.

FIGURE 1.3: Intergenerational Diet: RRS and IGE



(A) Rank-Rank Slope

(B) Intergenerational Elasticity

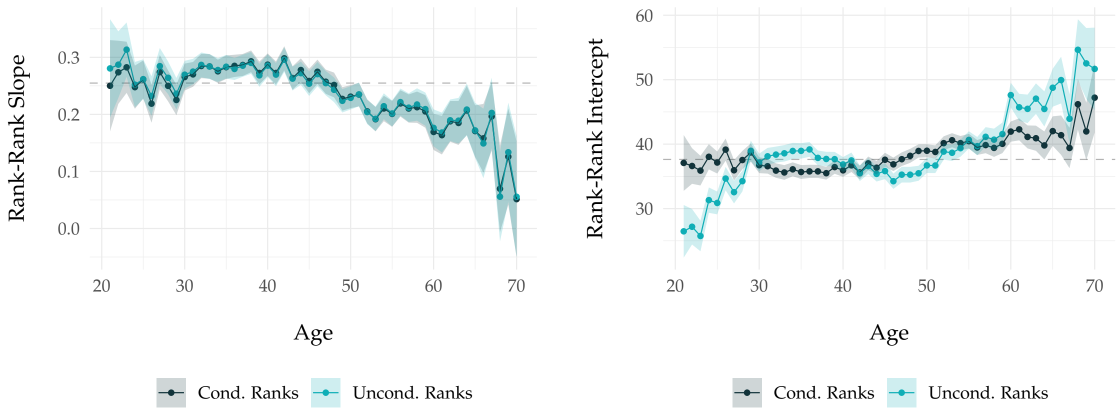
Notes: Figure 1.3a shows the estimated rank-rank regression line based on Equation (1.1) and Figure 1.3b shows the estimation results for the intergenerational elasticity in Equation (1.2). The dots in both graphs are the average children's ranks and values at each of the parents' percentiles.

slope to change over the lifecycle by saturating the model in children's age. While Figure 1.4a shows that the rank-rank slope is almost identical across both specifications, Figure 1.4b reveals that the intercept largely depends on the specification of the ranks, and in the specification using unconditional ranks, the intercept captures the lifecycle observed in Figure 1.2. This observation supports our expectation that conditional ranks are a better measure of dietary mobility than their unconditional counterparts. The rank-rank slope is large and roughly constant in early adulthood at around 0.27, showing that dietary habits acquired at an early age carry on far into adulthood. The slope starts declining at around age 45, which could be explained by habit adaptation, taking several periods to form. Yet, the relationship remains sizable until later in life.[24]

Intergenerational Elasticity – Panel (b) of Table 1.2 shows our estimates for the intergenerational elasticity in diet at different parental percentiles and Figure 1.3b shows that the estimated slope decreases as the parents' share increases and that the quadratic model fits the data well. The decreasing slope suggests that the intergenerational persistence in

[24]Note that the more noisy estimates for higher age groups are due to the smaller sample as for most individuals in these age groups, we cannot observe their parents' consumption since they might be deceased or live in a retirement home.

FIGURE 1.4: Rank-Rank Slope: Lifecycle



(A) Rank-Rank Slope (B) Rank-Rank Intercept

Notes: Figure 1.4a shows the rank-rank slope by age group. The grey line uses ranks for children and parents conditional on their age in a variation of Equation (1.1) fully saturated in the children's age. The blue line provides the results of the same estimation using unconditional ranks. Figure 1.4b shows the intercepts (the expected rank for a child with parents at rank zero) from the respective regressions. The dashed lines show the average RRS slope and intercept reported in Table 1.2. 95% confidence bands are computed using bootstrapped standard errors (1,000 replications).

diet is larger in the lower tail of the parents' distribution. For example, a one percentage point increase in the parents' fruit and vegetable consumption is associated with a 0.274 percentage point increase in child consumption for parents at the 25th percentile. This relationship decreases to 0.226 when the parents are at the 75th percentile. Therefore, targeted policy interventions might have the largest benefits for unhealthily eating families, resulting in more sizeable improvements in children's diets.

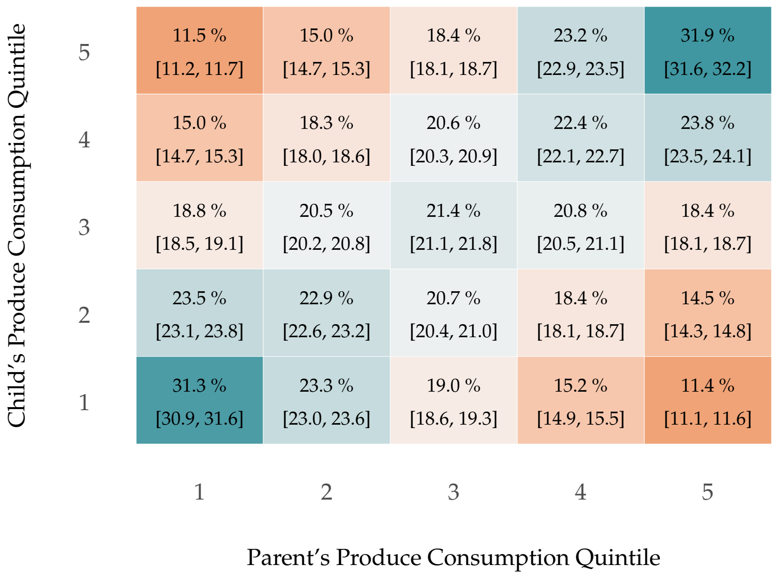
Conditional Expected Ranks – Panel (c) in Table 1.2 shows the nonparametric estimates of the conditional expected rank. We estimate a CER25 and CER75 of 45.0 and 56.1, respectively. Hence, a child with parents at the 25th percentile of the parents' distribution of fruits and vegetables is, on average, at the 45th conditional percentile of children. In contrast, children with parents at the 75th percentile can expect to reach the 56th percentile. Hence, although we observe strong persistence across generations in diet, there is still substantial reversion to the mean.

Transition Matrix – Figure 1.5 shows the estimated transition matrix with the corresponding confidence interval. We include selected key results of the transition matrix in Table 1.2 panel d). Without intergenerational persistence of diet across generations, the transition probabilities would not depend on parents' ranks, and we would observe 20% of children in each cell. The estimated transition probabilities reveal a strong persistence in diet between generations, as children are most likely to be in the same quintile as their parents. Focusing on the cells in the tails of the parents' distribution, we see that 31.3% of children whose parents buy the least fruits and vegetables are also in the lowest quintile of children (corresponding to a Q1Q1 transition), while only 11.5% move up to the highest quintile (Q1Q5). If, on the other hand, a household's parents are among their generation's top 20% fruits and vegetable consumers, their children are also most likely to be in the fifth quintile (Q5Q5). These particularly strong results in the "extreme" transition probabilities provide evidence that the so-called cycles of poverty and privileges are pronounced. At the same time, mobility appears larger around the center of the distribution.

1.4.2 Comparison to Income Mobility

To put the magnitude of our findings into perspective, we compare them to intergenerational mobility in income. More specifically, we focus on the relationship between children's and their fathers' income. To this end, we generate a data set for all Swiss children fulfilling the sample restriction criteria applied to the final data. To deal with lifecycle variation in income, we follow the procedure of the previous literature and focus on a subgroup of children and fathers with stable income (see, among others, Chetty et al., 2014, Corak, 2020, or Acciari, Polo, and Violante, 2022), and decide to restrict our analysis to children between the age of 30 and 40 with fathers between 50 and 62. This restriction ensures that most children are already participating in the labor market and fathers are not yet retired. Figure 1.2 shows that for these children, income only fluctuates slightly around the lifetime mean, and the fathers' income is also stable. Further, we average income over the years 2016-2021 to smooth out transitory fluctuations. We

FIGURE 1.5: Intergenerational Diet



Notes: The figure shows the transition probabilities for children's ranks conditional on their parents' ranks (Equation 1.3). We analyze transitions between quintiles and calculate the ranks conditional on age groups within the respective sub-sample of parents and children. 95% confidence intervals in parentheses are estimated using 1,000 bootstrap replications.

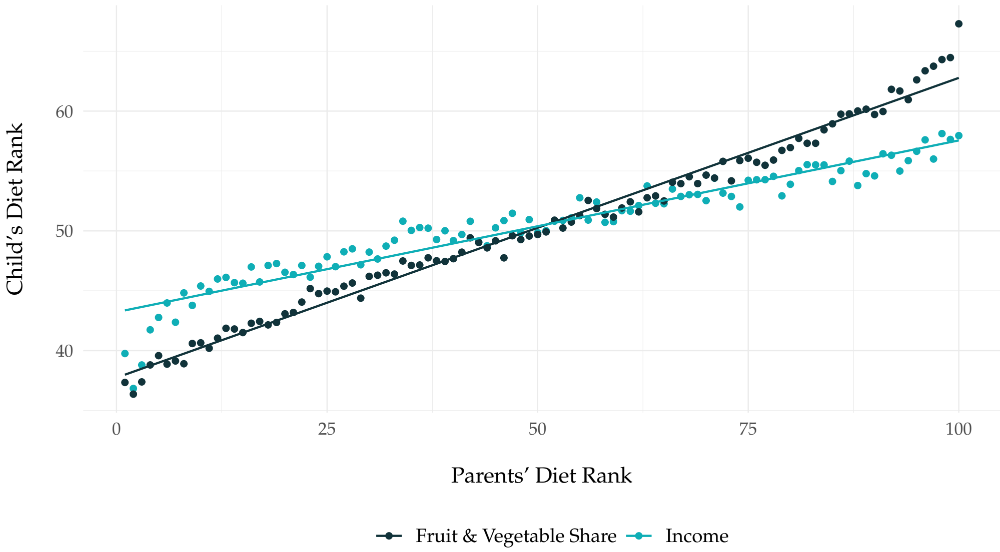
estimate the same measures for intergenerational income mobility we use for diet, again calculating the ranks within children and parents conditional on age. Table 1.2 shows an estimated RRS of 0.143 and an IGE of 0.120 at the 50th percentile.[25] The conditional expected ranks at the 25th and 75th percentile are 47.84 and 54.22. Also, more than one in four children with fathers' at the bottom quintile stay at the bottom, and 14.1% move up to the top.[26]

Comparing our estimated mobility measures between diet and income in Table 1.2, we observe that intergenerational transmission is more pronounced in the former across all the different metrics we consider. Figure 1.6 illustrates this graphically and shows that

[25]We measure the intergenerational elasticity in income with a classical log-log specification, however, including a quadratic term.

[26]Different sample selection procedures and income definitions (for example, using the average of parents' income) lead to comparable findings. Particularly, focusing on the sub-sample of households present in the diet as well as the income sample leaves our conclusions unchanged. Furthermore, our estimates on income mobility in Switzerland are in the range of those in Chuard and Grassi (2020), who measure the parental income as the average of the father's and mother's income when the child is between 15 and 20 years old. They find an RRS of 0.14 and an IGE of 0.22.

FIGURE 1.6: Intergenerational Diet vs. Income: RRS



Notes: The figure shows the estimation results for the rank-rank regression in Equation (1.1) for intergenerational diet and income. The dots in both graphs are the expected children's ranks at each of the parents' percentiles.

the slope of the rank-rank regression for diet is substantially steeper. This relationship suggests that the development of dietary habits during childhood is a persistent channel through which parents impact their children's future in a magnitude that exceeds the parental influence on the economic outcomes of their children. Nevertheless, it is important to note that income is particularly mobile in Switzerland in comparison with most other Western countries, and the relative persistence of diet and income may differ in other countries.[27]

1.5 Heterogeneities

Heterogeneities in the persistence of dietary habits across socioeconomic variables might enable dietary changes for some individuals while trapping others. This section unfolds heterogeneities between income classes, education levels, degrees of urbanization, and

[27]Previous literature estimates, for example, a rank-rank slope for income of 0.34 for the United States (Chetty et al., 2014), 0.24 for Canada (Corak, 2020), 0.22 for Sweden and Norway (Bratberg et al., 2017), 0.25 for Italy (Acciari, Polo, and Violante, 2022), and 0.21 for Australia (Deutscher and Mazumder, 2020).

the distance to parents. To correct for a possible mechanical result that children belonging to an unhealthy group have a higher chance of surpassing their parents, we use percentile ranks based on the entire sample but reweight the observations in each group such that the parents' distribution imitates the one in the entire sample.[28]

Table 1.3 shows the rank-rank slopes, conditional expected ranks, intergenerational elasticities, and selected transition probabilities for the different subgroups. The second column contains the P-value associated with a Wald test, testing for equality of the rank-rank slope between all the subgroups. Bootstrapped standard errors are in parentheses.

First, Panel (a) shows the results for the three education levels: primary, secondary, and tertiary. The rank-rank slopes lie around 0.23 in all groups and are not statistically different from each other. This suggests that higher children's education does not impact how parents transfer their diet. Instead, Figure 1.7a reveals that the intercept increases with education such that higher-educated children consume more fruits and vegetables. Therefore, education allows children to break out of unhealthy dietary habits, not through a change in the transmission of these habits but through the simple fact that higher-educated households systematically follow a healthier diet, independent of their parents. Multiple reasons may explain this observation. For example, higher-educated individuals may have a more profound nutritional knowledge, a better assimilation of dietary information, or a higher patience.

Second, Panel (b) digs into differences between income groups.[29] As shown, the rank-rank slope and intergenerational elasticity monotonically decrease as children's income increases. For children in the first income quartile, we find a rank-rank slope of 0.279 compared to 0.208 for individuals in the fourth quartile. These differences are also statistically significant, suggesting that percentile ranks are more persistent over generations

[28]This happens because, in unhealthy groups, children are more likely to surpass their parents' outcomes through mean reversion. The reweighting procedure gives equal weights to all percentiles in the rank-rank regression and the conditional expected rank. For the transition matrix, the reweighting changes the distribution of children conditional on their parents' bins and, therefore, also changes the children's ranks. For an extensive discussion of weighting approaches in these settings, see Deutscher and Mazumder (2023).

[29]To account for the lifecycle in income, we condition income quartiles on age and keep only working-age children (25-64). The results are not affected if we use all observations.

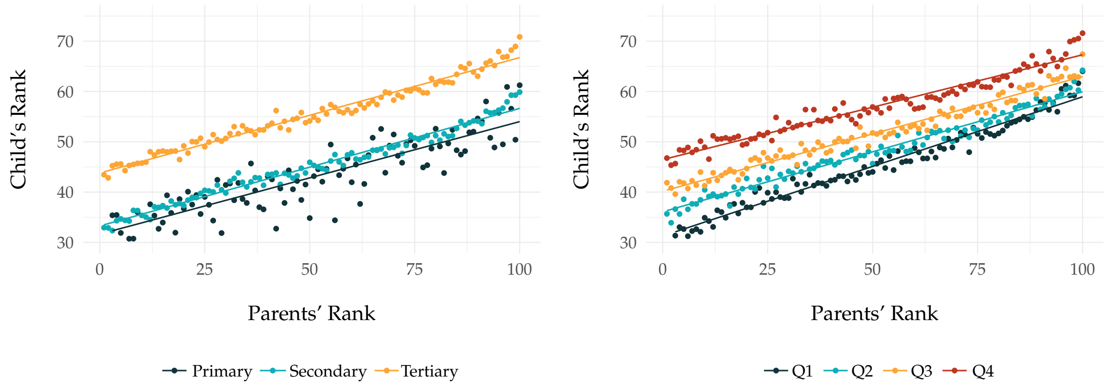
TABLE 1.3: Heterogeneities

	Rank-Rank		IGE			CER		Transition Prob.			
	RRS	P-value	25	50	75	25	75	Q1Q5	Q1Q1	Q5Q5	N
(a) Child's Education											
Primary	0.242	0.457	0.289	0.255	0.215	39.04	49.19	7.66	43.78	24.16	7,272
	(0.012)		(0.018)	(0.013)	(0.013)	(3.69)	(4.33)	(0.66)	(1.23)	(1.28)	
Secondary	0.234		0.249	0.226	0.200	40.35	50.77	8.18	37.68	24.61	82,763
	(0.004)		(0.005)	(0.004)	(0.004)	(1.09)	(1.30)	(0.20)	(0.33)	(0.35)	
Tertiary	0.229		0.242	0.226	0.208	49.00	60.65	15.46	23.43	36.66	103,676
	(0.003)		(0.005)	(0.004)	(0.003)	(1.08)	(1.07)	(0.24)	(0.30)	(0.31)	
(b) Child's Income											
1th Quartile	0.279	0.000	0.296	0.273	0.246	39.87	51.59	8.00	40.98	28.46	64,881
	(0.004)		(0.006)	(0.004)	(0.004)	(1.25)	(1.47)	(0.23)	(0.40)	(0.40)	
2nd Quartile	0.239		0.253	0.231	0.205	42.77	52.30	9.22	32.83	27.07	64,873
	(0.004)		(0.006)	(0.004)	(0.004)	(1.16)	(1.34)	(0.24)	(0.39)	(0.41)	
3rd Quartile	0.228		0.246	0.225	0.200	46.78	55.77	12.04	27.91	31.40	64,863
	(0.004)		(0.006)	(0.004)	(0.004)	(1.42)	(1.30)	(0.29)	(0.40)	(0.42)	
4th Quartile	0.208		0.230	0.217	0.201	51.81	61.90	18.18	20.86	38.73	64,852
	(0.004)		(0.007)	(0.005)	(0.004)	(1.47)	(1.22)	(0.35)	(0.37)	(0.40)	
(c) Child's Place of Residence											
Rural	0.236	0.013	0.260	0.234	0.203	42.02	51.84	8.02	35.43	24.67	58,732
	(0.004)		(0.006)	(0.004)	(0.004)	(1.34)	(1.46)	(0.21)	(0.38)	(0.43)	
Suburban	0.235		0.255	0.234	0.210	43.95	55.09	10.98	31.57	29.34	157,660
	(0.002)		(0.004)	(0.003)	(0.003)	(0.83)	(0.89)	(0.17)	(0.25)	(0.25)	
Urban	0.221		0.242	0.227	0.210	53.83	62.74	20.28	22.80	41.82	54,319
	(0.005)		(0.008)	(0.006)	(0.005)	(1.69)	(1.38)	(0.47)	(0.47)	(0.41)	
(e) Distance to Parents											
1th Quartile	0.281	0.000	0.298	0.277	0.254	41.55	55.70	8.57	34.77	30.60	63,842
	(0.004)		(0.006)	(0.004)	(0.004)	(1.28)	(1.47)	(0.24)	(0.39)	(0.42)	
2nd Quartile	0.252		0.273	0.249	0.222	44.07	54.82	10.44	31.43	30.71	63,841
	(0.004)		(0.007)	(0.005)	(0.004)	(1.37)	(1.28)	(0.27)	(0.41)	(0.40)	
3rd Quartile	0.225		0.246	0.226	0.203	43.80	53.99	12.88	29.90	30.77	63,841
	(0.004)		(0.006)	(0.004)	(0.004)	(1.38)	(1.35)	(0.29)	(0.40)	(0.39)	
4th Quartile	0.202		0.226	0.208	0.187	49.48	57.58	15.34	26.68	33.52	63,841
	(0.004)		(0.006)	(0.005)	(0.004)	(1.48)	(1.32)	(0.34)	(0.40)	(0.39)	

Notes: The table shows the results for different sub-samples defined by education, income, residence, and distance to their parents. The second column gives the P-value of the null hypothesis that the rank-rank slope is the same for all subgroups. Bootstrap standard errors in parentheses are computed using 1,000 replications. The number of observations in each subgroup is shown in the last column.

among low-income children. Figure 1.7b shows the rank-rank slope and expected ranks

FIGURE 1.7: Intergenerational Diet: Heterogeneous RRS



(A) RRS: Education (B) RRS: Income

Notes: This figure shows estimation results for the rank–rank regression in Equation (1.1) for different sub-populations, complementing the results in Table 1.3. Figure 1.7a displays the RRS for different education levels (primary, secondary, and tertiary) and Figure 1.7b for the four different income quartiles. The dots in both graphs are the expected children's ranks at each of the parents' percentiles.

for all four income quartiles. The differences in intercepts and slopes suggest that low-earning children are less successful at breaking unhealthy childhood habits and maintaining beneficial ones. For instance, a high-earning child with parents at the 10th percentile has the same expected rank as a low-earning child whose parents are at the 70th percentile.

These heterogeneities are also visible across geographical characteristics. Panel (c) shows that mobility is highest in urban areas and lowest in rural areas. The transition probabilities show that children living in urban areas have an outstanding likelihood of moving up in the distribution. Strikingly, a child born to parents in the first quintile of the distribution who lives in an urban area is more likely to find himself at the top of the distribution than in the first quintile.

Lastly, Panel (d) analyzes the role of the distance between the children's and parents' residences. We observe that nutritional persistence remains high even if children live far away from their parents. However, the further the children move away from their parents, the lower the persistence.[30] This result is not surprising as living away from one's family is often associated with moving away from one's childhood environment. However, it is striking that households are eight percentage points less likely to be trapped at the bottom if they live far away. This finding suggests that new social networks and environments might play a decisive role in breaking old habits and is consistent with previous findings on diminishing social interactions and responses to family-related shocks with increasing distance (see, e.g., Büchel et al., 2020 and Fadlon and Nielsen, 2019).

1.6 Mechanisms

The previous sections document a strong intergenerational persistence of diet across generations. In this part, we consider possible mechanisms driving our results. These factors influence children and their diet in many interconnected ways and could (partly) explain our findings. Assessing the importance of these mechanisms is crucial to designing well-targeted policies. Such mechanisms include the transmission of socioeconomic status across generations, location preferences, and unobserved family backgrounds, such as genetic variations in taste, genetic predispositions for diseases, or unobserved family shocks. In the following subsections, we analyze these factors in several ways.

First, we consider a counterfactual scenario in which we close down the indirect transmission of diet through income and education transmission. Second, we repeat this approach to look at the role of location. Third, we discuss the literature on the relationship between genes and diet and analyze the impact of the lifestyle-related death of a parent to assess whether information on genetic predisposition affects diet. Finding that the explanatory power of these factors is limited, we argue that habit formation is an important driver.

[30] We repeat this analysis for the sub-sample of children whose parents still live at the location their child grew up in. These individuals face a slightly higher rank-rank slope and higher transition probabilities in the Q1Q1 and Q5Q5 cells. Therefore, childhood networks beyond parents might play a role, but this role seems to be minor relative to parental diet.

1.6.1 Socioeconomic Status

This subsection isolates and quantifies the component of intergenerational transmission in diet that cannot be attributed to the transmission of two important socioeconomic characteristics: income and education. Isolating the influence of these channels is particularly important as Table 1.1 shows that better-earning and higher-educated individuals tend to consume more fruits and vegetables. Consequently, it is natural to ask whether and how much of the patterns that we document in this paper are due to the intergenerational transmission of these socioeconomic variables only. To this end, we compute counterfactual distributions in the spirit of Chernozhukov, Fernández-Val, and Melly (2013) to disentangle these socioeconomic drivers.[31] To identify the counterfactual distribution, we combine a population's cumulative distribution function (cdf) with an alternative covariate distribution. In this subsection, we are interested in the conditional distribution of the children's diet (conditional on their parents' diet) that we would observe if their income and education were independent of their parents' socioeconomic variables. Since the ranks are conditional on age, we include the children's and parents' age in the conditioning set. Once we have the counterfactual distribution, we can easily compute a counterfactual transition matrix, provided we observe the marginal distribution of the parents' diet conditional on age.

Let $F_{s_c|s_p,a_c,a_p}$ be the cdf of children's diet s_c conditional on the parents' diet s_p and the ages of children and parents, a_c and a_p. Let x_c denote a vector containing the children's income and education, and let x_p contain the corresponding parental variables. The main object of interest is the counterfactual distribution of the children's diet that we would observe if we change the covariate distribution $F_{x_c|s_p,a_c,a_p,x_p}(x_c|s_p,a_c,a_p,x_p)$ to a different

[31] A least squares regression of children's diet on parent diet controlling for socioeconomic variables does not disentangle this effect for several reasons. First, we need to model the distribution of children's diets to analyze directional mobility. Second, a least squares regression would fix a socioeconomic variable, whereas we want to consider a specific change in the covariate distribution. Third, comparing regressions that control for income and education with a regression without these controls provides meaningful results only under the strong assumptions of the correct specification. As we show in Section 1.5, diet transmission is heterogeneous across socioeconomic status, violating this assumption. While it would be possible to estimate a more flexible model that includes interactions between s_p and socioeconomic variables, such a model would become extremely tedious to compare. Instead, by estimating counterfactuals, even with a flexible model, results remain straightforward to interpret.

distribution $F_{x'_c|s_p,a_c,a_p,x_p}(x_c|s_p, a_c, a_p, x_p)$. We denote this counterfactual distribution as $F_{s_c|s_p,a_c,a_p}\langle x_c|x'_c\rangle\ (s_c|s_p, a_c, a_p)$.

Starting from the conditional cdf of the children's diet conditional on $(s_p, a_c, a_p, x_p, x_c)$, we can attain $F_{s_c|s_p,a_c,a_p,x_p}\langle x_c|x'_c\rangle(s_c|s_p, a_c, a_p, x_p)$ by integrating the conditional cdf over the alternative covariate distribution:

$$\begin{aligned} &F_{s_c|s_p,a_c,a_p,x_p}\langle x_c|x'_c\rangle(s_c|s_p, a_c, a_p, x_p) = \\ &\qquad \int_{\mathcal{X}'_c} F_{s_c|s_p,a_c,a_p,x_c,x_p}(s_c|s_p, a_c, a_p, x_c, x_p)dF_{x'_c|s_p,a_c,a_p,x_p}(x_c|s_p, a_c, a_p, x_p), \end{aligned} \tag{1.5}$$

where $\mathcal{X}_j$ denotes the support of the covariates x_j for $j = \{c, p\}$ conditional on the other variables. Then, integrating $F_{s_c|s_p,a_c,a_p,x_p}\langle x_c|x'_c\rangle(s_c|s_p, a_c, a_p, x_p)$ over the distribution of the parents' covariates yields the desired result:

$$\begin{aligned} &F_{s_c|s_p,a_c,a_p}\langle x_c|x'_c\rangle(s_c|s_p, a_c, a_p) = \\ &\qquad \int_{\mathcal{X}_p} F_{s_c|s_p,a_c,a_p,x_p}\langle x_c|x'_c\rangle(s_c|s_p, a_c, a_p, x_p)dF_{x_p|s_p,a_c,a_p}(x_p|s_p, a_c, a_p). \end{aligned} \tag{1.6}$$

In the counterfactual scenario that we consider, children's income and education are independent of the parental socioeconomic variables. Further, we assume that parents' age and parents' diet do not affect children's characteristics. Hence, the counterfactual covariate distribution is the conditional distribution of x_c given a_c:

$$F_{x'_c|s_p,a_c,a_p,x_p}(x_c|s_p, a_c, a_p, x_p) = F_{x_c|a_c}(x_c|a_c),$$

where the children's age in the conditioning set accounts for the lifecycle changes in income and different distributions of education over cohorts. Thus, this counterfactual scenario closes the path going from the parents' to the children's diet through the intergenerational transmission of education and income.

The estimation follows the plug-in approach. We obtain the conditional distribution function $F_{s_c|s_p,a_c,a_p,x_c,x_p}$ by inverting the conditional quantile function:[32]

$$F_{s_c|s_p,a_c,a_p,x_c,x_p}(s_c|s_p,a_c,a_p,x_c,x_p) = \int_{(0,1)} 1\left\{Q(u,s_c|s_p,a_c,a_p,x_c,x_p) \leq s\right\} du, \quad s \in \mathcal{S} \tag{1.7}$$

where $Q(\tau, s_c|s_p, a_c, a_p, x_c, x_p)$ is the τ conditional quantile function of s_c given the covariates. We estimate the conditional quantile function by fitting a flexible quantile regression model for $\tau = \{0.005, 0.015, \ldots, 0.995\}$. The regressions include a second-order polynomial of the parents' diet. Further, we include age and education dummies as well as household income (and its square) interacted with age and a dummy for age $\geq$ 65 for both parents and children. This last term allows income to have a different effect over the lifecycle, which is discontinuous after reaching retirement age.[33] All variables are also interacted with a second-order polynomial of the parents' diet.

For the estimation of the covariate distribution $F_{x_c'|a_c}$, we use the empirical distribution function:

$$\hat{F}_{x_c'|a_c=k} = \frac{1}{n_k}\sum_{i=1}^{n_k} 1\{x_{ci} \leq x\}, \tag{1.8}$$

where n_k is the number of children in a given age group.

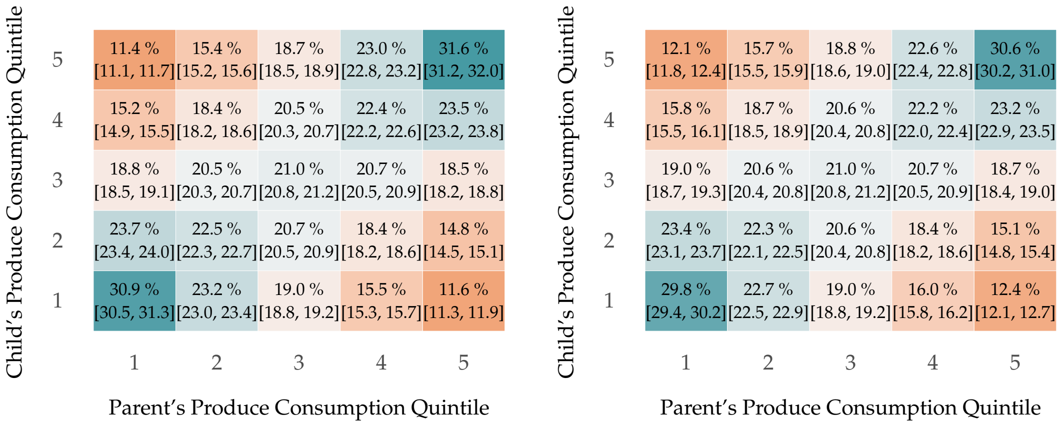
For this analysis, we restrict the sample to the 135,000 children for whom we observe their and their parents' education. The procedure in this section relies on the correct specification of the conditional quantile function. While we fit a flexible model, we re-estimate the baseline transition probabilities in this smaller sample using the same quantile model to further ensure a meaningful comparison.

Figure 1.8 shows the estimated transition probabilities with the corresponding bootstrap

[32]For this step, both a quantile regression or a distribution regression can be used (see Chernozhukov, Fernández-Val, and Melly, 2013). One of the main advantages of a distribution regression is that it does not require a continuous outcome and allows for mixed and discrete ones. However, this does not pose a problem in our case, as our outcome variable exhibits a smooth conditional density. On the other hand, the quantile regression coefficient provides a more natural interpretation.

[33]During the sample period, the retirement age in Switzerland is 65 for men and 64 for women.

FIGURE 1.8: Intergenerational Diet: the Role of Income and Education



(A) Transition Matrix

(B) Counterfactual Transition Matrix

Notes: Figure 1.8a shows the transition matrix and Figure 1.8b shows the counterfactual transition matrix. The counterfactual considers the case where the children's income and education are assigned independently from their parents' values. Bootstrap confidence intervals are in parentheses. The results are estimated using the sample of 135,213 children for which we observe their as well as their parents education.

confidence bands. Panel a) displays the transition probabilities estimated with the procedure described above; however, using the original covariates' distribution. These results are statistically indistinguishable from the transition probabilities computed nonparametrically for the entire sample in Figure 1.5. Panel b) shows the counterfactual transition probabilities. The transition matrix is similar to the one in Panel a). However, mobility is statistically significantly higher, mainly in the extremes. For example, the Q1Q1 and Q5Q5 probability decreases, and the Q1Q5 probability increases. Consider the Q5Q5 cell: In the original transition matrix, individuals whose parents are in the fifth quintile are 11.6 percentage points ($= 31.6 - 20.0$) more likely to be themselves in the fifth quintile than if there was no intergenerational transmission of diet. We refer to this as an excess probability. In the counterfactual scenario where we close the channel going through income and education, this number declines to 10.6 percentage points ($= 30.6 - 20.0$). This change suggests that the transmission of income and education over generations explains less than 9% of this excess probability. A similar calculation indicates that around 10% of the excess probability of remaining trapped at the bottom of the distribution can be attributed to income and education transmission.

In order to break down these transition matrices into a single number, we compute the normalized anti-diagonal trace similarly to Jäntti and Jenkins (2015). The normalization that we apply consists of subtracting the anti-diagonal trace of a completely mobile society. For the transition matrix in panel a), we find a normalized anti-diagonal trace of 28.4. In panel b), this statistic equals 25.9, suggesting that income and education drive only 8.8% of the intergenerational transmission of diet.[34]

Hence, these results suggest that only a small share of the intergenerational persistence of diet can be explained by the intergenerational transmission of income and education. This result is surprising and indicates that even if income and education were completely mobile across generations, we would still see a large intergenerational persistence of dietary habits. Hence, policies such as income redistribution or income benefits might only have a minimal impact on nutritional inequality. This finding is also in line with the small effect of monetary incentives in promoting healthier food choices among SNAP recipients (see, for example, Verghese, Raber, and Sharma (2019) and the references therein).

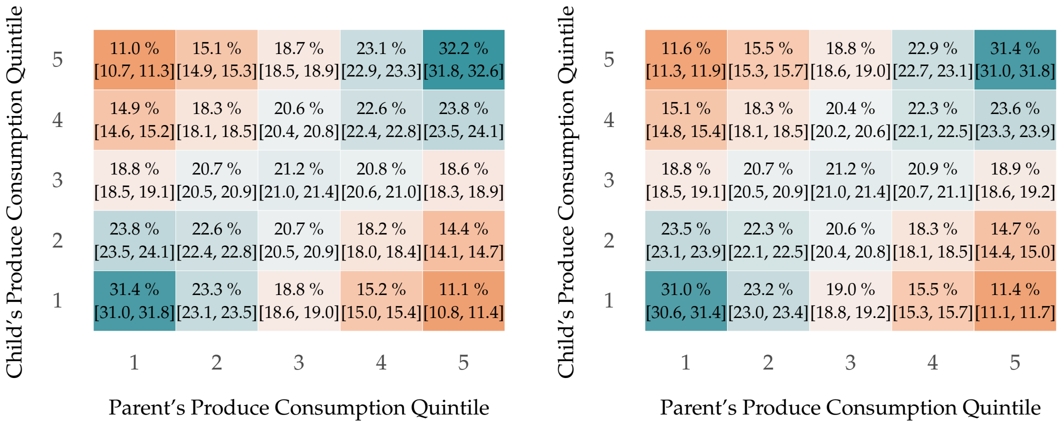
1.6.2 Current Location

Besides socioeconomic characteristics, also the transmission of location preferences might partly explain our results. Yet, these variables are more difficult to measure than income or education, and more importantly, it is unclear which characteristics of a location are meaningful in determining diet. In this analysis, we use population density as a broader measure of location characteristic that happens to be persistent across generations. For instance, children who grew up in rural (urban) areas are more likely to live in rural (urban) areas later in life.[35] Hence, the transmission of location preferences may partially drive dietary persistence as people in urban areas follow a healthier diet.

[34]The counterfactual analysis is not specific to the transition matrix. Instead, we can compute all mobility measures starting from the counterfactual distribution. The results are consistent across all mobility measures. To give an illustration, we find that after removing the transmission through socioeconomic variables, the IGE at the median parental rank decreases by 9.5%.

[35]55% of individuals in our sample whose parents live in rural areas also live in a rural area, while only 9% of them reside in an urban area. Similarly, 51% of individuals in our sample whose parents live in urban areas also live in an urban area, while only 12% of them reside in a rural area.

FIGURE 1.9: Intergenerational Diet: the Role of Location



(A) Transition Matrix

(B) Counterfactual Transition Matrix

Notes: Figure 1.9a shows the transition matrix and Figure 1.9b shows the counterfactual transition matrix. The counterfactual considers the case where the children's locations are assigned independently from their parents' values. Bootstrap confidence intervals are in parentheses. The results are estimated using the sample of 120,424 children for which we observe their and their parents' location.

To assess the share of dietary persistence attributable to the transmission of location preferences, we perform the same exercise we used to assess the role of income and education, where we now remove the link going through the transmission of location measured by the degree of urbanization. More precisely, we consider a counterfactual scenario where the probability that an individual lives in an urban, suburban, or rural environment is independent of their parents' location and other parental characteristics.

We again fit a flexible quantile regression model where we interact all variables with dummies for the degree of urbanization. Figure 1.9 displays the original and the counterfactual transition matrix.[36] Comparing the normalized anti-diagonal traces of the two matrices, we conclude that only 6.0% of the dietary transmission can be explained by children living in similar spatial environments as their parents (measured as urban, suburban, and rural areas). Notably, while the transmission of location plays a minor role as the two matrices are remarkably similar, some transition probabilities are statistically significantly lower in the counterfactual scenario.

[36] As before, we recompute the original transition matrix using the same flexible model. The marginal differences in the results are likely due to different samples.

Hence, this analysis suggests that while location is an important determinant of diet, the transmission of the level of urbanization plays a minimal role in the intergenerational persistence of diet. These results align with previous papers discovering limited adaptations in diets in response to changes in spatial environments (for example, Atkin, 2013, Atkin, 2016, or Allcott et al., 2019).

1.6.3 Genetic Family Background

Genetic family background can influence our diet in at least two ways. First, genetic variations may determine how we taste and appreciate different foods. Second, genetic predispositions to diseases could induce parents and children to adapt their diet. To give an illustration, a lifestyle-related death of a family member before the sample period could improve the diet for both parents and children. Here, we discuss these two channels, which could create a positive correlation between parents' and children's diets that is not explained by the direct transmission of dietary habits.

Taste – Genes determine how we perceive and interpret messenger signals sent from the taste receptors to the brain, and genetic variations in these taste receptor genes influence our individual sensitivity and preferences for flavors. Evidence is especially rich for receptor genes regulating the perception of bitter flavors (Mennella, Pepino, and Reed, 2005, Gervis et al., 2023), sweet flavors (Mennella, Pepino, and Reed, 2005, Mennella, Bobowski, and Reed, 2016, Søberg et al., 2017), alcohol (Allen, McGeary, and Hayes, 2014), and the olfactory perception of food in general (Cole, Florez, and Hirschhorn, 2020). These genetic variations shape food intake, and hundreds of genes are associated with our actual consumption of fruit, cheese, fish, tea, or alcohol, potentially affecting our results (Cole, Florez, and Hirschhorn, 2020).

To assess the importance of genetic variations in taste, we analyze the transmission of diet for the subsample of children with divorced parents who never remarried and live alone, observing, therefore, each parent's diet separately. Due to social norms, most of

TABLE 1.4: Divorced Parents

	Fruit & Vegetable Share Child					
Child Age at Divorce:	≤ 10	10–18	18–25	≤ 10	10–18	18–25
	(1)	(2)	(3)	(4)	(5)	(6)
Fruits & Vegetable Share Mother	0.225***	0.266***	0.238***			
	(0.019)	(0.015)	(0.016)			
Fruit & Vegetable Share Father				0.168***	0.146***	0.153***
				(0.024)	(0.026)	(0.028)
R^2	0.072	0.082	0.082	0.053	0.034	0.046
Observations	3,149	5,203	4,913	1,273	1,254	1,523

Notes: The table shows estimation results separately for divorced fathers and mothers who did not remarry and live alone. The regressions estimate the intergenerational elasticity (Equation 1.2), regressing the child's fruit and vegetable share on the parent's share s_{pi} and s_{pi}^2. Further, we control for the parent's and child's age as well as their squares. We report the slope coefficients at the 50th percentile of s_{pi}. Standard errors are computed using 1,000 bootstrap replications.

these children grew up with their mothers.[37] Hence, if the dietary transmission were mostly due to the genetic transfer of tastes, we should see no difference in the transmission of diet between their mother and their father, while a stronger link to the mother's diet indicates a stronger nurture channel. The estimation results in Table 1.4 show that the intergenerational link between children and their divorced mothers is substantially stronger than the link with divorced fathers. This relationship changes only slightly with the child's age at the divorce. Taken together, these results suggest an important role of nurture. Yet, we are not trying to rule out that *nature* – meaning, the transmission of taste across generations – drives a share of the correlation between children's and parents' diet. Instead, the relationship between taste receptors and genes is complicated, and taste receptors should not be regarded as an exogenous endowment. More precisely, as we explain later, what we eat can also alter the regulation of our genes.

Predispositions to Diseases – A revealed genetic predisposition for a lifestyle-related

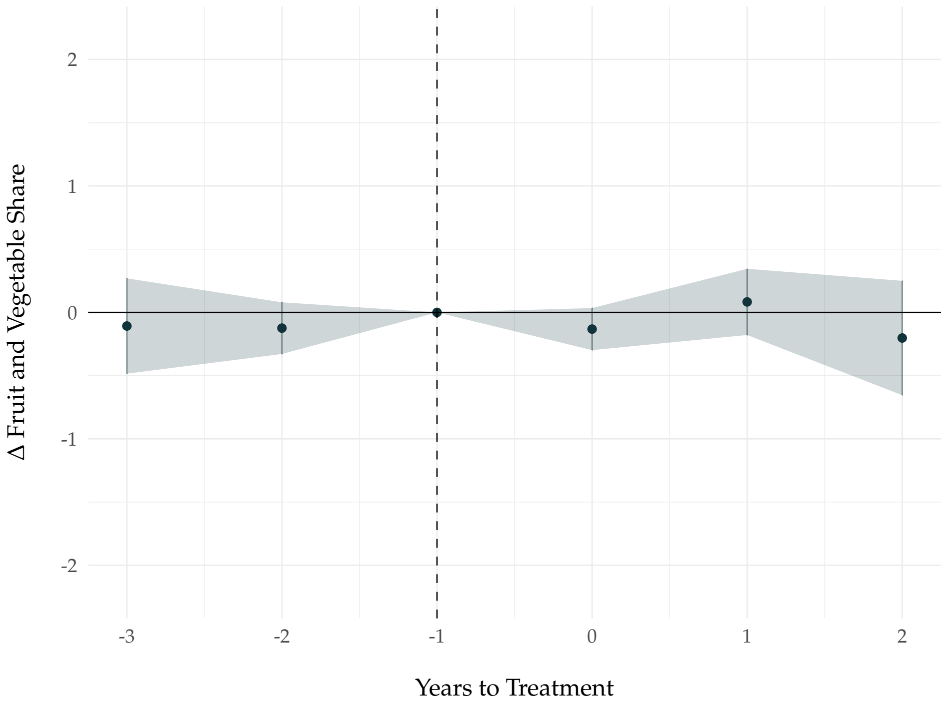
[37]We choose to focus on divorced parents who did not remarry to avoid possible contamination due to a new partner. Note that we do not observe who the child lived with after the divorce. Yet, a report by the *Federal Department of Home Affairs* (2022) shows that 46% of children spend at least two-thirds of their nights at their mother's place compared to only 10% who spend more than two-thirds at their father's place.

disease may drive family members to change their eating behaviors consciously. To assess the importance of this channel, we analyze the effect of the death of a parent due to lifestyle-related diseases on their children's diet. Such shocks might be informative for children, as individuals with a high genetic risk for heart disease almost double their risk for a stroke or heart attack, while a healthy lifestyle reduces this risk by half (Khera et al., 2016).

To conduct this analysis, we complement our data with the *Vital Statistics* administrative dataset for the years 2016-2021 that documents all deaths in Switzerland. The data includes the anonymized identifiers of all deceased residents and lists all underlying health conditions that either directly caused the death or may have contributed to it. If we find that children do not adjust their diet following the death of a parent, this channel is unlikely to play an important role in the transmission of diet. We use a staggered difference-in-differences design where we compare the diet of children whose parents die from a lifestyle-related disease (stroke and heart attack) to children who face the same shock in later years. We use the estimator proposed by Callaway and Sant'Anna (2021) and present the results in an event-study plot. Figure 1.10 shows that there is no change in fruit and vegetable intake for up to two years after the shock. This suggests that individuals might not perceive this shock as informative about their own risk for lifestyle diseases or simply do not respond to this information. Since genetic predispositions might already be known from unobserved non-fatal shocks or previous diagnoses, we alternatively focus on the deaths of individuals without any related pre-existing condition. We exclude, in this case, also deceased patients with a COVID-19 infection. This results in a data set of 22,500 observations, and the estimated coefficients remain insignificant.

One limitation of our analysis is that we consider only one shock, and other events, such as diagnoses, might be more informative about one's genetic predisposition. For instance, a diabetes or hypertension diagnosis could have a more substantial effect on diet as the affected person might receive or seek nutritional advice from a physician and pass the information to relatives. Nonetheless, Oster (2018) finds only minimal reductions in the caloric intake from unhealthy foods after a diabetes diagnosis, further suggesting that the

FIGURE 1.10: Lifestyle-Related Death of a Parent



Notes: Difference-in-differences estimates of the lifestyle-related death of a parent's effect on their children's annual fruit and vegetable intake using the estimator suggested by Callaway and Sant'Anna (2021). We use the *not-yet-treated* units as the comparison group. The estimation uses 38,177 observations, coefficients are normalized to the year before the treatment, and standard errors are clustered at the individual level.

predisposition to diseases is not a major channel.

1.6.4 Habits

We have seen that factors affecting both children's and parents' diets, such as income, education, location, and genes, do not explain much of the persistence in diet across generations. Based on this evidence, habit formation during childhood is potentially a sizable driver of our findings. These habits might capture many different *nurture* components, such as diet-related knowledge and skills that parents pass on to their children. This is consistent with the nutrition literature, which has long recognized the role of the family environment as a determinant of a child's diet (see, e.g., Birch, 1999, Scaglioni et al., 2018).

Supporting the importance of this habit mechanism, evidence shows that food intake –

even in utero and through breastfeeding – shapes a child's taste. For example, reducing sodium and sugar consumption sharpens the perception of saltiness and sweetness (Wise et al., 2016). At the same time, infants show a higher initial acceptance of fruits and vegetables if their mother eats them regularly during pregnancy (Mennella, Jagnow, and Beauchamp, 2001, Forestell, 2024) and breastfeeding (Forestell and Mennella, 2007). Hence, early over-consumption of unhealthy foods during childhood can reprogram our genes and numb our taste receptors, initiating a vicious cycle of bad habits, resulting in weight gain, obesity, and inflammation (May and Dus, 2021). Hence, parents shape children's taste preferences and consumption through many channels, which we summarize in this paper by habits.

While parental diet is likely a major determinant of the endowment habit stock of their children, many different factors, including childhood networks and location, might contribute to building and shaping this habit stock (see, for example, Story et al. (2008) for an overview). It is important to note that the presence of these factors does not invalidate the following framework. Eventually, understanding the determinants of these habits and separating nurture from nature components is necessary to implement the most effective policies, and future research should contribute in this direction.

1.7 Model Setup

To discuss potential mechanisms explaining the origins of our findings, we introduce a simple framework on habit formation. We model the persistence of diet between generations as the result of a habit stock built during childhood and adjusting over a lifetime (see, for example, Campbell and Cochrane (1999), Fuhrer (2000), and Carroll, Overland, and Weil (2000) for some early work on habit formation models). Habit formation has been used to explain a variety of economic behaviors. For instance, there is evidence of habit formation in voting behavior (Fujiwara, Meng, and Vogl, 2016), digital addition (Allcott, Gentzkow, and Song, 2022), health behaviors, or handwashing (Hussam et al., 2022). Related to nutrition, Atkin (2013) finds that higher relative prices in the past shape current tastes, providing evidence of habit formation.

In our setting, individuals are born into families whose diet, skills, and nutritional knowledge exogenously determine their initial stock of habits for their adult life, h_1. We think about the origin of h_1 as a Beckerian parental investment into their children's diet through the transfer of skills and knowledge (see, for example, Becker and Mulligan, 1997). Other unobserved factors outside the household, such as childhood networks, including extended family, friends, and school, also determine habits without invalidating the framework. Individuals enter adulthood and start their own household in period $t = 1$ and live on forever. They maximize their lifetime utility by choosing their relative intake of healthy foods $c_t \in [0, 1]$ for $t = 1, 2, \ldots$, given their initial endowment of habits h_1 and the degree of habit persistence mapping current consumption and habits into future habits:

$$h_{t+1} = h_t + \phi(c_t - h_t), \tag{1.9}$$

where $\phi \in [0, 1]$ measures the strength of habit formation. Hence, through their consumption behavior, agents continuously update their habits as a weighted average of current habits and consumption. Low values of ϕ imply a high degree of habit persistence and a low degree of learning, and deviations in c_t only have little effect on h_{t+1}. In the extreme case with $\phi = 0$, habits do not adapt, while with $\phi = 1$, the habit at time t equals consumption in the previous period, and there is no habit persistence.

Instantaneous utility in each period takes the form

$$u(c_t, h_t) = g(c_t - c^*) + h(c_t - h_t), \tag{1.10}$$

where c^* denotes the optimal (healthy) intake of fruits and vegetables, which is assumed to be the same and known for all agents, and the functions $g(\cdot)$ and $h(\cdot)$ have the following properties:

$$\frac{\partial g(c_t - c^{\star})}{\partial c} = \begin{cases} > 0, & if\ c_t < c^* \\ = 0, & if\ c_t = c^* \\ < 0, & if\ c_t > c^*, \end{cases} \tag{1.11}$$

and

$$\frac{\partial h(c_t - h_t)}{\partial c} = \begin{cases} > 0, & if\ c_t < h_t \\ = 0, & if\ c_t = h_t \\ < 0, & if\ c_t > h_t. \end{cases} \tag{1.12}$$

The two terms in Equation (1.10) account for two opposing forces. On the one hand, individuals want to eat healthily and be as close as possible to c^*. On the other hand, it is costly (painful) to deviate from one's habits h_t. Hence, any consumption different from $c_t = h_t$ causes disutility through adaptation costs.

To make the problem more concrete, we consider the following specification for the instantaneous utility function:

$$u(c_t, h_t) = -(c_t - c^*)^2 - \rho(c_t - h_t)^2, \tag{1.13}$$

where ρ is the importance of following one's habit relative to following a healthy diet. The quadratic specification means that small deviations from the optimal diet or one's habit cause little harm. However, large deviations are highly painful in utility terms. Intuitively, these deviations are costlier because they require additional preparation and shopping time, skills and information that need to be acquired (for example, by reading recipes), and new utensils.

Summarizing, each agent solves the following maximization problem:

$$\begin{aligned}
\max_{c_t, h_{t+1}} U(c_t, h_t) = \max_{c_t, h_{t+1}} \sum_{t=1}^{\infty} \beta^{t-1} u(c_t, h_t) \\
\text{s.t. } h_{t+1} = h_t + \phi(c_t - h_t), \\
u(c_t, h_t) = -(c_t - c^*)^2 - \rho(c_t - h_t)^2, \\
h_1 \text{ given,}
\end{aligned}$$

where β is the discount factor. Solving the model, we find that the policy function $c_t(h_t)$ is a weighted average of the optimal diet c^* and the current habit stock h_t:

$$c_t(h_t) = wc^* + (1 - w)h_t, \tag{1.14}$$

where the weight w is a function of the parameters (ϕ, β, ρ). Appendix A.C provides a detailed derivation of the solution and expression for w. The weight w given to healthy eating increases in β and ϕ and decreases in ρ. Hence, if households are forward-looking (meaning, they care about future consumption), have amenable habits, and derive significant utility from a healthy diet, then they give more weight to following a healthy diet relative to habits.

1.7.1 Identification and Estimation

To estimate the model, we rely on the same data we use in the rest of the paper and treat children of different ages as people in different periods of their lives. We use data on children between the ages of 30 and 60, calibrate $\beta = 0.95$, and set $c^* = 0.24$, which is the lowest fruit and vegetable share that meets the recommended consumption of five daily portions in Figure 1.1.

If we knew initial habits h_1, we could directly estimate $(1 - w)$ in Equation (1.14). Since we do not directly observe habits, we proxy them with parents' diet denoted $\tilde{h}_1$, introducing a measurement error. To deal with this challenge, we express h_t and c_t as functions of initial habits h_1 for $t \geq 2$ by iterating backwards the law of motions for habits in Equation (1.9) and the policy function for consumption in Equation (1.14):

$$h_t = h_1 \left(1 - w\phi\right)^{t-1} + c^* w\phi \sum_{j=0}^{t-2} \left(1 - w\phi\right)^j \tag{1.15}$$

$$c_t = h_1(1-w)\left(1 - w\phi\right)^{t-1} + c^* \left[(1-w)w\phi \sum_{j=0}^{t-2} \left(1 - w\phi\right)^j + w\right]. \tag{1.16}$$

A regression of c_t on $\tilde{h}_1$ interacted with age dummies identifies $\xi \cdot (1-w)\left(1 - w\phi\right)^{t-1} \forall t$, where the term $\xi \in (0, 1)$ arises from the measurement error. However, using data from different cohorts, we can identify $(1 - w\phi)$ and, therefore, the path for habits. We use a two-step estimator, where we first fit a saturated model of c_t on $\tilde{h}_1$ interacted with age fixed effects. Then, in the second step, we impose the structure $\xi \cdot (1-w)\left(1 - w\phi\right)^{t-1}$ on the coefficients by fitting a linear model in t on the logarithm of the first step slope coefficients. [38] We find a point estimate of

$$(1 - \hat{w}\hat{\phi}) = 0.988. \tag{1.18}$$

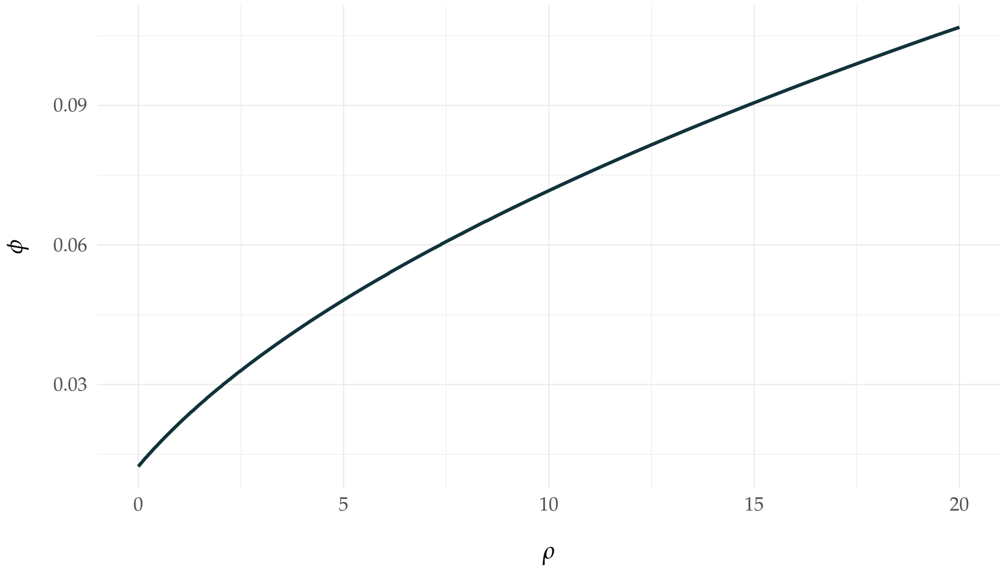
This expression does not separately identify ϕ and ρ because different values of the parameters are consistent with these results. As an example, consider an individual with $\rho = 1$ and $\phi = 0.021$, satisfying Equation (1.18). This individual values following her habits and a healthy diet equally, and gives a weight of $w = 0.57$ to healthy eating. Yet, the values $\rho = 2$ and $\phi = 0.028$ also satisfy Equation (1.18) and are, thus, observationally

[38] One potential worry of this analysis is that the measurement error is not constant over time. More precisely, if the measurement error increases with age, it would imply that ξ is decreasing over time, consequently affecting the estimation of $log(1 - w\phi)$. An alternative approach to estimate $(1 - w\phi)$ would deal with the ratios of adjacent cohorts' slope coefficients:

$$\frac{Cov(c_{t+1}, \tilde{h}_1)}{Cov(c_t, \tilde{h}_1)} = (1 - w\phi),\ \forall t > 2, \tag{1.17}$$

and we can take the average of these ratios. In this way, only the coefficients of adjacent cohorts are compared, making this estimator more robust to potential cohort effects. However, this procedure does not entirely exploit the relationship between the coefficients implied by the model. Using this alternative approach, we find a coefficient of 0.991, suggesting that cohort effects should not invalidate the results.

FIGURE 1.11: Habit Persistence Parameters



Notes: The figure shows the values of the habit persistence parameter ϕ and the relative utility weight ρ that are consistent with the result in Equation (1.18).

equivalent. While this second individual values following a healthy diet less and she assigns a lower weight to healthy eating ($w = 0.42$), she alters her habits faster. Hence, both of these individuals face the identical habit stock in the following periods, as a smaller deviation in consumption is coupled with more flexible habits such that Equation (1.18) holds.

Figure 1.11 pictures the continuum of compatible values for ϕ and ρ that satisfy Equation (1.18). We find that a higher valuation of a healthy diet (lower value of ρ) is consistent with our data if combined with stickier habits (lower ϕ). While, if individuals value a healthy diet less (higher ρ), then habits are more amenable (higher ϕ). However, what is striking is that even for extremely high values of ρ, our model still implies sticky habits, hence providing evidence for the important role of habit formation and giving an explanation as to why most individuals do not meet the dietary recommendations (for example, at $\rho = 20$, $\phi = 0.105$).[39]

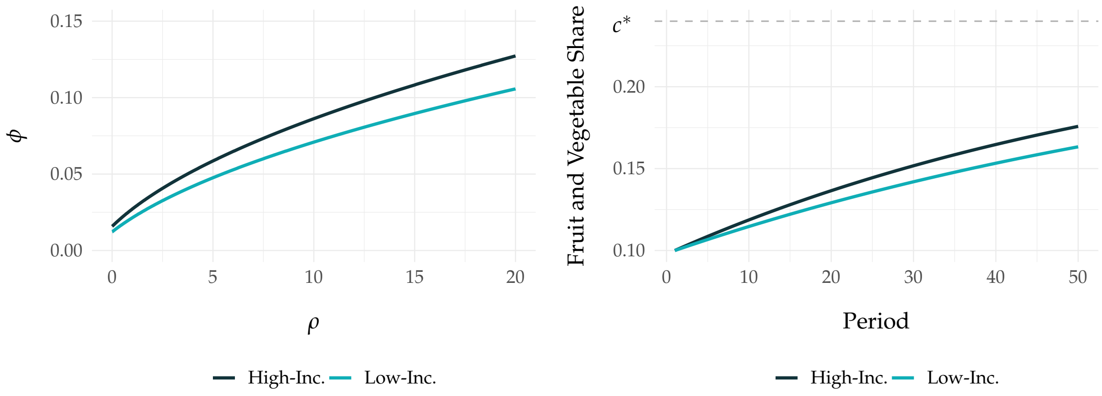
[39]Regarding the role of discounting, habits are less sticky if the discount rate β is low, as people have lower incentives to invest in future habits and assign more weight to following their habits.

Reconciling the model with the empirical heterogeneities we estimate in Section 1.5, we estimate our model for rich and poor households separately. Splitting the sample into income quartiles, we estimate $\hat{w}\hat{\phi} = 0.016$ for the top 25% and $\hat{w}\hat{\phi} = 0.012$ for the bottom quartile. Figure 1.12a shows the values of ϕ and ρ that are consistent with these results. The figure shows that as long as high-income individuals value healthy eating at least as much as low-income individuals, better-earning households face more amenable habits. If, however, low-income individuals value healthy eating more, it is possible that their habits adapt faster. Yet, this is unlikely to be the case as Lleras-Muney and Lichtenberg (2005) find that more educated individuals switch more easily to new drugs, suggesting their adaptation costs are lower. The difference in the estimated value of $w\phi$ for different income groups also implies that higher-income individuals have steeper habit trajectories. To give an illustration, Figure 1.12b shows the estimated habit trajectories of a low-income and a high-income individual, both with initial habits $h_1 = 0.10$. More affluent individuals build a habit stock that includes 1.25 percentage points more fruits and vegetables over fifty periods. All in all, these results are consistent with the finding of Cutler, Deaton, and Lleras-Muney (2006) that highly educated people are more likely to consume a healthy diet, exercise more, and take more preventive care. Also, evidence shows that a higher socioeconomic status might reduce adaptation costs in other areas.

1.8 Conclusion

The detrimental consequences of bad dietary habits are responsible for a sizeable social and economic burden, while the origins of these harmful eating habits are so far greatly understudied. This paper sheds light on the intergenerational transmission of dietary habits from parents to their children. We do so by combining unique supermarket transaction data with administrative records, including family linkages. We contribute to the literature with novel evidence showing that one's family background is a crucial determinant of persistent eating patterns, suggesting that the diet consumed early on in life at one's parents' dinner table shapes our nutritional tastes and preferences throughout our lives. Our results show that the intergenerational transmission of diet varies across

FIGURE 1.12: Income Heterogeneities in the Model



(A) Habit Persistence Parameters

(B) Evolution of the Habit Stocks

Notes: Figure 1.12a shows the values of the habit persistence parameter ϕ and the relative utility weight ρ for the best- and lowest-earning quartile of households in the sample. Figure 1.12b shows the evolution of the habit stock over 50 periods for the two income groups. The dashed grey line shows the optimal level of fruit and vegetable intake c^*.

observable covariates. Higher-educated and better-earning children generally eat better, independent of their parents. While the transmission mechanism (in terms of the rank-rank slope) does not vary between educational levels, it grows significantly weaker as income rises. Hence, low-income individuals are particularly vulnerable to getting stuck in a cycle of unhealthy diets. Further, upward mobility is larger among children living in urban areas, and the transmission becomes weaker as the distance between children and their parents increases, suggesting that breaking out of one's childhood environment can be a valid way to break unhealthy patterns.

We then test and discuss potential mechanisms driving our findings, including income, education, and family backgrounds. Isolating the part of dietary transmission going through education and income, we show that the transmission of these socioeconomic variables is responsible for only 10% of the intergenerational persistence in diet, and the transmission of location preferences explains around 6%. Further, we find that the unexpected death of a parent due to a lifestyle-related disease does not affect diet, suggesting that information about genetic predispositions is not an important determinant of diet, while there is substantial scientific evidence implying that diet affects our genes and taste perception. Similarly, the stronger persistence we observe in mother-children relationships compared to father-children relationships among children of divorcees further underscores the importance of the nurture component of intergenerational transmission. Although other unobserved variables of children likely influence eating habits throughout their lives, our results suggest that the direct effect of childhood diet is large. Thus, we argue that habit formation is an important mechanism, suggesting that not only does the apple not fall far from the tree but also that it does not roll far away afterward.

These findings have important implications for public health and policymakers. Recognizing the influence of family on dietary choices helps to design targeted interventions and formulate policy recommendations aimed at promoting healthier eating habits. By understanding the origins of unhealthy eating patterns and the mechanisms through which they are transmitted across generations, policymakers and healthcare professionals can develop effective strategies to combat the rising prevalence of diet-related diseases. Our results suggest that lump-sum transfers or SNAP benefits – which are not explicitly designed to improve diets – are potentially ineffective because they are unable to alter deeply anchored habits. Instead, policy interventions directly targeting the diet of young children while their habits are still forming might be more successful and cost-effective. Such policies may include, among others, healthy school lunch programs, nutritional education for children, and information campaigns at schools and doctors' offices. Future research should focus on disentangling specific mechanisms to optimally design such targeted policies. Houmark, Ronda, and Rosholm (2024) presents a promising approach in this direction, using genetic data to analyze the interaction of genes and parental investments in the formation of skills.

Chapter 2

Cross-Border Shopping: Evidence from Household Transaction Records

It is our choices, Harry, that show what we truly are, far more than our abilities.

Joanne K. Rowling, Harry Potter and the Chamber of Secrets

My life was narrated for me by others. Their voices were forceful, emphatic, absolute. It had never occurred to me that my voice might be as strong as theirs.

Tara Westover, Educated

Acknowledgments: I thank Martina Pons and Maximilian von Ehrlich for all their support. I thank Dino Collalti, Gilles Duranton, Jessie Handbury, Lukas Hauck, Costanza Naguib, Lindsay Relihan, and Tobias Seidel for their helpful comments. Support for this project from the Swiss National Science Foundation with grant ref. 187443 (NRP 77) is gratefully acknowledged. I thank Migros, the Swiss Federal Statistical Office, and the Central Compensation Office for providing the data.

2.1 Introduction

Cross-border shopping has been a growing phenomenon along national borders, where consumers from one nation can purchase goods and services at lower prices from neighboring countries. This outflow of customers puts pressure on domestic prices and increases product variety for households living close to the border, but it can also have adverse effects on local employment, consumption, sales, or tax collection (see Leal, López-Laborda, and Rodrigo, 2010, Knight and Schiff, 2012, or Baggs, Fung, and Lapham, 2018). Yet, while urban researchers understand the commuting behavior of workers well today, consumers' movement for shopping remains understudied, partially because suitable natural experiments are scarce.

However, numerous countries imposed rigorous travel restrictions at national borders in 2020 to contain the spread of COVID-19, providing such a natural experiment. This paper exploits the closure of the Swiss borders during the COVID-19 pandemic in order to examine patterns and heterogeneities in consumer mobility. On March 16, 2020, the Swiss government mandated the immediate closure of all national borders to neighboring countries to mitigate the spread of COVID-19. This policy was upheld until June 2020.[1] Additionally, the government announced the closing of all restaurants, bars, entertainment, and leisure facilities, with the exception of essential stores, including supermarkets and pharmacies.

Among countries introducing comparable policies, Switzerland is a unique case to study cross-border shopping for two reasons. First, Switzerland is surrounded by countries

[1] The borders to Liechtenstein remained open while crossing between Liechtenstein and Germany or Austria was prohibited. Nonetheless, crossings remained possible for work-related reasons for the 370,000 workers commuting from neighboring countries into Switzerland and the 29,000 Swiss residents working abroad.

with 28-39% lower grocery prices, allowing Swiss citizens to purchase comparable products at lower prices in Germany, Italy, Austria, or France.[2] These countries share a common currency, facilitating comparisons for Swiss households.[3] Hence, the relative attractiveness of these countries for Swiss consumers depends solely on the variety and prices of their grocery products. Second, the exact timing of the border closure was random for Swiss residents, and Burstein, Lein, and Vogel (2022) show that the policy was stringent and effective in achieving its purpose, as cross-border shopping shares almost fell to zero during the intervention.

I identify the causal effect of the border closure on expenditures at grocery stores in Switzerland by comparing Swiss households living close to a national border to Swiss households residing further inland within a difference-in-differences framework. The estimated increase in domestic grocery expenditures measures the magnitude of cross-border shopping during open borders as customers were forced by the shock to shift these expenditures to domestic retailers. To conduct this analysis, I merge unique grocery data featuring the universe of customer-linked transactions from the largest Swiss retailer for the year 2020 with individual-level administrative records on labor market income, commuting behavior, and household characteristics for the entire Swiss population. The final data set contains 40 million weekly shopping baskets for 750,000 households that I can uniquely link to residents in the administrative data. I use this setting to calculate a distance decay function (measuring the decline in cross-border shopping with distance) and analyze extensive heterogeneities across households' socioeconomic characteristics, cultural backgrounds, and commuting behavior.

My findings show that mobility patterns in consumption are persistent over time and vary strongly between different groups of customers. First, I find that the policy increases expenditures by 10.9% in border regions. This effect vanishes instantly and entirely once the border reopens, suggesting that behaviors in cross-border shopping are deeply rooted

[2]Imports into Switzerland are exempt from VAT for a total value below 300 Swiss francs, as long as certain limits for meat, tobacco, etc., are met. In addition, Switzerland also borders the Principality of Liechtenstein (40,000 inhabitants), which uses the Swiss franc as a currency and has almost identical grocery prices.

[3]The CHF/EUR exchange rate was stable throughout this period.

and resist temporary shocks. These estimated effects decay with distance, indicating that a household's probability of engaging in cross-border shopping decreases with travel time. Second, I document various heterogeneities and find larger effects among poorer, younger, and larger households in response to the policy. Third, I provide novel evidence that households combine their trips to work with cross-border shopping if they commute towards the border. Fourth, I find that cross-border shopping is more pronounced in areas with cheaper neighboring countries, suggesting a price elasticity of 0.61.

This paper relates to two strands of the literature. First, it contributes to the previous work on cross-border shopping, documenting that both consumers and retailers respond to changes in relative prices. For instance, a depreciation of the US dollar reduces the consumers' propensity to cross into Canada (Chandra, Head, and Tappata, 2014) while increasing US employment and the number of establishments close to the border (Campbell and Lapham, 2004). Similarly, Asplund, Friberg, and Wilander (2007) show that a cut in Danish spirits taxes reduces alcohol sales in Sweden, and Baker, Johnson, and Kueng (2021) find that customers in the United States use cross-border shopping to escape local sales taxes. Finally, Friberg, Steen, and Ulsaker (2022) demonstrate that the marginal customer further inland reacts stronger to foreign price changes while households close to the border shop abroad anyway. This implies that the response to relative price changes is an incomplete measure of the level of cross-border shopping. Therefore, I follow an alternative approach and use a natural experiment that restricts access to cross-border shopping completely rather than changing relative prices.

At least two other papers tackle the topic of cross-border shopping through COVID-19-related border closures, answering, however, different questions. First, Friberg, Halseth, Steen, and Ulsaker (2024) investigate the effect on taxes and find that Norwegian cross-border shopping reduces national tax revenues by 3.6% nationally and 27% in border regions. Second, Burstein, Lein, and Vogel (2022) study cross-border shopping in Switzerland using data from Nielsen and conclude that it lowers the cost of living by over 14% in some regions. In contrast to these papers, I focus on the customers' behaviors and the rich heterogeneities therein. My data – matching unique transaction records with administrative data – may be better suited for this analysis than the Nielsen data, whose self-recorded reporting errors are correlated with demographic variables (Einav, Leibtag, and Nevo, 2008).

In a broader context, this paper also links to the research on spatial shopping in general and trip chaining, showing that customers deliberately plan and adapt their grocery expenditures and shopping trips. For example, Agarwal, Jensen, and Monte (2022) suggest that consumers purchase products with a low storability within a shorter distance. Additionally, previous work on spatial trip-chaining demonstrates that customers strategically visit multiple non-tradable services along their daily travels. This travel behavior generates consumption externalities that explain one-third of the spatial concentration in non-tradable services (Oh and Seo, 2023) and Miyauchi, Nakajima, and Redding (2022) show that modeling trip-chaining is crucial to understanding the decreased demand for non-traded services following the shift to remote working during the COVID-19 pandemic. Furthermore, trip-chaining can cause complex adaptations in the spatial equilibrium with potentially winning and losing stores (Relihan, 2024). My paper contributes to this literature by showing that households strategically include their cross-border shopping trips into their daily commutes to work.

The remainder of this paper is structured as follows. Section 2.2 introduces the grocery and administrative data. Section 2.3 discusses the empirical strategy, while Section 2.4 presents my findings. Section 2.5 concludes.

2.2 Data

I combine unique transaction data from the largest Swiss retailer with administrative data from the Federal Statistical Office on a 100×100 meter spatial resolution.

The grocery data provides information on every customer-linked purchase at the retailer *Migros* in 2020, collected through their loyalty program in which customers identify themselves at the checkout with their loyalty card in exchange for exclusive offers and discounts. This loyalty program captures 79% of the retailer's total sales, and 2.4 million customers regularly participate in it (meaning. 33% of all Swiss residents above legal age). Furthermore, Migros charges the same prices throughout the country, independently of local purchasing power, wages, and costs. Hence, prices are not endogenously lower close to the border. Stores of similar size also generally offer similar goods, except for local products. The data set contains the universe of 600 million customer-linked purchases for the year 2020 and provides information on individual customer characteristics, including the location of their residence coded on a grid of 100 $\times$ 100 meter cells, their age, and household type.

I enrich the purchase data with individual-level administrative records for the entire Swiss population 8.7 million inhabitants in 2020). The *Population and Households Statistics* includes individual and household characteristics, including information on gender, age, household members, and residence location on the same 100 $\times$ 100 meter grid. The *Old Age and Survivors Insurance* provides annual gross labor market income, which I adjust by the square root of household size.[4] Finally, the administrative *Structural Surveys* add education and commuting behavior for the sub-sample of individuals participating in the survey.[5] Education is categorized as either primary, secondary, or tertiary education, and the commuting behavior is characterized by travel times in minutes, means of transport, and the municipality of the work location.[6]

Both data sets measure addresses on the same spatial grid spanning 350,000 cells over the entire country with a mean population of 25 residents. I merge the two data sets by identifying unique pairs of customers and residents using the common variables grid cell

[4]The calculation is income adjusted $= \frac{\text{income total}}{\sqrt{\#\text{household members}}}$, where I consider all household members, including small children. The adjustment follows one of the equivalence scales suggested by the OECD. I compute *income total* as the household's annual income by summing the income of all household members.

[5]This representative cross-sectional survey selects 200,000 people above age 15 every year. Individuals can be selected repeatedly, and participation is mandatory. To measure education, I use the highest-reported education between 2010 and 2021 and exclude individuals younger than 30 to capture students. For commuting, I only use the surveys since 2018 as workplaces are less stable than education.

[6]Primary (or compulsory) education ends at the latest after around eleven mandatory years of school (including kindergarten). Individuals who completed high school or an upper-secondary specialized school have a secondary education. Completing any degree at a university, university of applied sciences, or university of teacher education results in a tertiary degree.

and age. This approach matches 1.3 million customers in the grocery data uniquely to a citizen and her household in the administrative data. Hence, I can match 54% of the 2.4 million regular customers, corresponding to 20% of all adult Swiss residents. The outcome of interest throughout this analysis is a household's total grocery expenditures in a given week. I aggregate the individual shopping trips into weekly baskets and exclude customers who moved in 2020 as well as those spending less than 100 Swiss francs per capita a month before the shock (equalling 112 USD on July 29, 2024), as their baskets might not capture the overall consumption accurately. This procedure generates a final data set including 757,000 households and 40 million weekly consumption baskets.[7]

Table 2.1 shows summary statistics for the households and displays for how many of them I observe a given variable. The average matched household has an income of 60,000 Swiss francs (adjusted for the square root of household size), and the mean cardholder is 56.6 years old, while 44.4% have a tertiary education, and 80% live in multi-person households. Comparing these statistics to the entire administrative data shows that the matched sample represents the population well. Further, Table 2.2 shows summary statistics for the transactions. The average household makes 6.1 transactions and spends 92 Swiss francs (104 USD on July 29, 2024) per week. This corresponds to roughly 63% of the average household's grocery expenditures based on administrative consumption surveys. Looking at different subgroups, expenditures increase with household size and income, while they are hump-shaped for age. A comparison to the entire transaction data shows that the matched customers' shopping behavior matches expenditures in the full sample well.

Finally, I calculate car travel times to foreign shopping locations and workplaces. To this end, I scrape the location and Google review counts of all foreign supermarkets within 20 km of the Swiss border from *Google Maps*. This results in 117 cross-border locations and a total of 1,787 stores, of which 691 have at least 100 Google ratings. Table B.1 lists the largest identified cross-border locations, showing the number of stores with at least

[7] See Kluser and Pons (2024) and Kluser, Seidel, and von Ehrlich (2024) for additional information on the two data sources, the matching procedure, and the representativeness of the matched households for the general population.

TABLE 2.1: Household Summary Statistics

	Final Sample		Population	
Panel a)	Mean	SD	Mean	SD
Age	56.63	15.91	50.43	18.17
Income (1,000 CHF)	100.66	129.99	106.01	132.48
Income Adjusted (1,000 CHF)	60.09	80.29	64.90	78.96
Time Home to Work (min.)	28.21	23.02	29.12	23.70
Time Home to Border (min.)	57.69	24.27	56.13	25.28
Time Work to Border (min.)	58.28	31.75	56.08	23.81
Panel b)	Pct.	N	Pct.	N
Education		505,309		4,413,173
Primary	9.8	49,747	11.3	498,292
Secondary	45.8	231,237	44.3	1,954,810
Tertiary	44.4	224,325	44.4	1,960,071
Household Size		757,629		7,043,734
1	19.3	146,593	20.9	1,471,897
2	36.0	272,663	36.1	2,544,442
3-4	36.1	273,742	33.8	2,381,660
5+	8.5	64,631	9.2	645,735
Language		756,936		7,036,484
German	76.2	576,786	71.2	5,010,326
French	20.2	153,279	24.1	1,697,654
Italian	3.5	26,871	4.7	328,504
Population Density		756,936		7,036,484
Urban	24.4	184,556	30.2	2,122,190
Suburban	57.6	436,372	51.9	3,649,595
Rural	18.0	136,008	18.0	1,264,699
Nationality		757,568		7,042,341
Swiss	85.6	648,380	74.0	5,210,215
European	12.5	94,605	22.0	1,551,076
African	0.5	3,507	1.1	77,266
Asian	1.0	7,255	1.9	131,883
N.American	0.1	1,025	0.3	21,530
S.American	0.4	2,796	0.7	50,371
Commuting Mode		103,295		923,718
Car	59.0	60,973	55.4	511,779
Public Transport	24.8	25,595	27.8	256,869
Other	16.2	16,727	16.8	155,070
Observations		757,629		7,043,734

Notes: The table shows summary statistics for the customers uniquely matched to the administrative data and compares them to the entire Swiss population above legal age. *Income* equals the total annual labor market income of a household in 1,000 Swiss Francs, and *Income Adjusted* adjusts for the square root of household size. All *Time* variables measure the uncongested car travel time in minutes to the work location or the closest cross-border location. The variables *Commuting Mode* and *Education* are only available for the sub-sample participating in the *Structural Surveys*.

TABLE 2.2: Transactions Summary Statistics

Group	Mean	SD	p50	p1	p99
Weekly Grocery Purchases					
Expenditures in Matched Sample	92.5	64.1	75.5	12.9	300.7
Expenditures in Full Sample	88.7	62.3	72.0	12.2	293.1
Shop Visits in Matched Sample	6.1	3.5	5.5	0.8	17.5
Shop Visits in Full Sample	6.1	3.5	5.5	0.8	17.4
Expenditures by Age Group					
20–34	82.2	53.9	68.7	11.8	251.1
35–44	107.9	70.6	91.6	13.8	317.6
45–54	110.2	74.5	92.1	14.2	336.7
55–64	94.6	63.7	79.1	13.6	301.7
65–74	79.4	51.3	67.1	12.7	247.4
75+	68.3	44.4	57.4	11.2	217.4
Expenditures by Income Quintile					
25,000–73,000	79.3	53.1	65.7	12.7	255.5
73,001–106,000	90.7	59.7	75.5	13.4	280.5
106,001–137,000	104.0	66.6	89.6	14.2	302.4
137,001–181,000	111.9	71.0	97.6	14.3	321.4
181,001+	119.3	79.4	102.5	13.6	357.8
Expenditures by Education					
Primary	69.8	47.7	57.3	11.4	232.8
Secondary	90.5	60.2	75.6	13.3	284.2
Tertiary	107.9	71.9	91.3	13.7	328.8
Expenditures by Household Size					
1	60.0	37.3	51.8	11.2	191.0
2	83.2	51.5	72.5	12.6	244.0
3–4	111.5	71.0	97.1	14.5	319.9
5+	125.0	84.8	105.9	14.4	373.6
Transactions in Matched Sampled		40,179,519			
Transactions in Full Sampled		95,192,993			

Notes: The table shows summary statistics for the weekly expenditures and trip frequency of customers that I can match to residents in the administrative data. I compare these statistics to the full transaction data set, including the unmatched customers, and report statistics on sub-samples for the matched data. The statistics for the *Full Sample* apply the same sample selection criteria used for the matched sample to the 120 million weekly baskets (600 million shop visits) in the transaction data set.

100 and 500 Google ratings. A municipality with a large number of stores typically also has many larger stores with numerous Google reviews, and correlations between the population, the number of stores, and the number of stores with more than 100 and 500 Google ratings are very high, lying between 0.83 and 0.92. As cross-border shoppers likely focus on larger stores, I define a cross-border location as a foreign municipality with at least three stores that have more than 100 Google ratings.[8] Next, I scrape the car travel time from every raster cell to all these locations from a national online mapping service (*search.ch*) and select the shortest trip for each cell. One-fifth of all households reaches the closest cross-border location within a 30-minute car drive, while the maximum distance is three hours. Following the same approach, I calculate distances to workplaces. Table 2.1 shows the average car travel time to the closest cross-border location (57 minutes) and the work location (28 minutes). 59% commute to work by car, while 24.8% use public transportation.

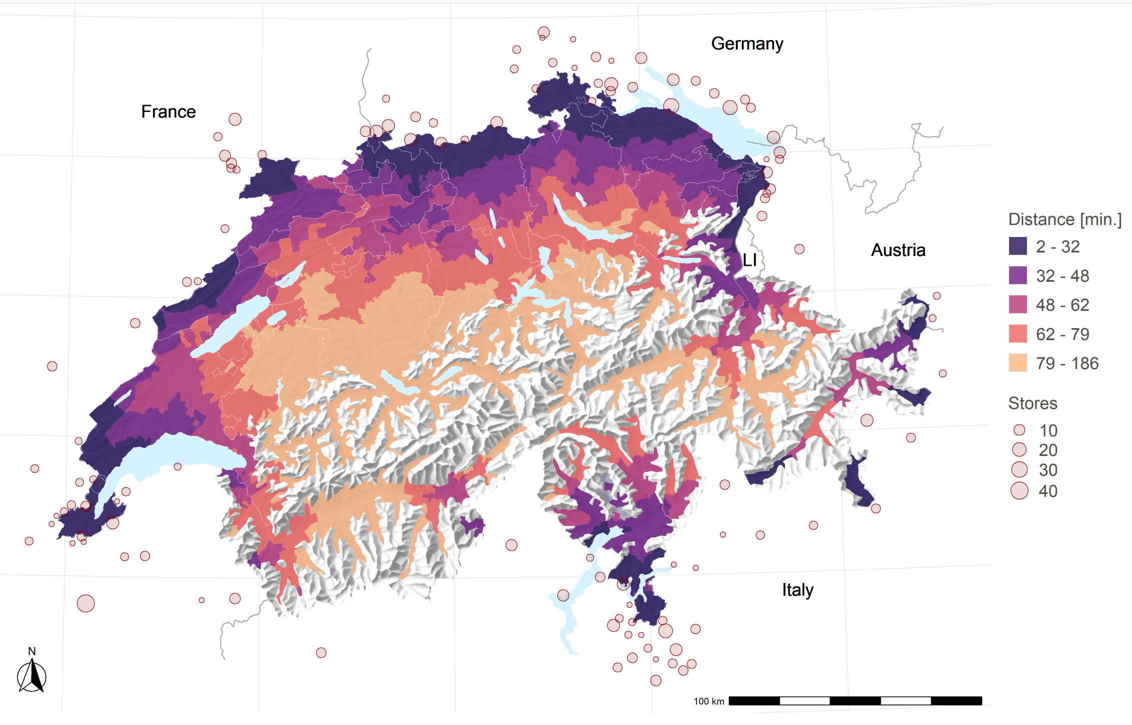
2.3 Empirical Strategy

I study the impact of the border closure on household expenditures by comparing households living within a half-hour car drive from a cross-border location (the first quintile) to those living far enough inland such that they typically do not shop abroad. Hence, I define the control group as households living more than 80 car minutes away (the fifth quintile) and drop all individuals residing within the doughnut area to ensure a clean control group. This results in a sample of roughly 150,000 treated and control households.[9] Figure 2.1 shows these travel distance bins to the closest foreign location across Switzerland. The figure further illustrates the importance of explicitly using travel times to cross-border locations rather than the Euclidean distance to the border due to the dispersion of these shopping locations and the morphology of the landscape.

[8] My results are robust if I define cross-border locations alternatively as (i) locations with at least one store with 500 Google reviews or as (ii) locations with at least three stores with 500 Google reviews.

[9] If a fraction of control units still reacted to the border closure, my estimates should be regarded as a lower bound. I will address this further in Section 2.4.2, showing that my results are robust if I use alternative comparison distances of 90 or 100 minutes.

FIGURE 2.1: Distance to the Closest Cross-Border Shopping Location



Notes: The figure shows the quintiles of car driving times to the closest cross-border shopping location on the municipality level. The dots show all 117 cross-border locations within 20 kilometers of the Swiss border, and the dots' size indicates the number of supermarkets at this location.

I use a difference-in-differences model to estimate the average treatment effect. Since all political regulations, grocery supply adaptations, and consumers' behavioral changes affect both the treatment and control group, I attribute any deviation after the intervention to cross-border shopping. As some households record zero expenditures in a given week, I follow Chen and Roth (2024) and Wooldridge (2023) and estimate the following QMLE-Poisson model:[10]

$$Y_{it} = \exp\left(\alpha_i + \gamma_t + \sum_{j=1}^{52} \beta_j (D_i \times T_j) + \tau z_{it}\right) \epsilon_{it}, \tag{2.1}$$

[10]Chen and Roth (2024) show that using a linear model with $log(Y+1)$ as a dependent variable does not allow interpreting the coefficients as percentage changes. Instead, estimating a QMLE-Poisson model and reporting the transformed coefficients $\hat{\beta}_{ATT\%} = exp(\hat{\beta} - 1)$ leads to the desired result.

where Y_{it} are the grocery expenditures of household i in week $t \in 1, \ldots, 52$. α_i and γ_t are the household- and week-specific fixed effects, controlling for unobserved heterogeneity. D_i is an indicator variable that equals one if household i is in the treatment group, the dummy variables T_j indicate the weeks of the year 2020, and β_j are the associated pre- and post-treatment coefficients for each period j. Finally, z_{it} measures the time-varying cantonally reported cases of COVID-19. Controlling for the COVID-19 cases accounts for the differential exposure to the pandemic over time, as the first wave of COVID-19 hit Switzerland in 2020 from the South, with the largest initial number of cases in the Italian-speaking region (Ticino). Therefore, these households were sooner and stronger affected by the outbreak than people in the North, and z_{it} controls for these varying exposures. Treatment starts in week twelve, and I normalize coefficients to the average in the pre-treatment period. I cluster standard errors in the QMLE Poisson regressions on the zip-code level and report the transformed coefficients $\hat{\beta}_{ATT\%} = exp(\hat{\beta} - 1)$, which gives the average proportional treatment effects and allows me to interpret the coefficients as percentage changes. I calculate the corresponding standard errors using the delta method.[11]

To analyze heterogeneities in the treatment effect, I use a static model and interact the treatment indicator with a categorical variable x_i:

$$Y_{ikt} = \exp\left(\alpha_i + \gamma_{tk} + \sum_{k \in \mathcal{K}} \beta_k (D_i \times Post_t \times x_{ik}) + \tau z_{it}\right) \epsilon_{ikt}, \tag{2.2}$$

where $Post_t = 1$ if $t \geq 12$, $k \in \mathcal{K}$ indexes the individual categories of x_i, $x_{ik} = \mathbb{1}(x_i = k)$, and β_k is the average treatment effect for each group k. In this specification, the time dimension of the treatment effect collapses to a single post-treatment coefficient. I allow the time fixed effect to vary between the different groups k by including week-group fixed effects γ_{tk} as the pandemic might affect the individual groups differently.

[11] Alternatively, I calculate standard errors from 1,000 clustered bootstrap replications for the main results. The bootstrapped standard errors give similar results.

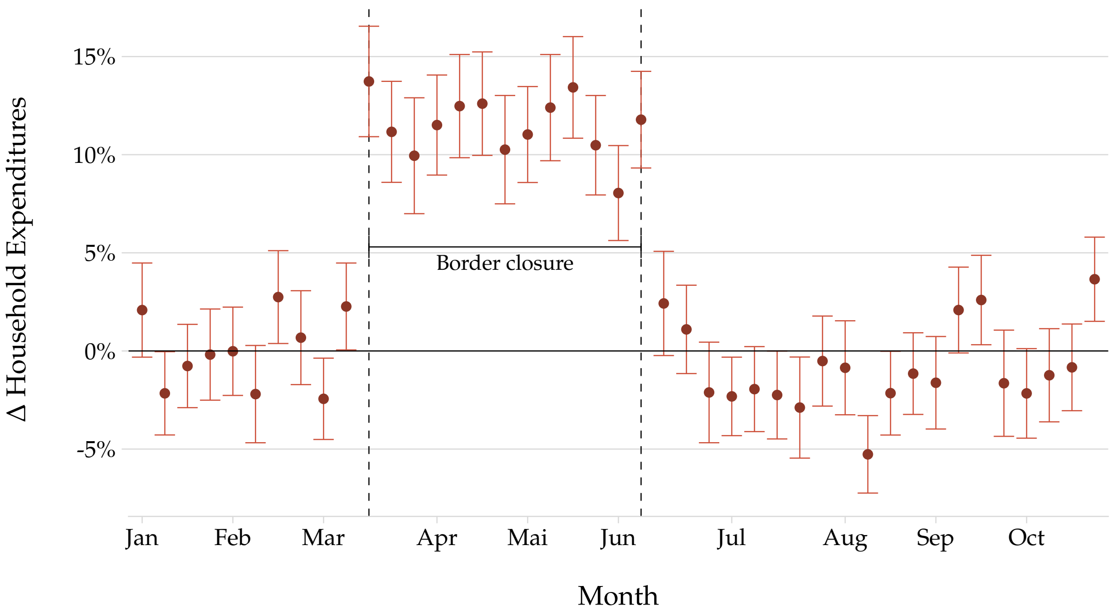
2.4 Results and Discussion

This section presents three sets of results. First, I study the treatment effects of the border-closing policy on grocery expenditures over time. Second, I examine the effect's decay with distance, assessing how far customers are willing to travel for lower prices. Third, I show diverse heterogeneities of the average treatment effect, including socioeconomic household characteristics, culture, and commuting behavior, as well as foreign grocery prices. This provides rich insights into the varying patterns of consumer mobility in space in response to price differences.

2.4.1 Dynamic Treatment Effects

Figure 2.2 shows the results for the dynamic difference-in-differences outlined in Equation (2.1). The borders close in week 12 and reopen in week 25, and vertical dashed lines indicate both events. Additionally, Table 2.3 reports the corresponding average treatment effects, grouping the periods during the border closure and after the reopening together. I find that the border closure temporarily increases domestic grocery expenditures by 10.9% at the border in comparison to households residing further inland, with week-specific effects ranging from 8% to 14%. These findings are in line with Burstein, Lein, and Vogel (2022), who estimate that Swiss households close to the border spend roughly 8% of their expenditures abroad. Further, this expenditure shift is immediate and remains constant as long as the border is impassable. After the reopening, expenditures immediately drop to the previous level. Hence, although households in border regions temporarily increased their spending at domestic supermarkets, they did not adjust their cross-border shopping behavior through the border closure and completely switched back to their old behavior as soon as possible. This result suggests that cross-border shopping follows deeply rooted routines that withstand major temporary shocks. Furthermore, Figure 2.2 shows that most coefficients in the initial weeks after the reopening are below zero with an average of -1.2%. This increase in cross-border consumption after the reopening is most likely due to a temporary catch-up or stockpiling effect.

FIGURE 2.2: Dynamic Treatment Effects



Notes: The figure shows the border closure's effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes. I indicate the period of border closure by vertical dashed lines. The regression estimates Equation (2.1) and uses 12 million observations. Coefficients are normalized to the pre-treatment periods' average, and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

One concern might be that consumers adapted their shopping behavior before the actual introduction of pandemic restrictions, especially in strongly affected areas (for example, in the form of stockpiling or by avoiding larger crowds). Yet, the insignificant pre-treatment coefficients in Figure 2.2 do not indicate a potential violation of the parallel trend assumption between treated and control units, suggesting that households living in the border region and further inland did not react differently to the pandemic's onset. This conclusion remains unchanged (and pre-treatment coefficients insignificant) if I do not control for the local number of COVID-19 cases.

2.4.2 The Distance Decay Function

Throughout this paper, I choose a doughnut–specification with control households living at least an 80-minute car drive from the closest cross-border shopping location. Yet,

TABLE 2.3: Average Treatment Effects

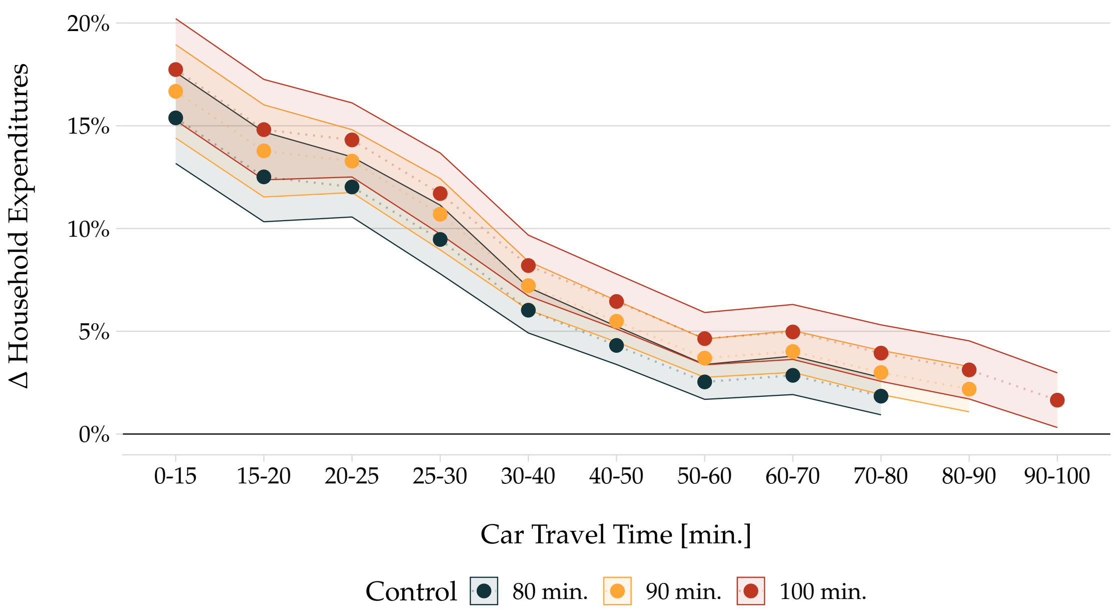
Dep. Var.: Household Expenditures	
Treat × Border Closed	0.109***
	(0.006)
Treat × Border Open	-0.012***
	(0.004)
n	12,030,579

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes. The regression follows Equation (2.1) but groups the periods during and after the border closure together (*border closed* and *border open*, respectively). Standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

choosing the radius of the inner doughnut defines the households left out in my analysis and features a trade-off between (i) ensuring that the treatment does not contaminate the control units and (ii) having a large and representative enough control group. If households living 80 minutes from a cross-border location are still affected, my results should be regarded as lower bounds.

To investigate this, I now consider larger doughnut areas. Figure 2.3 compares the distance decay function for my preferred specification to two alternative approaches based on control households with at least a 90-minute and 100-minute trip to the closest cross-border location. The results indicate that some control units in my baseline results are possibly still affected by the border closure, as the coefficient for the last distance bin is significant. As the alternative approaches consistently report higher point estimates, I likely underestimate the true effect. On the other hand, the size of the control group shrinks significantly from 150,000 to 68,000 and 28,000 households for the stricter definitions of control units. To balance this trade-off, I select the most conservative approach and present in the paper all estimates with a control group consisting of households living 80 minutes from the border. In the Appendix, Figure B.2 and Table B.2 to Table B.6 replicate all results for a control distance of 100 minutes and show that all conclusions remain the same.

FIGURE 2.3: Decay of the Treatment Effect



Notes: The figure shows the border closure's average treatment effect on household expenditures for households living within a certain distance bin. I compare these treated units to households living further away than 80, 90, and 100 minutes from the closest cross-border location, respectively. Standard errors are clustered at the zip code level. The regressions estimate Equation (2.2) and use 17.4 million observations in all three cases. Coefficients are exponentiated such that they equal proportional effects.

Focusing on the preferred specification of 80 minutes in Figure 2.3, I find that households living within a short distance of 15 minutes from the closest cross-border destination increase their expenditures by 16% during the border closure. This effect first declines linearly up to a distance of 50 minutes before flattening out, although remaining significant for at least 80 minutes. Note that these distances are potentially lower bounds of the actual travel distance as customers might prefer to shop at other foreign stores further away rather than at the closest location.

2.4.3 Variation Across Socioeconomic Characteristics

Consumers' preferences for cross-border shopping may vary based on their socioeconomic background. Hence, I analyze treatment effect heterogeneities for different household characteristics and Table 2.4 reports the estimation results of Equation (2.2) for the

TABLE 2.4: Treatment Effects by Socioeconomic Subgroups

Dep. Variable: Household Expenditures							
a) Household Size		b) Age		c) Income		d) Education	
Group	Coeff	Group	Coeff	Group	Coeff	Group	Coeff
1	0.068***	20–34	0.138***	Q1	0.150***	Primary	0.137***
	(0.006)		(0.010)		(0.008)		(0.010)
2	0.103***	35–44	0.142***	Q2	0.144***	Secondary	0.108***
	(0.007)		(0.009)		(0.008)		(0.006)
3-4	0.136***	45–54	0.134***	Q3	0.128***	Tertiary	0.108***
	(0.008)		(0.008)		(0.008)		(0.007)
$\geq$5	0.145***	55–64	0.122***	Q4	0.117***		
	(0.009)		(0.008)		(0.007)		
		65–74	0.130***	Q5	0.099***		
			(0.009)		(0.009)		
		75+	0.114***				
			(0.010)				
p-value	0.000	p-value	0.014	p-value	0.000	p-value	0.007
n	6,434,950	n	6,433,731	n	5,148,635	n	4,199,790

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes, separately for different household characteristics. These characteristics include the *household size*, *age* of the registered cardholder, household *income* adjusted by the square root of household size, and the highest *education* in the household. The regression estimates Equation (2.2), standard errors are clustered at the zip code level, and the reported p-values test the equality of all coefficients. Coefficients are exponentiated such that they equal proportional effects.

variables household size, age, income, and education in the panels a) to d). The table also reports p-values, testing the treatment effects' equality over the different groups (meaning, the null hypothesis is $\beta_k = \beta \; \forall k$).

First, I find that the effect increases in household size. While a one-person household increases their expenditures by 6.8% in response to the border closure, I document an increase by 10.3% for two-person households, and by 14% for households with at least three members. Hence, larger households engage in more cross-border shopping. Traveling abroad to shop at lower prices is particularly tempting if you buy large quantities, as it increases the trip's savings while the trip's traveling costs are fixed. Hence, relative costs decrease. Such economies of scale likely explain this finding, as the summary statistics in Table 2.2 show that larger households spend more money on groceries overall and consume larger quantities, making cross-border shopping more attractive for them.

Second, I find heterogeneous effects over age in the response to the border closure. The estimated effect lies around 14% for young households between age 20 and 44 and decreases slowly as households become older. Yet, even retired households after age 65 show a relatively high response of roughly 12%, while their total expenditures are markedly lower (see Table 2.2). This result is likely driven by the sharp decline in their income after retirement, which induces them to still shop abroad at lower prices. Furthermore, they presumably also face lower opportunity costs. Note that this heterogeneity can either be due to age or cohort effects, as the short sample period does not allow for disentangling them.

Third, I look at income. On the one hand, one should expect households with a lower income to engage in more cross-border shopping as they have higher import elasticities (see Auer, Burstein, Lein, and Vogel, 2023) and spend a higher share of their income on groceries. For instance, high-income households in my data (with a monthly income above 12,000 Swiss francs) spend 1.6% of their income on groceries compared to 3.5% for lower-income households (with a monthly income between 4,000 and 8,000 Swiss francs). On the other hand, lower car ownership might constrain the mobility of less affluent households. While 90% of high-income households (with a monthly income above 12,000 Swiss francs) own a car, this holds for only 77% of lower-income households (with a monthly income between 4,000 and 8,000 Swiss francs), according to the Federal Statistical Office. Similarly, lower-income households travel, on average, shorter distances on a given day (30.2 kilometers vs. 40.8 kilometers).
The results in panel c) show that the first argument dominates the narrative: the treatment effect decreases from 15.0% for the lowest-earning quintile to 9.9% for the highest-earning households. Hence, although traveling costs are relatively high for many of them, lower-income households still engage in more cross-border shopping activity.

TABLE 2.5: Treatment Effects by Cultural and Spatial Subgroups

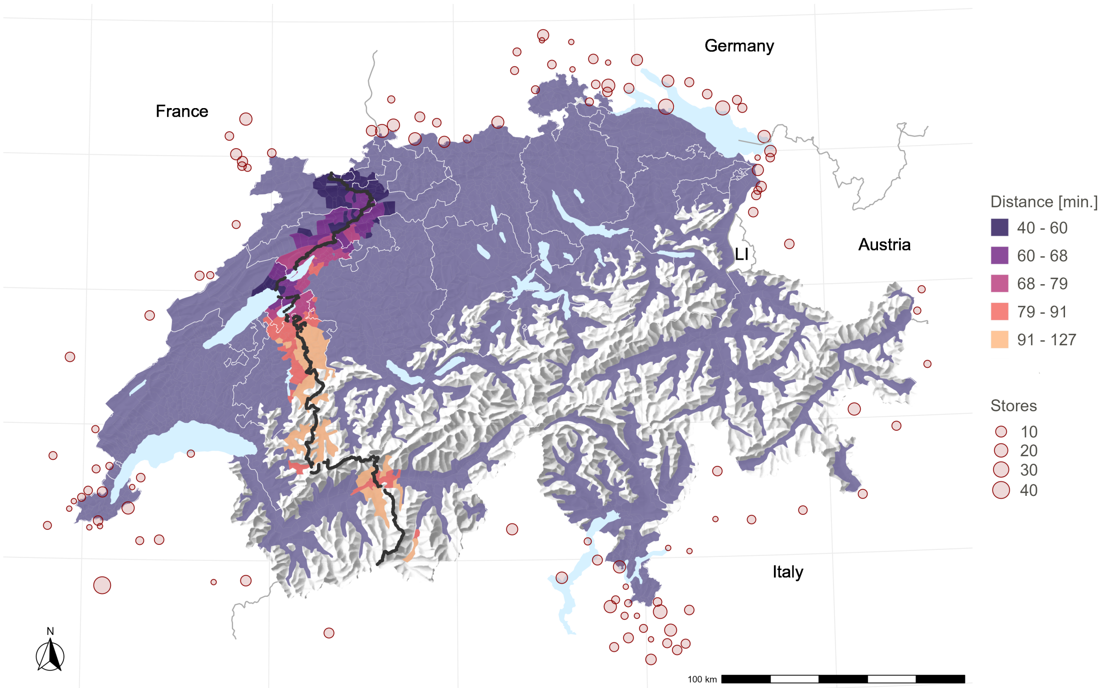
Dep. Variable: Household Expenditures			
a) Nationality		b) Country	
Group	Coeff	Group	Coeff
African	0.197***	AT	0.074***
	(0.038)		(0.013)
Asian	0.163***	GER	0.110***
	(0.025)		(0.008)
European	0.155***	FR	0.120***
	(0.012)		(0.009)
N.American	0.166**	IT	0.350***
	(0.062)		(0.040)
S.American	0.120**		
	(0.041)		
Swiss	0.105***		
	(0.006)		
p-value	0.000	p-value	0.000
n	6,434,398	n	6,235,192

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes, separately for different household characteristics. These characteristics include the cardholders' *nationality* and the *country* of their closest cross-border shopping location. The regression estimates Equation (2.2), standard errors are clustered at the zip code level, and the reported p-values test the equality of all coefficients. Coefficients are exponentiated such that they equal proportional effects.

Fourth, higher-educated individuals may have broader knowledge and access to more information to strategically optimize their consumption behavior while being less budget-constrained. Households with at least one member holding a tertiary education react less to the border closure than comparable households further inland. While high-educated households increase their expenditures by 10.8%, I estimate a higher effect of 13.7% for low-educated households.

Overall, these socioeconomic heterogeneities suggest that many households engage in cross-border shopping either (i) because of large potential savings relative to their low income or (ii) because they have high overall grocery expenditures and can, therefore, save more money in absolute terms.

FIGURE 2.4: German-French Language Border



Notes: The figure shows the quintiles of car driving times to the closest cross-border shopping location in a 20-kilometer-band around the French-German language border on the municipality level. The dots show all 117 cross-border locations within 20 kilometers of the Swiss border. The dots' size indicates the number of supermarkets at this location, and the black line is the language border.

2.4.4 Culture

Beyond the socioeconomic background of households, I address the role of cultural differences, as citizens from various cultural origins may have different shopping preferences. To this end, I analyze (i) a heterogeneity between customers of different nationalities and (ii) households living in close proximity but on opposite sides of the French-German language border within Switzerland.

To begin with, Panel a) in Table 2.5 shows the heterogeneous response of individuals from different nationalities, estimating again the regression Equation (2.2). I observe that

TABLE 2.6: Cultural Differences: Effect at Language Border

	Dep. Var: HH Expenditures		
Dist. to ntl. Border	German	French	p-value
Treat × 30-45 min.	0.101***	0.006	0.000
	(0.012)	(0.015)	
Treat × 45-55 min.	0.055***	0.025	0.175
	(0.016)	(0.017)	
Treat × 55-65 min.	0.041***	0.044***	0.842
	(0.011)	(0.011)	
n	1,158,263		

Notes: The table shows the border closure's average treatment effect on household expenditures for households living within 10 kilometers of the German-French language border. I compare these treated units to same-language households living further away than 80 minutes from the closest cross-border location. The regression estimates Equation (2.2), and standard errors are clustered at the zip code level. The reported p-values test the equality of the two coefficients in the same distance bin. Coefficients are exponentiated such that they equal proportional effects.

Swiss households are relatively less likely to shop abroad compared to foreign citizens. A Swiss citizen in the border region spent 10% more in response to the border closure, while other Europeans, North Americans, and Asians increased their expenditures by 15-16%, South Americans by 12%, and Africans by 19%, suggesting cultural differences in the preferences for foreign goods.

Furthermore, I use the intra-national Swiss language border between the French-speaking part of Switzerland in the West and the German-speaking part on the other side of this border to measure any cultural differences based on language. Figure 2.4 displays the language border crossing the entire country from North to South.[12] I use again Equation (2.2) to estimate the treatment effect separately for French- and German-speaking households living within 10 kilometers of the language border compared to households further inland speaking the same language. I estimate treatment effects separately for households living between 30-45, 45-55, and 55-65 minutes from the national

[12] I exclude in this analysis the German-Italian border in the South because very few people on both sides have comparable access to cross-border locations as this language border lies in the mountains.

border compared to households farther away than 80 minutes.[13] This empirical strategy relies on the testable assumption that households within this 20-kilometer band are comparable. Table 2.6 displays the estimation results for different distance bins to the border, and the reported p-value tests for equality of the coefficient in the two language regions. I find a stronger response for German-speaking households in the first distance bin but no significant difference for the other two bins further inland. One potential explanation for the difference in the first distance bin might be that Germany has lower grocery prices than France (see Table 2.8). An alternative reason could be that households prefer to shop in the country speaking their own language, and the German stores may be more attractive than the French ones.

2.4.5 Commuting and Trip Chaining

A key determinant of a household's shopping behavior may be her daily commute to work (see, for example, Miyauchi, Nakajima, and Redding, 2022). First, households can combine commuting and shopping through trip chaining if their workplace is closer to the border than their home. Second, frequent commuting trips to work may alter a household's perception of distance and traveling costs and influence her likelihood of traveling abroad, even if her workplace lies far away from the border. Hence, I use Equation (2.2) to estimate the treatment effect separately for households commuting either from home (i) towards foreign shopping locations or (ii) farther inland, away from cross-border locations. I focus on households that live 20 to 35 minutes from the border and report commuting by car.

Table 2.7 shows the estimation results. On the one hand, households with a commute taking them 5 to 15 minutes closer to the border increase their cross-border shopping by 14.5% in response to the border closure. For households whose workplace is 15-25 minutes closer to a cross-border location, I estimate an effect of 14.8%. On the other hand, I observe for households commuting away from the border lower effects of 8.8%

[13] I cannot report results for households living closer to the next cross-border location, as no household living along the language border can reach a cross-border location in less than 30 minutes.

TABLE 2.7: Treatment Effect for Different Commuting Behaviors

	Dep. Var: Household Expenditures		
Δ Border Access	Commute Towards Border	Commute Away f. Border	p-value
Treat × 5-15 min.	0.145**	0.088***	0.439
	(0.017)	(0.017)	
Treat × 15-25 min.	0.148***	0.107***	0.008
	(0.051)	(0.024)	
n	357,492		

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes for different household commuting trips. These trips include commutes by car for 0-15 minutes and 15-25 minutes, either towards the national border (bringing the commuter closer to a cross-border location) or further away from the border in comparison to the household's home. The regression estimates Equation (2.2) and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

and 10.7%, respectively. Therefore, these two observations provide conclusive evidence that households combine work commutes with cross-border shopping trips in the form of trip chaining.

2.4.6 Variation Across Cross-Border Locations

Finally, I look at the role of neighboring countries and their grocery prices. Panel b) of Table 2.5 shows the spatial variation of the effect by estimating heterogeneous treatment effects for the four neighboring countries Austria, Germany, France, and Italy.[14] The results show a large estimate for households living closest to Italy (35%), with smaller values for households living close to Germany, France, and Austria (12%, 11%, and 7.4%, respectively). To assess the role of prices behind these findings, I show in Table 2.8 national price level indices averaged over the period of 2015–2020 for different major product categories and how much these products are cheaper compared to Switzerland. While each

[14] For this spatial heterogeneity, I use week fixed effects compared to the week-group fixed effects in the case of socioeconomic variables.

TABLE 2.8: Prices in Neighboring Countries, 2015–2020

	Austria		France		Germany		Italy	
Category	PI	vs. CH	PI	vs. CH	PI	vs. CH	PI	vs. CH
Clothing and Footwear	102.83	-20%	105.53	-18%	98.80	-23%	100.52	-22%
Consumer Goods	106.37	-20%	107.02	-20%	103.12	-23%	105.18	-21%
Food and non-Alcoholic Beverages	120.47	-28%	112.38	-33%	102.52	-39%	109.30	-35%
Households Appliances	95.08	-21%	105.37	-12%	101.18	-16%	101.50	-15%
Recreation and Culture	113.27	-26%	107.28	-30%	104.57	-32%	100.10	-35%
Restaurants and Hotels	108.67	-35%	119.73	-28%	105.88	-36%	104.02	-38%

Notes: The table shows prices in neighboring EU countries averaged over the six years before and during the first wave of the COVID-19 pandemic, 2015–2020. Prices are shown as price indices (PI) for different product categories and relative to the category's price index in Switzerland. In each year, the EU27 average is set to 100.

product category is in every country cheaper than in Switzerland, relative prices between these neighboring countries vary for different product categories.

Using the price level index for consumer goods, the heterogenous coefficients are negatively correlated with the price index of the neighboring countries, meaning that higher foreign prices correspond to less Swiss cross-border shopping. Based on a back-of-the-envelope calculation, a 1% increase in the price index of a neighboring country is associated with a 0.61% decline in cross-border shopping expenditures. Note that any interpretation of this as a price elasticity assumes that all households assigned to a given neighboring country face the same price difference at home and abroad, which seems plausible as our retailer charges the same prices throughout the country. Yet, not all foreign retailers charge the same prices across the entire country, and foreign prices may be higher close to the Swiss border. Additionally, this calculation assumes that residential location choice does not depend on the households' cross-border shopping preferences and that customers buy the same products at home and abroad.

2.4.7 Robustness

Complementing the previous discussion of the doughnut design in Section 2.4.2, I discuss two additional robustness checks. First, I report in Figure B.3 the dynamic estimates for the full sample of transaction data rather than focusing on the sub-sample of customers matched to residents in the administrative data. The observed changes are negligible. Second, I also use another definition of cross-border locations where I only consider very large foreign stores that may be more attractive to travel to (Figure B.4). The changes in the coefficients are minimal, further supporting my findings.

2.5 Conclusion

Cross-border shopping provides researchers with a useful setting to analyze the households' heterogeneous willingness to travel for lower prices. While Friberg, Steen, and Ulsaker (2022) show that the traditional study of cross-border shopping through changes in relative prices does not measure cross-border shopping one-to-one, the Swiss COVID-19-related border closure (among others) provides a unique natural experiment that I exploit.

I find that cross-border shopping is a widespread and persistent phenomenon in Switzerland and that domestic sales would be 10.9% higher in border regions without it. I then investigate heterogeneities, indicating that larger, poorer, less-educated, and younger households engage in more cross-border shopping, and that the response is larger if the neighboring country has relatively low grocery price indices. In addition, I provide novel evidence that households commuting towards the border combine their trip to work with shopping abroad. Namely, commuting trips taking a household closer to the border correspond to an expenditure increase, while commuting to a workplace further inland has no effect.

These results have important implications for urban research. First, the uncovered heterogeneities may enhance normative analyses of the optimal spatial supermarket allocation, giving additional weight to households with a lower willingness to travel. Second, my findings might improve policies targeting the negative externalities of cross-border shopping on employment, consumption, sales, and tax collection (see again Leal, López-Laborda, and Rodrigo, 2010, Knight and Schiff, 2012, or Baggs, Fung, and Lapham, 2018). Ultimately, while numerous spatial models in economics incorporate trips to the agents' workplaces and a broad empirical literature uncovers patterns in commuting behavior, household mobility for shopping still needs to be studied more thoroughly. One notable exception is Miyauchi, Nakajima, and Redding (2022), who incorporate commuting and shopping trips jointly in a quantitative spatial model. Yet, as they cannot observe expenditures and focus on modeling the trips, they provide an incomplete picture, missing the intensive margin of spatial shopping. Future work could bridge this gap, incorporating the empirical findings on shopping in this and other papers into theoretical models. This would result in a more encompassing picture of the spatial equilibrium and allow for more credible counterfactual analyses.

Chapter 3

Spatial Frictions in Retail Consumption

joint with Maximilian von Ehrlich and Tobias Seidel

It's the possibility of having a dream come true that makes life interesting.

Paulo Coelho, The Alchemist

...things get broken, and sometimes they get repaired, and in most cases, you realize that no matter what gets damaged, life rearranges itself to compensate for your loss, sometimes wonderfully.

Hanya Yanagihara, A Little Life

Acknowledgments: I thank Gabriel Ahlfeldt, Costas Arkolakis, Pierre-Philippe Combes, Marcel Henkel, Christian Hilber, Gabriel Loumeau, Pierre Magontier, Raphaël Parchet, Martina Pons, Lindsay Relihan, and Frédéric Robert-Nicoud, for their helpful comments and suggestions. I thank Migros, the Swiss Federal Statistical Office, and the Central Compensation Office for providing the data, and acknowledge funding from the Swiss National Science Foundation's NFP 77 "Digital Transformation" (Project 407740 187443) and the Berne University Research Foundation.

3.1 Introduction

Spatial frictions in retail consumption significantly affect the competition between retailers and the residential appeal of different locations, as first emphasized by Hotelling (1929). On the one hand, distance can act as a barrier to competition, while on the other, proximity to other retailers can offer advantages such as trip-chaining and reduced search costs. From the consumer's perspective, models of spatial equilibrium highlight the importance of access to consumption amenities as a factor enhancing the attractiveness of locations (see, for example, Brueckner and Zenou, 1999, Couture and Handbury, 2020, and Handbury, 2021), while the concept of spatial equilibrium suggests that households in locations with fewer amenities require higher real wages as compensation (see, for example, Rosen, 1979, Roback, 1982, Caliendo, Dvorkin, and Parro, 2019, and Ahlfeldt et al., 2022).

To understand the spatial dynamics of non-tradable consumption and the value of consumption access across different areas, reliable estimates of spatial frictions are essential. These estimates are crucial not only for understanding household choices but also for informing policy decisions. For instance, zoning laws differentiate land into commercial and residential zones to manage spillovers and coordinate infrastructure effectively. Understanding these consumption-related spatial frictions is vital for planning and has significant implications for welfare and spatial disparities.

However, identifying the degree of spatial frictions in retail consumption is complex due to the endogenous nature of store locations, which can render traditional gravity model estimates unreliable. Retailers often choose locations close to customer bases that are likely to spend more or have a preference for their products. Moreover, spatial frictions and the value of consumption access may vary significantly across different household groups, potentially obscuring valuable insights about diverse residential location choices. Factors such as varying transport costs, demand elasticity, and expenditure shares of consumption, particularly grocery consumption, can differ between household types.

In this paper, we achieve a causal identification of spatial frictions by exploiting quasi-experimental variation from store openings. We collect information on hundreds of store openings, which we link to individual-level data on detailed consumption spending and sociodemographic characteristics such as income, age, and household size. The expenditure data includes expenditures for food and household products of more than 3 million Swiss households (85% of the population) and 1.5 billion daily transactions collected through the loyalty program of Migros, the largest Swiss retailer, for the period 2019Q1-2021Q2.[1] Households live in 315,000 grid cells, measuring 100 $\times$ 100 meters, and we have coordinate-level precision for stores. Together with hand-collected data on store openings for all major retail chains and administrative individual-level data, we are able to estimate consumption decay functions at a high spatial resolution.

To this end, we apply a staggered difference-in-differences approach to the geo-referenced household-store-linked consumption data. This allows us to isolate expenditure shifts to the new stores from incumbent stores within the same retail chain and from different chains. The *expenditure shift* within the chain is fully caused by distance reductions as, for a given size, stores of the same chain offer the same product variety at the same price. The *competitor shift* reflects variety substitutions as well as distance reductions. Variations in these two types of expenditure shifts enable us to estimate flexible distance gradients of consumption. In the second step, we estimate the shop-specific attraction parameters. These estimates from the store-opening experiments provide insights into the parameters of distance frictions, substitution elasticities, and quality-adjusted prices of stores in a spatial model of consumption activities. Building on the model structure and parameter estimates, we compute local and type-specific measures of consumption access. Finally, we demonstrate that our local measures of consumption access exhibit significant variation across and within cities and explain a substantial portion of regional variation in housing rents, consistent with spatial equilibrium theory.

Our results show that conventional gravity estimates yield biased estimates of distance

[1] Our result are still valid if we exclude the COVID-19 pandemic.

frictions. This is not surprising as residential choice and store locations are highly interdependent. Correcting for the endogenous nature of distances between stores and consumers, we find a distance elasticity of about 15 percent. Non-parametric estimates show that the marginal effect of distance ceases to be significant at around 14 minutes of travel time on average. Using detailed sociodemographic and location data, we document that distance frictions vary across heterogeneous households and locations. Based on the distance gradients and observed expenditures, we can recover shop attraction terms and compute consumption access measures. Consumption access varies significantly across regions and also within urban areas. Comparing the degree of disparities in income to the one in market access, we find that market access displays a much more pronounced variation. Combined with the observed positive correlation between income and market access, this underscores the relevance of consumption access for spatial disparities in real income. We further link the estimated market access measures to local rents. Consistent with spatial equilibrium theory, better market access capitalizes in higher rents. We further learn from event studies that households adjust their spatial consumption pattern quickly after the opening of a new store. For same-chain openings close to a customer's home, expenditures at incumbent stores decline by 30% within the first month and remain persistent after ten months. For the entry of competitors, the effect is about half as big.

This paper contributes to a recent strand of research that examines the role of consumption in space. Previous literature has documented a positive link between store openings and house prices, which suggests that households value consumption access positively (Pope and Pope, 2015; Hausman et al., 2023). Agarwal, Jensen, and Monte (2022) find that household expenditures decay more in distance for goods with lower storability, while Eizenberg, Lach, and Oren-Yiftach (2021) use credit card data at the neighborhood level for Jerusalem to document that residents from areas with a higher average income shop in more distant stores with lower product prices. Marshall and Pires (2018) use household-store-level data to show how customers trade off travel costs with prices and variety, and

Miyauchi, Nakajima, and Redding (2022) build a quantitative spatial model to disentangle consumption access from other local amenities. Hoelzlein and Miller (2024) study openings of Whole Foods Markets and document capitalization effects in house prices as well changes in neighborhood dynamics. Handbury and Weinstein (2015) show that price levels for food products fall with city size. In line with these results, we document that market access for grocery products improves significantly with population density. We relate to these papers by identifying consumption areas conditional on geographical and sociodemographic characteristics, where we employ an identification strategy that allows for quantification of the causal distance gradient.

A second line of research explores spatial consumption at the store level. For example, store entry reduces revenues of incumbent supermarkets (Arcidiacono et al., 2020) and facilitates access to cheaper goods, implying positive welfare effects (Hausman and Leibtag, 2007). Looking at endogenous location decisions, restaurants in Milan cluster close to each other (Leonardi and Moretti, 2023) and Big Box stores in the U.S. tend to locate close to complementary stores (Schuetz, 2015). In contrast to this literature, our analysis is not carried out at the store level. Instead, it focuses on changes in household-level expenditures in response to store openings within a certain distance. This turns out to be relevant, as the impact of store entry on incumbents depends largely on the location relative to the residence of potential consumers rather than on the distance to competitor stores.

Third, and more broadly, we relate to the amenity literature highlighting, among other things, sorting across heterogeneous agents (Diamond, 2016, Ahlfeldt et al., 2022, Almagro and Domınguez-Iino, 2024), access to workplaces (Monte, Redding, and Rossi-Hansberg, 2018), pollution (Heblich, Trew, and Zylberberg, 2021), noise (Ahlfeldt, Nitsch, and Wendland, 2019) or the value of leafy streets (Han et al., 2024).

The structure of the paper is as follows. First, Section 3.2 introduces the data sources and shows summary statistics before we present our conceptual framework on spatial grocery shopping in Section 3.3. Section 3.4 then discusses the empirical identification strategy to causally estimate the model's key parameter, and Section 3.5 examines these empirical findings (followed by robustness checks in Section 3.6). Finally, we bring the model and estimation results together in Section 3.7 and discuss the spatial distribution in market access as well as potential individual-level heterogeneities in spatial consumer behavior. Section 3.8 concludes.

3.2 Data

We combine (i) individual transaction data from the largest Swiss retailer with (ii) administrative data from the Swiss Federal Statistical Office on a high spatial resolution of 100×100 meters. To introduce the data, we refer to individuals in the grocery data as *customers* and those in the administrative data as *residents*. This section introduces the different data sets, explains the matching procedure based on residential location and age, and presents corresponding summary statistics of the final data set.

Transaction Data – The consumption data stems from the loyalty program of the largest Swiss grocery retailer *Migros* (holding a market share of 32.7% in 2020). We observe expenditures on 41 product groups for the universe of 1.5 billion customer-store-linked purchases between 2019Q1 and 2021Q2, and customer characteristics include their residence location, age, and household type. Locations are coded on a grid of 350,000 100×100 meter cells with a mean population of 25 residents.[2] In this program, participants identify themselves at the checkout with their loyalty cards in exchange for exclusive offers and discounts. The program has substantive coverage, tracking expenditures of 2.1 million active users (32% of all Swiss residents above legal age), spending on average at least 50 Swiss francs monthly (USD 56 on July 29, 2024), and capturing 79% of the retailer's total sales. Importantly, the chain charges the same prices throughout the country, independently of local purchasing power, wages, and costs. Stores of similar size also generally offer a similar assortment of goods, except for local products.

[2]The major product groups include, among others, *fruits and vegetables*, *meat and fish*, *milk products and eggs*, and *bakery and convenience*. The household types include the categories *small households*, *young families*, *established families*, *golden agers*, and *pensioners*. To be a family, consumers have to register their children. This registration gives access to additional benefits related to family products.

Administrative Data – We enrich this unique consumption data with administrative records for the entire Swiss population (8.7 million inhabitants in 2020). Pseudo social security numbers allow linking residents across three different administrative data sets. The *Population and Households Statistics* provides socio-demographic characteristics for each resident for the years 2016–2021. This includes, among others, information on gender, age, marital status, residence location, and household identifiers. The residence locations are coded on the same 100×100 meter grid as in the grocery transaction data. The *Old-Age and Survivors Insurance* dataset contains annual gross labor market income for every resident for the years 2016 to 2021.[3] We average annual household income for the years 2016–2021 to reduce biases in permanent income from transitory shocks and adjust, in most cases, average household income by the square root of household size.[4] Finally, the *Structural Survey* gives information on the highest completed education in a household for the years 2010–2021.[5]

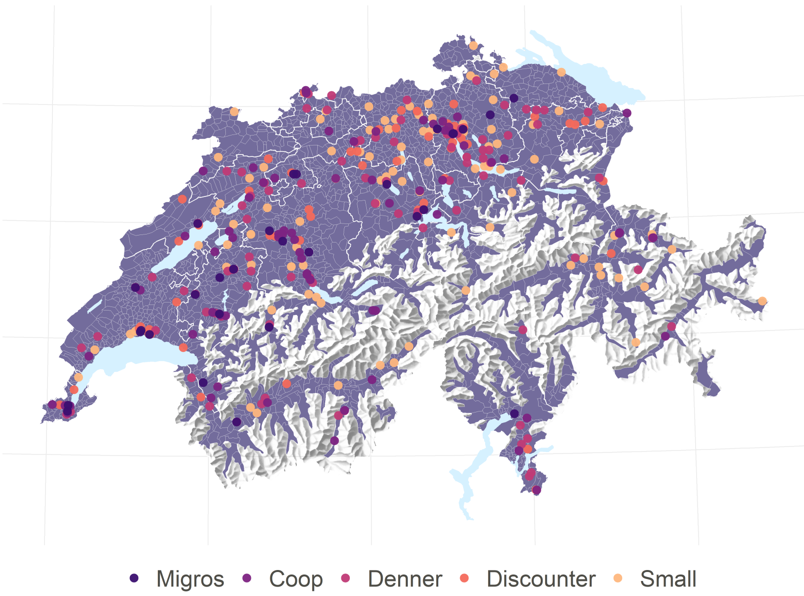
Supermarket Entries – Finally, we collect data on supermarket openings between 2019Q1 and 2021Q2. To this end, we use a web-scraped and geo-coded monthly panel on supermarkets' locations and define a store's emergence in the panel as a potential opening. We observe the true openings for a subset of chains – including our data provider *Migros*, their discounter *Denner*, and one of the competitors – to validate the high accuracy of the scraped data and cross-check the scraped opening dates with newspaper announcements

[3]Contribution to this insurance is mandatory for everyone except for individuals younger than 25 with an annual income below 750 Swiss francs. The contributions amount to a fixed share of the gross labor market income, including official awards, gifts, and bonuses, and are also mandatory for self-employed individuals.

[4]The calculation is $income\_adjusted = \frac{income\_total}{\sqrt{\#household\_members}}$, where we consider all household members, including small children. The adjustment follows one of the equivalence scales suggested by the OECD. We compute *income_total* as the household's annual income by summing the income of all household members but excluding grown-up children who still live with their parents, as they likely do not contribute to the household's budget.

[5]The survey questions a representative sample of 200,000 people above age 15 every year on housing, employment, mobility, and education. Participation is mandatory. Education is categorized as either primary, secondary, or tertiary education. Primary (or compulsory) education ends at the latest after eleven mandatory school years (including kindergarten). Individuals who completed high school or an upper-secondary specialized school have a secondary education. The completion of any degree at a university, university of applied sciences, or university of teacher education results in a tertiary degree. As education stabilizes for most individuals after a certain age, we use educational variables only for individuals above age 25 at the time of the survey.

FIGURE 3.1: Spatial Distribution of Grocery Shop Openings



Notes: The figure shows the spatial distribution of store openings in Switzerland between 2019Q1 and 2021Q2. We show openings for Migros, as well as the main competitor Coop, the discounters Denner, Lidl and Aldi, and smaller chains operating mostly in rural areas.

on *Factiva*, a global database of more than 400 news agencies. Finally, we manually exclude gas stations and stores that are too small to matter in their neighborhood and select 351 entries between 2019Q1 and 2021Q2 as treatments.[6] Figure 3.1 shows the geographical distribution of all 351 openings across Switzerland. Seventy-five stores entered the market in urban areas (corresponding to 21% of entries for 30% of the population), and all administrative regions, except for two, received at least one new supermarket.[7] The correlation between the regional number of entrants and the population is 0.91.

[6]Our analysis will mostly focus on the 31 same-chain openings by Migros. Additionally, we identify 69 openings for the main competitor Coop, 159 for discounters, and 96 for smaller chains that mainly operate in rural areas.

[7]Switzerland consists of 26 federal units called *cantons*. The ones without any opening are Appenzell Innerrhoden and Obwalden.

TABLE 3.1: Transactions Summary Statistics

	Spending		No. of Visits		Road Dist.		Car Travel	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Overall	146	189	10.3	10.9	8.80	13.9	14.2	13.2
By Age Group								
<34	108	137	8.3	8.7	10.6	15.8	15.8	14.6
35–44	156	199	9.8	10.3	9.5	14.2	14.8	13.4
45–54	165	217	10.0	10.7	9.5	14.1	14.9	13.4
55–64	146	194	9.8	10.5	9.6	14.7	15.1	13.9
65-74	134	170	10.3	11.0	8.5	13.9	13.9	13.2
75+	137	160	11.9	12.0	6.6	11.6	11.8	11.5
By Income Quintile								
< 4,530	128	156	10.9	11.5	7.4	12.6	12.6	12.2
4,530–6,717	127	164	9.6	10.4	9.1	14.1	14.4	13.4
6,718–9,288	145	185	9.9	10.6	9.5	14.1	14.8	13.5
9,289–12,855	162	206	10.1	10.7	9.7	14.2	15.1	13.5
12,856+	175	229	9.8	10.4	9.9	15.2	15.5	14.2
By Education								
Primary	125	154	11.1	11.8	7.1	11.8	12.2	11.6
Secondary	146	183	10.5	11.0	8.8	13.3	14.1	12.8
Tertiary	162	212	10.0	10.7	9.5	15.1	15.0	14.0
By Household Size								
1	103	123	9.8	10.1	7.8	13.5	13.1	13.0
2	140	170	10.4	11.0	8.7	13.8	14.0	13.2
3–4	169	217	10.3	11.1	9.4	14.2	14.9	13.4
5+	192	250	10.5	11.3	9.5	14.2	14.9	13.5
Number of Monthly Visits	23,155,515							
Number of Households	780,429							

Notes: The table shows summary statistics for the 23 million monthly shopping trips in the final data: monthly expenditures, number of visits, and travel distances across different groups of household characteristics. *Spending* is measured in Swiss francs, *Road Dist.* in kilometers, and *Car Travel* in minutes.

Sample Construction – We determine the closest supermarket entry for each household in terms of car travel time and concentrate on households who receive a new supermarket within less than 30 minutes.[8] Our analysis focuses on *customers* that we can uniquely match to a *resident* based on the common variables of age and location. Appendix C.A describes the individual steps of the matching procedure. We focus on treated customers

[8]We calculate car travel times in minutes and road distance in meters between stores and customers using the API of *search.ch*, a Swiss mapping service.

TABLE 3.2: Summary Statistics for Households

	Final sample		Population	
	Mean	SD	Mean	SD
Age	61.00	15.19	54.84	17.53
Income Total	94.05	127.23	88.63	119.73
Income Adjusted	57.60	78.39	59.22	77.03
Panel b)	Pct.	N	Pct.	N
Gender		780,429		3,991,230
Female	40.4	315,233	39.6	1,578,660
Male	59.6	465,196	60.4	2,412,570
Marriage		780,429		3,991,230
Married	62.0	483,780	46.8	1,866,832
Not Married	38.0	296,649	53.2	2,124,398
Highest Education		514,297		2,311,993
Primary	10.9	56,036	13.4	308,754
Secondary	46.2	237,460	45.2	1,045,440
Tertiary	42.9	220,801	41.4	957,799
Language Region		779,407		3,987,127
French	23.5	183,343	25.3	1,007,039
German	70.8	551,637	67.9	2,705,434
Italian	5.7	44,427	6.9	274,654
Pop. Density		779,407		3,987,127
Rural	18.0	140,168	17.4	693,093
Suburban	56.3	438,614	51.1	2,038,383
Urban	25.7	200,625	31.5	1,255,651
Household Size		780,429		3,991,230
1	23.2	180,694	37.0	1,475,101
2	38.5	300,839	32.7	1,306,748
3-4	31.2	243,496	24.8	991,735
5+	7.1	55,400	5.5	217,646
Observations		780,429		3,991,230

Notes: The table shows summary statistics for the characteristics of households in the final sample and compares them to the households in the population. *Income Total* is the total sum of annual labor market income in a household, and *Income Adjusted* adjusts this by the square root of household size.

who did not move during the sample period and whose average monthly grocery expenditures lie between CHF 20 and 1,000 per capita in the year before the treatment (between 23 and 1,126 USD on June 18, 2024). This restriction is important because too-small monthly baskets might not accurately capture the overall consumption, while too-large monthly baskets are unlikely to suit personal use but are from business customers. Eventually, we aggregate the remaining transactions into monthly expenditures and visits per household, which yields 23 million observations for 780,000 households.

Summary Statistics – Table 3.1 shows summary statistics for the 23 million shopping trips. First, individuals visit a store ten times a month on average, and the number of visits is relatively stable across income, age groups, education, and household size. Second, the average household spends 150 Swiss francs a month (169 USD on July 29, 2024). Although these expenditures increase monotonically with income, the share of grocery expenditure relative to income declines. This observation suggests the presence of non-homothetic preferences, and we incorporate this in our theoretical framework. Finally, shopping tends to be quite local, with an average road distance of about 9 minutes and only minor variation across household characteristics.

Table 3.2 presents summary statistics for household characteristics. Our matched data includes more than a quarter of Swiss households and is highly representative of the total population. Notably, average income, gender composition, and education levels are very close to the corresponding population values. Households in our final data are slightly larger, live more often in suburban areas, and have older household heads.[9]

3.3 Conceptual framework

To inform our empirical analysis, we use a simple model of store choice: Household ω resides in location i and shops at store $s \in n$.[10] We assume non-homothetic preferences

[9]See further discussion of the data and its representativeness in Kluser and Pons (2024), analyzing the intergenerational persistence of consumption, and Kluser (2024), studying cross-border shopping.

[10]Note that we can restrict the choice-set of stores for each household ω in location i, meaning, n_i can be i-specific and include all stores within, for example, a half-hour car drive, s.t. $Distance_{is} \leq 30$ minutes.

of the form:

$$U_{\omega i} = \left(\frac{G_{\omega i} - \bar{G}}{\alpha}\right)^{\alpha} \left(\frac{h_{\omega i} - \bar{h}}{\beta}\right)^{\beta} \left(\frac{x_{\omega i}}{1 - \alpha - \beta}\right)^{1-\alpha-\beta}, \tag{3.1}$$

where $G_{\omega i} = \left[\int_0^n g_{\omega i}(s)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$ is a composite good of store-specific grocery varieties, and $\bar{G}$ denotes subsistence consumption. $h_{\omega i}$ represents housing consumption with $\bar{h}$ being the subsistence level and $x_{\omega i}$ captures the consumption of all other freely tradable goods. The latter goods specify a numéraire with the price of $x_{\omega i}$ being unity, and we denote housing cost (rents) by r_i. This yields utility-maximizing demands $G_{\omega i} = \bar{G} + \alpha\tilde{w}_{\omega i}/P_{\omega i}$, $h_{\omega i} = \bar{h} + \beta\tilde{w}_{\omega i}/r_i$, and $x_{\omega i} = (1-\alpha-\beta)\tilde{w}_{\omega i}$, where $\tilde{w}_{\omega i} = w_{\omega i} - P_{\omega i}\bar{G} - r_i\bar{h}$. The location- and household-specific price index for groceries is

$$P_{\omega i} = \left[\int_0^n (p(s)\tau_{\omega is})^{1-\sigma}\right]^{\frac{1}{1-\sigma}}, \tag{3.2}$$

where $p(s)$ is the producer price (optimally chosen as a fixed markup over marginal cost) and $\tau_{\omega is}$ are the distance costs, which may vary with the household type and location. With this setting, indirect utility is given by $V_{\omega i} = \frac{w_{\omega i} - P_{\omega i}G_{\omega i} - r_i h_{\omega i}}{P_{\omega i}^{\alpha} r_i^{\beta}}$, the price index $P_{\omega i}$ is decreasing in n and increasing in $p(s)$ and $\tau_{\omega is}$, such that the relative indirect utility between locations is decreasing in the relative price index. Note that even with price indices being not individual-specific (i.e., $P_{\omega i} = P_i \; \forall\omega$), the relative indirect utility within locations between a high-income (w_{1i}) and a low-income (w_{2i}) household is increasing in P_i:

$$\frac{\partial(V_{1i}/V_{2i})}{\partial P_i} = \frac{\bar{G}\,(w_1 - w_2)}{(\tilde{w}_2)^2} > 0.$$

Ceteris paribus, better consumption access is more valuable for low-income households. Solving for the expenditure per variety we obtain

$$p(s)g_{\omega is}\tau_{\omega is} = \frac{(\tau_{\omega is}p(s))^{1-\sigma}}{P_{\omega i}^{1-\sigma}}\left(P_{\omega i}G_{\omega i} + \alpha\tilde{w}_{\omega i}\right). \tag{3.3}$$

In the following, we estimate this expression to identify the role of distance frictions. The distance costs to travel to the store are specified as a function of distance, which we either measure in minutes of travel time or kilometers of road distance: $ln(\tau_{\omega is}) \equiv \kappa_{\omega} ln(Distance_{is})$.

In the empirical analysis, producer prices are captured by a store (or chain) fixed effect, and the remaining components of Equation (3.3) are collected in a location-household-type fixed effect. Accordingly, this yields the following estimation equation:

$$\ln(Y_{\omega ist}) = \gamma_{\omega i} + \gamma_t + \lambda_s - \beta ln(Distance_{is}) + \epsilon_{ist}, \tag{3.4}$$

where $Y_{\omega ist}$ captures either expenditures $p(s)g_{\omega is}\tau_{\omega is}$ or the number of shop visits. The gravity coefficient β is identified from is-specific variation and reflects the product of distance costs (κ) and the elasticity of substitution (σ).
With consistent estimates of β and λ_s as well as the set of all store locations, we can compute the individual and location-specific market access measure:[11]

$$\Phi_{\omega i} = \sum_{s \in n} \left[\exp(\lambda_s) \times Distance_{is}^{-\beta} \right]. \tag{3.5}$$

This measure reflects the utility contributions of access to stores with different weights depending on the income groups and their expenditure shares. According to spatial equilibrium theory, we expect market access to be relevant for residential location choice and accordingly to be capitalized in local housing prices.

Isolating spatial friction parameters from the estimation of Equation (3.4) is complicated by the fact that stores choose their location according to the expected expenditures that they can attract at a certain location. This implies that the $\hat{\beta}$ is likely biased if the above equation is estimated. In the following, we discuss our empirical strategy addressing this endogeneity bias, and we compare the 'conventional estimates' of Equation (3.4) with our approach.

[11]Note that $\Phi_{\omega i} = P_{\omega i}^{1-\sigma}$. Furthermore, β here is constant across all individuals. We relax this assumption in the last section.

3.4 Empirical Strategy

To identify a causal estimate of β, we have to overcome the challenge that stores and residents do not locate randomly. Specifically, customers likely attract stores and vice versa, leading to a simultaneity issue. Also, unobserved determinants of a place's attractiveness for retailers to open a store and for customers to locate lead to an omitted variable bias. The ideal experiment to resolve these issues and isolate a causal effect would randomly allocate stores across space as an exogenous supply-shifter. To get as close as possible to such a supply-shifter, we exploit the quasi-experimental variation in the exact timing of supermarket openings to estimate the response of household consumption patterns.

Particularly, we are interested in two distinct explanatory variables. First, the change of weekly expenditures at the incumbent same-chain stores isolates the expenditure shift from incumbent stores to the new store in response to the opening. Second, the change in total weekly expenditures at all stores of the grocery chain measures the expenditure shift from competitors to the new store induced by the opening.[12] Taken together, these two channels reflect the overall change in consumer behavior.

To estimate the model's key Equation (3.4), we rely on a staggered difference-in-differences design where the treatment interacts with a logarithmic distance function. This model isolates the causal effects of interest if the parallel-trend assumption holds. Yet, as the probability of receiving a treatment likely depends on location characteristics, untreated households may not be a valid comparison. Therefore, we exploit the variation in the exact timing of openings and use not-yet-treated control units, as the retailers' strategic planning cannot explain short-term differences between opening dates. Instead, the exact opening date is due to administrative and bureaucratic delays, and locations treated within a short time span are comparable.

We report our main findings for a conventional TWFE model but also take into account the recent advances in the theoretical difference-in-differences literature, considering the

[12]This effect also includes a general income effect, which may change grocery spending. For groceries, this effect is likely much less relevant than the shift from competitors, so we refer to this part as the competition shift.

potential heterogeneity of treatment effects across periods and cohorts.[13] For our context, we require an estimator that allows the treatment effect to vary with a distance covariate. Furthermore, weekly expenditures have a mass point at zero, and we follow Chen and Roth (2024) by estimating a Quasi-MLE Poisson model in this case to recover the proportional treatment effect. Wooldridge (2022) and Wooldridge (2023) proposes a robust staggered difference-in-differences estimator fulfilling these requirements – allowing for (i) a Poisson model and (ii) interacted covariates – that will complement our conventional estimates and assure their validity.

3.4.1 Staggered DiD: Average Effect of Entry

To start with, we estimate a baseline parametric specification of the following form:[14]

$$Y_{it} = \exp\left(\alpha_i + \gamma_t + \beta(T_{it} \times ln(Dist_i)) + \delta T_{it}\right) \epsilon_{it}, \tag{3.6}$$

where Y_{it} is either the sum of incumbent same-chain expenditures (number of visits) or total chain expenditures (number of visits) for the incumbent and competitor shifts, respectively. The treatment indicator T_{it} equals one if the store assigned to household i as a treatment opened in period $z \leq t$. In particular, we are interested in the distance gradients for both expenditure shifts (*incumbent* and *competitor*), captured by β. We control for unobserved time-invariant household-specific characteristics α_i. These fixed effects capture idiosyncratic characteristics such as workplace location, school location of children, or other routine trips. The period fixed effect γ_t absorbs common time trends and seasonality. While β in Equation (3.6) is constant across all households, we provide household-type specific estimates of the distance gradient by allowing β to vary with income, age, and household size in Section 3.7.3, allowing for heterogeneities in our model.

[13]See, for example, de Chaisemartin and D'Haultfœuille (2020) and Callaway and Sant'Anna (2021). In our context, our results may be biased if an opening affects the same household differently depending on the shock's timing. Examples include the COVID-19 pandemic, openings during holiday seasons, etc. While we can exclude these specific periods from our estimation sample, other less apparent heterogeneities within groups may remain over time. In addition, households may adjust their consumption habits slowly over time, leading to a dynamic build-up in the effects. This would violate the heterogeneity across time.

[14]We run a robustness check where we allow for non-parametric bins of car travel time to assess the validity of the parametric form the model imposes.

Our baseline approach in Equation (3.6) estimates a conventional two-way fixed effects (TWFE) model with a QMLE-Poisson regression. Following Chen and Roth (2024), we report in our tables and figures the proportional treatment effects $\hat{\beta}_{\%} = \exp(\hat{\beta}) - 1$, allowing for a percentage change interpretation of the coefficients, and calculate standard errors of $\hat{\beta}_{\%}$ using the Delta method.

3.4.2 Staggered DiD: A Robust Approach (Wooldridge, 2022)

Alternatively, we use a robust estimator accounting for the recent advances in difference-in-differences models. Wooldridge (2022) suggests the following flexible extension of the TWFE estimator, allowing the coefficient of interest to vary across periods and cohorts:[15]

$$Y_{it} = \exp\left(\alpha_i + \gamma_t + \delta_{g,t}(T_{it} \times g_i \times \gamma_t \times ln(Dist_i)) + \beta_{g,t}(T_{it} \times g_i \times \gamma_t) + \xi_t(\delta_t \times ln(Dist_i))\right)\epsilon_{it}, \tag{3.7}$$

where the time-constant g_i denotes the period household i is treated (meaning, it indicates which cohort or group household i belongs to). Hence, $\beta_{g,t}$ reflects the average treatment effect and $\delta_{g,t}$ the parametric distance coefficient for the respective cohort-period combination. We aggregate the weighted coefficients for $\beta_{g,t}$ and $\delta_{g,t}$ to get an average marginal treatment and interaction effect for all period-cohort pairs:[16]

$$\beta_{ATT} = \sum_{t \times g, t \geq g}^{T,G} W_g \beta_{g,t}, \qquad \delta_{ATT} = \sum_{t \times g, t \geq g}^{T,G} W_g \delta_{g,t}.$$

[15]Wooldridge (2023) discusses the extension to Poisson regressions in more detail.

[16]It is unclear whether the suggestions of Chen and Roth (2024) to recover proportional treatment effects recover an ATT in the staggered intervention case. To make the obtained marginal effects (in Swiss francs) comparable to the proportional treatment effects from Equation (3.6), we relate the estimates to the average expenditures in the data. Standard errors can be obtained by bootstrapping. This approach is computationally very expensive, so we estimate this robust estimator with quarterly data (instead of monthly) to reduce the number of coefficients. The point estimates remain almost unaffected by this change. We use the Wooldridge (2022) approach as a robustness check for the conventional TWFE main results and stick to the conventional method for all other analyses.

3.4.3 Dynamic DiD: An Event-Study Style Approach

Additionally, we report dynamic event-study style estimates in the robustness section. There are two additional benefits of this approach. First, these estimates are informative *per se*, as one might expect a gradual build-up of the effect over time due to an incremental adaptation of consumer behaviors. Second, the dynamic estimation results allow for placebo tests of the parallel-trend assumption. Hence, we want to estimate a coefficient for every pre- and post-treatment period of interest. We write this model in the following form:

$$Y_{it} = \exp\left(\alpha_i + \gamma_t + \sum_{\substack{k=-12 \\ k\neq -1}}^{12} \beta_k T_{i,t}^k\right)\epsilon_{it}, \tag{3.8}$$

where $T_{i,t}^k$ is a set of dummies indicating that at time period t household i got a treatment $k \in [-12, 12]$ months ago. The exclusion of $k = -1$ normalizes the coefficients to the period preceding the treatment, and we stick here again to the standard TWFE estimator.[17]

3.5 Empirical Results

We next present our empirical results. We first investigate the role of distance frictions in conventional gravity-type specifications. Then, we estimate the impact of a store entry on shifts of average expenditures and the number of shopping trips from same-chain incumbent stores as well as from competitor shops. Based on these results, we quantify the geographical size of consumption areas and the distance gradients in consumption by exploiting the distances between households' residences and store entries. Most of our discussion focuses on car travel times as our preferred measure of distance.

[17]We assign the event periods of 12 and -12 to any observation lying outside this window. To consider an alternative approach, we can aggregate the estimated coefficients from model Equation (3.7) in an event-study fashion. However, Wooldridge (2022) does not compute pre-treatment coefficients, and therefore, the visual check for the parallel trend assumption is not possible. Yet, the post-treatment coefficients provide very similar estimates to the TWFE results.

3.5.1 Conventional Gravity Estimates

In the first step, we run conventional gravity regressions to estimate the decline of shopping activity with distance, ignoring the potential endogeneity of β. Since we are interested in the combination of the extensive and intensive margin effects, we estimate the Poisson QMLE model in Equation (3.4) and report proportional effects in Table 3.3 for total expenditures and the number of visits.[18] We study parametric log functions of distance as measured by Euclidean distance, road travel distances in kilometers, and car travel times in minutes. The estimates indicate a significant decline in expenditure by 6.9 to 8.7 percent for a ten percent increase in distance. The number of visits responds very similarly to distance changes. Overall, distance frictions implied by the conventional gravity model are substantial. The linearized coefficients of the Poisson model imply that expenditures fall to zero at a distance of only about 4 to 5 minutes of travel time.

3.5.2 Store-Opening Effects

Distance between stores and households is an endogenous variable that may partly but not fully be addressed by including store and household fixed effects as in the specification in Table 3.3. Therefore, we exploit the store openings as quasi-experimental shocks to identify causal distance frictions. The total expenditures at a new store consist of two parts: (i) the expenditure shift from same-chain incumbent stores (*incumbent shift*) and (ii) the expenditure shift from competitor stores (*competitor shift*).[19] We first analyze both shifts individually and discuss then the implications for the distance parameter β in our model. We estimate the role of distance costs separately for both parts, where we use the expenditures at the incumbent same-chain stores as the dependent variable to measure the incumbent shift and the total chain expenditures as the dependent variable to measure the competitor shift. In both cases, we estimate the difference-in-differences model in Equation (3.6) with a Poisson QMLE model and report the exponentiated proportional

[18] We include zero-values for all stores household ω every visited within the sample period. Additionally, we perform robustness checks including 'irrelevant alternatives', meaning we sample ten shops that each household has never visited. The results remain qualitatively unchanged.

[19] Note that the latter part also includes potential income effects.

TABLE 3.3: Conventional Gravity (Poisson Model)

Proportional Effects	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Euclid. Dist. in km)	-0.683***			-0.665***		
	(0.000)			(0.000)		
ln(Road Dist. in km)		-0.813***			-0.796***	
		(0.000)			(0.000)	
ln(Car Dist. in min)			-0.877***			-0.861***
			(0.000)			(0.000)
Household and Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.54	0.55	0.55	0.48	0.48	0.48
Observations	1.62e+08	1.62e+08	1.62e+08	1.62e+08	1.62e+08	1.62e+08

Notes: The table shows conventional two-way gravity regression results, estimating Equation (3.4). The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. We use the same sample as in the following opening experiments, where we focus on households who live within 30 minutes of a store opening. The panel includes all shops ever visited by a customer, including month-shop observations with zero expenditures/visits.

treatment effects. As robustness checks, we complement our findings with conventional specifications using logarithmic dependent variables instead of Poisson models in Appendix C.C.1 (This approach ignores the mass point at zero in the dependent variables).

Incumbent Expenditure Shift – Table 3.4 reports the results for the incumbent shift, estimating the effect of a store opening on a household's expenditures at the incumbent stores. Note that the *incumbent shift* is the inverse of the estimated coefficients. Table 3.4 shows that the more distant the new store is, the lower the household's response. Accordingly, the more distant the new store, the smaller the incumbent shift. An opening within one minute of car travel time leads to a 22.6% reduction in expenditures at incumbents and reduces the number of visits by 22.4%, leaving, therefore, the average spending per trip unchanged. We assume that the new store is comparable to the incumbents for all characteristics beyond location (and, thus, distance to the household's residence). This

TABLE 3.4: Incumbent Expenditure Shift (Poisson Model)

Proportional Effects	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.073***	-0.098***	-0.226***	-0.068***	-0.091***	-0.224***
	(0.003)	(0.003)	(0.006)	(0.008)	(0.003)	(0.005)
Treat × ln(Euclid. Dist. in km)	0.037***			0.039***		
	(0.002)			(0.003)		
Treat × ln(Road Dist. in km)		0.045***			0.049***	
		(0.002)			(0.002)	
Treat × ln(Car Dist. in min)			0.099***			0.104***
			(0.003)			(0.003)
Marginal Effect at Mean	-0.017***	-0.015***	-0.016***	-0.003	-0.001	-0.003
	(0.002)	(0.002)	(0.002)	(0.006)	(0.002)	(0.002)
Observations	2,599,180	2,601,098	2,601,098	2,599,180	2,601,098	2,601,098
Mean Distance	5.27	7.30	12.64	5.27	7.30	12.64
Mean ln(Distance)	1.66	1.99	2.54	1.66	1.99	2.54
Squared Correlation	0.757	0.757	0.757	0.754	0.754	0.754

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores, estimating Equation (3.6). This captures expenditures shifted from incumbent stores to the new store. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

assumption seems plausible as the retailer charges the same prices throughout the country and offers similar products for a given store size. Additionally, the estimated slope coefficient shows how the household's response declines with distance to the new store. Namely, a doubling of distance corresponds to a ten percentage point lower reallocation of expenditures and store visits.

Competitor Expenditure Shift – Households shift a second part of their grocery expenditures from competing chains to the new store. To isolate the impact of distance on this *competitor shift*, we use the same empirical strategy with the total expenditure of household ω at the supermarket chain as a dependent variable (meaning, not only expenditures at incumbent stores but including the expenditures at the new store). Since we expect the

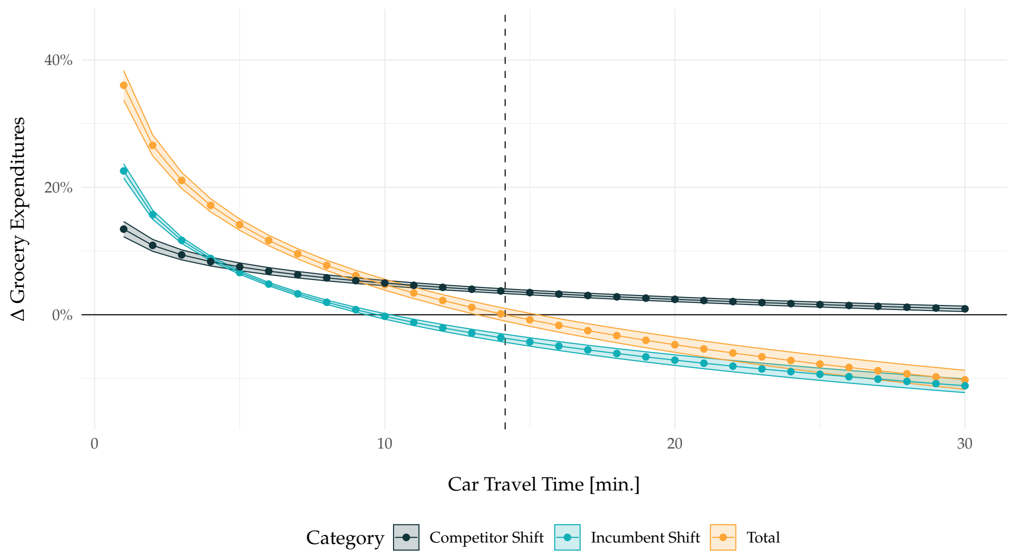
TABLE 3.5: Competitor Expenditure Shift (Poisson Model)

Proportional Effects	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.059***	0.070***	0.134***	0.081***	0.101***	0.195***
	(0.002)	(0.003)	(0.006)	(0.008)	(0.003)	(0.006)
Treat × ln(Euclid. Dist. in km)	-0.016***			-0.024***		
	(0.001)			(0.003)		
Treat × ln(Road Dist. in km)		-0.018***			-0.027***	
		(0.001)			(0.001)	
Treat × ln(Car Dist. in min)			-0.037***			-0.052***
			(0.002)			(0.002)
Marginal Effect at Mean	0.031***	0.031***	0.031***	0.042***	0.041***	0.043***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.002)	(0.002)
Observations	2,599,180	2,601,098	2,601,098	2,599,180	2,601,098	2,601,098
Mean Distance	5.27	7.30	12.64	5.27	7.30	12.64
Mean ln(Distance)	1.66	1.99	2.54	1.66	1.99	2.54
Squared Correlation	0.760	0.760	0.760	0.752	0.752	0.752

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at all same-chain stores, estimating Equation (3.6). This captures expenditures shifted from competitors to the new store. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

supermarket chain to be profit-maximizing, a store opening should, on average, increase total expenditures for the chain. Thus, in contrast to the main effect of the incumbent shift, the main effect of the competitor shift should be positive. Table 3.5 reports the results, following the same structure as for the incumbent expenditure shift. Overall, we find that total expenditures at the chain increase by 13.4 percent if a new store opens in close proximity to a household. For the number of visits, the corresponding effect amounts to about 19.5 percent. Distance costs are significant, as a doubling of distance to the new store reduces the competitor expenditure shift by about 4 percent. Distance frictions are more pronounced for the number of visits, where the elasticity for travel times amounts to about 5 percent. The fact that the distance elasticity is smaller for the

FIGURE 3.2: Distance Gradients



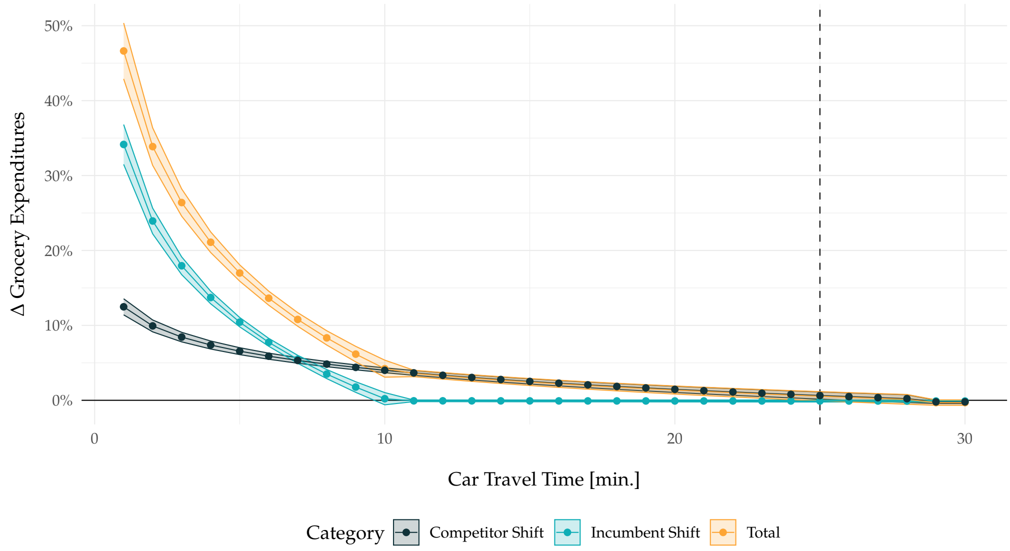
Notes: The figure shows distance gradient functions, indicating how the treatment effects from the difference-in-difference analyses decline over time. The *incumbent shift* is based on the results in Table 3.4 (mirrored along the horizontal axis) and the *competitor shift* in Table 3.5. We calculate standard errors for the individual fitted points using the delta method. The *total expenditure shift* is the sum of the two curves, and the corresponding confidence bands are the aggregate of the two other bands. The vertical dashed line indicates the insignificance of the total shift.

competitor shift than for the incumbent shift reflects imperfect substitutability between the product ranges of the different chains.

Total expenditure shift – With the expenditure shifts from the same chain incumbents and competitor shops at hand, we can compute the total distance gradient. For a new shop opening at a one-minute distance to the customer, Table 3.4 and Table 3.5 imply that 22 percent of the previous same-chain incumbent expenditures are shifted to the new store and 15.1 percent of the total same-chain expenditures are shifted from competitor stores. As distance increases, expenditures shifted to the new store via both channels decline according to the corresponding slope coefficients.

Figure 3.2 depicts graphically the marginal effects of both parametrically estimated distance gradients as well as the total distance gradient. We observe a steep decline for the total gradient, which, however, is much less pronounced than for the conventional gravity specifications. The marginal effect of distance starts to become insignificant at a

FIGURE 3.3: Log-Kink Distance Gradients



Notes: The figure shows distance gradient functions, indicating how the treatment effects from the difference-in-difference analyses decline over time. Compared to the baseline results in Figure 3.2, we add a kink in the parametric functions once the *incumbent shift* and the *competitor shift* turn insignificant. The *incumbent shift* is based on the results in Table 3.4 (mirrored along the horizontal axis and adding the kink) and the *competitor shift* on Table 3.5 (with the additional kink). We calculate standard errors for the individual fitted points using the delta method. The *total expenditure shift* is the sum of the two curves, and the corresponding confidence bands are the aggregate of the two other bands. The vertical dashed line indicates the insignificance of the total shift.

distance of 14 minutes of car travel time. We interpret this point as the maximum spatial scope of average consumption areas as a household's consumption behavior is not significantly impacted by an entry of a shop at a distance beyond that.

Robust Estimator (Wooldridge, 2022) – Accounting for the recent advances in the theoretical difference-in-differences literature on staggered interventions, we apply the novel approaches suggested by Wooldridge (2022) and Wooldridge (2023), which seem to be best-suited for our case. Figure C.C3 shows the estimation results for the distance decay functions analogously to the main results in Figure 3.2. The point estimates are slightly higher compared to the conventional estimator but support qualitatively all our previous statements. As the competitor effect seems more persistent, the average customer will respond to an opening for up to 24 minutes of car travel time.

Non-parametric estimation and distance gradients with a kink – While the log-specification of the distance decay derived in our model section follows the standard approach for gravity models, Figure 3.2 displays a function that might potentially be mis-specified as the coefficients become negative instead of converging to zero. One explanation might be that the log specification fits well for short-distance shopping trips but may be an inappropriate measure for longer-distance traveling. Then, the log-specification of distance may put a too rigid structure on the distance frictions. Such non-linearities are, for example, documented by Hillberry and Hummels (2008) for shipments of manufacturing firms in the United States.

Therefore, we assess the suitability of our baseline specification and display in Figure C.C4 and Figure C.C5 a non-parametric function where we estimate a coefficient for travel time bins with a width of two minutes. The logarithmic specification and non-parametric alternative are very similar for the competitor shift. Yet, although the logarithmic specification captures the non-parametric alternative initially very well for the incumbent shift at short distances, the curves diverge for longer-distance trips, consistent with our hypothesis (while the non-parametric specification converges to zero as expected).

Hence, we want to ensure the reliability of our estimates. To do this, we estimate a log function including a kink, meaning, we allow the slope to change after the baseline function crosses the zero line. Figure C.C4 and Figure C.C5 show that this functional form follows closely the non-parametric estimation, therefore likely capturing the true decay more accurately.[20] Yet, choosing the kink in this way is an arbitrary choice by the authors, and we show in Table C.C5 and Table C.C6 that the estimated slopes of the left part of the function only change slightly for different values of the kink within a reasonable distance range.

Figure 3.3 then shows the incumbent and competitor shift as well as the total distance gradient for our preferred kink-cutoffs, complementing the baseline results in Figure 3.2

[20]Non-linear gravity specifications are widely used in the trade literature (see, for example, Eaton and Kortum, 2002, Henderson and Millimet, 2008, or Hillberry and Hummels, 2008).

and providing qualitatively identical results. Nonetheless, the added flexibility allows for an initially stronger household response and a steeper decline of the decay function.

3.6 Robustness and Sensitivity

In this section, we address three concerns about the validity of our empirical results and discuss potential violations of the parallel trend assumption that would invalidate our identification.

Dynamic Responses to Store Openings – Figure C.D6 shows the results for the dynamic difference-in-differences model outlined in Equation (2.1) for the store-opening experiment on household expenditures. The pre-treatment coefficients show no apparent pre-trend and violation of the parallel trend assumption. In a distance of less than two minutes, households reduce their expenditures by more than 40%, while the response quickly decreases with distance below 20% and 10% for distances bins of 2-5 minutes and 5-10 minutes, respectively. Furthermore, the reaction is immediate, and we do not observe a dynamic build-up of the effect.

COVID-19 Pandemic – Additionally, we might worry that the COVID-19 pandemic led to a fundamental shift in grocery shopping behavior that our empirical strategy cannot capture. We address these concerns by restricting our analysis to a sub-sample that only includes observations before the start of the pandemic in Switzerland (namely, for the period January 2019 to February 2020). Table C.D7 and Table C.D8 show the corresponding estimation results. We find qualitatively identical results across all distance bins that further ensure the credibility of our findings.

Multiple-Treated Units – We focus in Section 3.5 on a binary treatment and ignore multiple ones, and we might be concerned that additional openings bias our coefficients. To analyze this potential bias, we focus on a sub-sample of individuals who were only once treated during the observed sample period. Table C.D9 and Table C.D10 show these estimation results, which are again qualitatively identical to our previous findings.

3.7 Consumption Areas

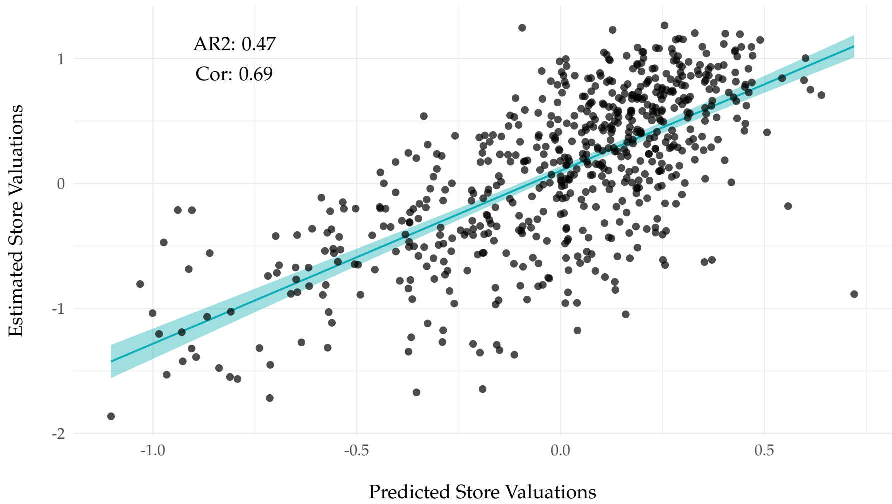
Next, we bring together our causal estimates of the distance decay functions with our model of spatial shopping. This section first discusses how we recover the fixed effects of unobserved stores to calculate a utility-based value for access to grocery shopping across different household locations. Then, we present our resulting market access measures and relate them to spatial equilibrium theory.

3.7.1 Recovering Store Valuations

Observed supermarkets (Constrained gravity regressions) – First, we recover the average store valuations λ_s for all the stores we observe according to Equation (3.4), where we use the causal estimates of the distance gradient. Since the kinks in the distance costs turn out significant and relevant, we use our preferred kink specifications shown in Figure 3.3 to calculate the market access measures. Specifically, we denote by $\overline{dist}$ the critical distance level beyond which an opening does not yield any significant expenditure shifts. Up to the distance level $dist_k$ at kink k, we identify a steeper gradient β_1, whereas, for higher distance levels, the gradient flattens to β_2. Accordingly, the distance costs are $dist^{\beta_1}$ for $dist <= dist_k$ and $(dist^{\beta_2}/dist_k^{\beta_2})dist_k^{\beta_1}$ for $dist_k < dist <= \overline{dist}$. Following Table C.C5 and Table C.C6, we set $\beta_1 = 0.2$, $\beta_2 = 0.05$, $dist_k = 14\ min$, and $\overline{dist} = 30\ min$. Hence, all stores beyond a distance of 30 minutes receive no expenditures and are not considered for recovering the store valuations.

Unobserved Supermarkets (Lasso Regressions) – As we have only data for one chain, we need to impute the store valuations of other chains by implementing a second-step regression inferring the store valuations of other supermarkets. If a set of observed characteristics sufficiently determines the valuation of the observed stores, then we can use these variables to infer the value of unobserved competing stores. Characteristics that might be useful include, first, the store size and quality that we approximate with review

FIGURE 3.4: Estimated vs. Predicted Store Fixed Effects



Notes: The figure plots the estimated store valuations for our main retailer against the predicted store valuations from the Lasso regression as a cross-check of Lasso's predictive power. The store valuations are recovered as the fixed effects λ_s from Equation (3.4) using the estimated distance gradients.

counts and average ratings from Google Maps. Second, the store's value likely also depends on the surrounding neighborhood in terms of the neighborhood's residents, the local labor market, and local amenities, including other potentially complementary stores. Hence, we infer the residents' characteristics using our administrative data sets to calculate local averages of income, household size, education, age, etc., as well as counts of sector-specific employees and firms around the supermarkets. Regarding local amenities, we use additional administrative data measuring the walking distance in meters to the closest store, pharmacy, bank, restaurant, etc.

To learn which of these 716 observed variables determine store valuations without overfitting the model, we apply a Lasso variable selection approach. Specifically, we use ten-fold cross-validation and feed the model with the above determinants that explain the variation in store valuation. Then, we let the algorithm choose the best predictors for store valuations of the retail chain.[21] Figure 3.4 depicts the store valuations estimated from Equation (3.4) for the stores we observe against the recovered store valuations from

[21] Alternatively, we use a Ridge regression and different versions of elastic nets. Our results remain qualitatively unchanged.

the Lasso regression for the same stores. We observe a high correlation of 0.69 between the recovered and the predicted values, and the model explains about 50 percent of the variation.

3.7.2 Spatial Consumption Access

Therefore, we can now derive a measure for market access. In order to compute the local values of consumption access, we need to combine the locations of stores, their valuations in terms of quality-adjusted prices, and the estimated distance frictions. We compute market access according to Equation (3.5) at a granular level of 100 $\times$ 100 meter grid cells across the whole country. Figure 3.5 shows the spatial distribution of market access for the city of Zurich and the entire country (note that we focus only on inhabited grid cells that are within the construction areas). We observe substantial differences in market access, with a mean of 40 and a standard deviation of 46. The interquartile range for the total country ranges from 21 to 66. Across the largest 10 Swiss cities, we observe a range between 55 and 98.[22] Even within the city of Zurich, the differences are pronounced with an interquartile range between 113 and 160. Hence, we observe substantial variations in market access both within cities and across the country.

Furthermore, Figure C.D7 maps, in addition to our measure of grocery market access, graphically the spatial distribution of income and population and the distance to the closest supermarket in the city of Zurich. Looking at correlations between these variables, we see that access correlates positively with population density with a value of 0.22 and negatively with income and the distance to the closest supermarket (-0.22 and -0.11, respectively). This means that grocery market access within the city is higher in denser and lower-earning neighborhoods with faster access to supermarkets.

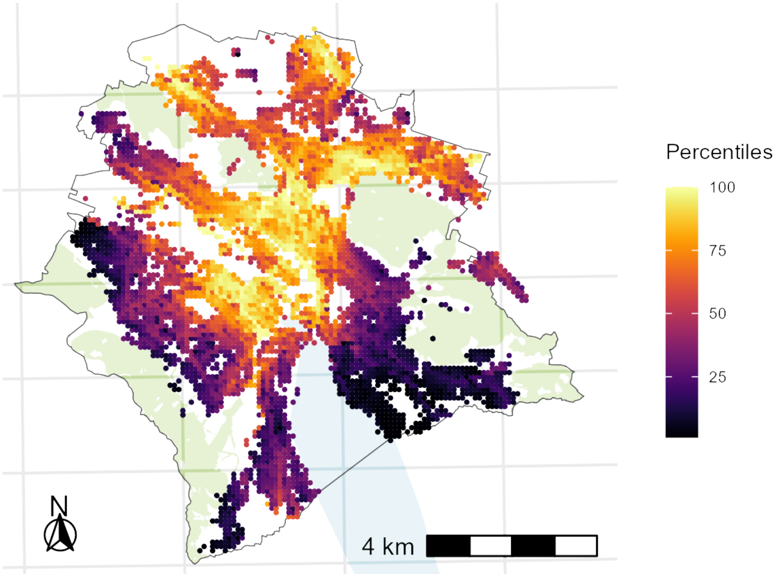
Population Density and Rents

In order to relate our access measures to the spatial equilibrium, we depict in Figure 3.6 market access against the percentiles of population density as well as against hedonic

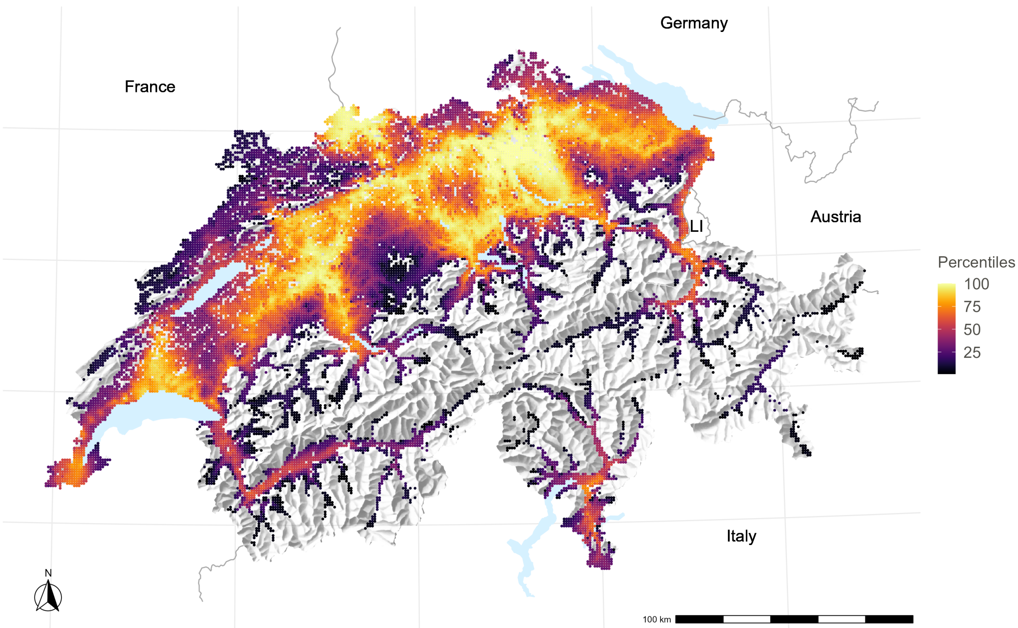
[22] Zurich, Geneva, Basel, Lausanne, Bern, Winterthur, Lucerne, St. Gallen, Lugano, Biel/Bienne.

FIGURE 3.5: Spatial Distribution of Grid-Level Retail Market Access

(A) Percentiles of Market Access in the City of Zurich

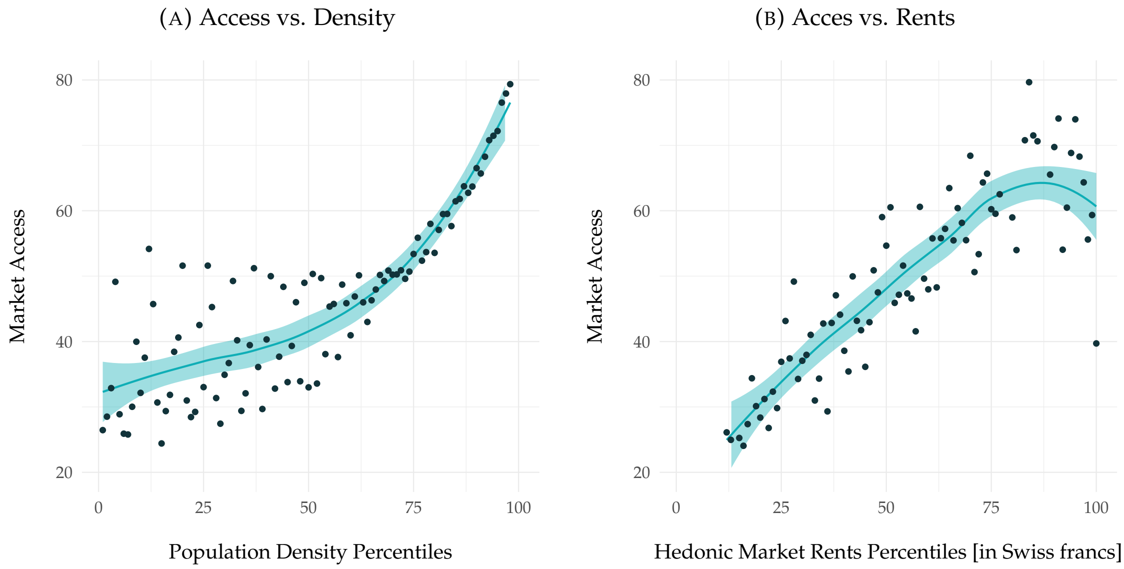


(B) Percentiles of Market Access in Switzerland



Notes: The figure plots for each populated 100 × 100 meter grid cell in Switzerland our utility-based valuation of market access as in Equation (3.5). We consider all stores of major grocery retailers in Switzerland, recover the unobserved store valuations with a Lasso approach, and use our causal estimates of travel costs. Travel distances between households and stores are measured as car travel times in minutes. Figure 3.5a zooms into the market access for the city of Zurich, while Figure 3.5b shows the market access for the entire country (aggregated to 1,000 × 1,000 meter cells).

FIGURE 3.6: Relating to the Spatial Equilibrium



Notes: The figure plots the average cell-level market access (as in Equation 3.5) against the percentiles of the cell-level population count in Figure 3.6a and against hedonic market rents

market-rate rents. The figure shows a strong correlation in both cases (0.83 and 0.90, respectively). Notably, the correlation with population density is very strong among the densest areas, while we observe way more variation for less dense areas. Similarly, neighborhoods with higher rents typically have better grocery market access, with the weakest link observed in the most expensive areas. Note that the strong link between rents and market access also holds when population density is conditioned out (the slope coefficient declines from 0.040 to 0.037). These observations suggest that grocery market access represents an important amenity that capitalizes into local housing rents.

Market Access Across Different Groups and Locations

As evident from the summary statistics in Table 3.1, the households' expenditure shares vary with income levels such that higher market access might, *ceteris paribus*, be more valuable for lower-income groups and for larger household sizes. Therefore, we next analyze patterns in market access based on household-level characteristics, and Table 3.6 shows the variation in market access across major household variables, including income,

TABLE 3.6: Grocery Market Access Across Household Characteristics

Dependent Variable	Market Access			
	Total	Urban	Suburban	Rural
	(1)	(2)	(3)	(4)
Household Income				
Q1	63.427***	81.123***	52.391***	28.022***
	(0.049)	(0.085)	(0.063)	(0.083)
Q2	65.175***	83.789***	56.516***	31.998***
	(0.058)	(0.085)	(0.063)	(0.083)
Q3	64.981***	84.72***	58.056***	33.462***
	(0.055)	(0.085)	(0.063)	(0.083)
Q4	64.545***	84.305***	58.982***	35.79***
	(0.053)	(0.085)	(0.063)	(0.083)
Q5	68.017***	81.972***	60.836***	38.513***
	(0.05)	(0.085)	(0.063)	(0.083)
Age				
<34	69.228***	85.208***	57.693***	32.517***
	(0.050)	(0.078)	(0.064)	(0.089)
35-44	67.724***	84.226***	58.545***	34.178***
	(0.046)	(0.075)	(0.058)	(0.079)
45-54	63.545***	81.751***	57.13***	33.968***
	(0.044)	(0.075)	(0.054)	(0.071)
55-64	61.542***	81.653***	56.356***	33.328***
	(0.043)	(0.075)	(0.053)	(0.068)
65-74	60.554***	81.561***	56.053***	31.688***
	(0.05)	(0.089)	(0.061)	(0.078)
75+	61.219***	81.272***	55.565***	30.576***
	(0.047)	(0.079)	(0.058)	(0.078)
Household Size				
1	65.364***	83.23***	55.646***	30.123***
	(0.031)	(0.049)	(0.040)	(0.054)
2	63.672***	83.244***	57.808***	33.779***
	(0.033)	(0.058)	(0.041)	(0.053)
3-4	62.76***	81.297***	57.283***	34.487***
	(0.038)	(0.066)	(0.047)	(0.061)
5+	60.586***	80.549***	57.13***	34.205***
	(0.081)	(0.146)	(0.100)	(0.122)
n	3,989,077	869,932	1,711,803	570,697

Notes: The table shows the dispersion of our market access measure in Equation (3.5) across all 4 million Swiss households and grouped by the degree of urbanization. Across the three panels, we regress in three independent regressions market access on household income quintiles, the age of the household's head, and household size. Income quintiles are recalculated for each urbanization group.

age, and household size. We report estimation results for the entire country as well as for urban, suburban, and rural regions separately. First, we observe considerable differences across income quintiles, where higher-income households typically benefit from better market access. This is especially the case in rural and suburban areas. Second, we document that market access is especially advantageous for young households in urban areas, with only minor patterns for other regions. Third, smaller households across the country live, on average, in places with better market access, a pattern that reverses in rural municipalities. Note that, so far, this dispersion in market access is calculated with distance decay functions and costs that are identical across different household characteristics, and we will discuss this assumption in the following subsection.

Spatial Variation in Market Access vs. Income

Finally, we relate the spatial variation in market access to spatial income disparities and observe that market access displays much higher variation. Across the entire country, the ratio of the 75th to the 25th percentile is 3.05 for local market access and 1.83 for local household income. Similarly, the coefficients of variation (meaning the ratio of the standard deviation to the mean) are 1.43 for market access and 1.06 for household income. In urban areas, the coefficient of variation in market access is even twice as high as that for income. Given the positive correlation between income and market access, this underscores the importance of price variation in measuring real income disparities across different regions.

3.7.3 Type-Specific Spatial Frictions

So far, we have estimated and used a distance decay parameter β that is constant across locations and different household characteristics. However, if distance costs vary across an important dimension that is spatially segregated, this might imply significant welfare costs for the disadvantaged group that we would miss so far. Therefore, we re-estimate the incumbent and competitor shift as well as the total distance gradient for different

TABLE 3.7: Heterogeneous Distance Costs

	Incumbent Shift		Competitor Shift		Total Shift				
Group	Intercept	Slope	Intercept	Slope	Intercept	Slope	Mean Dist	Cons. Area	n
Household Income									
<4,530	-0.234***	0.107***	0.101***	-0.028***	0.335***	-0.135***	12.3 min.	11.9 min.	1,053,897
	(0.009)	(0.005)	(0.009)	(0.003)	(0.013)	(0.006)			
4,530-6,717	-0.176***	0.071***	0.146***	-0.043***	0.323***	-0.114***	12.7 min.	17.0 min.	266,912
	(0.018)	(0.009)	(0.020)	(0.007)	(0.027)	(0.011)			
6,717-9,288	-0.242***	0.104***	0.128***	-0.034***	0.369***	-0.137***	13.1 min.	14.8 min.	357,721
	(0.015)	(0.008)	(0.019)	(0.006)	(0.024)	(0.010)			
9,289-12,855	-0.230***	0.097***	0.162***	-0.046***	0.392***	-0.143***	13.1 min.	15.5 min.	438,179
	(0.014)	(0.008)	(0.014)	(0.005)	(0.020)	(0.009)			
12,856+	-0.205***	0.088***	0.191***	-0.055***	0.397***	-0.143***	12.6 min.	16.0 min.	484,389
	(0.013)	(0.007)	(0.014)	(0.004)	(0.019)	(0.008)			
Age									
<34	-0.184***	0.070***	0.156***	-0.055***	0.340***	-0.125***	12.1 min.	15.2 min.	84,625
	(0.033)	(0.017)	(0.027)	(0.011)	(0.042)	(0.021)			
35-44	-0.209***	0.083***	0.166***	-0.057***	0.375***	-0.140***	12.6 min.	14.6 min.	321,972
	(0.017)	(0.009)	(0.015)	(0.006)	(0.023)	(0.011)			
45-55	-0.241***	0.106***	0.141***	-0.042***	0.382***	-0.148***	12.8 min.	13.2 min.	542,698
	(0.012)	(0.007)	(0.012)	(0.005)	(0.017)	(0.008)			
55-64	-0.195***	0.082***	0.150***	-0.045***	0.345***	-0.128***	12.9 min.	14.8 min.	585,039
	(0.013)	(0.007)	(0.011)	(0.004)	(0.017)	(0.008)			
65-74	-0.224***	0.100***	0.113***	-0.032***	0.336***	-0.132***	12.7 min.	12.7 min.	494,905
	(0.012)	(0.007)	(0.012)	(0.005)	(0.017)	(0.008)			
75+	-0.243***	0.115***	0.088***	-0.023***	0.332***	-0.138***	12.3 min.	11.0 min.	571,859
	(0.012)	(0.007)	(0.012)	(0.004)	(0.017)	(0.008)			
Household Size									
1	-0.210***	0.096***	0.110***	-0.034***	0.320***	-0.130***	11.9 min.	11.7 min.	610,431
	(0.012)	(0.007)	(0.012)	(0.005)	(0.016)	(0.008)			
2	-0.234***	0.105 ***	0.107***	-0.030***	0.341***	-0.135***	12.8 min.	12.5 min.	970,711
	(0.009)	(0.005)	(0.009)	(0.003)	(0.013)	(0.006)			
3-4	-0.218***	0.092***	0.148***	-0.047*	0.366***	-0.139***	12.8 min.	13.9 min.	833,765
	(0.010)	(0.005)	(0.009)	(0.004)	(0.013)	(0.006)			
5+	-0.157***	0.061***	0.159***	-0.047***	0.316***	-0.108***	13.2 min.	18.7 min.	43,500
	(0.050)	(0.025)	(0.034)	(0.013)	(0.060)	(0.028)			
Pop. Density									
Rural	-0.503***	0.269***	0.410***	-0.100***	0.913***	-0.369***	16.3 min.	11.8 min.	658,270
	(0.014)	(0.013)	(0.023)	(0.005)	(0.020)	(0.018)			
Suburban	-0.236***	0.102***	0.132***	-0.039***	0.368***	-0.141***	12.9 min.	13.6 min.	1,214,021
	(0.008)	(0.004)	(0.009)	(0.003)	(0.011)	(0.006)			
Urban	-0.249***	0.134***	0.069***	-0.021*	0.318***	-0.155***	9.0 min.	7.8 min.	728,807
	(0.011)	(0.007)	(0.010)	(0.004)	(0.015)	(0.010)			

Notes: The table shows for different characteristics heterogeneous difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores and all stores, estimating in both cases Equation (3.6). This captures the *incumbent shift* and *competitor shift* respectively. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

groups of household characteristics – income, age, household size, and population density – allowing β to vary across the groups.

Table 3.7 shows the estimation results. We display the intercepts and slopes of the incumbent and competitor effects for each group, respectively, as well as the total expenditure shift. Across the household-level socioeconomic and sociodemographic characteristics, we find rather little variation in the two coefficients of interest. Yet, looking at the implied consumption area measured in minutes – reflecting the distance beyond which expenditures are predicted to decline to zero – interesting patterns emerge. With regard to income, there is no clear pattern, as the smallest consumption area is predicted for the lowest income quintile and the largest consumption area for the second quintile. This suggests that spatial frictions are rather homogeneous across different income groups. Yet, looking at age, we observe more pronounced patterns as consumer areas decline with age, and accordingly, spatial frictions increase with age. Similarly, consumption areas grow more extensive as household size increases, especially for households with at least five members. Lastly, we observe substantial differences in spatial frictions across different locations. The distance elasticity in rural areas is more than twice as large as the distance elasticity in urban areas. Accordingly, we observe a significantly smaller consumption area in urban than in rural areas. This seems logical, as newly constructed supermarkets in areas with potentially insufficient grocery supply likely attract customers from further away. In principle, variable coefficients for any combination of attributes are possible, but the computational burden increases substantially. Additionally, Table C.D12 and Table C.D13 in the Appendix consider the influence of sociodemographic attributes for urban and rural locations separately.

3.8 Conclusions

This paper provides causal estimates of distance costs in grocery shopping. We exploit the quasi-random variation induced by openings of new supermarkets with a unique and large representative data set of households' transaction records from the largest retailer in Switzerland (1.5 billion transactions of more than 2 million households). Our empirical results show parametric distance elasticities of expenditure of roughly 0.15, while conventional gravity regressions suggest an elasticity of 0.85. Therefore, conventional estimates are largely biased upwards because stores locate endogenously close to households with high potential sales. Including our causal estimates into a simple conceptual framework of spatial grocery shopping, we show that grocery market access strongly varies in space – between regions as well as within cities. Our measure of grocery market access is consistent with predictions from standard spatial equilibrium theory, and better access correlates with higher population density and housing rents. Analyzing potential heterogeneities in the distance decay parameter, we find evidence for differences between socioeconomic and sociodemographic groups and particularly strong differences between rural and urban areas as consumption areas appear much larger in rural areas with worse store access.

Appendix A

The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits

A.A Data: Matching Procedure

This section describes how we match the *customers* in the grocery transaction data with the *residents* in the administrative data. To begin with, we select all combinations of residents and customers with the same location grid cells and age. This generates 4.5 million matches between customers and residents, and we refer to them as *pairs*.[1] We take some additional steps to isolate the unique matches between *residents* and *customers*, proceeding as follows.

1. First, we want to exclude pairs where the customer's shopping behavior does not fit the resident's past locations of residence, as these residents are likely not the owners of the loyalty card they link to. So, we calculate the median annual road distance traveled between a resident's home location and the stores visited by the customer (weighted by trip expenditures). Then, we exclude customer-resident

[1] Note that some customers do not match any resident, which is most likely because their addresses in the grocery data are outdated. This is the case for 380,000 of the 2.8 million customers (13.5%), of which 260,000 are active customers (spending more than 50 Swiss francs monthly over our sample period).

pairs with median shopping trips exceeding 20 kilometers in any year. This step excludes 191,000 pairs.

2. Customers can register in the loyalty program as a family if they have at least one child younger than 25. Hence, we delete all pairs where the customer is registered as a family and the resident does not fulfill this criterion. This excludes 355,000 pairs.

3. Then, we select all customers that link to exactly one household (multiple residents can live in this household). This gives 1,585,204 unique customer-resident matches.

4. Although households can own multiple loyalty cards, the minimum age to register is 18. Hence, we exclude pairs with more customers than adult residents, eliminating 77,935 pairs.

5. We recover some additional unique matches by identifying consumers who have moved recently without notifying the retailer. To this end, we check whether these movers uniquely match a resident at their old location. This procedure identifies 47,571 additional unique pairs.

6. Removing the customers and residents matched in the previous step, we find an additional 3,845 unique matches at current locations. Steps (1) to (6) result in 1.55 million customers uniquely linked to a resident, accounting for 73% of active customers and 21% of Swiss adult residents.

7. For households owning multiple loyalty cards, we then aggregate expenditures within the household before calculating the relative fruit and vegetable share over the sample period at the household level.

8. Additionally, some children moved out recently. In this case, we exclude their expenditures in the periods they still lived with their parents when aggregating the expenditures over time, as these children may contaminate our measure of diet for their parents in the periods before they moved out.[2]

[2]Excluding them entirely leaves our estimates unchanged.

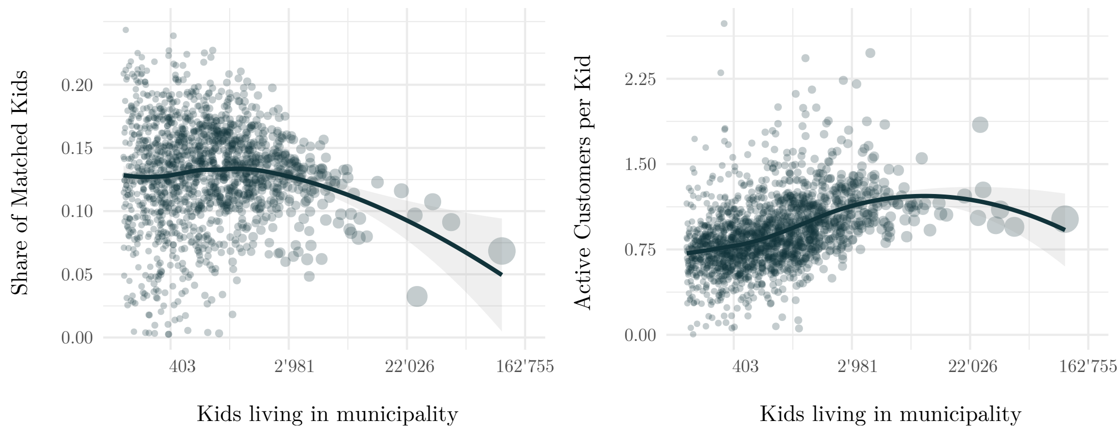
9. We assign the aggregated transaction data to all adult residents in the household. This provides grocery expenditures for 2,248,059 million residents living in 1.17 million different households.

10. Finally, we select the 337,950 children for whom we observe at least one of their parents in the final data set.

A.B Data: Additional Figures and Tables

FIGURE A.B1: Match Rate

(A) Children in the Final Data (B) Children in the Population



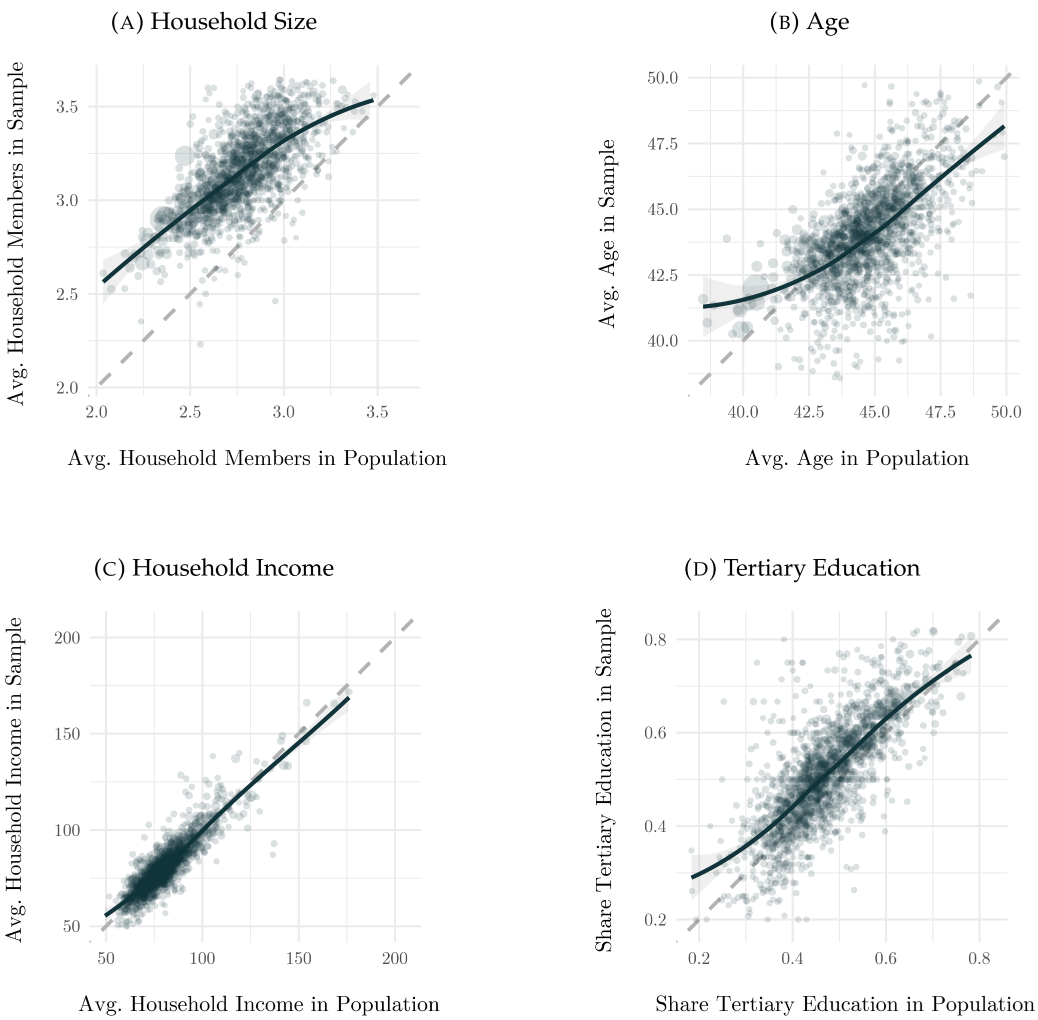
Notes: The figure illustrates the representativeness of the retailer's loyalty program. Figure A.B1a shows the share of matched children as a function of the number of children living in a municipality. Figure A.B1b shows the number of active customers in the full customer data as a function of the number of children living in this municipality. Each dot represents a municipality, while the size is proportional to its population. The solid line is estimated by a local regression.

TABLE A.B1: Summary Statistics for Children

	Final Sample		Population	
Panel a)	Mean	SD	Mean	SD
Age	43.72	10.69	43.70	11.70
Age Father	71.87	9.66	71.05	10.35
Age Mother	71.03	10.35	70.85	11.36
HH Income Total	142.37	137.04	129.68	109.09
HH Income Adjusted	83.25	87.01	81.60	64.88
Panel b)	Pct.	N	Pct.	N
Gender		270,957		2,276,806
Female	53.9	146,148	50.8	1,155,646
Male	46.1	124,809	49.2	1,121,160
Marriage		270,957		2,276,806
Married	62.3	168,776	50.3	1,145,736
Not Married	37.7	102,181	49.7	1,131,070
Highest Education		193,711		1,554,739
Tertiary	53.5	103,676	50.0	777,901
Secondary	42.7	82,763	44.6	694,110
Primary	3.8	7,272	5.3	82,728
Language Region		270,711		2,274,341
German	76.9	208,283	72.3	1,644,202
French	19.1	51,643	22.0	500,133
Italian	4.0	10,785	5.7	130,006
Pop. Density		270,711		2,274,341
Rural	21.7	58,732	21.6	490,681
Suburban	58.2	157,660	52.2	1,186,301
Urban	20.1	54,319	26.3	597,359
Household Size		270,957		2,276,806
1	10.2	27,715	21.0	478,435
2	26.9	72,900	33.2	754,928
3-4	51.1	138,377	37.2	846,201
5+	11.8	31,965	8.7	197,242
Observations		270,957		2,276,806

Notes: This table shows summary statistics for the households in the final data. *HH Income Total* is a household's average gross labor market income 2016-2020 in 1,000 CHF, and *HH Income Adjusted* adjusts it by the square root of household size. *Highest Education* is the highest education completed by anyone within the household, and *Pop. Density* is the municipality's population density.

FIGURE A.B2: Municipality Averages: Sample vs. Population



Notes: The figure illustrates the representativeness of the final data by comparing municipality averages using the final data and the administrative data. Each dot represents a municipality's average, while the dot's size is proportional to the municipality's population. The solid line is estimated using a local regression. The dashed line is the 45-degree line. *Household Size* is the count of members living in an average household, *Age* is the average age of all children in this municipality, *Household Income* is the average household labor market income, and *Tertiary Education* is the average share of households with at least one member having a tertiary degree.

TABLE A.B2: Comparison of Mobility Measures

	(a) Rank-Rank Reg.		(b) IGE			(c) CER		(d) Transition Prob.		
	Intercept	Slope	25	50	75	25	75	Q1Q1	Q1Q5	Q5Q5
Diet	36.1	0.270	0.293	0.265	0.232	46.5	54.3	32.2	10.9	32.0
	(0.28)	(0.005)	(0.008)	(0.006)	(0.005)	(1.80)	(1.91)	(0.50)	(0.37)	(0.50)
Income	43.6	0.131	0.115	0.123	0.131	46.7	52.7	24.7	14.1	28.0
	(0.3)	(0.006)	(0.007)	(0.008)	(0.009)	(2.14)	(1.91)	(0.50)	(0.42)	(0.54)

Notes: The diet results are estimated using 32,168 observations. The income results are estimated using 29,098 observations as we restrict the sample to children between 32 and 38 and fathers between 50 and 62. The IGE uses the log of the father's income as an explanatory variable and the log of the children's income as a dependent variable. Standard errors are from 1,000 bootstrap replications.

A.C Model: Derivations

The Bellman equation $V_t(h_t)$ of the dynamic programming optimization problem takes the following form:

$$\begin{aligned} V_t(h_t) &= \max_{c_t} -(c_t - c^*)^2 - \rho\,(c_t - h_t)^2 + \beta V_{t+1}(h_{t+1}) \quad \text{ s.t. } h_{t+1} = h_t + \phi(c_t - h_t) \\ &= \max_{c_t} -\left(\frac{h_{t+1}}{\phi} - \frac{h_t}{\phi} + h_t - c^*\right)^2 - \rho\left(\frac{h_{t+1}}{\phi} - \frac{h_t}{\phi} + h_t - h_t\right)^2 + \beta V_{t+1}(h_{t+1}) \end{aligned} \tag{A.1}$$

with the resulting optimality conditions:

$$0 = -\frac{2}{\phi}(c_t - c^*) - \frac{2\rho}{\phi}(c_t - h_t) + \beta V'_{t+1}(h_{t+1}), \tag{A.2}$$

$$V'_t(h_t) = -\frac{2(\phi - 1)}{\phi}(c_t - c^*) - \frac{-2\rho}{\phi}(c_t - h_t). \tag{A.3}$$

Shifting the second FOC one period ahead and combining it with Equation (A.2) gives the following Euler equation:

$$(c_t - c^*) + \rho(c_t - h_t) = \beta(1 - \phi)(c_{t+1} - c^*) + \beta\rho(c_{t+1} - h_{t+1}). \tag{A.4}$$

Based on our setting with a quadratic utility function and a linear constraint, we can use a guess-and-verify approach. We guess that the policy function for $c_t(h_t)$ is a weighted average of the optimal healthy diet c^* and the current habit stock h_t ($w \in [0, 1]$):

$$c_t(h_t) = wc^* + (1 - w)h_t. \tag{A.5}$$

Inserting the guess into the Euler equation yields

$$[wc^* + (1 - w)h_t](1 + \rho + \beta\rho\phi) =$$
$$c^*[1 - \beta(1 - \phi)] + h_t\,[\rho - \beta\rho(1 - \phi)] + [c^*(w + \phi w - \phi w^2) + h_t(1 - w - \phi w + \phi w^2)](\beta(1 - \phi) + \beta\rho).$$

The method of undetermined coefficients provides the following two quadratic equations:

$$\begin{aligned}
0 &= \phi\beta(1-\phi)w^2 + \phi\beta\rho w^2 + (1+\rho+\beta\rho\phi - \beta(1-\phi) - \beta\rho - \phi\beta(1-\phi) - \phi\beta\rho)w \\
&\quad - 1 + \beta(1-\phi) \qquad \text{(A.6)} \\
0 &= \phi\beta(1-\phi)w^2 + \phi\beta\rho w^2 + (1+\rho+\beta\rho\phi - \beta(1-\phi) - \beta\rho - \phi\beta(1-\phi) - \phi\beta\rho)w \\
&\quad + \rho - \beta\rho(1-\phi) + \beta(1-\phi) + \beta\rho - 1 - \rho - \beta\rho\phi, \qquad \text{(A.7)}
\end{aligned}$$

which both simplify to:

$$0 = (\phi\beta(1-\phi) + \phi\beta\rho)w^2 + (1+\rho-\beta-\beta\rho+\beta\phi^2)w - 1 + \beta(1-\phi). \tag{A.8}$$

Solving this equation, we find that for any calibration, there is a single root satisfying the requirement $w \in [0,1]$:

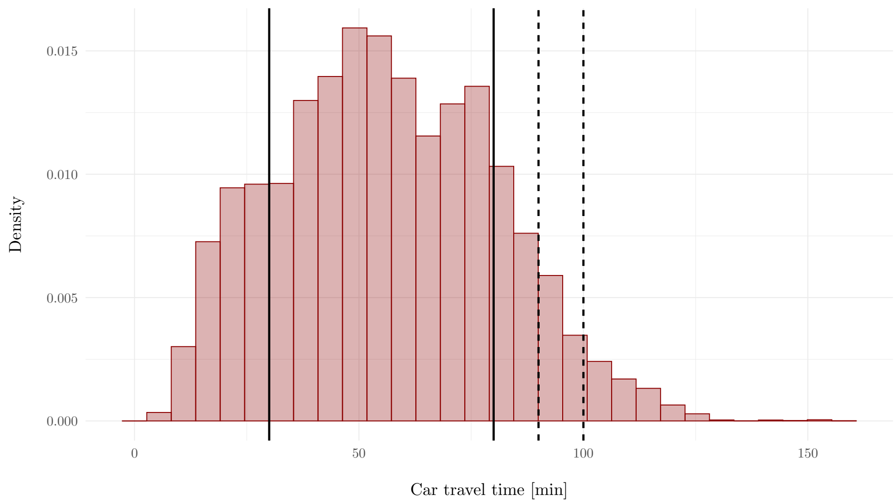
$$w = \frac{-\phi^2\beta + (1+\rho)(\beta-1) + \sqrt{-4\phi\beta(-1+\beta-\phi\beta)(1-\phi+\rho) + (-\phi^2\beta + (1+\rho)(\beta-1))^2}}{2\phi\beta(1-\phi+\rho)}. \tag{A.9}$$

Under this value of w, the Euler equation and the resource constraint hold, justifying our initial guess.

Appendix B

Cross-Border Shopping: Evidence from Household Transaction Records

FIGURE B.1: Distribution of Travel Times



Notes: The figure shows the distribution of car travel times from a household's home to the closest cross-border shopping location. The subsamples of control units used in the different robustness checks of the dynamic results are marked by vertical dashed lines.

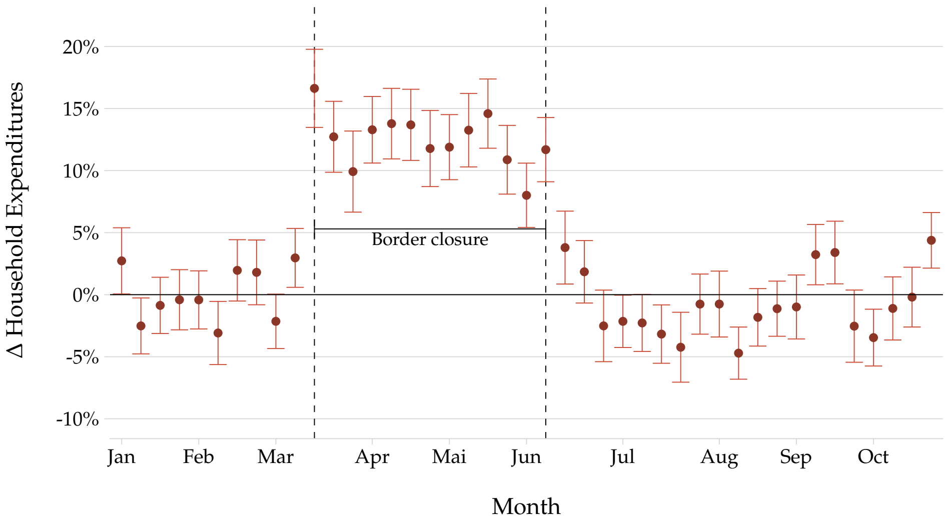
TABLE B.1: Cross-Border Locations

	Location	Country	Population	Number of Stores			Rank		
				Google Reviews			*Google Reviews*		
				-	100	500	-	100	500
1	Annecy	FR	131,766	79	29	11	1	1	3
2	Como	IT	84,808	76	21	14	2	4	1
3	Konstanz	GER	84,446	71	29	14	3	1	1
4	Singen	GER	48,033	50	18	10	4	5	4
5	Annemasse	FR	36,582	49	13	5	5	13	15
6	Aosta	IT	34,052	47	7	3	6	30	34
7	Livigno	IT	6,363	47	14	5	6	12	15
8	Varese	IT	80,588	46	15	7	8	8	8
9	Friedrichshafen	GER	61,561	45	23	10	9	3	4
10	Sondrio	IT	21,457	40	3	1	10	67	67
11	Cantù	IT	40,031	39	12	6	11	16	10
12	Belfort	FR	45,458	37	15	4	12	8	22
13	Lindau	GER	25,547	36	15	9	13	8	6
14	Domodossola	IT	17,930	35	11	4	14	18	22
15	Lörrach	GER	49,295	33	15	7	15	8	8
16	Weil am Rhein	GER	30,009	31	18	9	16	5	6
17	Saronno	IT	39,332	30	9	6	17	24	10
18	Waldshut-Tiengen	GER	24,067	30	13	6	17	13	10
19	Stockach	GER	17,118	29	11	5	19	18	15
20	Radolfzell	GER	31,582	28	7	4	20	30	22
21	Überlingen	GER	22,684	27	13	4	21	13	22
22	Rheinfelden	GER	32,919	26	16	5	22	7	15
23	Bad Säckingen	GER	17,510	25	11	4	23	18	22
24	Bregenz	AT	29,806	25	12	5	23	16	15
25	Montbéliard	FR	25,806	25	10	3	23	22	34
			...						
Overall									
117			1,980,614	1,787	691	304			

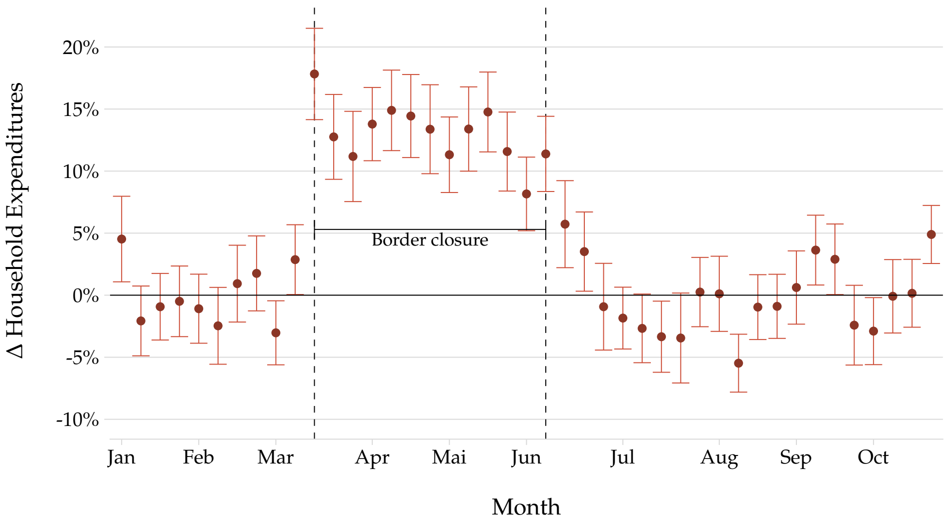
Notes: The table shows the 25 largest cross-border locations for grocery shopping. *Number of Stores* counts the municipality's stores for a given minimum of Google reviews, while *Rank* ranks the locations according to the number of stores. All store locations are scraped from Google Maps.

FIGURE B.2: Robustness of the Dynamic Treatment Effects: Different Control Distance

(A) Control Group: More than 90 min. Distance



(B) Control Group: More than 100 min. Distance



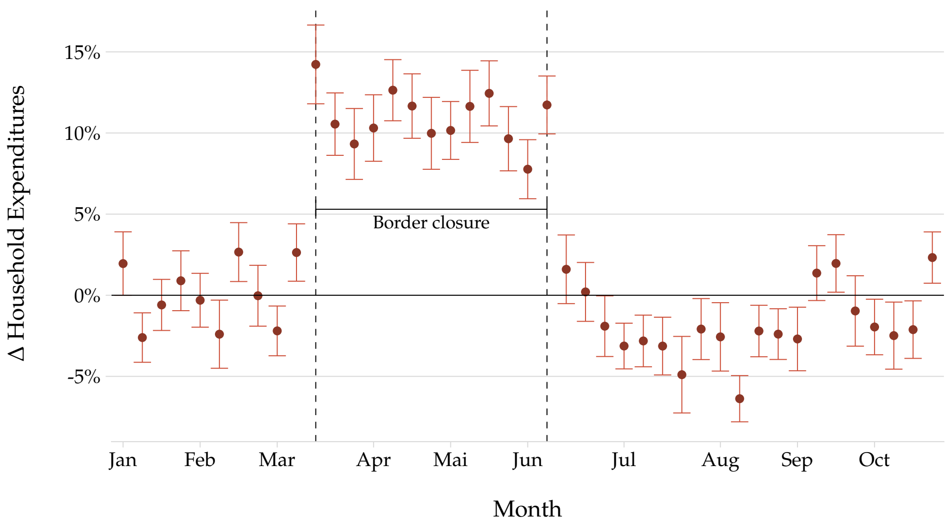
Notes: Figure B.2a shows the border closure's effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 90 minutes. I indicate the period of border closure by vertical dashed lines. The regression estimates Equation (2.1) and uses 8.8 million observations. Figure B.2b also estimates Equation (2.1) for a distance of 100 minutes using 7.1 million observations. Coefficients are normalized to the pre-treatment periods' average, and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

TABLE B.2: Average Treatment Effects (With a 100 min. Control Group)

Dep. Var.: Household Expenditures	
Treat × Border Closed	0.126***
	(0.008)
Treat × Border Open	-0.008
	(0.005)
n	7,051,422

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 100 minutes. The regression follows Equation (2.1) but groups the periods during and after the border closure together (*border closed* and *border open*, respectively). Standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

FIGURE B.3: Robustness of the Dynamic Treatment Effects: the Full Grocery Transaction Data



Notes: The figure shows the border closure's effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes. I indicate the period of border closure by vertical dashed lines. The regression estimates Equation (2.1) and uses all the 28.1 million observations in the full grocery transaction data. Coefficients are normalized to the pre-treatment periods' average, and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

TABLE B.3: Treatment Effects by Socioeconomic Subgroups (With a 100 min. Control Group)

Dep. Variable: Household Expenditures							
a) Household Size		b) Age		b) Income		b) Education	
Group	Coeff	Group	Coeff	Group	Coeff	Group	Coeff
1	0.095***	20–34	0.152***	Q1	0.155***	Primary	0.134***
	(0.011)		(0.016)		(0.010)		(0.016)
2	0.117***	35–44	0.164***	Q2	0.145***	Secondary	0.111***
	(0.008)		(0.012)		(0.011)		(0.009)
3-4	0.152***	45–54	0.153***	Q3	0.133***	Tertiary	0.130***
	(0.010)		(0.011)		(0.011)		(0.011)
≥5	0.162***	55–64	0.140***	Q4	0.132***		
	(0.014)		(0.011)		(0.011)		
		65–74	0.147***	Q5	0.132***		
			(0.011)		(0.014)		
		75+	0.131***				
			(0.013)				
p-value	0.000	p-value	0.220	p-value	0.199	p-value	0.062
n	3,771,701	n	3,770,827	n	2,979,910	n	2,509,512

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 100 minutes, separately for different household characteristics. These characteristics include the *household size*, *age* of the registered cardholder, household *income* adjusted by the square root of household size, and the highest *education* in the household. The regression estimates Equation (2.2), standard errors are clustered at the zip code level, and the reported p-values test the equality of all coefficients. Coefficients are exponentiated such that they equal proportional effects.

TABLE B.4: Treatment Effects by Cultural and Spatial Subgroups (With a 100 min. Control Group)

Dep. Variable: Household Expenditures			
a) Nationality		b) Country	
Group	Coeff	Group	Coeff
African	0.169**	AT	0.097**
	(0.059)		(0.034)
Asian	0.174***	GER	0.129***
	(0.044)		(0.010)
European	0.168***	FR	0.131***
	(0.017)		(0.015)
N.American	0.159*	IT	0.412***
	(0.083)		(0.042)
S.American	0.132*		
	(0.065)		
Swiss	0.124***		
	(0.008)		
p-value	0.071	p-value	0.000
n	3,771,425	n	3,573,599

Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 100 minutes, separately for different household characteristics. These characteristics include the cardholders' *nationality* and the *country* of their closest cross-border shopping location. The regression estimates Equation (2.2), standard errors are clustered at the zip code level, and the reported p-values test the equality of all coefficients. Coefficients are exponentiated such that they equal proportional effects.

TABLE B.5: Cultural Differences: Effect at Language Border (With a 100 min. Control Group)

	Dep. Var: HH Expenditures		
Dist. to ntl. Border	German	French	p-value
Treat × 30-45 min.	0.111***	0.014	0.000
	(0.015)	(0.017)	
Treat × 45-55 min.	0.064***	0.034	0.184
	(0.018)	(0.019)	
Treat × 55-65 min.	0.049***	0.053***	0.812
	(0.014)	(0.014)	
n	695,593		

Notes: The table shows the border closure's average treatment effect on household expenditures for households living within 10 kilometers of the German-French language border. I compare these treated units to same-language households living further away than 100 minutes from the closest cross-border location. The regression estimates Equation (2.2) and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

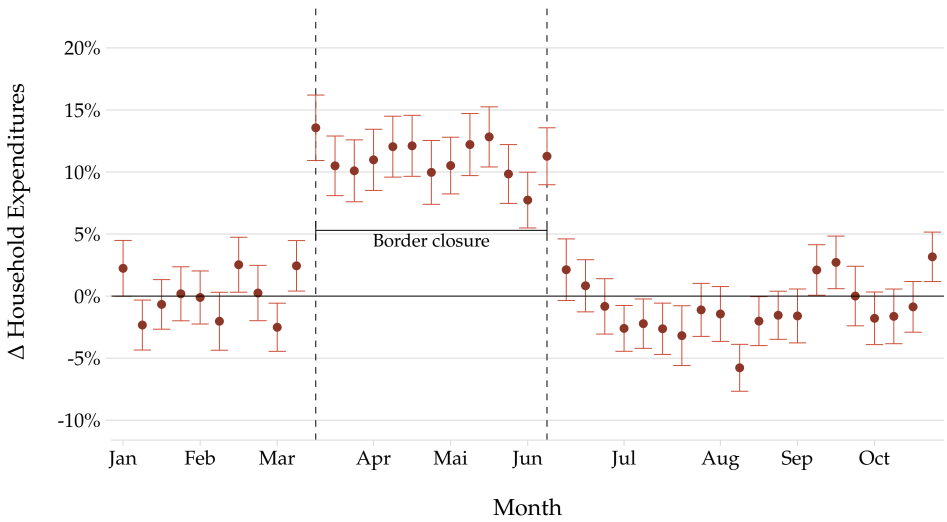
TABLE B.6: Treatment Effect for Different Commuting Behaviors (With a 100 min. Control Group)

	Dep. Var: Household Expenditures		
Δ Border Access	Commute Towards Border	Commute Away f. Border	p-value
Treat × 5-15 min.	0.157**	0.099***	0.459
	(0.020)	(0.020)	
Treat × 15-25 min.	0.158***	0.118***	0.008
	(0.052)	(0.027)	
n	174,180		

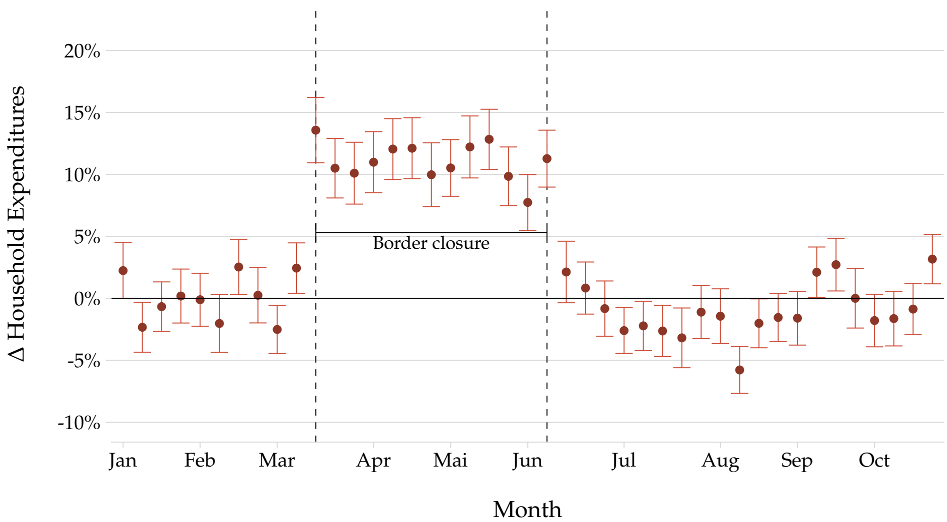
Notes: The table shows the border closure's average treatment effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 100 minutes for different household commuting trips. These trips include commutes by car for 5-15 minutes and 15-25 minutes, either towards the national border (bringing the commuter closer to a cross-border location) or further away from the border in comparison to the household's home. The regression estimates Equation (2.2) and standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

FIGURE B.4: Robustness of the Dynamic Treatment Effects: Different Definitions of Cross-Border Locations

(A) At Least One Store With More Than 500 Google Reviews



(B) At Least Three Stores With More Than 500 Google Reviews



Notes: Figure B.4a shows the border closure's effect on household expenditures within a 30-minute car ride from a cross-border location compared to households living further away than 80 minutes. I consider all cross-border locations with at least one store with more than 500 Google reviews. In comparison, Figure B.4b shows the same results but considers locations with at least three stores with more than 500 Google reviews. Both regressions estimate Equation (2.1) and use 12 million observations. Standard errors are clustered at the zip code level. Coefficients are exponentiated such that they equal proportional effects.

Appendix C

Spatial Frictions in Retail Consumption

C.A Data: Matching Procedure

This section describes how we match the *customers* in the grocery transaction data with the *residents* in the administrative data. To begin with, we select all combinations of residents and customers with the same location grid cells and age. This generates 4.5 million matches between customers and residents, and we refer to them as *pairs*.[1] We take some additional steps to isolate the unique matches between *residents* and *customers*, proceeding as follows.

1. First, we want to exclude pairs where the customer's shopping behavior does not fit the resident's past locations of residence, as these residents are likely not the owners of the loyalty card they link to. So, we calculate the median annual road distance traveled between a resident's home location and the stores visited by the customer (weighted by trip expenditures). Then, we exclude customer-resident pairs with median shopping trips exceeding 20 kilometers in any year. This step excludes 191,000 pairs.

[1] Note that some customers do not match any resident, which is most likely because their addresses in the grocery data are outdated. This is the case for 380,000 of the 2.8 million customers (13.5%), of which 260,000 are active customers (spending more than 50 Swiss francs monthly over our sample period).

2. Customers can register in the loyalty program as a family if they have at least one child younger than 25. Hence, we delete all pairs where the customer is registered as a family, and the resident does not fulfill this criterion. This excludes 355,000 pairs.

3. Then, we select all customers that link to exactly one household (multiple residents can live in this household). This gives 1,585,204 unique customer-resident matches.

4. Although households can own multiple loyalty cards, the minimum age to register is 18. Hence, we exclude pairs with more customers than adult residents, eliminating 77,935 pairs.

5. We recover some additional unique matches by identifying consumers who have moved recently without notifying the retailer. To this end, we check whether these movers uniquely match a resident at their old location. This procedure identifies 47,571 additional unique pairs.

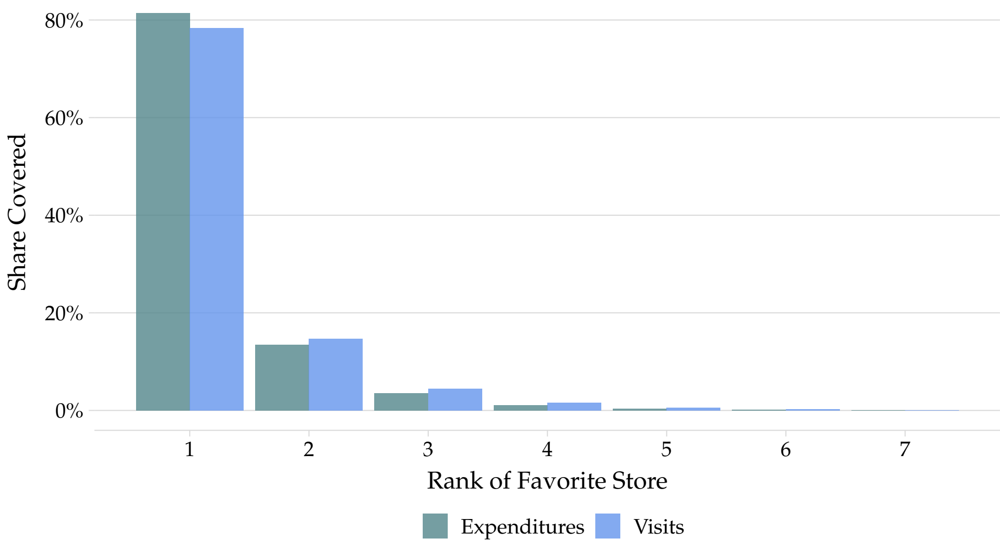
6. Removing the customers and residents matched in the previous step, we find an additional 3,845 unique matches at current locations. Steps (1) to (6) result in 1.55 million customers uniquely linked to a resident, accounting for 73% of active customers and 21% of Swiss adult residents.

7. For households owning multiple loyalty cards, we then aggregate expenditures within the household.

8. We assign the aggregated transaction data to all adult residents in the household. This provides grocery expenditures for 2,248,059 million residents living in 1.17 million different households.

C.B Summary Statistics

FIGURE C.B1: Ranking of Favorite Stores



Notes: The figure shows the share of households' visits and expenditures at their ranked favorites. The figure aggregates all 1.5 billion transactions between 2019Q1 and 2021Q2 that have non-negative amounts and are at supermarkets within 30 minutes by car from the household's residence.

C.C Additional Results

C.C.1 Staggered DiD: OLS With log(expenditures) Instead of Poisson

In the paper, we report the treatment effect for store openings using a QMLE-Poisson model – see Equation (3.6) – to take into account the mass point at zero in the dependent variable. Here, we ignore this mass point and estimate a more standard TWFE model with OLS, where we take the logarithm of the dependent variables. Therefore, we ignore all weekly expenditures with a value of zero:

$$log(Y_{it}) = \alpha_i + \gamma_t + \beta(T_{it} \times ln(Dist_i)) + \delta T_{it} + \epsilon_{it}, \tag{C.1}$$

where Y_{it} will again be (i) the expenditures at incumbent stores, capturing the incumbent shift, and (ii) total expenditures at any same-chain store, capturing the competitor shift. Therefore, we focus on the intensive margin of the store opening intervention, and zero-valued observations drop out in this case.

As a naive alternative, trying to incorporate the zero-valued observations, we report additional estimation results by adding the value 1 to expenditures Y_{it} in Equation (C.1). In this way, zero-valued observations do not drop out, and we use the entire balanced panel. However, note that Chen and Roth (2024) show that the resulting coefficients cannot be interpreted as a proportional treatment effect (meaning, a proportional change in percentage points).

TABLE C.C1: Incumbent Expenditure Shift – Intensive Margin (Log Model)

	log(Expenditures)			log(No. of Visits)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.098***	-0.129***	-0.307***	-0.077***	-0.105***	-0.264***
	(0.003)	(0.004)	(0.008)	(0.008)	(0.003)	(0.007)
Treat × ln(Euclid. Dist. in km)	0.043***			0.041***		
	(0.002)			(0.004)		
Treat × ln(Road Dist. in km)		0.053***			0.049***	
		(0.002)			(0.002)	
Treat × ln(Car Dist. in min)			0.111***			0.100***
			(0.003)			(0.003)
Observations	2,333,704	2,335,343	2,335,343	2,333,704	2,335,343	2,335,343
Squared Correlation	0.664	0.664	0.665	0.659	0.659	0.660

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores, estimating Equation (C.1). This captures expenditures shifted from incumbent stores to the new store. The dependent variable is log(expenditures). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

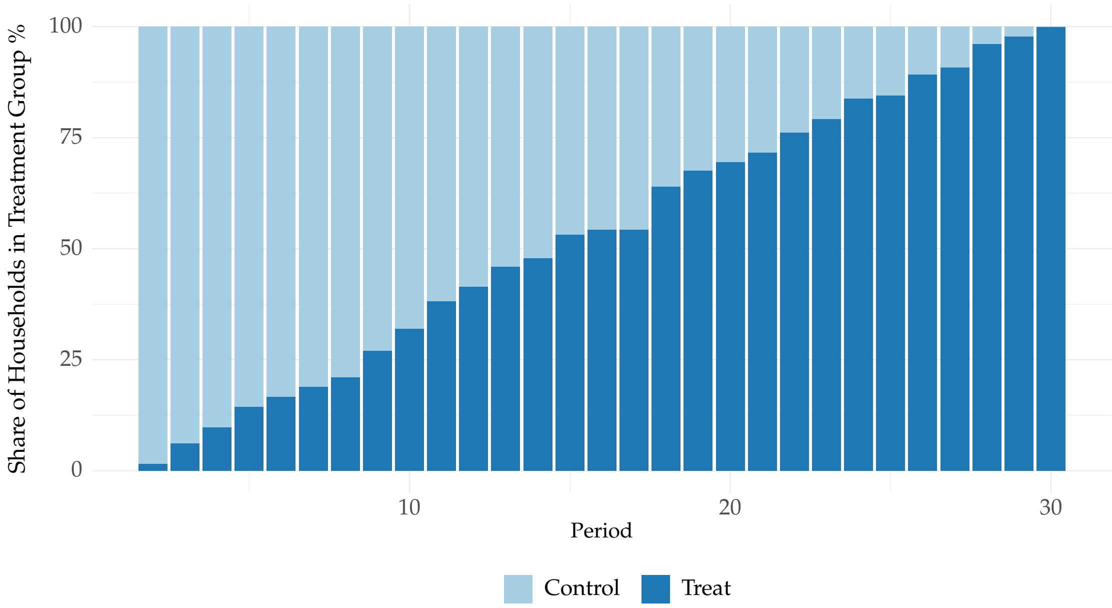
TABLE C.C2: Incumbent Expenditure Shift (Log+1 Model)

	log(Expenditures + 1)			log(No. of Visits + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.211***	-0.269***	-0.602***	-0.123***	-0.162***	-0.376***
	(0.006)	(0.007)	(0.016)	(0.016)	(0.004)	(0.009)
Treat × ln(Euclid. Dist. in km)	0.080***			0.056***		
	(0.003)			(0.006)		
Treat × ln(Road Dist. in km)		0.098***			0.068***	
		(0.004)			(0.002)	
Treat × ln(Car Dist. in min)			0.208***			0.136***
			(0.006)			(0.004)
Observations	2,599,180	2,601,098	2,601,098	2,599,180	2,601,098	2,601,098
Squared Correlation	0.572	0.572	0.573	0.653	0.653	0.654

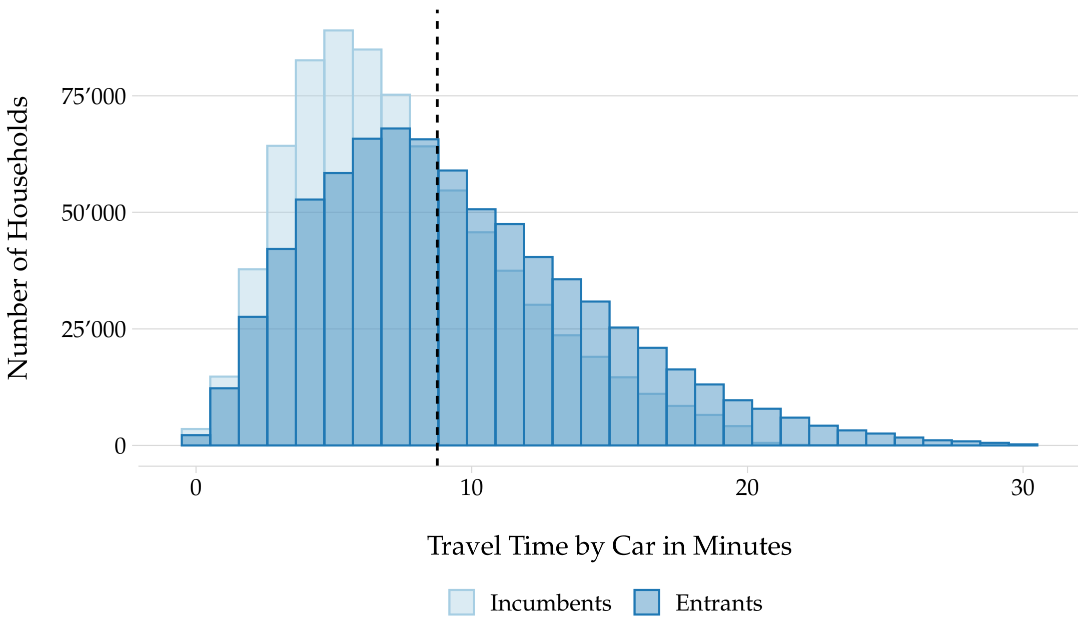
Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores, estimating Equation (C.1). This captures expenditures shifted from incumbent stores to the new store. The dependent variable is log(expenditures+1), where we add the value 1 to each household's expenditures. We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

FIGURE C.B2: Descriptives of the Treatments

(A) Composition of Treatment and Control Groups



(B) Distance to Closest Treatment



Notes: Figure C.B2a shows how the composition of the treatment and control groups changes over time as more and more households switch to the treatment group. Figure C.B2b presents the distribution of the car travel time to the closest opening against the average distance a household travels to spend 1 CHF. The vertical line shows the mean distance for the entrants.

TABLE C.C3: Competitor Expenditure Shift – Intensive Margin (Log Model)

	log(Expenditures)			log(No. of Visits)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.058***	0.070***	0.137***	0.087***	0.103***	0.193***
	(0.003)	(0.003)	(0.006)	(0.008)	(0.003)	(0.006)
Treat × ln(Euclid. Dist. in km)	-0.020***			-0.027***		
	(0.002)			(0.004)		
Treat × ln(Road Dist. in km)		-0.022***			-0.031***	
		(0.002)			(0.001)	
Treat × ln(Car Dist. in min)			-0.043***			-0.059***
			(0.003)			(0.002)
Observations	2,380,182	2,381,977	2,381,977	2,380,182	2,381,977	2,381,977
Squared Correlation	0.673	0.673	0.673	0.662	0.662	0.662

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at all same-chain stores, estimating Equation (C.1). This captures expenditures shifted from competitors to the new store. The dependent variable is log(expenditures). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

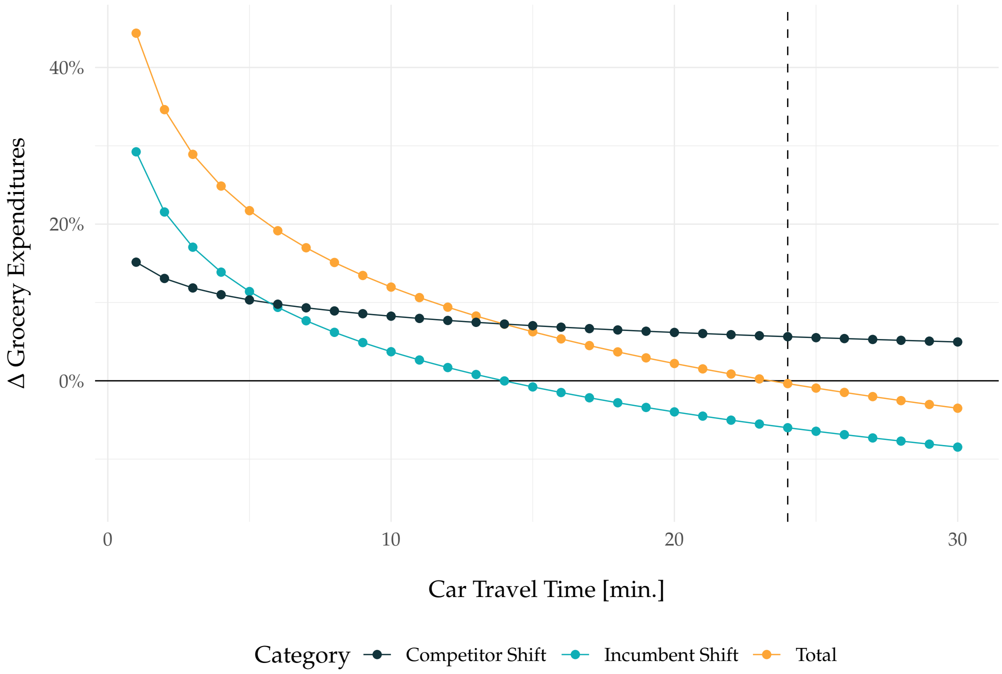
TABLE C.C4: Competitor Expenditure Shift (Log+1 Model)

	log(Expenditures + 1)			log(No. of Visits + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.131***	0.162***	0.352***	0.111***	0.134***	0.272***
	(0.006)	(0.007)	(0.014)	(0.015)	(0.004)	(0.009)
Treat × ln(Euclid. Dist. in km)	-0.046***			-0.035***		
	(0.003)			(0.005)		
Treat × ln(Road Dist. in km)		-0.055***			-0.042***	
		(0.004)			(0.002)	
Treat × ln(Car Dist. in min)			-0.118***			-0.087***
			(0.006)			(0.003)
Observations	2,599,180	2,601,098	2,601,098	2,599,180	2,601,098	2,601,098
Squared Correlation	0.557	0.557	0.557	0.639	0.639	0.640

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at all same-chain stores, estimating Equation (C.1). This captures expenditures shifted from competitors to the new store. The dependent variable is log(expenditures), where we add the value 1 to each household's expenditures. We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

C.C.2 Robust DiD Estimator and DiD With Kinks in the Log-Distance

FIGURE C.C3: Distance Gradients (Robust Estimator)



Notes: The figure shows distance gradient functions, indicating how the treatment effects from the difference-in-difference analyses decline over time. Compared to Figure 3.2, this figure uses the robust DiD estimator proposed by Wooldridge (2022) and Wooldridge (2023) as in Equation (3.7).

TABLE C.C5: Log Specifications With Kinks (Incumbent Shift)

	Expenditures					
Kink at Travel Time Distance	10 min.	11 min.	12 min.	13 min.	14 min.	15 min.
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.2893***	-0.2846***	-0.2723***	-0.2688***	-0.2640***	-0.2621***
	(0.0096)	(0.0091)	(0.0082)	(0.0079)	(0.0076)	(0.0073)
Treat × log(Car Dist) × 1(Below Kink)	0.1588***	0.1527***	0.1383***	0.1347***	0.1299***	0.1281***
	(0.0083)	(0.0075)	(0.0062)	(0.0057)	(0.0053)	(0.0050)
Treat × log(Car Dist) × 1(Above Kink)	0.0003***	0.0003***	0.0005***	0.0005***	0.0006***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	2,601,098	2,601,098	2,601,098	2,601,098	2,601,098	2,601,098
Squared Correlation	0.75687	0.75686	0.75684	0.75684	0.75683	0.75682

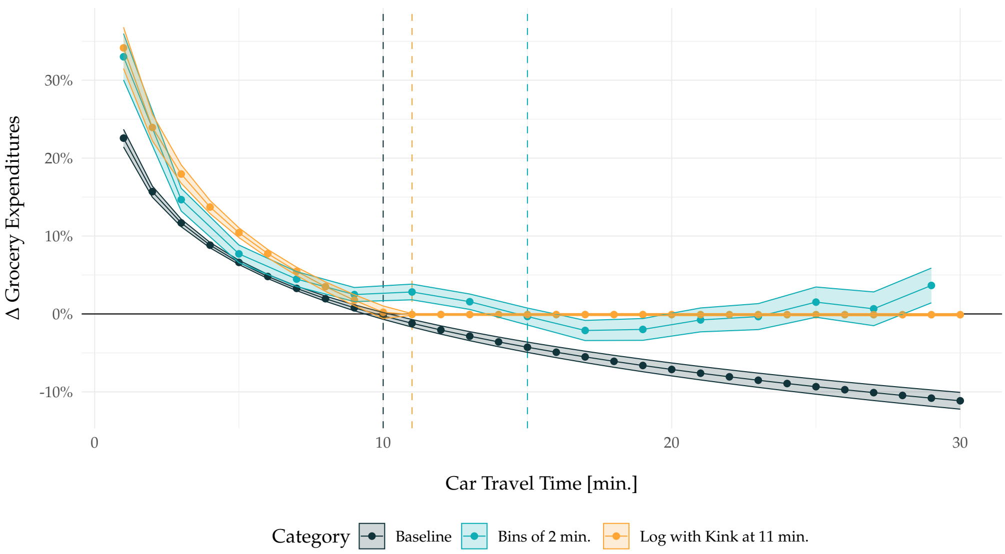
Notes: The table shows alternative model estimates for the distance decay functions in Figure C.C4 with a kink in the logarithmic specifications. Compared to Figure C.C4, the table displays intercepts and slope-coefficients of interest for kinks between 10 and 15 minutes.

TABLE C.C6: Log Specifications With Kinks (Competitor Shift)

	Expenditures					
Kink at Travel Time Distance	23 min.	24 min.	25 min.	26 min.	27 min.	28 min.
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.1315***	0.1316***	0.1329***	0.1325***	0.1329***	0.1329***
	(0.0065)	(0.0064)	(0.0063)	(0.0063)	(0.0062)	(0.0062)
Treat × log(Car Dist) × 1(Below Kink)	-0.0355***	-0.0355***	-0.0361***	-0.0359***	-0.0361***	-0.0361***
	(0.0023)	(0.0022)	(0.0022)	(0.0022)	(0.0021)	(0.0021)
Treat × log(Car Dist) × 1(Above Kink)	-0.0001	-0.0002	-0.0001	-0.0003	-0.0004	-0.0006**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Observations	2,601,098	2,601,098	2,601,098	2,601,098	2,601,098	2,601,098
Squared Correlation	0.76008	0.76008	0.76008	0.76008	0.76008	0.76008

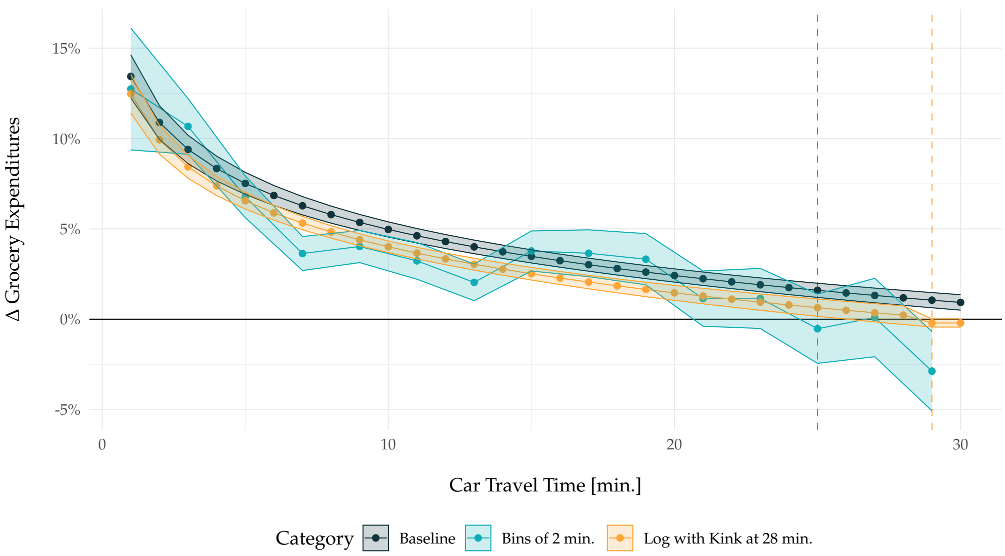
Notes: The table shows alternative model estimates for the distance decay functions in Figure C.C5 with a kink in the logarithmic specifications. Compared to Figure C.C5, the table displays intercepts and slope-coefficients of interest for kinks between 23 and 28 minutes.

FIGURE C.C4: Non-Parametric and Log-Kink Specification of the Incumbent Shift



Notes: The figure shows different specifications for the distance gradient functions, indicating how the treatment effects from the difference-in-difference analyses decline over time for the baseline *incumbent shift* estimates in Table 3.4 (mirrored along the horizontal axis). We calculate standard errors for the individual fitted points using the delta method. The dark *baseline* specification corresponds to the results displayed in Figure 3.2. The blue *bins of 2 min.* specification uses non-parametric travel time bins of 2 minutes. The orange *log with kink at 11 min.* estimates the baseline logarithmic model in Equation (3.6) but allows the slope to change after 11 minutes (when the baseline gravity function becomes insignificant).

FIGURE C.C5: Non-Parametric and Log-Kink Specification of the Competitor Shift



Notes: The figure shows different specifications for the distance gradient functions, indicating how the treatment effects from the difference-in-difference analyses decline over time for the baseline *competitor shift* estimates in Table 3.5. We calculate standard errors for the individual fitted points using the delta method. The dark *baseline* specification corresponds to the results displayed in Figure 3.2. The blue *bins of 2 min.* specification uses non-parametric travel time bins of 2 minutes. The orange *log with kink at 28 min.* estimates the baseline logarithmic model in Equation (3.6) but allows the slope to change after 28 minutes (when the baseline gravity function becomes insignificant).

C.D Robustness

TABLE C.D7: Incumbent Expenditure Shift Pre-COVID-19 (Poisson Model)

Proportional Effects	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.068***	-0.089***	-0.202***	-0.061***	-0.079***	-0.188***
	(0.003)	(0.003)	(0.006)	(0.011)	(0.003)	(0.006)
Treat × ln (Euclid. Dist. in km)	0.034***			0.031***		
	(0.002)			(0.004)		
Treat × ln (Road Dist. in km)		0.040***			0.038***	
		(0.002)			(0.002)	
Treat × ln (Car Dist. in min)			0.085***			0.081***
			(0.003)			(0.003)
Observations	823,046	823,367	823,367	823,046	823,367	823,367
Squared Correlation	0.802	0.802	0.802	0.807	0.807	0.808

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores, estimating Equation (3.6). This captures expenditures shifted from incumbent stores to the new store. Here, we focus on the period before the start of the COVID-19 pandemic, 2019/01 - 2020/02. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

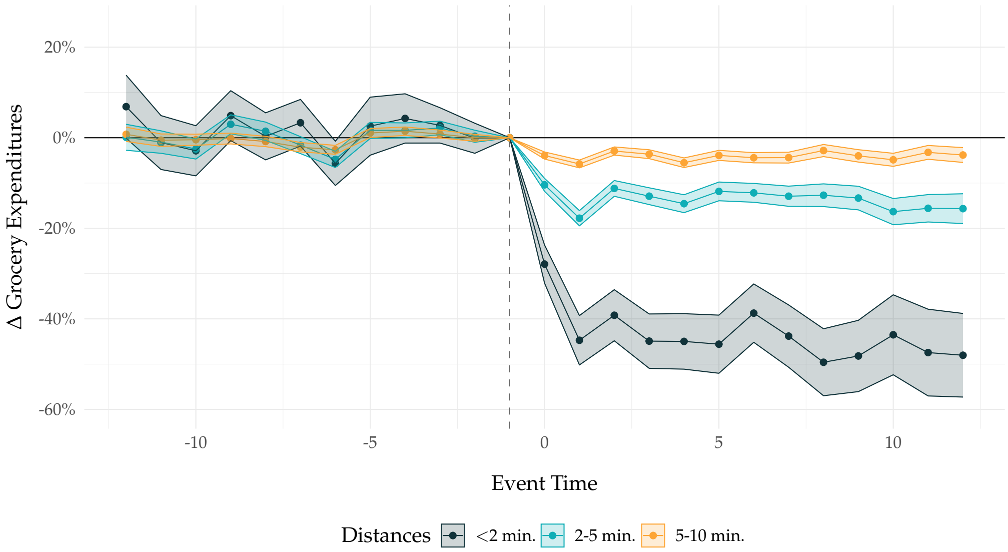
TABLE C.D8: Competitor Expenditure Shift Pre-COVID-19 (Poisson Model)

Proportional Effects	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.045***	0.052***	0.105***	0.058***	0.069***	0.141***
	(0.003)	(0.003)	(0.007)	(0.014)	(0.003)	(0.007)
Treat ln(Euclid. Dist. in km)	-0.007***			-0.014***		
	(0.001)			(0.004)		
Treat $\times$ ln(Road Dist. in km)		-0.009***			-0.016***	
		(0.001)			(0.001)	
Treat $\times$ ln(Car Dist. in min)			-0.026***			-0.037***
			(0.002)			(0.002)
Observations	823,046	823,367	823,367	823,046	823,367	823,367
Squared Correlation	0.803	0.803	0.803	0.804	0.804	0.805

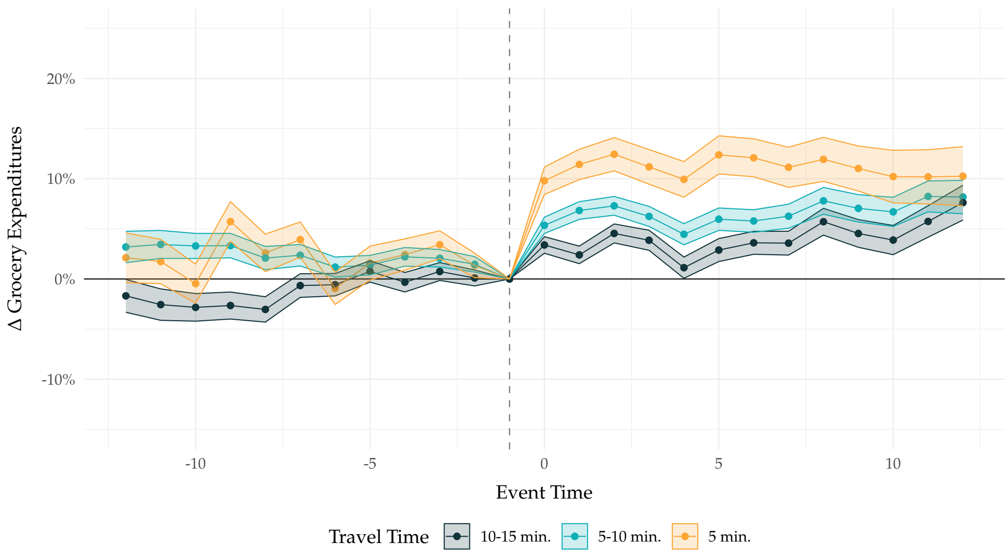
Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at all same-chain stores, estimating Equation (3.6). This captures expenditures shifted from competitors to the new store. Here, we focus on the period before the start of the COVID-19 pandemic, 2019/01 - 2020/02. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

FIGURE C.D6: Dynamic Treatment Effects

(A) Dynamic Effects: Incumbent Shift



(B) Dynamic Effects: Competitor Shift



Notes: The figure shows dynamic difference-in-differences estimates for the effect of a store opening on expenditures in an event-study fashion, estimating Equation (3.8). Figure C.D6a shows the incumbent shift and Figure C.D6b shows the competitor shift. As in the static results, coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

TABLE C.D9: Incumbent Expenditure Shift, Only Once-Treated Households (Poisson Model)

Proportional Effect	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.153***	-0.219***	-0.290***	-0.180***	-0.216***	-0.288***
	(0.029)	(0.033)	(0.048)	(0.060)	(0.028)	(0.039)
Treat × ln(Euclid. Dist. in km)	0.051***			0.067***		
	(0.014)			(0.023)		
Treat × ln(Road Dist. in km)		0.074***			0.084***	
		(0.016)			(0.014)	
Treat × ln(Car Dist. in min)			0.097***			0.106***
			(0.023)			(0.019)
Observations	198,354	198,354	198,354	198,354	198,354	198,354
Squared Correlation	0.735	0.735	0.735	0.755	0.755	0.755

Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores, estimating Equation (3.6). This captures expenditures shifted from incumbent stores to the new store. Here, we focus on households who were only treated once, meaning they only received one opening within 30 minutes from 2019Q1 to 2021Q2. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

TABLE C.D10: Competitor Expenditure Shift, Only Once-Treated Households (Poisson Model)

Proportional Effect	Expenditures			No. of Visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.075**	0.144***	0.259***	0.080***	0.144***	0.193***
	(0.034)	(0.045)	(0.080)	(0.025)	(0.038)	(0.061)
Treat × ln(Euclid. Dist. in km)	-0.042***			-0.031***		
	(0.012)			(0.009)		
Treat × ln(Road Dist. in km)		-0.057***			-0.045***	
		(0.013)			(0.011)	
Treat × ln(Car Dist. in min)			-0.078***			-0.052***
			(0.018)			(0.015)
Observations	198,354	198,354	198,354	198,354	198,354	198,354
Squared Correlation	0.741	0.741	0.741	0.757	0.757	0.757

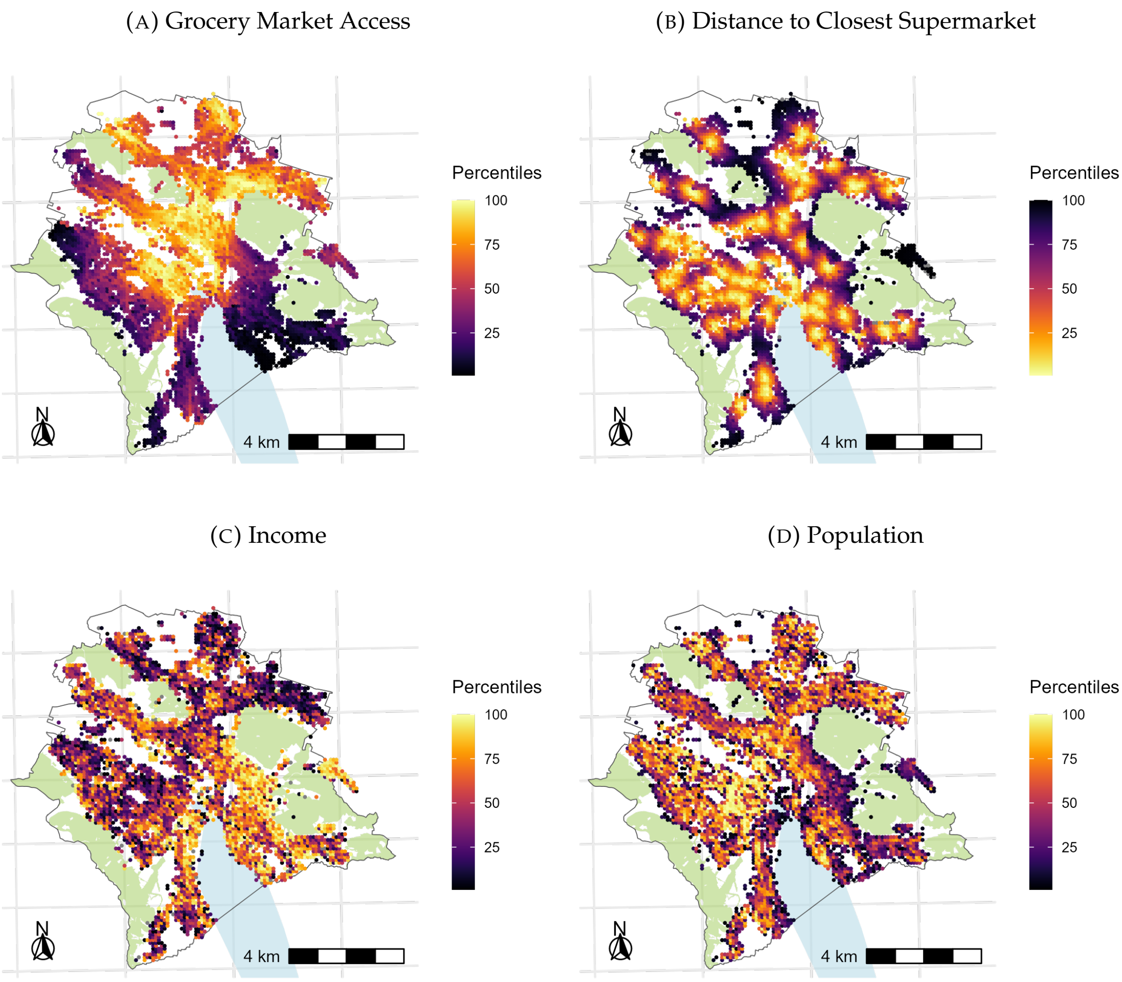
Notes: The table shows difference-in-differences estimates for the effect of a store opening on expenditures at all same-chain stores, estimating Equation (3.6). This captures expenditures shifted from competitors to the new store. Here, we focus on households who were only treated once, meaning they only received one opening within 30 minutes from 2019Q1 to 2021Q2. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

TABLE C.D11: Correlations for Figure C.D7

	Access	Inc.	Pop.
Income	-0.22		
Pop.	0.22	-0.15	
Dist. Store	-0.11	0.06	-0.26

Notes: This table shows the correlation matrix for four spatial variables, shown in Figure C.D7 for the City of Zurich.

FIGURE C.D7: Spatial Distribution: City of Zurich



Notes: The figure plots for each populated 100 × 100 meter grid cell in the city of Zurich the percentiles of (a) our utility-based valuation of market access, (b) the distance to the closest supermarket, (c) average household labor market income, and (d) population. The market access is based on Equation (3.5). We consider all stores of major grocery retailers in Switzerland, recover the unobserved store valuations with a Lasso approach, and use our causal estimates of travel costs. Travel distances between households and stores are measured as car travel times in minutes. Blue areas are water bodies, and green areas indicate forest areas.

TABLE C.D12: Heterogeneous Distance Costs (Rural Areas)

	Incumbent Shift		Competitor Shift		Total Shift				
Group	Intercept	Slope	Intercept	Slope	Intercept	Slope	Mean Dist	Cons. Area	n
Household Income									
<4,530	-0.827***	0.283***	0.356***	-0.108***	0.991***	-0.430***	16.4	10.0	224,761
	(0.050)	(0.018)	(0.029)	(0.010)	(0.047)	(0.025)			
4,530-6,717	-0.599***	0.198***	0.369***	-0.114***	0.897***	-0.327***	16.4	15.5	66,028
	(0.119)	(0.042)	(0.059)	(0.021)	(0.108)	(0.055)			
6,718-9,288	-0.812***	0.272***	0.335***	-0.099***	0.954***	-0.406***	16.6	10.5	105,178
	(0.083)	(0.029)	(0.040)	(0.014)	(0.067)	(0.040)			
9,289-12,856	-0.734***	0.252***	0.391***	-0.122***	0.999***	-0.401***	16.3	12.1	141,209
	(0.050)	(0.018)	(0.032)	(0.011)	(0.053)	(0.025)			
12,856+	-0.530***	0.182***	0.278***	-0.088***	0.732***	-0.283***	15.8	13.3	121,094
	(0.049)	(0.018)	(0.033)	(0.012)	(0.052)	(0.024)			
Age									
<34	-0.790***	0.262***	0.363**	-0.114*	0.985***	-0.408***	15.7	11.2	19,747
	(0.162)	(0.057)	(0.128)	(0.046)	(0.198)	(0.085)			
35-44	-0.559***	0.184***	0.359***	-0.115***	0.859***	-0.311***	16.2	15.8	90,255
	(0.082)	(0.029)	(0.041)	(0.014)	(0.075)	(0.038)			
45-55	-0.698***	0.240***	0.344***	-0.105***	0.913***	-0.371***	16.1	11.7	150,839
	(0.047)	(0.017)	(0.032)	(0.011)	(0.050)	(0.023)			
55-64	-0.706***	0.239***	0.321***	-0.097***	0.885***	-0.363***	16.4	11.5	161,584
	(0.050)	(0.018)	(0.031)	(0.011)	(0.050)	(0.024)			
65-74	-0.768***	0.262***	0.325***	-0.098***	0.920***	-0.393***	16.4	10.4	124,419
	(0.058)	(0.021)	(0.038)	(0.013)	(0.059)	(0.029)			
75+	-0.781***	0.273***	0.398***	-0.121***	1.030***	-0.428***	16.5	11.1	111,426
	(0.071)	(0.025)	(0.043)	(0.015)	(0.071)	(0.036)			
Household Size									
1	-0.661***	0.226***	0.334***	-0.098***	0.880***	-0.347***	16.3	12.6	113,633
	(0.114)	(0.041)	(0.043)	(0.015)	(0.084)	(0.053)			
2	-0.719***	0.245***	0.335***	-0.102***	0.911***	-0.375***	16.5	11.4	254,463
	(0.042)	(0.015)	(0.027)	(0.010)	(0.043)	(0.021)			
3-4	-0.668***	0.226***	0.309***	-0.097***	0.850***	-0.345***	16.1	11.7	231,827
	(0.039)	(0.014)	(0.024)	(0.009)	(0.039)	(0.019)			
5+	-0.815***	0.288***	0.498***	-0.150***	1.203***	-0.473***	16.5	12.7	58,347
	(0.082)	(0.029)	(0.048)	(0.017)	(0.087)	(0.041)			

Notes: The table shows for different characteristics in rural areas heterogeneous difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores and all stores, estimating in both cases Equation (3.6). This captures the *incumbent shift* and *competitor shift* respectively. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

TABLE C.D13: Heterogeneous Distance Costs (Urban Areas)

	Incumbent Shift		Competitor Shift		Total Shift				
Group	Intercept	Slope	Intercept	Slope	Intercept	Slope	Mean Dist	Cons. Area	n
Household Income									
<4,530	-0.326***	0.141***	0.028*	-0.009	0.307***	-0.161***	8.8	6.7	328,510
	(0.021)	(0.010)	(0.014)	(0.006)	(0.021)	(0.013)			
4,530-6,717	-0.205***	0.092***	0.088**	-0.031*	0.278***	-0.128***	8.6	8.8	71,365
	(0.039)	(0.018)	(0.028)	(0.013)	(0.044)	(0.023)			
6,718-9,288	-0.262***	0.119***	0.069*	-0.019	0.301***	-0.145***	8.8	8.0	85,111
	(0.042)	(0.019)	(0.028)	(0.013)	(0.045)	(0.025)			
9,289-12,856	-0.280***	0.124***	0.096***	-0.033**	0.346***	-0.165***	8.7	8.1	97,735
	(0.036)	(0.016)	(0.024)	(0.011)	(0.039)	(0.021)			
12,856+	-0.246***	0.109***	0.136***	-0.045***	0.364***	-0.159***	9.7	9.9	146,086
	(0.037)	(0.015)	(0.024)	(0.010)	(0.039)	(0.020)			
Age									
<34	-0.249***	0.093**	0.083·	-0.043*	0.306***	-0.140***	8.7	8.9	26,399
	(0.071)	(0.032)	(0.044)	(0.022)	(0.073)	(0.041)			
35-44	-0.243***	0.115***	0.107***	-0.033**	0.329***	-0.154***	8.8	8.5	86,220
	(0.044)	(0.020)	(0.026)	(0.012)	(0.045)	(0.025)			
45-55	-0.302***	0.132***	0.081***	-0.026**	0.345***	-0.167***	9.1	7.9	142,122
	(0.034)	(0.015)	(0.022)	(0.010)	(0.035)	(0.019)			
55-64	-0.250***	0.112***	0.077***	-0.023**	0.301***	-0.142***	9.0	8.3	150,992
	(0.033)	(0.015)	(0.019)	(0.009)	(0.033)	(0.019)			
65-74	-0.273***	0.123***	0.066**	-0.019*	0.308***	-0.149***	8.9	7.9	133,567
	(0.027)	(0.012)	(0.021)	(0.010)	(0.031)	(0.016)			
75+	-0.356***	0.150***	0.008	-0.003	0.308***	-0.165***	9.0	6.5	189,507
	(0.028)	(0.013)	(0.019)	(0.009)	(0.027)	(0.017)			
Household Size									
1	-0.266***	0.112***	0.054**	-0.022*	0.289***	-0.141***	8.7	7.8	210,790
	(0.025)	(0.011)	(0.019)	(0.009)	(0.028)	(0.015)			
2	-0.325***	0.145***	0.039**	-0.007	0.317***	-0.163***	9.0	7.0	257,345
	(0.022)	(0.010)	(0.015)	(0.007)	(0.022)	(0.013)			
3-4	-0.273***	0.121***	0.098***	-0.033***	0.342***	-0.161***	9.0	8.4	214,856
	(0.027)	(0.012)	(0.018)	(0.008)	(0.029)	(0.015)			
5+	-0.186**	0.084**	0.141***	-0.051**	0.321***	-0.137***	9.1	10.4	45,816
	(0.066)	(0.029)	(0.037)	(0.016)	(0.069)	(0.035)			

Notes: The table shows for different characteristics in urban areas heterogeneous difference-in-differences estimates for the effect of a store opening on expenditures at incumbent stores and all stores, estimating in both cases Equation (3.6). This captures the *incumbent shift* and *competitor shift* respectively. The coefficients are Poisson estimates where we report the exponentiated Poisson coefficients $exp(\beta) - 1$, following Chen and Roth (2024). We focus on households with average monthly expenditures between 20 and 10,000 Swiss francs with full information on income and household size who did not move during our sample period. Households are treated if they live within 30 minutes of a store opening.

Bibliography

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante. 2022. "And Yet It Moves: Intergenerational Mobility in Italy." *American Economic Journal: Applied Economics* 14 (3):118–163.

Adermon, Adrian, Mikael Lindahl, and Daniel Waldenström. 2018. "Intergenerational Wealth Mobility and the Role of Inheritance: Evidence from Multiple Generations." *The Economic Journal* 128 (612):F482–F513.

Afshin et al. 2019. "Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017." *The Lancet* 393 (10184):1958–1972.

Agarwal, Sumit, J. Bradford Jensen, and Ferdinando Monte. 2022. "Consumer Mobility and the Local Structure of Consumption Industries." *NBER Working Paper* 23616.

Aguilar, Arturo, Emilio Gutierrez, and Enrique Seira. 2021. "The effectiveness of sin food taxes: Evidence from Mexico." *Journal of Health Economics* 77:102455.

Ahlfeldt, Gabriel M., Fabian Bald, Duncan Roth, and Tobias Seidel. 2022. "Quality of Life in a Dynamic Spatial Model." *Working Paper* .

Ahlfeldt, Gabriel M., Volker Nitsch, and Nicolai Wendland. 2019. "Ease vs. noise: Long-run changes in the value of transport (dis)amenities." *Journal of Environmental Economics and Management* 98:102268.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *The*

Quarterly Journal of Economics 134 (4):1793–1844.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2022. "Digital Addiction." *American Economic Review* 112 (7):2424–2463.

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence." *Journal of Economic Perspectives* 33 (3):202–227.

Allen, Alissa L., John E. McGeary, and John E. Hayes. 2014. "Polymorphisms in TRPV and TAS2Rs Associate with Sensations from Sampled Ethanol." *Alcoholism: Clinical and Experimental Research* 38 (10):2550–2560.

Almagro, Milena and Tomas Domınguez-Iino. 2024. "Location Sorting and Endogenous Amenities: Evidence from Amsterdam." *Working Paper* .

Andersen, Carsten. 2021. "Intergenerational health mobility: Evidence from Danish registers." *Health Economics* 30 (12):3186–3202.

Araya, Sebastián, Andrés Elberg, Carlos Noton, and Daniel Schwartz. 2022. "Identifying Food Labeling Effects on Consumer Behavior." *Marketing Science* 41 (5):982–1003.

Arcidiacono, Peter, Paul B. Ellickson, Carl F. Mela, and John D. Singleton. 2020. "The Competitive Effects of Entry: Evidence from Supercenter Expansion." *American Economic Journal: Applied Economics* 12 (3):175–206.

Asher, Sam, Paul Novosad, and Charlie Rafkin. 2024. "Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups." *American Economic Journal: Applied Economics* 16 (2):66–98.

Asplund, Marcus, Richard Friberg, and Fredrik Wilander. 2007. "Demand and distance: Evidence on cross-border shopping." *Journal of Public Economics* 91 (1-2):141–157.

Atkin, David. 2013. "Trade, Tastes, and Nutrition in India." *American Economic Review* 103 (5):1629–1663.

———. 2016. "The Caloric Costs of Culture: Evidence from Indian Migrants." *American Economic Review* 106 (4):1144–1181.

Auer, Raphael, Ariel Burstein, Sarah Lein, and Jonathan Vogel. 2023. "Unequal expenditure switching: Evidence from Switzerland." *Review of Economic Studies* forthcoming.

Baggs, Jen, Loretta Fung, and Beverly Lapham. 2018. "Exchange rates, cross-border travel, and retailers: Theory and empirics." *Journal of International Economics* 115:59–79.

Bailey, Martha J, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker. 2024. "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program." *Review of Economic Studies* 91 (3):1291–1330.

Baker, Scott R., Stephanie Johnson, and Lorenz Kueng. 2021. "Shopping for Lower Sales Tax Rates." *American Economic Journal: Macroeconomics* 13 (3):209–250.

Barahona, Nano, Cristóbal Otero, and Sebastián Otero. 2023. "Equilibrium Effects of Food Labeling Policies." *Econometrica* 91 (3):839–868.

Becker, G. S. and C. B. Mulligan. 1997. "The Endogenous Determination of Time Preference." *The Quarterly Journal of Economics* 112 (3):729–758.

Belloc, Marianna, Francesco Drago, Mattia Fochesato, and Roberto Galbiati. 2024. "Multigenerational Transmission of Wealth: Florence, 1403–1480." *American Economic Journal: Applied Economics* 16 (2):99–129.

Berry, James, Saurabh Mehta, Priya Mukherjee, Hannah Ruebeck, and Gauri Kartini Shastry. 2021. "Crowd-out in school-based health interventions: Evidence from India's midday meals program." *Journal of Public Economics* 204 (C):104552.

Birch, Leann L. 1999. "Development of Food Preferences." *Annual Review of Nutrition* 19:41–62.

Black, Sandra, Neil Duzett, Adriana Lleras-Muney, Nolan Pope, and Joseph Price. 2024. "Intergenerational Transmission of Lifespan in the US." *NBER Working Paper* 31034.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital." *American Economic Review* 95 (1).

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. 2017. "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US." *The Scandinavian Journal of Economics* 119 (1):72–101.

Bronnenberg, Bart J, Jean-Pierre H Dubé, and Matthew Gentzkow. 2012. "The Evolution of Brand Preferences: Evidence from Consumer Migration." *American Economic Review* 102 (6):2472–2508.

Brueckner, Jan K and Yves Zenou. 1999. "Why is central Paris rich and downtown Detroit poor? An amenity-based theory." *European Economic Review* 43 (1):91–107.

Bruze, Gustaf. 2018. "Intergenerational mobility: New evidence from consumption data." *Journal of Applied Econometrics* 33 (4):580–593.

Burstein, Ariel, Sarah Lein, and Jonathan Vogel. 2022. "Cross-border shopping: evidence and welfare implications for Switzerland." *Working Paper* .

Büchel, Konstantin, Maximilian V. Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2020. "Calling from the outside: The role of networks in residential mobility." *Journal of Urban Economics* 119:103277.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87 (3):741–835.

Callaway, Brantly and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2):200–230.

Campbell, Jeffrey and Beverly Lapham. 2004. "Real Exchange Rate Fluctuations and the Dynamics of Retail Trade Industries on the U.S.-Canada Border." *American Economic*

Review 94 (4).

Campbell, John Y. and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2):205–251.

Carroll, Christopher D, Jody Overland, and David N Weil. 2000. "Saving and Growth with Habit Formation." *American Economic Review* 90 (3):341–355.

Chandra, Ambarish, Keith Head, and Mariano Tappata. 2014. "The Economics of Cross-Border Travel." *Review of Economics and Statistics* 96 (4):648–661.

Charles, Kerwin, Sheldon Danziger, Geng Li, and Robert Schoeni. 2014. "The Intergenerational Correlation of Consumption Expenditures." *American Economic Review* 104 (5):136–140.

Charles, Kerwin and Erik Hurst. 2003. "The Correlation of Wealth across Generations." *Journal of Political Economy* 111 (6):1155–1182.

Chen, Jiafeng and Jonathan Roth. 2024. "Logs with Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics* 139 (2):891–936.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6):2205–2268.

Chetty, Raj and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics* 133 (3):1163–1228.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and Economic Opportunity in the United States: an Intergenerational Perspective." *The Quarterly Journal of Economics* 135 (2):711–783.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4):855–902.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104 (5):141–147.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, and al. 2022a. "Social capital I: measurement and associations with economic mobility." *Nature* 608 (7921):108–121.

———. 2022b. "Social capital II: determinants of economic connectedness." *Nature* 608 (7921):122–134.

Chuard, Patrick and Veronica Grassi. 2020. "Switzer-Land of Opportunity: Intergenerational Income Mobility in the Land of Vocational Education." *Working Paper* .

Clark, Gregory and Neil Cummins. 2015. "Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility." *The Economic Journal* 125 (582):61–85.

Cole, Joanne B., Jose C. Florez, and Joel N. Hirschhorn. 2020. "Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations." *Nature Communications* 11 (1):1467.

Cook, Philip J, Jan Ostermann, and Frank A Sloan. 2005. "The Net Effect of an Alcohol Tax Increase on Death Rates in Middle Age." *American Economic Review* 95 (2):278–281.

Corak, Miles. 2020. "The Canadian Geography of Intergenerational Income Mobility." *The Economic Journal* 130 (631):2134–2174.

Couture, Victor and Jessie Handbury. 2020. "Urban revival in America." *Journal of Urban Economics* 119:103267.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives* 20 (3):97–120.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–2996.

Deutscher, Nathan and Bhashkar Mazumder. 2020. "Intergenerational mobility across Australia and the stability of regional estimates." *Labour Economics* 66:101861.

———. 2023. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature* 61 (3):988–1036.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3):479–524.

Dickson, Alex, Markus Gehrsitz, and Jonathan Kemp. 2023. "Does a Spoonful of Sugar Levy Help the Calories Go Down? An Analysis of the UK Soft Drinks Industry Levy." *Review of Economics and Statistics* :1–29.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2020. "How Well Targeted Are Soda Taxes?" *American Economic Review* 110 (11):3661–3704.

Eaton, Jonathan and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5):1741–1779.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2008. "On the Accuracy of Nielsen Homescan Data." *USDA Report* 69:34.

Eizenberg, Alon, Saul Lach, and Merav Oren-Yiftach. 2021. "Retail Prices in a City." *American Economic Journal: Economic Policy* 13 (2):175–206.

Fadlon, Itzik and Torben Heien Nielsen. 2019. "Family Health Behaviors." *American Economic Review* 109 (9):3162–3191.

Fernandez, R., A. Fogli, and C. Olivetti. 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *The Quarterly Journal of Economics* 119 (4):1249–1299.

Forestell, Catherine A. 2024. "Does Maternal Diet Influence Future Infant Taste and Odor Preferences? A Critical Analysis." *Annual Review of Nutrition* 44.

Forestell, Catherine A. and Julie A. Mennella. 2007. "Early Determinants of Fruit and Vegetable Acceptance." *Pediatrics* 120 (6):1247–1254.

Friberg, Richard, Emil M. S. Halseth, Frode Steen, and Simen A. Ulsaker. 2024. "The effect of cross-border shopping on commodity tax revenue: results from Norway's COVID-19 border closings." *The Scandinavian Journal of Economics* forthcoming.

Friberg, Richard, Frode Steen, and Simen A. Ulsaker. 2022. "Hump-Shaped Cross-Price Effects and the Extensive Margin in Cross-Border Shopping." *American Economic Journal: Microeconomics* 14 (2):408–438.

Frimmel, Wolfgang, Martin Halla, and Jörg Paetzold. 2019. "The Intergenerational Causal Effect of Tax Evasion: Evidence from the Commuter Tax Allowance in Austria." *Journal of the European Economic Association* 17 (6):1843–1880.

Fuhrer, Jeffrey C. 2000. "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review* 90 (3):367–390.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4):160–188.

Gervis, Julie E., Jiantao Ma, Kenneth K.H. Chui, Nicola M. McKeown, Daniel Levy, and Alice H. Lichtenstein. 2023. "Bitter- and Umami-Related Genes are Differentially Associated with Food Group Intakes: the Framingham Heart Study." *The Journal of Nutrition* 153 (2):483–492.

Goldin, Jacob, Tatiana Homonoff, and Katherine Meckel. 2022. "Issuance and Incidence: SNAP Benefit Cycles and Grocery Prices." *American Economic Journal: Economic Policy* 14 (1):152–178.

Halliday, Timothy J., Bhashkar Mazumder, and Ashley Wong. 2020. "The intergenerational transmission of health in the United States: A latent variables analysis." *Health Economics* 29 (3):367–381.

Han, Lu, Stephan Heblich, Christopher Timmins, and Yanos Zylberberg. 2024. "Cool Cities: The Value of Urban Trees." *NBER Working Paper* 32063.

Handbury, Jessie. 2021. "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities." *Econometrica* 89 (6):679–2715.

Handbury, Jessie and Sarah Moshary. 2021. "School Food Policy Affects Everyone: Retail Responses to the National School Lunch Program." *NBER Working Paper* 29384.

Handbury, Jessie and David E Weinstein. 2015. "Goods Prices and Availability in Cities." *Review of Economic Studies* 82 (1):258–296.

Hastings, Justine, Ryan Kessler, and Jesse M. Shapiro. 2021. "The Effect of SNAP on the Composition of Purchased Foods: Evidence and Implications." *American Economic Journal: Economic Policy* 13 (3):277–315.

Hausman, Jerry and Ephraim Leibtag. 2007. "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart." *Journal of Applied Econometrics* 22 (7):1157–1177.

Hausman, Naomi, Peleg Samuels, Maxime Cohen, and Roy Sasson. 2023. "Urban Pull: The Roles of Amenities and Employment." *Working Paper* .

Heblich, Stephan, Alex Trew, and Yanos Zylberberg. 2021. "East-Side Story: Historical Pollution and Persistent Neighborhood Sorting." *Journal of Political Economy* 129 (5):1508–1552.

Henderson, Daniel J. and Daniel L. Millimet. 2008. "Is gravity linear?" *Journal of Applied Econometrics* 23 (2):137–172.

Hillberry, Russell and David Hummels. 2008. "Trade responses to geographic frictions: A decomposition using micro-data." *European Economic Review* 52 (3):527–550.

Hoelzlein, Matthias and Sam Miller. 2024. "The Gentrification Effect of Retail Amenities: Evidence from Whole Foods Market." *Working Paper* .

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal* 39 (153):41.

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm. 2024. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." *American Economic Review* 114 (2):385–425.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2022. "Rational Habit Formation: Experimental Evidence from Handwashing in India." *American Economic Journal: Applied Economics* 14 (1):1–41.

Hut, Stefan. 2020. "Determinants of Dietary Choice in the US: Evidence from Consumer Migration." *Journal of Health Economics* 72:102327.

Hut, Stefan and Emily Oster. 2022. "Changes in household diet: Determinants and predictability." *Journal of Public Economics* 208:104620.

Jäntti, Markus and Stephen P. Jenkins. 2015. "Income Mobility." *Handbook of Income Distribution* 2:807–935.

Khera, Amit V., Connor A. Emdin, Isabel Drake, Pradeep Natarajan, Alexander G. Bick, Nancy R. Cook, Daniel I. Chasman, Usman Baber, Roxana Mehran, Daniel J. Rader, Valentin Fuster, Eric Boerwinkle, Olle Melander, Marju Orho-Melander, Paul M Ridker, and Sekar Kathiresan. 2016. "Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease." *New England Journal of Medicine* 375 (24):2349–2358.

Kluser, Frederic. 2024. "Cross-Border Shopping: Evidence from Household Transaction Records." *Working Paper* .

Kluser, Frédéric and Martina Pons. 2024. "The Apple Does Not Fall Far From the Tree: Intergenerational Persistence of Dietary Habits." *Working Paper* .

Kluser, Frédéric, Tobias Seidel, and Maximilian von Ehrlich. 2024. "Spatial Frictions in Retail Consumption." *Working Paper* .

Knight, Brian and Nathan Schiff. 2012. "Spatial Competition and Cross-Border Shopping: Evidence from State Lotteries." *American Economic Journal: Economic Policy* 4 (4):199–229.

Law, Cherry, Andrés Sánchez Pájaro, Richard Smith, and Laura Cornelsen. 2024. "Do healthier food baskets cost more?" *Working Paper* .

Leal, Andrés, Julio López-Laborda, and Fernando Rodrigo. 2010. "Cross-Border Shopping: A Survey." *International Advances in Economic Research* 16 (2):135–148.

Leonardi, Marco and Enrico Moretti. 2023. "The Agglomeration of Urban Amenities: Evidence from Milan Restaurants." *American Economic Review: Insights* 5 (2):141–57.

Lleras-Muney and Lichtenberg. 2005. "Are the More Educated More Likely to Use New Drugs?" *Annals of Economics and Statistics* 79/80:671–69.

Marshall, Guillermo and Tiago Pires. 2018. "Measuring the Impact of Travel Costs on Grocery Shopping." *The Economic Journal* 128 (614):2538–2557.

May, Christina E. and Monica Dus. 2021. "Confection Confusion: Interplay Between Diet, Taste, and Nutrition." *Trends in Endocrinology & Metabolism* 32 (2):95–105.

Mennella, Julie A., Nuala K. Bobowski, and Danielle R. Reed. 2016. "The development of sweet taste: From biology to hedonics." *Reviews in Endocrine and Metabolic Disorders* 17 (2):171–178.

Mennella, Julie A., Coren P. Jagnow, and Gary K. Beauchamp. 2001. "Prenatal and Postnatal Flavor Learning by Human Infants." *Pediatrics* 107 (6):e88–e88.

Mennella, Julie A., M. Yanina Pepino, and Danielle R. Reed. 2005. "Genetic and Environmental Determinants of Bitter Perception and Sweet Preferences." *Pediatrics* 115 (2):e216–e222.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding. 2022. "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data." *Working Paper* .

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review* 108 (12):3855–3890.

Oh, Ryungha and Jaeeun Seo. 2023. "What Causes Agglomeration of Services? Theory and Evidence from Seoul." *Working Paper* .

Oster, Emily. 2018. "Diabetes and Diet: Purchasing Behavior Change in Response to Health Information." *American Economic Journal: Applied Economics* 10 (4):308–348.

Pope, Devin G. and Jaren C. Pope. 2015. "When Walmart comes to town: Always low housing prices? Always?" *Journal of Urban Economics* 87:1–13.

Rabbitt, Matthew P., Laura J. Hales, Michael P. Burke, and Alisha Coleman-Jensen. 2023. "Household food security in the United States in 2022." *U.S. Department of Agriculture, Economic Research Service* Report No. ERR-325.

Relihan, Lindsay E. 2024. "Clicks and Bricks: The Complementarity of Online Retail and Urban Services." *Working Paper* .

Ritchie, Hannah. 2021. "Three billion people cannot afford a healthy diet." URL `https://ourworldindata.org/diet-affordability`.

Ritchie, Hannah, Pablo Rosado, and Max Roser. 2023. "Hunger and Undernourishment." URL `https://ourworldindata.org/hunger-and-undernourishment`.

Ritchie, Hannah and Max Roser. 2024. "Obesity." URL `https://ourworldindata.org/obesity`.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6):1257–1278.

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban, Quality of Life." *Current issues in urban economics* :74–104.

Rothstein, Jesse. 2019. "Inequality of Educational Opportunity? Schools as Mediators of the Intergenerational Transmission of Income." *Journal of Labor Economics* 37 (1):85–123.

Scaglioni, Silvia, Valentina De Cosmi, Valentina Ciappolino, Fabio Parazzini, Paolo Brambilla, and Carlo Agostoni. 2018. "Factors Influencing Children's Eating Behaviours." *Nutrients* 10 (706).

Schuetz, Jenny. 2015. "Why are Walmart and Target Next-Door neighbors?" *Regional Science and Urban Economics* 54:38–48.

Springmann, Marco, Gary Sacks, Jaithri Ananthapavan, and Peter Scarborough. 2018. "Carbon pricing of food in Australia: an analysis of the health, environmental and public finance impacts." *Australian and New Zealand Journal of Public Health* 42 (6):523–529.

Story, Mary, Karen M. Kaphingst, Ramona Robinson-O'Brien, and Karen Glanz. 2008. "Creating Healthy Food and Eating Environments: Policy and Environmental Approaches." *Annual Review of Public Health* 29 (1):253–272.

Søberg, Susanna, Camilla H. Sandholt, Naja Z. Jespersen, Ulla Toft, Anja L. Madsen, Stephanie Von Holstein-Rathlou, Trisha J. Grevengoed, Karl B. Christensen, Wender L.P. Bredie, Matthew J. Potthoff, Thomas P.J. Solomon, Camilla Scheele, Allan Linneberg, Torben Jørgensen, Oluf Pedersen, Torben Hansen, Matthew P. Gillum, and Niels Grarup. 2017. "FGF21 Is a Sugar-Induced Hormone Associated with Sweet Intake and Preference in Humans." *Cell Metabolism* 25 (5):1045–1053.e6.

Verghese, Anila, Margaret Raber, and Shreela Sharma. 2019. "Interventions targeting diet quality of Supplemental Nutrition Assistance Program (SNAP) participants: A scoping review." *Preventive Medicine* 119:77–86.

Waldkirch, Andreas, Serena Ng, and Donald Cox. 2004. "Intergenerational Linkages in Consumption Behavior." *The Journal of Human Resources* 39 (2):355.

Wise, Paul M, Laura Nattress, Linda J Flammer, and Gary K Beauchamp. 2016. "Reduced dietary intake of simple sugars alters perceived sweet taste intensity but not perceived pleasantness." *The American Journal of Clinical Nutrition* 103 (1):50–60.

Wooldridge, Jeffrey. 2023. "Simple approaches to nonlinear difference-in-differences with panel data." *The Econometrics Journal* 26 (3):C31–C66.

Wooldridge, Jeffrey M. 2022. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." *Working Paper* .