Essays in Industrial Organization of Spatially Differentiated Markets

Alina Ozhegova
Norwegian School of Economics

Acknowledgements

Embarking on the journey of a Ph.D. is often accompanied by moments of uncertainty and challenge. However, I have been incredibly fortunate to have had the unwavering support and guidance of individuals whose impact on my academic and personal growth has been profound.

First and foremost, I extend my heartfelt gratitude to my brilliant advisors, Morten Sæthre and Mateusz Myśliwski. Morten's exceptional mentorship, keen insights, and support have been my anchor throughout these five years. His profound understanding of the research field, coupled with our lengthy meetings and his invaluable feedback, have played an indispensable role in shaping this thesis. Morten's contagious passion for research ignited my interest in the field and propelled me forward. Similarly, I am immensely grateful to Mateusz for his open-door policy and boundless support, from career advice to paper details. With his sense of humor, Mateusz has made even the most daunting challenges seem surmountable. I am profoundly grateful for his guidance and friendship.

I want to express my sincere gratitude to Pierre Dubois for his warm welcome in Toulouse and invaluable advice on my job market paper. Despite his busy schedule, Pierre always made time for our discussions. These conversations enriched my knowledge and fueled my curiosity, leaving a lasting impact on me.

To Fedor Iskhakov, I am grateful for his guidance, for warmly welcoming me to Canberra, and for introducing me to the world of dynamic programming, sailing, and orienteering. Our discussions, though at times as complex as the algorithms we dissected, have been both enlightening and entertaining.

I had the privilege of receiving thoughtful and helpful feedback from Frode Steen. His expertise in retail markets proved invaluable, and I am grateful for the time he took to review my work. In addition, the use of Norwegian data helped to significantly enrich the depth and quality of my research.

My gratitude also extends to the Department of Economics. I would like to thank Aline for supporting me from my first day at the Department and being a great role model. My gratitude also extends to Kurt for everything he has helped me during my time at NHH. A big thank you to Lars and the BECCLE group for their dedicated efforts in cultivating Bergen as a thriving hub for industrial organization. The atmosphere at the Department has consistently been welcoming, encouraging, and intellectually stimulating. Perhaps it is the abundance of free food, but I have always found joy in coming to the office.

To my fellow Ph.D. candidates, I want to express my gratitude for your support, the dinners we shared, and the memorable trips we took to the Småbruket cabin. Special appreciation goes to Emil and Paula, my companions in the IO group, for not only being exceptional colleagues but also for the unforgettable adventures we shared during conference trips. I am especially grateful to Akshay and Daniel for being a safe place throughout these five years. Our talks have been a big help and a comfort during tough times.

I am grateful to the graduate students at the Toulouse School of

Economics for their exceptional warmth and hospitality. I extend a special thanks to Johanna and Oscar for our weekly Zoom discussions, which have been incredibly helpful throughout my PhD journey.

Additionally, I want to thank everyone who supported me during the challenging job market season. In particular, I want to thank Anders Munk-Nielsen, Mario Samano, Anna Ignatenko, Isis Durrmeyer, and Daniel Ershov for their time, effort, and our engaging discussions. I am grateful to Alon Eizenberg, Jessie Handbury, Matthew Backus, Saul Lach, Kevin Remmy, Andreas Haller, Katrine Løken, Bertel Tungodden, and Lukasz Grzybowski for the time they spent reading and giving invaluable feedback on my job market paper. I also want to express my gratitude to Andreas Haller, Sam Hirshman, and Katrine Løken for their constructive feedback and support throughout this stressful period. Their contributions made a significant difference, and I truly appreciate their collaborative efforts.

I would like to express my gratitude to my co-authors: Anders Munk-Nielsen, Alexander Tarasov, Philip Ushchev, Sergey Kokovin, and Shamil Sharapudinov, for the countless hours dedicated to our collaborative efforts during numerous Zoom calls.

I want to thank the people who inspired me and set me on a path of research. In particular, I thank Sergey Kokovin, Dmitriy Potapov, and Victor Chernozhukov for igniting my interest in research and in the field of industrial organization and for encouraging me to start a Ph.D. journey.

I am incredibly grateful for my family's unwavering love and support throughout this journey. To my parents, thank you for standing by me, for your support, and for believing in my dreams even when they seemed out of reach. To my brother, your presence and belief in me have been a constant source of strength and motivation.

Lastly, I am endlessly grateful to Zhenya for his practical, moral and always unselfish support. His constant presence throughout the entire journey has been a source of immeasurable comfort. I cherish every shared moment and feel truly blessed to have him by my side.

This thesis would not have been possible without the collective efforts and support of all those mentioned above. Thank you for being a part of this journey.

Alina Ozhegova
Bergen, April 2024

Contents

Introduction

The market offers a vast selection of products to cater to the diverse tastes of consumers. This diversity in taste can be categorized into two levels. First, individuals may prefer to consume different products on different occasions, showing a preference for variety over time. Second, individuals often have unique preferences for specific product variants, such as their preferred brand of yogurt or cereals. Thus, the population appears heterogeneous, and this diversity creates a demand for a wide range of products.

To appeal to heterogeneous tastes, firms offer differentiated products. The decisions regarding what products to produce, how much to produce, and at what price to sell are crucial factors that determine a firm's profit. These decisions are also at the core of strategic interactions since a firm's profit relies not only on its own decisions but also on those of its competitors. In a homogeneous good market, a firm's demand will reach zero as soon as a competitor undercuts its price. In an industry where products are differentiated, the demand for a firm's product decreases as other firms lower their prices. The extent of the decrease in demand will depend on the level of product differentiation. Therefore, understanding the degree of product differentiation has first-order importance to comprehending competition and market outcomes in most industries.

The typical situation in most industries is that consumers face a choice of products that vary along different dimensions. The most common typology of product differentiation is the horizontal-vertical distinction (Waterson, 1989). Vertical differentiation refers to the differences in quality of products, where if two products are offered at the same price, all consumers prefer the same higher-quality product (Sutton, 1986). Examples of vertical differentiation can be found across various industries, such as the performance of a product in terms of speed or efficiency (e.g., the operating speed of a computer), service quality in service industries, or the reliability of a product (e.g., more reliable cars would be perceived as superior to others).

In the case of horizontal product differentiation, there is no natural ranking of products. If all products are sold at the same price, there is a positive demand for each of them (Sutton, 1986). In most industries, brands offer various flavors, textures, and packaging designs to appeal to diverse consumer tastes. For example, in the cereal market, some consumers may prefer healthier options with whole grains, while others may seek sweeter varieties.

Geographical location is another common way to differentiate products horizontally. Consumers have a preference for a product based on where they live. If two products are identical and priced the same but have different locations, consumers will most likely buy from the seller closest to them. Then, two consumers who have different locations may have distinct preferences over the product variants that are being offered.

Since consumers are heterogeneous and willing to pay more for products that cater to their tastes, this creates a source of market power for firms. Despite consumer heterogeneity, the increasing returns to scale of production and distribution make it difficult to sustain a large number of products. As each firm must capture a large share of demand to cover their fixed costs, only a few firms offering a limited number of products can survive in the market. This creates a market distortion resulting in non-optimal market outcomes, including price and product selection Spence (1976*b*). While the oligopoly distortion in the pricing decision is well studied in the literature, firms often also compete via the products they choose to offer and other non-price product characteristics, leading to another possible market failure, which is the primary focus of this thesis.

Theoretical groundwork to study product differentiation was established by landmark papers from Hotelling (1929*a*) and Chamberlin (1933). Building on their foundation, significant contributions have been made to understanding the implications of horizontal and vertical product differentiation on market structure, considering both single- (Shaked and Sutton, 1982, 1987) and multi-product firms (Schmalensee, 1978; Eaton and Lipsey, 1989; Judd, 1985).

One of the key themes explored in these papers is the welfare aspects of product differentiation (Spence, 1976*a*). Product differentiation is often seen as a way to improve efficiency by providing consumers with more suitable product variants that match their preferences. However, product differentiation can also soften price competition (Shaked and Sutton, 1982). In addition, there is typically a divergence between firms' profits and product differentiation's contributions to surplus (Spence, 1976*b*). In particular, firms have the incentive to choose their product characteristics to exploit market power and maximize profits rather than maximize social surplus, which may also lead to inefficiency (Spence, 1976*b*). Theory alone cannot predict the ultimate result of these competing forces which depends on the structure of demand and costs. Hence, product differentiation presents a trade-off that requires understanding both consumer preferences and companies' strategic interactions and has a clear policy implication, particularly in terms of promoting efficiency.

In the traditional marginal analysis of consumer behavior, it is assumed that consumer behavior is continuous. However, there are instances where consumer choices are mutually exclusive, meaning that choosing one product rules out the possibility of choosing another (Domenchich and McFadden, 1975). Even though, in this case, a consumer's response to a firm's price change may be discontinuous, there is a way to generate continuous demand. This can be done by assuming that firms cannot observe some idiosyncratic taste parameters that affect consumer choice. If we know the distribution from which these taste parameters are drawn, we can forecast demand using a discrete choice model of consumer behavior. Discrete choice models start from the assumption that each consumer chooses one option that brings the highest utility. The utility is described as a random variable that depends on observable product characteristics and the properties of the taste distribution (McFadden, 1974; Anderson, De Palma and Thisse, 1992).

Starting with Bresnahan (1987), it became common to study oligopoly models with product differentiation using the joint empirical analysis of demand and supply using discrete-choice models of product differentiation. Some seminal studies in this field include Berry (1994), which sets the theoretical benchmark for addressing unobserved product characteristics, and Berry, Levinsohn and Pakes (1995), which carries out an equilibrium analysis using the theoretical framework proposed by Berry (1994). Following McFadden (1974), the demand framework is based on the consumer's utility, which is a function of product and consumer characteristics. Then, demand for each product is an aggregation of demands of individual consumers who prefer this product to all others. On the supply side, multi-product firms act as oligopolists and compete by prices, and other product characteristics are assumed to be exogenous. The approach by Berry, Levinsohn and Pakes (1995) has built the foundation for empirical studies of product differentiation in many industries, where along with price, firms compete by other product characteristics (Matsa, 2011; Crawford, Shcherbakov and Shum, 2019; Fan and Yang, 2020), including the choice of location (Bresnahan and Reiss, 1991; Seim, 2006; Igami and Yang, 2013; Zheng, 2016; Richards, Chenarides and Çakir, 2022).

In this thesis, I explore questions related to strategic interaction in various imperfectly competitive industries, building upon previous studies that established the foundations for studying differentiated product markets. In Chapter 1, I take a closer look at how multi-store firms compete through non-price attributes. While much attention has been given to market power over price, there is less focus on market power over non-price product characteristics. As previously mentioned, in imperfectly competitive markets, firms have the incentive to distort non-price characteristics from socially optimal levels (Spence, 1976*b*). As theory cannot predict the ultimate direction of this distortion, the answer to this question remains empirical. In particular, the chapter focuses on how firms choose which products to offer depending on local competition. This is an important topic as retailers often provide numerous products or services in a single store. While it is widely recognized that the selection of products to offer is a strategic decision for retailers Argentesi et al. (2021), the effects of such choices on consumers residing in different competitive environments have yet to be explored in the existing literature.

The second chapter addresses the question of how multi-store firms compete by choosing geographic locations for new stores. A conventional economic assumption is that free entry into markets is socially desirable. However, Mankiw and Whinston (1986) argues that, in certain circumstances, free entry can lead to outcomes that are not socially optimal. For instance, free entry can result in too little or too much entry, and the level of product differentiation is a crucial factor in determining the direction of entry bias. Berry and Waldfogel (2001) further empirically shows that free entry in differentiated product markets can result in a different number of products than socially optimal. The paper aims to provide new evidence on what drives incumbent chains to enter the same local market, particularly exploring the role of informational advantage in making entry decisions.

Finally, the third chapter studies product differentiation in a monopolistic competition setting. Building on the seminal work of Dixit and Stiglitz (1977), we develop a novel theory of monopolistic competition that takes into account the horizontal heterogeneity of consumers in their spatial locations and the vertical heterogeneity of firms in their productivities. Despite the extensive literature on firm heterogeneity in monopolistic competition, the role of consumer heterogeneity remains relatively unexplored. This paper aims to address this gap and contribute to understanding the consequences of this two-sided heterogeneity in a free-entry equilibrium framework.

Throughout my thesis, I combine theory with an empirical approach. In particular, I use game-theoretic models of competition together with modern structural econometrics to better understand market outcomes observed in real-world data. In the following section, I provide a summary of each of the three chapters.

This first chapter, titled **Assortment Choice and Market Power under Uniform Pricing**, sheds light on the importance of market power in non-price attributes in imperfectly competitive industries. While previous studies have mainly focused on price, firms can also distort other non-price attributes from socially optimal levels. Such attributes include delivery time in online shopping, product downsizing in the retail industry, or product selection in the grocery industry. The study focuses on the Norwegian grocery industry.

To begin with, I establish two key stylized facts. First, I show that pricing decisions for individual products are made at the national level. This observation aligns with the established fact in the literature that in many countries, retailers employ uniform pricing, meaning they charge the same price for a product in stores across markets. While this fact is puzzling in itself, it also raises the question of whether firms leverage their market dominance through non-price characteristics. Second, I show that, unlike prices, product selection decisions appear to vary locally. Then, based on the stylized facts, I develop and estimate an equilibrium model, where I model how consumers decide where to do grocery shopping and chains decide on store-level product selection. In this chapter, I aim to distinguish the impact of market power on assortment choice from other market forces, such as logistics costs.

Using the model, I show that firms adjust their product selection based on the level of competition in the local markets. In areas with stronger market power, stores offer a smaller range of products at higher prices. In the counterfactual analysis, I also show that government intervention in the form of subsidies to consumers or retailers in remote areas can help improve the overall market welfare.

In the second chapter titled **Preemption in Spatial Competition: Evidence from the Retail Pharmacy Market**, the focus shifts to how multi-store firms make entry decisions. This study is a joint work with Anders Munk-Nielsen and Morten Sæthre. It focuses on the prescription pharmaceutical market in Norway. Motivated by the deregulation of the industry in 2001, we analyze the dynamics of entry among pharmacy chains in Norway, with a specific focus on spatial competition. While theory suggests that firms tend to open new outlets close to competitors under business-stealing motives, in this study, we aim to explain the phenomenon that incumbent firms frequently open new outlets close to their existing ones. Notably, the study capitalizes on the unique characteristics of the prescription pharmaceutical market, where demand is largely unaffected by pharmacy entry and price competition is nonexistent due to strict price regulation. Therefore, pharmacies mainly compete based on their location. By utilizing detailed transaction data on prescription pharmaceuticals and household locations disaggregated into demographic groups, we aim to provide novel insights into the drivers of entry decisions among pharmacy chains.

In the study, we employ a rich demand model inspired by Ellickson, Grieco and Khvastunov (2020), allowing for overlapping consumer choice sets across space. Unlike conventional studies that primarily observe aggregated sales, this research benefits from finely disaggregated store-level sales data, enabling a precise examination of cannibalization versus business-stealing motives for entry. We propose that incumbents have a unique advantage in learning about local market demand through their sales observations, which allows them to identify areas where consumer demand exceeds expectations based on demographic characteristics. Essentially, we argue that repeated entries into a market may be attributed to access to private information. While previous literature, such as Igami and Yang (2016), has discussed the potential influence of common information shocks on preemptive entry, this study proposes a distinction between common and private information shocks. It suggests that if the shocks are private information, incumbent firms are more likely to respond to positive demand residuals compared to competing firms, leading to a clearer understanding of the underlying dynamics driving entry behavior and location choice.

The third chapter turns attention to a monopolistic competition setting. The chapter **A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers** is a collaborative work with Sergey Kokovin, Shamil Sharapudinov, Alexander Tarasov, and Philip Ushchev. Recent developments in the monopolistic competition model have not paid much attention to the interaction between consumer heterogeneity and supply-side heterogeneity. The third chapter aims to fill this gap by developing a novel theory of monopolistic competition with bilateral heterogeneity, considering both horizontal heterogeneity of consumers in spatial locations and vertical heterogeneity of firms in productivities.

The proposed theory diverges from traditional monopolistic competition models by allowing active firms to choose price and location in the product space, where space can be interpreted either as a geographical space or as a space of characteristics of a differentiated good. Then, we explore the trade-off each firm faces between accessing a larger local market and encountering softer local competition, shedding light on the equilibrium outcomes of monopolistic competition models.

We discuss patterns of equilibria that can arise in a setting of fully localized competition when firms serve only those consumers for whom their products are the most preferred ones. In particular, we find that under certain conditions, the equilibrium exists and exhibits positive assortative matching, with more productive firms opting for larger local markets. This matching between firms and market niches has significant implications for sales distribution, prices, and markups, offering insights into firms' market niche choices and spatial distribution of consumer welfare.

Next, we calibrate the model using cross-sectional data from the hairdressing industry in Bergen, Norway, which closely aligns with the assumptions of monopolistic competition. By observing population distribution, hairdresser locations, turnovers, and profits, the model effectively captures the relationship between firms' prices, markups, and productivities, including potential non-monotonic markup patterns.

Finally, we conduct two counterfactual experiments to explore how a proportional increase in population density and setting fixed production costs to zero affect market entry, competition levels, and consumer welfare. We find that more firms enter the market, increasing competition. While consumers gain from these changes, not all benefit equally. Those who live closer to the city center benefit more than those who live in remote locations. This result highlights the importance of the sorting mechanism in creating heterogeneity in consumer welfare.

Chapter 1

Assortment Choice and Market Power under Uniform Pricing

Abstract: This paper studies how retailers strategically use product assortment to respond to local market conditions when prices are set at the national level. When firms cannot increase the price of a product that is particularly popular in a local market, they can instead replace the product with a more expensive substitute. The profitability of these assortment substitutions depends on the degree of market competition. This study uses extensive receipt and store-level data and a structural equilibrium model to distinguish the impact of market power on assortment choice from other market forces, such as logistics costs. The findings confirm that firms make use of assortment choices, offering fewer and pricier products in markets with stronger local market power. I show that a uniform assortment would benefit consumers but would reduce firm profits. Counterfactual policy experiments reveal that government intervention can improve total market welfare through subsidies to consumers or retailers in remote areas.

1.1 Introduction

Unlike market power over price, there has been much less focus on market power regarding non-price characteristics. Similar to prices, firms operating within imperfectly competitive industries have the ability to distort non-price attributes from socially optimal levels. Examples include delivery time in online shopping (Ater and Shany, 2021), product downsizing in the retail industry (Yonezawa and Richards, 2016), or, as the central focus of this paper, product selection in the grocery industry, where firms can deliberately restrict consumer choice in stores.

The importance of this issue has recently become apparent, as there is increasing evidence that multi-store retailers follow uniform pricing. Uniform pricing refers to the practice of charging the same prices for products across markets with different demographics, preferences, and levels of competition (DellaVigna and Gentzkow, 2019; Adams and Williams, 2019; Hitsch, Hortacsu and Lin, 2019). This study focuses on the Norwegian grocery industry and demonstrates the use of uniform pricing in this context, even though many supermarkets have substantial local market power. In Norway, 22% of grocery stores are considered local monopolies with no competitors within a 5 km radius. This raises the question of whether grocery chains leverage their market dominance through non-price channels when prices are fixed. In this paper, I show that the choice of product assortment offered is one such possible channel and the strategic decisions regarding product assortment made by these firms can significantly affect consumer welfare.

When deciding what products to offer, store managers consider the following trade-off. Removing cheap products from a store may cause some consumers to switch to another store, while the remaining consumers are more likely to purchase expensive, higher-margin products. If local competition is intense, then the first effect prevails. However, if competition is weak, reducing product assortment may be profitable. This example highlights how store managers can strategically make assortment decisions to maximize profits based on the level of local competition they face.

Informal discussions with industry experts indicate that the decision-making process occurs at two levels. First, each chain decides on product-level prices, and then regional and store managers make decisions regarding product selection. This two-tiered decision-making process provides informal evidence for the importance of the assortment channel and allows me to focus solely on assortment decisions while considering product-level prices as given. However, to rigorously investigate this process, I establish two key stylized facts. Firstly, I provide evidence that pricing decisions for individual products are made at the national level. Secondly, I show that product selection decisions appear to be made at the local level.

To study assortment decisions, I use data from multiple sources. The primary data source is weekly sales at the product and store level for all stores belonging to a large Norwegian retail group. The secondary source is a database provided by Geodata, the primary Norwegian spatial data provider. The database contains information on yearly store-level revenue, location, and other characteristics for all grocery stores in Norway. Next, I collected data on the location of distribution centers and driving distance between stores and distribution centers. Finally, I use detailed information on demographic distribution from Geodata.

Based on the stylized facts, I develop and estimate an equilibrium model for the grocery market. On the demand side, I specify a spatial model where consumers decide which store to visit and how much to spend on groceries. In particular, I model how consumers weigh the travel distance against store characteristics, including assortment. On the supply side, chains decide on assortment in each store. The key tradeoff for a firm is that removing cheap products might discourage some consumers but increases the marginal profit from the remaining consumers. Since local market power tends to be more pronounced in certain areas, e.g. rural areas, where the distance between stores is large, so that consumers are unwilling to switch to a different store, this can result in substantial assortment differences across markets. Therefore, based on the model estimates, I quantify the welfare effects of assortment differences driven by local market power for consumers in different markets.

To measure assortment at a store level, I aggregate individual product items into a composite good. Each store is modeled as making choices regarding two assortment measures characterizing a composite good: *price*, which represents the price level of assortment offered, and *variety*, which quantifies the breadth of assortment. In particular, when designing the price of the composite good, I calculate the average expenditures on a typical shopping basket in each store, similar to Eizenberg, Lach and Oren-Yiftach (2021). When measuring assortment breadth, I count the number of unique products offered in a particular store, consistent with the approach in previous studies (see, e.g., Argentesi et al., 2021; Kim and Yeo, 2021). Using a composite good, I can simplify the assortment analysis and capture the main factors influencing consumer's store choices, such as shopping costs and product selection.

The structural model builds on the novel approach of Ellickson, Grieco and Khvastunov (2020), which allows spatially heterogeneous consumers to have location-specific choice sets and extends it by introducing an unobserved demand shifter. This framework differs from the conventional isolated markets' approach used in previous literature (Bresnahan and Reiss, 1991; Zheng, 2016). In particular, I employ a spatial discrete choice model that explicitly accounts for the distance between consumers and stores, allowing me to measure local competitive pressure more accurately. In the model, the set of available stores and the degree of substitution depend on how consumers trade off travel distance and store characteristics, including price level and breadth of assortment. Additionally, I extend the model to allow for structural unobserved store-level component, which is a significant improvement as it allows to separate unobserved store quality from the preferences of consumers residing in a particular location.

On the supply side, each chain makes store-level assortment decisions, determining the price and variety of composite goods to maximize chain profit. In order to account for the higher costs associated with offering a wider variety of products, I specify a cost function that accounts for logistics costs and store characteristics, including assortment breadth.

Since assortment variables could be correlated with the unobserved demand shifter, I have to address the endogeneity problem. To obtain consistent estimates of the model parameters, I employ instrumental variables and use the generalized method of moments (GMM) for estimation. In particular, I follow Houde (2012) and bring the Berry (1994) approach for inverting market shares to the spatial model of Ellickson, Grieco and Khvastunov (2020). As instruments, I leverage differentiation and BLP instruments along with exogenous cost shifters. These instruments aim to isolate variation that drives the assortment decisions from the unobserved demand determinants while capturing competitive pressure. BLP instruments are designed to exploit observed characteristics of competing stores, while differentiation instruments are based on the distance between a store and its competitors in the characteristics space.

Based on the estimates of spatial demand model, I can revisit the market concentration discussion. Dealing with aggregated data, I do not observe grocery expenditure flows between consumers and stores and cannot evaluate the level of competition for all possible consumer locations. However, the model allows me to overcome this limitation and calculate market concentration for each consumer based on their specific location without making strict assumptions about the geographic boundaries of the market. This approach allows me to more accurately quantify local competition, even in small rural areas that would typically be aggregated into larger geographic regions, leading to potentially inaccurate competition assessment. In particular, spatial concentration is calculated based on choice probabilities predicted by the demand model. These localized concentration measures show that most markets in Norway are moderately concentrated (56%) or highly concentrated (41%), and only 3% are considered competitive.[1] Additionally, the market concentration is higher in rural areas compared to urban areas.

Next, the spatial demand model uncovers assortment inequality across different regions. Residents of large cities have access to more affordable groceries and greater variety, while consumers in

[1]Standard cutoffs are used here. A market with an HHI of less than 1,500 is considered a competitive marketplace, an HHI of 1,500 to 2,500 is moderately concentrated, and an HHI of 2,500 or greater is highly concentrated.

remote markets face a more limited and pricier assortment.

Using the model, I can separate the impact of local market power from other factors that may affect assortment choice, such as logistics costs, local tastes, and store characteristics. In particular, the model allows me to estimate store-level margins that illustrate stores' ability to raise prices above the marginal cost or limit variety, thus reducing marginal costs - both are indicative of local market power. Conversely, factors other than local market power are reflected in the marginal cost. Furthermore, by connecting the choice-weighted margin per person to the localized degree of market concentration, I quantify the variance of margins that can be associated with differences in market concentration. In the most concentrated markets, consumers spend up to 25% more than in the most competitive markets, which amounts to EUR 1,500 annually.

Then, I use the model to conduct three counterfactual experiments. The first experiment is a synthetic one aimed at better understanding the current market equilibrium and quantifying the overall effects of assortment differences across markets. Specifically, I simulate a scenario where grocery chains adopt a uniform assortment strategy, meaning that stores of one chain provide the same composite good. I show that the uniform assortment scenario leads to an increase in the variety and price of the composite good. Additionally, I analyze the effects across different markets in detail. Interestingly, markets with higher concentration experience a relatively smaller price increase but a more significant increase in variety. This finding highlights the current assortment and welfare inequality between competitive and concentrated markets. While, on average, consumers benefit from the uniform assortment, the policy has only a minor effect on consumer inequality across various locations, primarily because consumers in remote areas continue to face higher transportation costs compared to urban residents.

Next, I show that varying assortment across markets is a profitable strategy for firms. If firms were to provide an equal assortment instead, it would result in lower profits for firms and negative profits for some stores. Thus, imposing a uniform assortment is not a feasible solution due to store closures and reduced competition in some markets.

Therefore, I run counterfactual experiments designed to mimic realistic policies that could mitigate the distortionary effects of assortment choices. In the next counterfactual experiment, I explore the implications of reducing travel costs for consumers in remote areas. Reducing travel costs facilitates better access to stores, consequently enhancing competition in remote areas. Market concentration changes, leading to a lower number of concentrated markets and a higher number of competitive areas, putting downward pressure on prices and upward pressure on variety. As a result, consumer welfare and firms' profits increase by 11.4% and 5.6%, respectively. Considering the cost of implementation, the policy has a positive net welfare effect.

In the last counterfactual scenario, I examine the potential impact of subsidies to retailers in remote areas to compensate for higher logistics costs. The results show that such subsidies lead to a modest reduction in the price of the composite good by 1.92% and a slight increase in variety by 0.69%. This, in turn, leads to a 1.8% increase in consumer welfare and a 6.8% rise in firms' profits. Furthermore, the policy generates a positive net welfare effect, taking into account the costs of its implementation.

The paper speaks to the empirical literature that explores the effects of competition on non-price attributes. Although there is extensive literature on price-setting under imperfect competition, much less attention has been paid to the impact of competition on quality and non-price attributes in a more general sense. As with prices, firms in imperfectly competitive industries tend to deviate from socially optimal levels of quality, but unlike prices, the direction of this distortion is not clear (Spence, 1975). For instance, Crawford, Shcherbakov and Shum (2019) and Fan and Yang (2020) show that under competitive pressure, firms tend to provide higher quality and higher prices than socially optimal. The literature also includes studies exploring the effects of mergers on product offerings in the market, such as the work by Mazzeo, Seim and Varela (2018) and Sweeting (2010). Additionally, Matsa (2011) studies how competition affects quality in a grocery context, where quality is measured as the number of stockouts.

This study is closely related to the work of Argentesi et al. (2021), which examines the effect of a merger between two chains on prices and product assortment. The authors find that after the merger, chains tend to adjust their assortment rather than prices, suggesting that product selection is a strategic variable for retail chains. Similar to Argentesi et al. (2021), I find empirical evidence that product selection can vary locally. However, this paper differs from theirs in several aspects. First, I use a structural model to separate the impact of local competition from other forces that can impact product assortment decisions. Second, the structural model allows me to examine the effects of these assortment differences on consumers across various markets. Lastly, using the model, I simulate counterfactual experiments and propose policy insights on improving assortment in remote areas.

This paper also contributes to the growing literature on food price and assortment inequality between markets with different socio-demographic and economic characteristics (Dubois, Griffith and Nevo, 2014; Handbury and Weinstein, 2015; Allcott et al., 2019; Handbury, 2019; Eizenberg, Lach and Oren-Yiftach, 2021). The findings in Handbury (2019) suggest that low-income households face different assortment and prices than high-income households mainly due to their income-specific tastes. In this vein, a higher degree of heterogeneous local tastes can be beneficial for all consumers in a market, leading to increased variety (Quan and Williams, 2018). Additionally, Eizenberg, Lach and Oren-Yiftach (2021) study price differences within a city's stores and attribute them mainly to spatial frictions. In this paper, I show how, in the context of uniform pricing, firms resort to other strategies to imperfectly segregate the market. Furthermore, I explore how this assortment strategy creates spatial inequalities and affects consumers in urban and rural markets. Similar to Eizenberg, Lach and Oren-Yiftach (2021), I show that urban residents have better access to a cheaper assortment than residents of rural areas. Using the counterfactual analysis, I also provide policy insights on how to reduce welfare distortions associated with assortment inequality.

Lastly, the paper relates to a growing literature on uniform pricing (Adams and Williams, 2019; DellaVigna and Gentzkow, 2019; Hitsch, Hortacsu and Lin, 2019). In a seminal paper (DellaVigna and Gentzkow, 2019), the authors document the use of uniform pricing by a number of US retailers. Adams and Williams (2019) study welfare effects and find that uniform pricing can shield consumers from higher prices in less competitive markets. Similarly, this study confirms the practice of uniform pricing among retailers in Norway. Moreover, this study complements this strand of literature by showing that when prices are fixed nationally, firms use other non-price channels, in this case product selection, to respond to changes in market structure.

The paper proceeds as follows. In the next section, I describe the data used in the analysis. Section 3 presents stylized facts. In section 4, I describe the equilibrium demand and supply framework underlying my empirical model. Section 5 describes the identification of structural parameters. Section 6 presents the estimation results of the demand and supply models. Section 7 presents the results from the counterfactual experiments. Section 8 concludes.

1.2 Data

I begin by describing the Norwegian grocery landscape and the data sources used in the study. Next, I explain how I utilize the data to construct the price and variety measures of the composite good.

The Norwegian grocery industry consists of four retail groups: NorgesGruppen (NG), Rema1000, Coop, and Bunnpris. As of 2018, these four corporations control 99.9% of the market.[2] Table 1.1 presents selected statistics for the Norwegian grocery market. Some of the retail groups have multiple chains representing different grocery formats. For example, the market leader NorgesGruppen has a discount format (Kiwi), a convenience store format (Joker), supermarkets (Spar), and high-quality supermarkets (Meny). Such differentiation allows for serving various consumer segments. Independent stores not belonging to the four listed retail groups constitute a small part of the market (less than 0.1%). Most of them are located in large cities and usually provide a spe-

[2] Nielsen, Grocery report 2017

cific assortment, such as imported products targeted at consumers with non-Norwegian backgrounds.

Table 1.1: Market structure of the grocery industry, 2018

	Market share	Revenue	Number of stores
NorgesGruppen	42.5	72,614	1,734
Kiwi	20.4	34,892	646
Meny	10.8	18,428	193
Spar/Eurospar	7.1	12,054	282
Joker	3.6	6,156	448
Coop	29.6	50,469	1,114
Coop Extra	13.3	22,726	424
Coop Obs	5.6	9,523	30
Coop Prix	4.4	7,456	254
Coop Mega	3.9	6,716	75
Coop Marked	1.7	2,949	227
Rema 1000	24.1	41,153	589
Bunnpris	3.8	6,510	246
Total	100	79,215	3683

Note: Market shares are in percent, revenues are in million Norwegian krones. Numbers were retrieved from companies' annual reports.

The data comes from multiple sources. The primary data source is receipt data from one large Norwegian retail group, which operates throughout the entire country and has stores of all existing formats in a market, such as discounters, convenience stores, and supermarkets. The data contains item-level transactions in all individual shopping receipts for March 2018 across all stores belonging to the retail group. Each item is a unique stock keeping unit (SKU). The dataset contains information about prices with and without discount for individual items in a receipt, quantity purchased, store, and product IDs. The information about prices and products offered in stores obtained from this dataset serves as the foundation for constructing store-level assortment measures, which will be used in subsequent analyses.

The second data source is a geocoded store-level panel provided by Geodata, a Norwegian spatial data provider. Geodata's database contains yearly information on store-level revenue for 2010-2021. Additionally, it includes information on location, store ID, store opening date, size, and the number of employees. Table 1.2 shows descriptive statistics for stores.

Table 1.2: Store-level descriptives, 2018

	Mean	SD	Min	Median	Max
Revenue (mln NOK)	48.39	51.43	0.07	39.71	1249.5
Number of employees	25.21	73.02	1	17	2304
Open hours	13.95	2.56	3	15	24
Open on Sunday	0.16	-	0	-	1
Location in mall	0.16	-	0	-	1

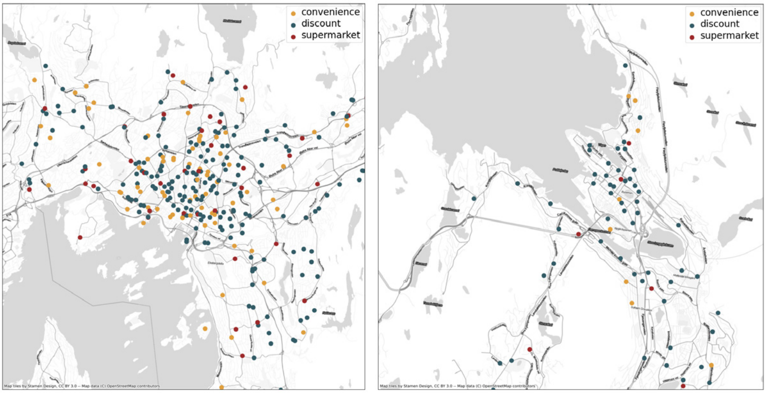
Source: Geodata.

Geodata's database covers the whole grocery market in Norway, providing a comprehensive overview of the industry. Figure 1.1 illustrates the spatial distribution of stores in the two largest cities of Norway. I use the information on store locations to measure the degree of spatial competition and to construct choice sets of consumers residing in different locations in the spatial demand model. The unique store ID allows to link Geodata's database on revenues with the receipt data.

Additionally, I use a detailed demographic database provided by Geodata. I use this data at the most granular statistical geographic unit known as a basic unit (BU).[3] To illustrate the spatial distribution, Figure 1.2 demonstrates how the two largest cities in Norway are divided into basic units. Table 1.3 reports descriptive statistics of demographic data at the basic unit level.

Similar to other scanner datasets, the receipts do not contain

[3]Basic units are generally geographically smaller than zip codes. Basic units are similar to census blocks in the US.



(a) Oslo (b) Bergen

Figure 1.1: Location of stores

information on the residential location of consumers. Therefore, I need to assume which stores consumers can visit. Since it is likely that consumers residing in a particular basic unit shop in stores, located in different basic units, I do not adopt the conventional isolated markets' approach inspired by Bresnahan and Reiss (1991). Instead, I link the store-level aggregate revenues to consumer choices using the spatial demand model, exploiting data on store locations and the distribution of consumer demographics. Section 5 provides details of the modeling procedure.

Table 1.3: Descriptive statistics of demographics data by basic units

	Mean	SD	Min	Median	Max
Area (km^2)	22.98	67.62	0.03	3.44	1805.21
Population	283.7	314.6	1	179	4272
Population density (people/km^2)	1248	29366	0.09	41.9	3472394
Average income (thou. NOK)	659.5	546.9	78.8	546.7	18000

Source: Geodata.

(a) Oslo (b) Bergen

Figure 1.2: Division into basic units

Composite Good

To document assortment differences across stores in Norway, I aggregate individual product items into a *composite good* representing a basket of groceries purchased by an average consumer. The composite good is characterized by price and variety measures at the store level. Using a composite good is common in industrial organization (Handbury, 2019; DellaVigna and Gentzkow, 2019; Eizenberg, Lach and Oren-Yiftach, 2021; Duarte, Magnolfi and Roncoroni, 2020) and urban economics literature (MacDonald and Nelson Jr, 1991) when one needs to compare multi-products stores by relative shopping costs and product selection.

To construct a composite good, I focus on fourteen popular product categories that most households consume daily. The categories are selected based on their sales revenues, excluding fruits and vegetables, as they are not subject to uniform pricing.[4] The

[4]The suppliers of fruits and vegetables can vary across regions.

final set of product categories comprises cheese, eggs, fresh bread, juice, frozen fish, chocolate bars, beer, jam, dry bread, coffee, milk, yogurt, frozen pizza and canned fish. Each category includes from 10 to 162 products, where a product is identified by a stock-keeping unit ID which is a consistent identifier across all stores in Norway.

Information about products offered in each store and individual product-level prices are collected from the receipt data. As the receipt data records a product's price, quantity purchased, and package size, it allows calculating a price for a standardized product unit (for example, a kilogram of cheese or a liter of milk).

Following Eizenberg, Lach and Oren-Yiftach (2021), I define the price of the composite good as the revenue-weighted average across the chosen categories. In the notation below, i represents a product, c denotes a category, and j is the subscript for a store. To aggregate product-level prices p_i into a category-level price p_{cj}, I calculate a revenue-weighted average for products within category c and store j, denoted as Ω_{cj}. I use relative total product revenue in the retail group as weights, so more popular products have higher weights in the category-level price. To estimate category costs, I multiply the revenue-weighted average by the average purchased units in the category or the *average basket*. Thus, the revenue-weighted average price for category c in store j is given by:

$$p_{cj} = \text{average basket}_c \times \left(\frac{\sum_{i \in \Omega_{cj}} w_i p_i}{\sum_{i \in \Omega_{cj}} w_i} \right). \tag{1.1}$$

Note that since product-level prices p_i are fixed, and weights w_i are determined globally and do not vary across stores, variations in the composite good price solely arise from the differences in the product set Ω_{cj} across stores. This difference plays a crucial role in the analysis as it allows to investigate strategic assortment decisions made by retailers.

Finally, I calculate the price of a single unit of the composite good p_j by averaging category-level prices p_{cj} across chosen categories:

$$p_j = \frac{1}{C} \sum_{c=1}^{C} p_{cj}, \tag{1.2}$$

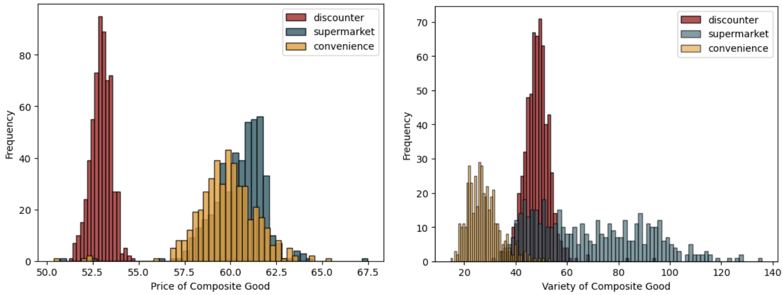
where C is the total number of categories.

To measure the breadth of assortment, I first calculate ν_{cj} as the number of unique products offered in category c of store j. Then following Argentesi et al. (2021), I define variety ν_j of store j as an average number of unique products across chosen categories:

$$\nu_j = \frac{1}{C} \sum_{c=1}^{C} \nu_{cj}. \tag{1.3}$$

Figures 1.3a and 1.3b show the distribution of price and variety across different retail formats. First, they reveal notable differences in assortment across different retail formats. As expected, discount stores offer a cheaper assortment than supermarkets and convenience stores. Furthermore, the assortment within discount stores is more uniform in terms of price and variety measures compared with other formats. Convenience stores offer expensive but a more limited range of products. Finally, supermarkets exhibit greater variation in the assortment breadth compared to other formats.

Table 1.4 presents descriptive statistics for the price and variety of composite good across stores. Given that the receipt data is available only for one retail group, each format corresponds to a single chain. Additionally note that stores of one chain have the same prices for products. Hence, any differences in the price of composite good originates only from the difference in the prod-



(a) Price (b) Variety

Figure 1.3: Distribution of price and variety across chains

uct selection. Further notice that this price variation measured in the 95% confidence interval accounts for 10% of the average price of the composite good for convenience stores, 7% for discounters, and 9% for supermarkets, which can result in significant welfare effects. Variety differs noticeably across stores of one format, too. Aside from market power, this variation could be explained by many confounding factors, including the size of a store and local tastes. I will explore these differences further in the following section.

Table 1.4: Price and variety summary statistics

	Mean	SD	Min	Median	Max
Price					
Convenience store	59.85	1.44	56.08	59.78	65.48
Discount	53.02	0.89	51.21	52.98	61.44
Supermarket store	60.6	1.41	52.14	60.79	67.49
Variety					
Convenience store	27.21	6.15	14.64	26.43	53.36
Discount	48.43	5.19	16.64	48.43	94.36
Supermarket store	69.45	22.04	30.57	66.71	135.93

It should be noted that assortment information is inferred from the transaction data. Given the limited shelf space in stores, it is plausible to assume that each product displayed in a store has been purchased at least once during the observed month; otherwise, it would not be stocked. Since the transaction data captures one month of purchase activity, any short-term stockouts are assumed to occur randomly.

Furthermore, in Norway, retailers have three periods per year, so-called *launch windows* (in February, in June, and October), when chain managers can centrally introduce changes in the assortment. The data available for this study covers the period between these *launch windows*, leading me to assume that the chains did not alter their assortment during a given month.[5]

1.3 Stylized Facts

This section uses the data described before to present two stylized facts that support my model assumptions presented in the next section. First, I show that retail chains indeed follow uniform pricing. Second, I document that product selection can vary locally depending on local market conditions.

Retail Chains Follow Uniform Pricing

Studies by DellaVigna and Gentzkow (2019) and Hitsch, Hortacsu and Lin (2019) show that national pricing is an industry norm among grocery chains in the US. In contrast, Eizenberg, Lach and

[5] The standardization committee for the Norwegian grocery industry: `https://stand.no/prosess/sortiment/grunndataregistrering-og-produktpresentasjon/`

Oren-Yiftach (2021) reveal significant local price differences in grocery prices in Israel. Given extensive receipt data available, I investigate whether there is variation in product prices within chains.

To begin, I visualize price variation both across all chains and within stores of one chain. Figure 1.4 illustrates that price deviations from the mean product price within stores of the same chain are concentrated around zero. Conversely, there is substantial variation in prices for the same product across different chains. Figures A.1 and A.2 in the Appendix present similar plots for product price variation in separate product categories. This result supports the fact that product prices do not vary across stores of one chain.

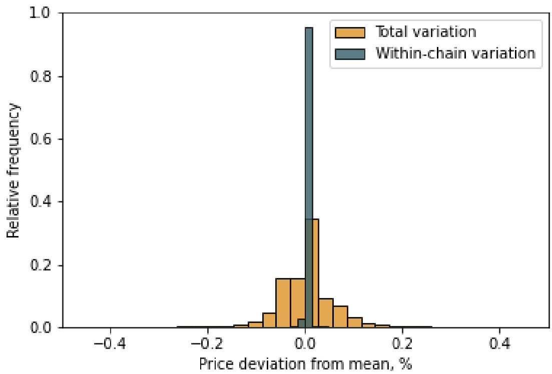


Figure 1.4: Price variation within and across chains
Note: One observation is one SKU in one store in one day

Additionally, I calculate how often product prices deviate from the mean price both within and between chains. In particular, I look at the share of observations when prices deviate from the mean by more than 1%. The results are summarized in Table 1.5. The share of non-identical prices within stores of the same chain varies across categories and on average amounts to 2.2%. On the other hand, the share of non-identical prices within all stores is 67.7% on average. While product prices within chains might differ due to store-specific sales or personal discounts, this variation remains relatively small.

Table 1.5: Share of non-identical prices within and between chains

Category	# of obs.	% non-identical prices within SKU-chain-time	% non-identical prices within SKU-time
Milk	107425	4.9	91.9
Fresh bread	81185	0.7	64.5
Beer	41188	0.8	52.3
Chocolade bars	33600	1.9	66.4
Dry bread	29109	1.0	48.4
Cheese	21944	1.1	61.6
Coffee	19046	6.0	78.4
Juice	18545	1.3	72.1
Frozen pizza	18483	0.8	47.5
Jam	15321	0.7	41.6
Frozen fish	13359	0.3	42.8
Yogurt	13327	2.1	60.9
Canned fish	8054	0.7	67.9
Eggs	3559	2.7	53.2
Total	424145	2.2	67.7

Note: One observation is price for one SKU in one store in one day. Non-identical price refers to deviation from the mean price for more than 1%.

Finally, I explore whether the potential variation in product prices within a chain responds to local market conditions. In particular, I run a regression of product-level prices p_{ijt} on market characteristics z_j, where the store j is located, while controlling for store attributes x_j and including fixed effects for the combination of chain g, product i, and day t. After accounting for chain, product, and day fixed effects, the remaining variation in product-level prices pertains to the differences between stores of the same chain. The regression looks as follows:

$$p_{ijt} = z_j\alpha + x_j\gamma + \kappa_{igt} + \epsilon_{ijt}, \quad (1.4)$$

Columns I-III in Table 1.6 show results for different specifications, which vary by the size of the market. In particular, I define a market as the area within 5, 10 or 30 km driving distance from a store. For each market definition, I calculate the market-specific income as the average income of consumers residing within that distance from a store. Additionally, I calculate a market-specific dummy variable for a store if it belongs to a chain that has no competitors within the given radius.

Regardless the size of the market, I find no evidence that prices at the product level respond to local market conditions. Moreover, more than 99% of the variation in p_{ijt} is explained by κ_{igt}. This finding provides further support to the notion that pricing decisions are predominantly made at the national level.

Following DellaVigna and Gentzkow (2019), the decision to employ uniform pricing can be attributed to several factors. While setting optimal prices for thousands of products is simply costly for a company, reputation and fairness concerns are often mentioned as an explanation for charging equal prices and seem the most plausible for the Norwegian context (Merker, 2022; Friberg, Steen and Ulsaker, 2022).

Table 1.6: Assortment choice and competition

	I	II	III	IV	V	VI	VII	VIII	IX
	SKU price			**Average store price**			**Average store variety**		
	5 km	10 km	30 km	5 km	10 km	30 km	5 km	10 km	30 km
Local monopoly (in radius)	-0.001	-0.001	-0.001	2.27***	2.30***	2.73***	-11.64***	-11.06***	-8.35***
	(0.001)	(0.001)	(0.001)	(0.189)	(0.236)	(0.566)	(0.886)	(1.12)	(2.68)
Average income, 100,000 NOK (in radius)	0.000	0.000	0.000	0.089**	0.126***	0.114	0.797***	1.02***	1.36***
	(0.000)	(0.000)	(0.000)	(0.037)	(0.048)	(0.07)	(0.174)	(0.228)	(0.331)
Location in mall	-0.002	-0.002	-0.002	-0.008	-0.095	-0.212	10.75***	11.26***	11.84***
	(0.001)	(0.001)	(0.001)	(0.22)	(0.223)	(0.229)	(1.036)	(1.06)	(1.09)
Location in city center	0.001	0.001	0.001	-0.063	-0.33*	-0.522***	3.39***	4.52***	5.31***
	(0.001)	(0.001)	(0.001)	(0.181)	(0.182)	(0.186)	(0.852)	(0.861)	(0.882)
Open on Sunday	-0.000	-0.000	-0.000	2.64***	2.65***	2.63***	-5.43***	-5.42***	-5.07***
	(0.001)	(0.001)	(0.001)	(0.186)	(0.189)	(0.193)	(0.874)	(0.894)	(0.916)
Distance to distribution center, km	-0.000	-0.000	-0.000	0.002	0.003**	0.006***	-0.051***	-0.052***	-0.061***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)	(0.009)
Const	-0.001	-0.001	-0.001	53.97***	54.00***	54.26***	39.14***	36.20***	32.75***
	(0.001)	(0.001)	(0.001)	(0.274)	(0.339)	(0.485)	(1.29)	(1.61)	(2.30)
FE	Chain-Day-SKU			Chain			Chain		
# of obs.	424145	424145	424145	1524	1524	1524	1524	1524	1524
R^2	0.99	0.99	0.99	0.47	0.45	0.42	0.61	0.59	0.56

Note: Significance levels are: *** - 1%, ** - 5%, * - 10%.

Assortment Responds to Changes in Local Market Conditions

Existing literature provides evidence that food assortment can differ among various markets. For instance, Handbury (2019) indicates that retailers tailor product offerings to income-specific preferences. Similarly, Quan and Williams (2018) find that diverse local tastes contribute to an enhanced variety of products within a market. When retailers set prices nationally, product selection can serve as a means to adapt to local market conditions.

To explore the potential variation in assortment within a chain due to local market conditions, I run a regression similar to Equation 1.4. Specifically, I estimate the following regression equation for the composite good at the store level:

$$y_j = z_j\alpha + x_j\gamma + \kappa_g + \epsilon_j, \tag{1.5}$$

where y_j denotes either price p_j or variety ν_j of the composite good, z_j represents market characteristics of store j, x_j is a vector of store attributes, and κ_g captures chain fixed effects.

The results are reported in columns IV-IX of Table 1.6. As the price of the composite good can vary only due to the assortment changes, these results indicate that assortment can differ within stores of the chain. In particular, after controlling for chain fixed effects, product selection responds to differences in local market conditions. Similar to the findings in Handbury (2019), I find that assortment decisions are correlated with income. Furthermore, product selection is influenced by store characteristics, such as location in the city center and location in a mall. Importantly, product selection is associated with the distance to the distribution center. Finally, local market power tends to play a role in product selection as well. For instance, when the chain has a local monopoly, it tends to offer a more expensive and narrower assortment.

In summary, this section provides evidence that variation in product-level prices across stores of the same chain is minimal and does not respond to changes in local market competition, indicating the presence of uniform pricing. At the same time, there is evidence that assortment can vary across markets, and that local competition might play a role in these differences. In particular, stores operating in more concentrated markets tend to offer a pricier and narrower assortment. However, determining whether these assortment differences stem from local market power or other factors, such as logistics costs, requires further investigation beyond the ad hoc price and variety measures studied earlier. The structural analysis below aims to disentangle the role of market power in choosing product offerings and quantify how this strategic product selection affects consumers residing in urban and remote areas.

1.4 Model of Spatial Demand and Assortment Choice

In this section, I develop a framework for investigating the role of local market power in assortment decisions. In particular, I specify an empirical model of consumer and firm behavior suitable for analyzing the grocery sector and the available data. In the model, spatially heterogeneous consumers choose a store to visit, taking into account store attractiveness based on its characteristics and the associated travel costs. Firms compete in the market for consumers via assortment decisions.

Demand

Before introducing the demand framework, I discuss the main features of the model and provide the reasoning behind them. Given that competition in the grocery industry is localized and market power is confined to a specific geographic area, it is important to incorporate a spatial dimension into the demand model. As consumers choose which store to visit, travel distance appears to be an important factor influencing their decisions. In this study, I use travel distance between consumers and stores to determine the relevant choice set of stores. In spatial competition, available stores and the degree of substitution depend on how consumers tradeoff factors such as travel distance and store characteristics, particularly product variety and price. To address these considerations, I leverage the flexible demand approach of Ellickson, Grieco and Khvastunov (2020). This framework allows working with overlapping markets where each consumer has her own choice set instead of isolated markets as in Zheng (2016), Handbury (2019) or Argentesi et al. (2021).

I extend the approach of Ellickson, Grieco and Khvastunov (2020) to allow for endogenous unobserved demand shifters. Although the inclusion of the unobserved store-level demand component complicates the computation, it is necessary to incorporate factors determining consumer choices that are unobserved to researchers and may also impact firms' strategic decisions. Examples of such factors may include the overall appearance or the presence of additional amenities or services within or nearby the store, such as a postal office or parking lot. By explicitly addressing these considerations, I account for the potential endogeneity problem, which in turn enables modeling firms' strategic incentives regarding optimal assortment.

Finally, to model individual consumer expenditures and map them to observed store revenues, I build on previous research on the grocery industry (Duarte, Magnolfi and Roncoroni, 2020; Eizenberg, Lach and Oren-Yiftach, 2021) and use a discrete-continuous choice demand model initially proposed by Hanemann (1984) and later adopted to the aggregate discrete choice framework by Bjornerstedt and Verboven (2016). The discrete-continuous choice model offers a more suitable framework for modeling demand in the grocery shopping context than the standard unit demand specification. It allows consumers first to decide which store to shop at and then how many units of the composite good to buy. Further details about this model are discussed later in this section.

Each consumer i residing in a location l has Cobb-Douglas preferences over $z_{i(l)}$ units of the numeraire and $q_{i(l)j}$ units of groceries. Since the actual place of residence for each consumer is not observed, the centroid of the basic unit is used as the consumer's location. Each store j offers a basket of groceries characterized by p_j and ν_j. Consumer choice generates aggregate demand $q_j(p_j, \nu_j)$, representing the total quantity of the composite good sold in a store j. I assume that the demand arises from a discrete-continuous choice model in which consumers allocate a constant budget share $\varphi_{i(l)}$ of their income $y_{i(l)}$ to grocery shopping. Then, consumers decide in which store $j \in \mathcal{J}_{i(l)}$ to purchase a continuous quantity of grocery goods $q_{i(l)j}$. As highlighted in other studies of the grocery industry (Duarte, Magnolfi and Roncoroni, 2020; Eizenberg, Lach and Oren-Yiftach, 2021), this assumption appears to be more realistic for the grocery shopping setting as opposed to a unit-good assumption.

The conditional direct utility function when choosing store j is defined as:

$$u_{i(l)j} = (1 - \varphi_{i(l)}) \ln z_{i(l)} + \varphi_{i(l)} \ln q_{i(l)j} + \varphi_{i(l)} \ln \psi_{i(l)j}, \tag{1.6}$$

where $\psi_{i(l)j}$ is the parameter that governs the preferences of consumer i for store j and specified as:

$$\psi_{i(l)j} = e^{\frac{\theta_j + \rho d_{lj} + \epsilon_{i(l)j}}{\alpha}}. \tag{1.7}$$

Here, θ_j represents the utility from store characteristics other than price, d_{lj} denotes the distance between location l and store j, $\epsilon_{i(l)j}$ accounts for the consumer-store specific shock with a type-I extreme value distribution, and α governs the relative importance of the utility from chosen alternative j and the utility from numeraire.

Then maximization of the conditional direct utility under a budget constraint $p_j q_{i(l)j} + z_i = y_{i(l)}$ gives demand functions:

$$q_{i(l)j}(p_j, y_{i(l)}) = \frac{\varphi_{i(l)} y_{i(l)}}{p_j}, \quad z(p_j, y_{i(l)}) = (1 - \varphi_{i(l)}) y_{i(l)}. \tag{1.8}$$

When substituting demand functions into the direct utility function, I derive the indirect utility function:

$$v_{i(l)j} = \frac{\alpha}{\varphi_{i(l)}} \ln y_{i(l)} - \alpha \ln p_j + \theta_j + \rho d_{lj} + \epsilon_{i(l)j}, \tag{1.9}$$

with θ_j being a linear function of variety ν_j, a vector of observed store characteristics $\mathbf{x}_j$ and an unobserved component of a store's utility ξ_j that captures factors that are not directly accounted for by the observed characteristics of the store.

Finally, I define mean utility δ_j is a linear function of price p_j, variety ν_j, a vector of observed store characteristics $\mathbf{x}_j$ and an unobserved component ξ_j:

$$\delta_j = -\alpha \ln p_j + \theta_j = -\alpha \ln p_j + \gamma \nu_j + \mathbf{x}_j \beta + \xi_j. \tag{1.10}$$

Inclusion of the structural error ξ_j into the indirect utility function extends the spatial demand approach proposed by Ellickson, Grieco and Khvastunov (2020). This extension allows me to address the endogeneity issue that arises when retailers strategically choose certain characteristics, such as, in this case, the price and variety of assortment, that enter the utility function. Introducing the structural error makes the estimation process computationally demanding due to the need to solve for ξ_j to evaluate the estimation objective function. However, this extension allows me to account for the strategic decision-making of retailers and obtain consistent estimates of the model parameters.

To complete the specification of the demand system, I incorporate an outside option to account for the possibility that some consumers may choose to spend a portion of their grocery budget outside of the observed stores:

$$u_{i(l)0} = \frac{\alpha}{\varphi_{i(l)}} \ln y_{i(l)} + \xi_0 + \epsilon_{i(l)0}, \tag{1.11}$$

where $\epsilon_{i(l)0}$ is a zero-mean individual store specific shock. The term ξ_0 is normalized to zero.

Finally, the probability that a consumer residing in location l decides to buy groceries from store j takes the usual logit form:

$$\mathbb{P}_{lj}(p_., \nu_., \xi_., d_{l.}; \theta_d) = \frac{\exp(\delta_j(p_j, \nu_j, \xi_j; \theta_d) + \rho d_{lj})}{1 + \sum_{k \in \mathcal{J}} \exp(\delta_k(p_k, \nu_k, \xi_k; \theta_d) + \rho d_{lk})}. \tag{1.12}$$

The constant expenditure model assumes that a consumer's grocery budget is defined as a constant share of their income. Thus, the total grocery budget of location l is denoted as B_l and defined as:

$$B_l = \int \varphi_{i(l)} y_{i(l)} dF(\varphi, y), \tag{1.13}$$

where $y_{i(l)}$ represents the consumer's income and $\varphi_{i(l)}$ denotes the fraction of income that the consumer allocates to grocery spending.

Since individual data on grocery expenditure is unavailable, I approximate B_l as the weighted average over the distribution of consumer types in each location defined by income y_l and the proportion of individual budgets spent on groceries φ_l:

$$B_l \approx \varphi_l \cdot y_l \cdot N_l. \tag{1.14}$$

Note that data on y_l and N_l are immediately available from the demographics data. Meanwhile, the value for parameter φ_l I infer from the Survey of Consumer Expenditures published by Statistics Norway.[6] The survey provides information on the percentage of household income allocated to food expenditures across various income deciles. Since these food expenditures do not include restaurant spending, they serve as a suitable proxy for grocery expenses. Then, I assign each basic unit to an income decile based on its average income and utilize the corresponding φ_l value associated with that decile. By incorporating this information, I can account for the variations in consumer behavior and expenditure patterns across different income levels without estimating φ_l.

Estimating φ_l would require an additional structural error at the basic unit level and an additional set of moment conditions. However, an unobserved component driving grocery expenditures in a specific location would contradict the assumption that con-

[6] `https://www.ssb.no/statbank/table/10444/`

sumers spend a constant fraction of their income on groceries. In the constant expenditure specification, consumers can have different grocery expenditures across basic units, but these differences should be explained either by differences in income y_l or the fraction of income allocated to grocery spending φ_l. As a result, the model does not incorporate an unobserved component in grocery expenditures, ensuring that the assumption of constant expenditure holds.

As data on grocery expenditure flows between basic units and stores are not available, I aggregate over the model-implied individual choices to connect basic unit-level consumer demographics to store-level market shares. Next, I describe the steps required to transition from individual choices to observed store-level market shares.

Equation 1.12 allows me to predict store choice probabilities for a consumer residing in location l for each store in her choice set. Then the grocery expenditure flow between store j and location l is computed as the total grocery budget of location l multiplied by the probability of visiting store j:

$$\hat{R}_{lj}(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d) = B_l \cdot \mathbb{P}_{lj}(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d). \tag{1.15}$$

To connect the observed store market shares and the grocery expenditure flows between locations and stores, I aggregate the grocery expenditure flows $\hat{R}_{jl}$ over locations to formulate the revenue of each store as a function of model parameters:

$$\hat{R}_j(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d) = \sum_{l \in L_j} \hat{R}_{lj}(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d), \tag{1.16}$$

where L_j is a group of locations that could potentially visit store

j. Then, dividing store revenue by the total grocery budget of locations L_j, I obtain a store-level market share:

$$\hat{s}_j(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d) = \frac{\hat{R}_j(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d)}{\sum_{l \in L} B_l}. \tag{1.17}$$

I assume that the consumers' choice set consists of all stores within a 30 km radius from the centroid of the basic unit and the outside option. Since the demand model has an explicit disutility of distance, which should account for consumers' preferences to shop in closer stores, the choice of a particular radius is not critical here. Rather, it has to be no less than how consumers are willing to travel.

Finally, I solve the implicit system of equations with respect to $\xi_{\cdot}$:

$$s_j = \hat{s}_j(p_{\cdot}, \nu_{\cdot}, \xi_{\cdot}, d_{l\cdot}; \theta_d). \tag{1.18}$$

Note that the current specification of the model does not account for unobserved heterogeneity beyond standard logit error. While theoretically, it is possible to incorporate random coefficients into the model to address this limitation, the practical implementation becomes computationally burdensome due to the large number of locations involved (more than 13 thousand) and numerous stores.

Supply

The entire decision-making process of a retailer can be seen as a two-stage game. In the first stage, multi-store retailers set product-level prices at the national level. Then, in the second stage, they select the assortment for each store, taking product-level prices as given. The supply model in this study focuses on the second stage of this decision process.[7]

Considering the large number of products typically offered by retailers, explicitly modeling each product choice would be computationally complex. Therefore, the problem is simplified to focusing on the two strategic variables: price level of assortment p_j and assortment breadth ν_j. The marginal cost of a store j of providing a bundle of goods characterized by p_j and ν_j is defined as:

$$mc_j = mc(\nu_j, \boldsymbol{\omega}_j; \theta_s), \tag{1.19}$$

where $\boldsymbol{\omega}_j$ denotes a vector of cost shifters, θ_s is a vector of supply-side cost function parameters. Note that in the given specification, I assume that the marginal costs do not change with the quantity of the composite good consumed, indicating no economies of scale. However, I allow the marginal costs to vary with the assortment breadth ν_j to make providing more items on a shelf costly.

Then the multi-store firm's maximization problem can be represented as follows:

$$\max_{\{p_j, \nu_j\}_{j \in \mathfrak{J}_f}} \sum_{j \in \mathfrak{J}_f} q_j(p_., \nu_., \xi_., d_{.j}; \theta_d)(p_j - mc(\nu_j, \boldsymbol{\omega}_j; \theta_s)), \tag{1.20}$$

where $\mathfrak{J}_f$ is a set of stores belonging to chain f and q_j denotes the demand for store j aggregated over locations, measured in units of

[7]It is important to note common ownership among some retail chains. Some retail chains are part of a retail group with access to the same producers and shared distribution channels. However, despite this joint ownership, each chain negotiates different purchase prices. Moreover, each chain has its own management and operates independently, treating other group chains as competitors rather than as own-firm stores. Therefore, in the supply model, each chain maximizes its profit independently from other chains within the retail group.

the composite good and calculated as follows:

$$q_j = \sum_{l \in L} \frac{\hat{R}_{lj}}{p_j}, \tag{1.21}$$

with $\hat{R}_{lj}$ being the revenue of store j generated by consumers of location l defined in Equation 1.15.

The first-order conditions for profit-maximizing firm over price and variety are:

$$F.O.C.[p_j] : q_j + \sum_{r \in \mathfrak{J}_f} (p_r - mc_r) \frac{\partial q_r}{\partial p_j} = 0, \tag{1.22}$$

$$F.O.C.[v_j] : -\frac{\partial mc_j}{\partial v_j} q_j + \sum_{r \in \mathfrak{J}_f} (p_r - mc_r) \frac{\partial q_r}{\partial v_j} = 0. \tag{1.23}$$

Firms engage in Bertrand competition simultaneously choosing price and variety of the composite good.

1.5 Identification and Estimation

In this section, I describe the identification and estimation of demand and supply-side parameters. Estimating demand-side parameters can be problematic due to the endogeneity issue, which is here related to price and variety measures of assortment. Since demand-side shocks realize before the decision on assortment is made, price and variety can be correlated with unobserved demand shocks. Therefore, instruments are needed to account for the endogeneity issue. To estimate the structural parameters governing consumer preferences $\{\alpha, \gamma, \beta, \rho\}$, I employ the two-step approach developed in Berry (1994) and incorporate the observed spatial consumer heterogeneity similar to Davis (2006). By solving the supply-side first-order conditions for a particular set of demand-side parameters, I can estimate $\widehat{mc}_j$ and $\partial \widehat{mc}_j / \partial v_j$. Finally, I estimate the supply-side parameters θ_s. Similarly to the demand model, supply-side shocks can potentially correlate with cost-shifters. Therefore, I need to account for potential endogeneity issue in the supply model by employing instrumental variables and using the GMM procedure for estimation. The rest of this section provides details of the estimation procedure.

Demand

To estimate demand-side parameters $\theta_d = \{\alpha, \gamma, \beta, \rho\}$, I begin by selecting an initial value for ρ. Then, I iteratively update the store's mean utility vector, δ, until it converges, using a similar process to the BLP inner loop. In particular, I use the fixed point iterator for the random vector of starting values of δ and iterate the expression: $\delta_j' = \delta_j + \ln(s_j) - \ln(\hat{s}_j(\delta_., \rho))$, where $\hat{s}_j(\delta_., \rho)$ is calculated according to Equation 1.17. I update the vector of δ until the difference between two consecutive iterations falls below a predetermined tolerance level.

Once the vector δ is obtained, the parameters $\{\alpha, \gamma, \beta\}$ governing preferences for price and variety of the composite good and other observed store characteristics can be identified. Here, I assume that not only price but also variety can correlate with the unobserved store quality. Therefore, I use differentiation instruments proposed by Gandhi and Houde (2019) to address this endogeneity issue.

Differentiation instruments are variants of the common BLP instruments and represent differences between own and rival store characteristics. The basic idea is to use each product's exogenous degree of differentiation — in this case, each store in a market — as instruments for price and variety. In particular, for a continuous characteristic, the difference for a pair of stores (j,k) is constructed as $\tilde{x}_{jk} = x_j - x_k$. For each store j, I aggregate these differences across competing stores in a 2 km and 5 km radius. Then under the assumption $\mathbb{E}[\xi_j|Z_j^d] = 0$, parameters $\{\alpha, \gamma, \beta\}$ are identified, where Z_j is a vector of instruments and ξ_j is obtained as:

$$\xi_j(\delta, \theta_d) = \delta_j(\rho) + \alpha \ln p_j - \gamma \nu_j - x_j \beta. \tag{1.24}$$

Assortment information is derived from the receipt data available only for one retail group. To address this, I define a missing indicator d_j that equals one if store j has information about price and variety and zero otherwise similar to Duarte, Magnolfi and Roncoroni (2020). Then, the model is identified under the assumption $\mathbb{E}[\xi_j|Z_j^d, d_j] = \mathbb{E}[\xi_j|Z_j^d] = 0$. This assumption implies that stores with available data are not more or less attractive to consumers than other stores with similar characteristics. This is a plausible assumption as the retail group that provides the data has stores of all types across the country, making it representative of the broader population of stores.

In the last step, I recover the distance cost parameter ρ. Since store location is simply a product characteristic, the estimates will suffer from the standard endogeneity problem if retailers choose it strategically. If, for instance, stores with high ξ_j are located closer to densely populated areas, such that the average travel distance is low, then $\mathbb{E}[d_j \xi_j] < 0$. To correct for this source of endogeneity, one needs to find an instrument that is correlated with the store location or distance to competitors and is not influenced by the store's unobserved factor. For this purpose, I use the average distance to consumers weighted by population for neighboring stores. I define neighboring stores as those within a 1-km radius that can be perceived as immediate competitors. Then under the assumption $\mathbb{E}[\xi_j|Z_j^d] = 0$, parameter ρ is identified.

These steps describe one iteration of the outer loop, and the procedure is repeated with the updated value of ρ until convergence is achieved.

Supply

Following the approach of Crawford, Shcherbakov and Shum (2019), I specify a function for marginal costs:

$$mc_j = \exp(c_{0j} + c_1 \nu_j). \tag{1.25}$$

The exponential functional form is chosen to reflect the nature of the retail industry, where store capacity is limited. In the context of limited capacity, the cost per unit of the composite good is expected to be convex. As the assortment breadth increases, the additional cost incurred for providing more items on the shelf becomes progressively higher. By incorporating this convexity in the marginal cost function, the model accounts for the cost implications of expanding the assortment.

Finally, I allow the marginal costs to depend on observed and unobserved cost shifters. In particular, I specify the coefficient c_0 as a linear function of cost shifters $\boldsymbol{\omega}_j$ and a structural error ζ_j:

$$c_{0j} = \boldsymbol{\omega}_j \theta_s + \zeta_j. \tag{1.26}$$

The vector $\boldsymbol{\omega}_j$ includes characteristics that can potentially affect the costs of running a store, such as the number of employees and whether the store is located within a mall. Marginal costs are allowed to depend on the retail group of a store, as different retail groups might have different input prices. The retail group also determines the distance of a store to a distribution center, which is relevant in counterfactual experiments where the market structure can change.

One also needs to control the assortment's quality in the marginal costs as, for example, better products tend to have higher input prices. Since direct data on assortment quality is unavailable, I infer the assortment quality from the unobserved component of the demand model ξ_j.

It is worth noting that the unobserved component ξ_j may capture not only assortment-related characteristics but also other factors that make consumers more likely to choose a particular store, such as unobserved store amenities. I recognize that ξ_j serves more as a proxy and might not perfectly capture the true quality of the assortment. However, despite the potential noise in ξ_j, it remains important to account for assortment quality when modeling the cost of operating a store.

Equation 1.22 solely allows to back out the marginal costs mc_j. Having a functional form for mc_j in Equation 1.25 and first-order conditions for variety v_j in Equation 1.23, I can obtain estimates for $\partial\widehat{mc}_j/\partial v_j$, which are used to compute c_{0j} and c_{1j}:

$$\hat{c}_{0j} = \ln(\widehat{mc}_j) - \frac{\partial\widehat{mc}_j/\partial v_j}{\widehat{mc}_j} v_j, \tag{1.27}$$

$$\hat{c}_1 = \frac{\partial\widehat{mc}_j/\partial v_j}{\widehat{mc}_j}. \tag{1.28}$$

I estimate the vector of supply-side parameters θ_s using GMM, which accounts for the fact that the unobserved store characteristics ξ_j included in $\boldsymbol{\omega}_j$ might be correlated with the unobserved cost component ζ_j. I employ BLP instruments constructed based on ξ_j's of neighboring stores belonging to the same chain, having the same format, or being part of the same retail group. Then, the identification of parameters relies on a GMM procedure where equations 1.27-1.28 serve as constraints for the minimization problem.

1.6 Estimation Results

In this section, I present the estimation results of the demand model. Based on the demand estimates, I compute the market concentration for each consumer location. Additionally, I leverage the demand estimates to calculate the Average Assortment Consumed (AAC) for each consumer location. This allows me to explore the relationship between assortment differences and variations in market concentration.

Next, I discuss the findings of the supply model. Specifically, the model provides estimates of marginal costs and markups for each store. Moreover, I show the spatial distribution of markups across the country, providing insights into how different areas are affected by the assortment strategies of grocery retailers.

Demand

Table 1.7 summarizes results for the spatial demand model. Both the price and variety coefficients have the expected sign and are statistically significant. As expected, consumers are averse to traveling long distances to stores, reflecting the costliness and inconvenience associated with longer travel. Consumers show a strong preference for supermarkets over discounters and favor stores located in shopping malls.

Table 1.7: Demand parameters estimates

Variable	**Estimate**
Log price	-4.612***
	(1.302)
Variety	0.171*
	(0.008)
Distance	-0.235***
	(0.000)
Supermarket	3.782***
	(0.000)
Number of employees	0.154***
	(0.000)
Mall	11.57***
	(0.000)
Open on Sunday	39.75***
	(0.000)
# of obs.	3718

Note: Significance levels are: *** - 1%, ** - 5%, * - 10%.

Localized Concentration and Assortment Measures

The empirical framework of the demand model allows to calculate localized concentration measures. Typically, concentration measures require a predetermined market definition, which has often played a decisive role in antitrust cases. The spatial model employed in this study overcomes this limitation by defining markets based on consumers and their choice sets rather than the geographic locations of stores. This approach measures concentration at a localized level, providing a more accurate representation of local market power.

Based on the demand model, I predict the probability that a consumer residing in location l visits store j $\mathbb{P}_{lj}$, which is not observed in the data and can be recovered only from the model. Then, I use $\mathbb{P}_{lj}$ to calculate HHI for each location. The distribution of these localized concentration measures across basic units is illustrated in Figure 1.5. The analysis reveals that most markets in Norway are moderately concentrated (56%), 41% are highly concentrated, and only 3% are considered competitive. Figure 1.6 shows the spatial distribution of market concentration for Vestland, a region in Norway. The key finding is that the area around Bergen is predominantly competitive, with a lower concentration level. However, as we move away from Bergen towards more rural areas, the level of concentration gradually increases.

In Table 1.8, I compare the classification of basic units based on the HHI calculated using a predefined market definition, which in this case is the municipality, and based on localized HHI. While the overall composition of markets remains almost the same, there are changes in the level of competition when considering local competition at the basic unit level instead of aggregating them to municipalities. For example, more than half of the competitive markets are estimated to be moderately or highly concentrated. Similarly, some markets that were initially attributed to highly concentrated municipalities have access to more competitive markets when not imposing strict geographical boundaries on the market definition.

Table 1.8: Market concentration comparison

		Localized HHI			
		Competitive	Moderately Concentrated	Highly Concentrated	**Total**
Municipality based HHI	Competitive	331	283	85	699 (5.2%)
	Moderately concentrated	69	5405	1466	6940 (51.5%)
	Highly concentrated	29	1857	3950	5836 (43.3%)
	Total	430 (3.2%)	7552 (56.0%)	5502 (40.8%)	

Note: One observation is one basic unit.

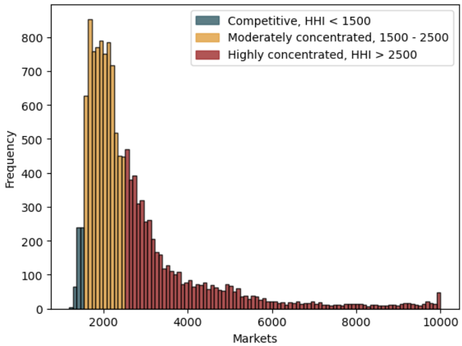


Figure 1.5: Distribution of localized concentration measures

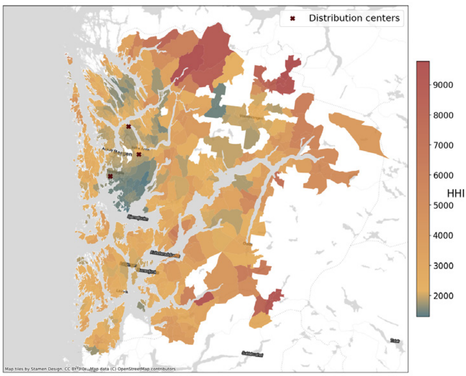
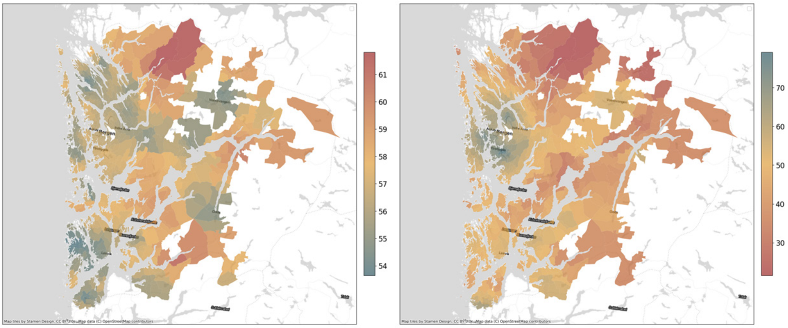


Figure 1.6: Spatial distribution of market concentration

Additionally, the estimated demand model allows revisiting assortment inequality across different regions. As before, the demand model allows me to compute the probability that a resident of location l visits store j, $\mathbb{P}_{jl}$. Then, I can calculate the Average Assortment Consumed for each location l in terms of price (AAC_l^p) and variety (AAC_l^v). Specifically, AAC_l^p is calculated as an average price of stores j in the choice set $\mathcal{J}_l$, weighted by the probabilities $\mathbb{P}_{jl}$: $AAC_l^p = \sum_{j \in \mathcal{J}_l} \mathbb{P}_{jl} \cdot p_j$. Similarly, AAC_l^v is obtained as an average variety of stores weighted by $\mathbb{P}_{jl}$: $AAC_l^v = \sum_{j \in \mathcal{J}_l} \mathbb{P}_{jl} \cdot v_j$. Therefore, both AAC_l^p and AAC_l^v represent weighted averages that take into account the shopping behavior of consumers. Figure 1.7 illustrates assortment differences across locations. The primary finding is that residents of urban areas, such as Bergen, have access to a more affordable assortment with a greater variety. At the same time, residents of rural areas have a limited assortment and lack access to cheap products. These results, along with the localized concentration measures, demonstrate that consumers residing in concentrated markets face higher prices and a narrower range of choices.

Lastly, I explore the relationship between the basic unit market concentration and the average assortment consumed in the basic units. As illustrated in Figure 1.8, the relationship between the HHI and AAC_l^p is not strictly monotone. However, one can notice that more concentrated markets have more expensive assortment (the correlation between HHI and AAC_l^p is 0.12). Conversely, the plot shows a negative monotonic relationship for variety: consumers in competitive markets enjoy a higher variety of products (correlation between HHI and AAC_l^v is -0.33).



(a) Price (b) Variety

Figure 1.7: Average assortment consumed

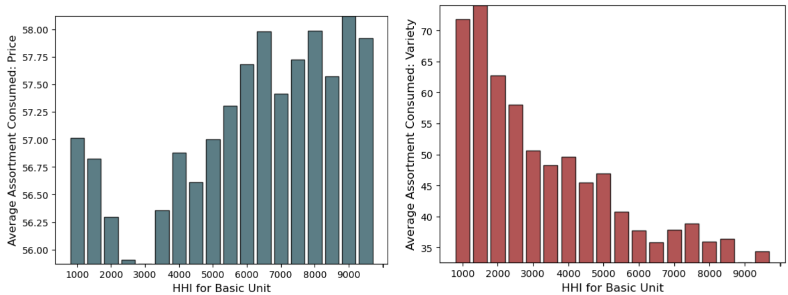


Figure 1.8: Average assortment consumed and market concentration

Supply

The descriptive statistics of the marginal costs and markups are reported in Table 1.9. Figure 1.9 shows the distribution of marginal costs across formats. As a format providing higher quality and variety, supermarkets have higher marginal costs on average. In contrast, discounters have the lowest marginal costs. As for markups, there is no noticeable difference between stores of different formats. The estimates of markups are similar to what other studies obtained when dealing with a composite good (Duarte, Magnolfi and Roncoroni, 2020; Eizenberg, Lach and Oren-Yiftach, 2021).

Table 1.9: Summary statistics for costs and margins

	Price	MC	Markup
Mean (all)	56.47	44.54	0.21
Median (all)	55.75	43.95	0.19
By formats			
Median (discounter)	54.15	42.74	0.20
Median (convenience)	58.73	47.02	0.19
Median (supermarket)	60.67	48.66	0.19

Note: Markups are calculated at the store level. Officially reported markups are typically 2-4% and include management and other fixed costs of running a retail group.

Table 1.10 reports the marginal cost function estimates. As expected, providing higher variety and quality is costly for a retailer. Other estimates of the supply-side function also have expected signs. The further the distance to the distribution center, the more expensive it is to transport goods. It is more costly to have a store in a shopping mall. Stores open on Sundays have higher marginal costs, as by Norwegian legislation, they must pay higher taxes. Supermarkets have higher marginal costs than discounters and convenience stores as they usually have more em-

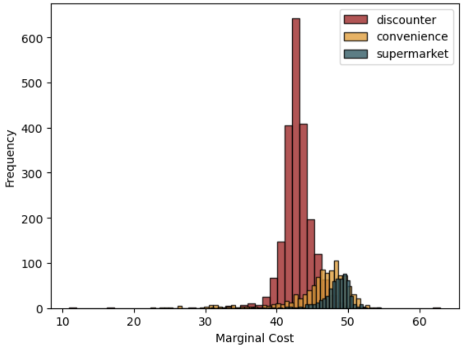


Figure 1.9: Distribution of marginal costs across formats

ployees. Larger retail groups have lower marginal costs, which lower input prices and economies of scale could explain. The negative effect of store size and the number of employees could also be attributed to economies of scale.

Once the marginal costs are estimated, it is possible to calculate the profit of each store. The demand model allows for a more detailed analysis and allows to calculate the contribution of each location to each store's profit. Then summing over stores, one can calculate the total profit of grocery stores generated by consumers of location l:

$$\Pi_l = \sum_{j \in \mathcal{J}_l} (p_j - mc_j) \cdot q_{jl}, \tag{1.29}$$

where q_{jl} represents the number of composite goods purchased by consumers of location l in store j and is defined as:

$$q_{lj} = \frac{\mathbb{P}_{lj} B_l}{p_j}. \tag{1.30}$$

Table 1.10: Marginal Cost Function Parameters

Variable	**Estimate**
Const (c_0)	3.645
	(0.137)
Variety (c_1)	0.037
	-
Other observed cost shifters	
Quality of assortment	0.029^{***}
	(0.004)
Supermarket	0.291^{***}
	(0.039)
Number of employees	-0.019^{***}
	(0.001)
Mall	0.276^{***}
	(0.056)
Liquor store	-3.465^{***}
	(0.429)
Open hours	0.008
	(0.006)
Sunday	1.278^{***}
	(0.145)
Costs of toll roads to dist.center	0.002^{**}
	(0.001)
Store size	-0.626^{***}
	(0.026)
Retail group A	0.378***
	(0.043)
Retail group B	-0.029
	(0.025)
Retail group C	0.159***
	(0.032)
# of obs.	3639

Note: Retail group D is taken as a base category. Significance levels are: *** - 1%, ** - 5%, * - 10%.

Figure 1.10 displays the spatial distribution of profit Π_l scaled by the number of consumers in location l. The plot suggests that the per capita profits are higher in less densely inhabited areas and lower in large cities. Finally, I examine how profit per capita is related to market concentration. As shown in Figure 1.11, it is evident that more concentrated markets are charged higher profits per capita.

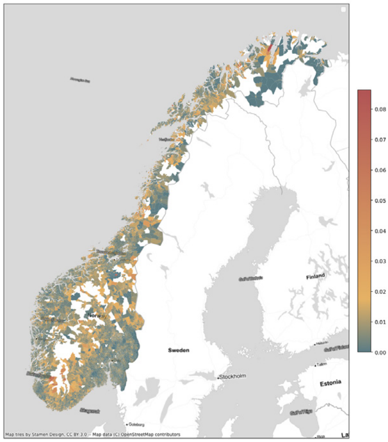


Figure 1.10: Spatial distribution of profit per person

1.7 Counterfactual Analysis

The counterfactual analysis begins by summarizing the results concerning assortment inequality. Then, I examine the role of local assortment in generating welfare inequality and consider policies that could improve assortment, such as reducing consumer travel costs and providing cost subsidies to retailers in remote areas.

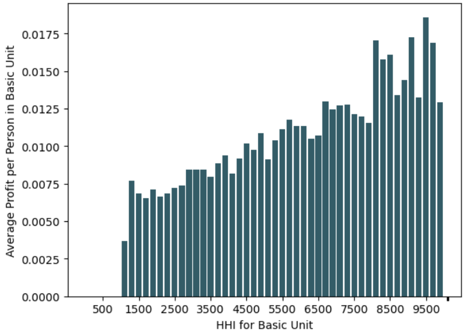


Figure 1.11: Profit per person and market concentration in basic unit

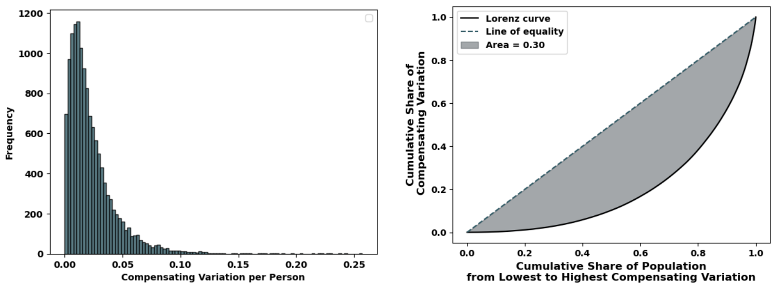
Assortment Inequality

In the spatial demand model, Figure 1.7 sheds light on the assortment inequality across different locations. It indicates that consumers in concentrated areas face limited and more expensive product variety. Figure 1.10 further emphasizes assortment inequality by illustrating that firms charge higher margins in less populated areas even after controlling for logistics costs. These findings suggest that assortment choice could serve as a strategic channel for firms to maximize their profits.

Further, I use a compensating variation metric to compare consumer welfare across different locations. To measure consumer welfare in the benchmark equilibrium, I calculate the compensating variation between the benchmark equilibrium and an alternative environment where only the outside option is available. Following the approach by Atal, Cuesta and Sæthre (2022), I define compensating variation for consumer i residing in location l as:

$$\max_{j} u\left(y_i, \delta_j, d_{lj}, \epsilon_{i(l)j}\right) = \max_{j'} u\left(y_i - CV_i, \delta_{j'}, d_{lj'}, \epsilon_{i(l)j'}\right). \quad (1.31)$$

Figure 1.12a displays the distribution of consumer welfare per person across basic units. To quantify the extent of assortment inequality, I employ the Gini index, computed based on consumer welfare. Figure 1.12b presents the Lorenz curve for the consumer welfare per person, where the cumulative share of the population is plotted against the cumulative share of consumer welfare. The calculated Gini index of 0.3 quantitatively measures assortment inequality and serves as a basis for comparing the benchmark equilibrium with equilibria in counterfactual policies.



(a) Consumer welfare (b) Lorenz curve

Figure 1.12: Inequality in consumer welfare across locations

Note: In the left panel, one observation corresponds to compensating variation for one person in a basic unit measured in MNOK.

Counterfactual Policies

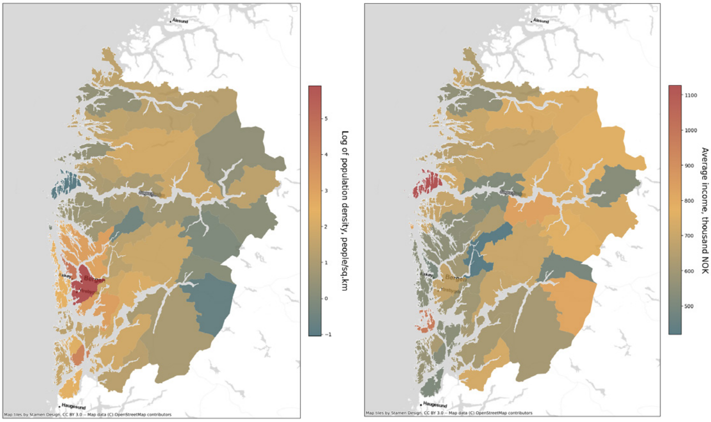
For illustrative purposes, the counterfactual analysis focuses on the Vestland region with the center in Bergen. Vestland is a relatively isolated market, and Bergen serves as a central hub for various retail chains, as evidenced by the presence of their distribution centers on the outskirts of the city. As the distance from Bergen increases, the costs associated with logistics for serving stores in remote areas also rise. Regarding consumer distribution, Bergen is classified as an urban and densely populated area, with a population density of 650.2 people per square kilometer as of 2023. Conversely, there are rural neighborhoods in Vestland where the population density can be as low as 0.69 people per square kilometer. Figure 1.13a illustrates the population density of Vestland.

Additionally, Vestland has relatively low income inequality, measured in average income across basic units, similar to the overall trend in Norway. Figure 1.13b shows the spatial distribution of income across municipalities in Vestland, with most municipalities having similar income levels. Thus, Vestland presents a relevant setting for studying assortment decisions across different markets.

Welfare Analysis of Local Assortment. To quantify the welfare effects of the local assortment, I compare the observed assortment with a counterfactual scenario where chains adopt a unified assortment strategy, offering the same bundle of groceries across all their stores. Then the maximization problem for a multi-store firm f looks as follows:

$$\max_{p_f, \nu_f} \sum_{j \in \mathfrak{J}_f} q_j(p_\cdot, \nu_\cdot, \xi_\cdot, d_{\cdot j})(p_f - mc(\nu_f, \boldsymbol{\omega}_j; \theta_s)). \tag{1.32}$$

Using the first-order conditions for the problem 1.32, I calcu-

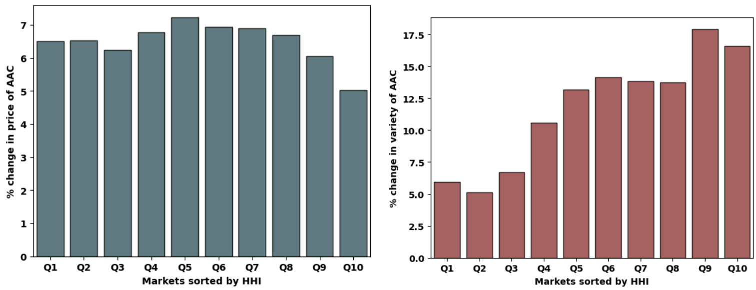


(a) Population density (b) Average income

Figure 1.13: Vestland

late each firm's new equilibrium price and variety of the composite good. Under uniform assortment, stores offer a wider range of products, resulting in an 11.1% increase in variety. However, this also leads to an average 5.5% increase in the price of goods. Consumers' shopping behavior reflects similar changes. The average assortment consumed (AAC) experiences a 6.4% increase in price and a 11.6% increase in variety, taking into account changes in both price and variety as well as the probability of visiting stores.

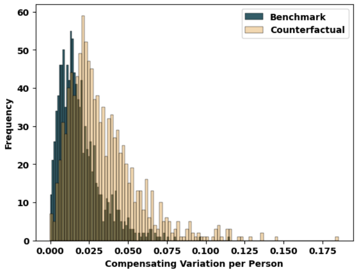
To further understand the welfare implications, I explore how the uniform assortment policy affects markets with different market concentration. Figure 1.14 provides a summary of the results, with basic units sorted by the baseline HHI. Across all markets, there is a rise in both the price and variety of AAC. However, markets with higher concentration experience a smaller increase in price and a more significant increase in variety compared to competitive markets. This result indicates that in the benchmark equilibrium, retailers offer limited and pricier assortment in concentrated markets.



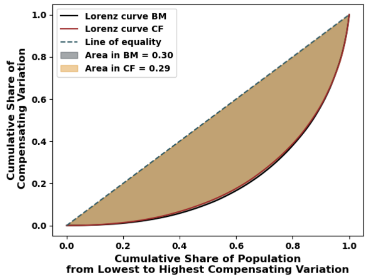
(a) Effect on price (b) Effect on variety

Figure 1.14: Average assortment consumed and market concentration

To measure consumer welfare, I use compensating variation between the counterfactual scenario and the benchmark equilibrium. As anticipated, the uniform assortment positively affects consumers, resulting in a remarkable increase in total consumer welfare, amounting to 7756 MNOK. The impact of the policy intervention on the distribution of consumer welfare per person is illustrated in Figure 1.15a. Additionally, Figure 1.15b illustrates that while the policy benefits consumers, it does not significantly reduce consumer inequality. Although grocery chains offer an equal assortment across stores, the policy does not address the limited availability of stores in remote markets. Consequently, consumers in these areas continue to face a limited choice of stores and higher transportation costs compared to residents of urban areas. This highlights that different interventions would be necessary to address the disparities in consumer welfare across locations.



(a) Consumer welfare



(b) Lorenz curve

Figure 1.15: Change in consumer welfare due to uniform assortment

The implementation of the uniform assortment policy has a detrimental effect on firms. The industry's total profit declines significantly by 8417 MNOK, and a substantial portion of stores, 28%, experience negative profits in the counterfactual equilibrium. This indicates that the policy adversely affects the profitability and viability of some retail outlets.

While consumers benefit from the uniform assortment in the short run, the overall impact on welfare is negative, with a reduction of 660 MNOK, representing a decrease of 4.5%. The decline in profits and the risk of stores becoming unprofitable could lead to store closures in the long run, which would further exacerbate market concentration. With fewer active stores, consumers in certain regions may face even more limited options and potentially higher prices, ultimately deepening disparities in consumer welfare among different regions. This reinforces the need for a more nuanced approach to tackle assortment inequality.

Reducing travel disutility. In the previous counterfactual experiment, despite grocery chains providing an equal assortment, consumers in remote areas still have to travel farther than those in urban areas. In this counterfactual policy, I address disparities in travel disutility across different regions. The counterfactual policy aims to improve the accessibility and availability of stores for residents of remote areas, which could positively affect consumer welfare. In particular, I investigate the effects of halving the distance disutility for markets that lack stores within a 3 km radius. In reality, this policy could be implemented by reimbursing fuel or electricity costs or reducing public transportation fees for individuals living in remote regions.

First, I examine how the reduction in travel disutility affects market concentration. Table 1.11 summarizes changes in market concentration at the basic unit level. Notably, the number of highly concentrated markets decreases by approximately ten percentage points, while the count of moderately concentrated and competitive markets increases by eight and three percentage points, respectively. These findings indicate that reducing travel disutility fosters competition among retailers.

Table 1.11: Change in Market Concentration

		HHI Counterfactual			
		Competitive	Moderately Concentrated	Highly Concentrated	**Total**
HHI	Competitive	33	0	0	33 (3.1%)
	Moderately concentrated	27	668	2	697 (65.1%)
	Highly concentrated	4	115	222	341 (31.6%)
	Total	64 (6.0%)	783 (73.1%)	224 (20.9%)	

Note: One observation is one basic unit.

As a result, the price change varies from -9.3% to 1.3% across stores, with an average decrease of 0.14%. The variety change varies from -0.83% to 4.3% with an average increase of 0.06%. The reduction in travel costs leads to increased competition in most markets, leading to downward pressure on prices and upward pressure on variety.

However, contrary to standard economic intuition, some stores change prices and variety in the opposite direction. This results from a change in demand composition. As travel costs decrease, consumers who continue shopping in expensive stores are those for whom reduced travel costs offer little benefit. Even though traveling becomes less costly, their choice set does not expand.

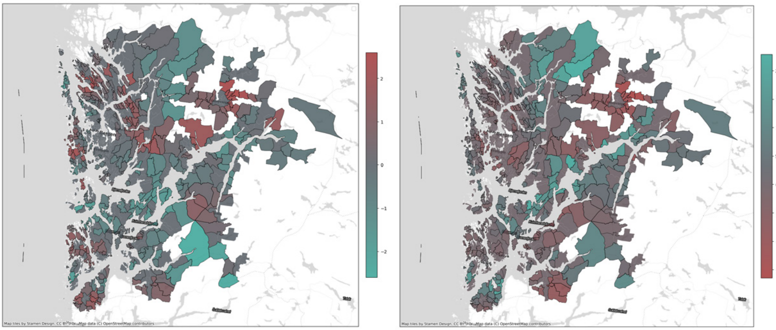
To explore this idea, I compare each store's average choice-weighted traveled distance between the benchmark equilibrium and the counterfactual scenario. To compute the average choice-weighted traveled distance, I aggregate the distances traveled from different markets to the store weighted by the choice probabilities derived from the demand model and the share of consumers from each market. The negative correlation of -0.3 confirms the intuition that stores experiencing an increase in prices are those for which the catchment area decreases in the counterfactual scenario. Moreover, as a result of the policy, expenditures by a representative consumer in grocery stores increase as they obtain compensation of transportation costs. Therefore, in these markets, the retailers encounter a less elastic demand with higher grocery budgets, leading them to raise prices and reduce variety.

Additionally, I investigate how the average choice-weighted HHI at the store level changes as a result of the policy intervention. The average choice-weighted HHI is computed by aggregating HHIs weighted by the share of consumers from each market across locations in the store catchment area. The positive correlation of 0.55 indicates that stores that raise prices in the counterfactual experience an increase in the average weighted HHI. This suggests these stores now cater to consumers from more concentrated markets with limited choices. This further reinforces the observation that, supermarkets face less elastic consumers with higher grocery budgets in these markets, leading them to raise prices and reduce variety. This creates a counterbalancing effect that reduces, and sometimes even neutralizes, the competitive pressure exerted on price and variety.

To explore the changes in consumers' shopping behavior, I calculate changes in Average Assortment Consumed, the weighted average of price and variety consumed by residents of each basic unit, taking into account the probability of shopping in each particular store. The change in the price of AAC varies from -2.6% to 2.6% with an average increase of 0.2%. The change in the variety of AAC varies to a greater extent, from -16.5% and 22.9% with an average increase of 1.3%. Figure 1.16 visually presents the changes in AAC across different basic units in Vestland. The green-colored areas receive a better assortment in the new equilibrium, characterized by lower prices and higher variety.

It is important to note that for some residents, the price and variety of Average Assortment Consumed may rise. As travel costs decrease, consumers can reach more competitive areas, such as Bergen, that offer a greater variety with higher prices. To examine this idea deeper, I investigate whether consumers are more inclined to choose stores with lower average choice weighted HHI in the counterfactual scenario. By aggregating HHIs, weighted by the share of consumers from each market within a store's catchment area, I find a negative correlation of 0.1, indicating that market share increases for stores with lower HHI in the new equilibrium.

Finally, for some areas, AAC might change in the opposite direction. This occurs in those regions where retailers face a less elastic demand, as discussed earlier, leading them to raise prices and reduce variety.

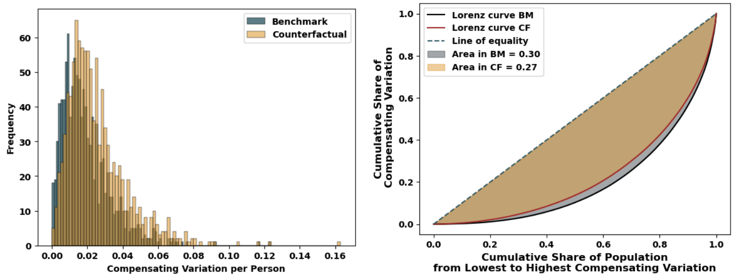


(a) Effect on Price (b) Effect on Variety

Figure 1.16: Counterfactual changes in average assortment consumed due to reduced travel disutility

As expected, the policy positively impacts consumer welfare, resulting in a substantial increase of 11.4% or 1261 MNOK. Figure 1.17a demonstrates how the distribution of consumer welfare per person changes due to the policy intervention. The Gini index for the counterfactual scenario illustrates a modest improvement in consumer inequality. The changes are visually depicted with the Lorenz curve in Figure 1.17b.

The policy also has a positive impact on firms. The industry's total profit increases by 215 MNOK, equivalent to an improvement of 5.6%. The total welfare gain from the policy calculated as a sum of the change in consumer welfare and change in profits amounts to 1476 MNOK, equivalent to an increase of 9.9% compared to the benchmark equilibrium.



(a) Consumer welfare (b) Lorenz curve

Figure 1.17: Change in consumer welfare due to reduced travel disutility

Furthermore, I compute the policy cost as the sum of transfers the government needs to provide to consumers residing in remote regions to offset fifty percent of their travel disutility. In other words, for consumers in remote locations, the transfer is defined as follows:

$$u\left(y_{i(l)} + T_{i(l)}, \delta_j, d_{lj}, \rho^{BM}, \epsilon_{i(l)j}\right) = u\left(y_{i(l)}, \delta_j, d_{lj}, \rho^{CF}, \epsilon_{i(l)j}\right), \tag{1.33}$$

where $j = \arg\max_k u(y_{i(l)}, \delta_k, d_{lk}, \rho^{BM})$, ρ^{BM} represents the parameter for travel disutility in the benchmark equilibrium, and ρ^{CF} is the parameter for travel disutility in the counterfactual scenario. After aggregating the transfers across markets, the total cost accounts to 1198 MNOK.

Finally, I calculate the net welfare effect of the counterfactual policy as follows:

$$\Delta W = \sum_i CV_{i(l)} + \sum_j \Delta\Pi_j - \sum_i T_{i(l)} \times MCPF, \tag{1.34}$$

which includes the compensating variation for consumers $CV_{i(l)}$ and the change in firms' profits $\Delta\Pi_j$. The last term stands for the cost of the policy, which is the total amount of transfers to consumers $T_{i(l)}$ adjusted by the Marginal Cost of Public Funds (MCPF) specific to Norway. By multiplying the transfers by the MCPF, I account for the deadweight loss that may arise due to government interventions leading to inefficient allocation of resources. The value of MCPF is adopted from the guidelines outlined in the Principles for profitability assessments in the public sector (NOU 1997:27).[8] As a result, the net welfare effect sums up to 38.4 MNOK. The policy demonstrates promising outcomes for consumers and firms, contributing to an overall improvement in total welfare.

Although this counterfactual experiment is rather conceptual and not meant to simulate specific policies, it bears some policy relevance. In 2022, a similar policy was implemented in France as a way to support residents of remote regions who were particularly affected by the energy crisis.[9] The government introduced an energy cheque scheme aimed at compensating for increased travel costs. The policy was specifically targeted at the residents of remote areas.

Subsidies for Stores Located in Remote Areas. In the experiment on uniform assortment, some stores become unprofitable as they provide the same range of products in all locations, including remote areas. This leads to higher prices as firms must compensate for higher logistics costs. To address this issue, in this counterfactual policy, stores in less populated areas receive subsidies to offset lo-

[8]NOU 1997:27, Nyttekostnadsanalyser – Prinsipper for lønnsomhetsvurderinger i offentlig sektor (Utredninger, 1997)

[9]`https://www.intereconomics.eu/contents/year/2023/number/1/article/exiting-the-energy-crisis-lessons-learned-from-the-energy-price-cap-policy-in-france`

gistics costs. This financial aid aims to incentivize chains to offer better and more affordable products in these regions.

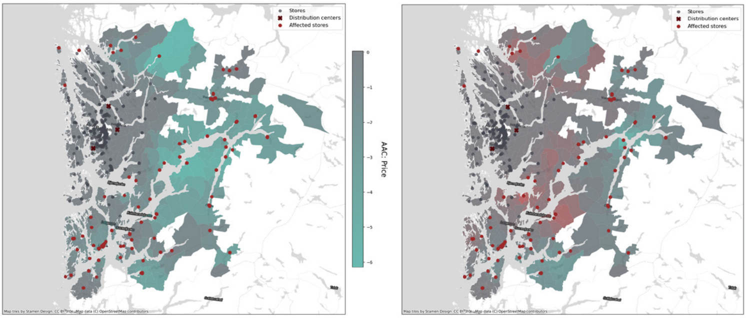
As shown in Figures 1.6 and 1.7, regions with limited assortment tend to be farther away from distribution centers. In this counterfactual experiment, I examine stores whose distribution centers are located further than 70 km of driving distance, corresponding to the 70th percentile of the driving distance distribution for stores in Vestland. These selected stores receive subsidies to compensate 10% of their marginal costs. The idea behind this analysis is reminiscent of an actual policy implemented in Sweden, which aimed at incentivizing stores in rural areas to offer a diverse range of products.[10]

The results from this policy indicate that retailers involved in the policy improve assortment by reducing prices by 1.9% and increasing variety by 0.69%. On the consumer side, the price of the Average Assortment Consumed declines by -0.9%, while variety increases by 0.11%. Figure 1.18 illustrates the spatial distribution of the changes in AAC.

The policy exhibits a modest positive impact on consumer welfare, resulting in a slight increase of 1.8% or 199 MNOK. Figures 1.19a and 1.19b show that the policy's effectiveness in addressing inequality is limited. Despite the positive changes in consumer welfare, the policy does not significantly contribute to reducing income inequality within the affected markets, as evidenced by the unchanged Gini index.

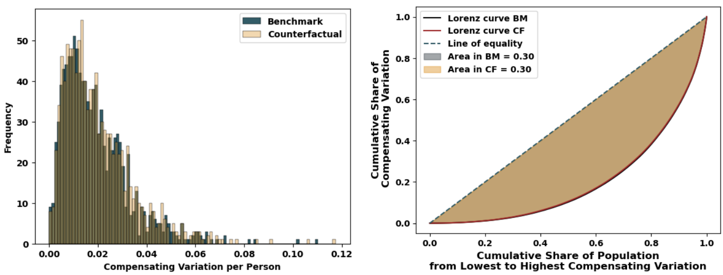
The policy has a notable positive impact on firms, resulting in a total profit increase of 262 MNOK, equivalent to 6.8%. Summing over the change in consumer welfare and firms' profit, I calculate that the welfare gain from the policy amounts to 461 MNOK,

[10]Bill 2001/02:4 A policy for growth and viability for the whole country



(a) Effect on price (b) Effect on variety

Figure 1.18: Counterfactual changes in average assortment consumed due to subsidies to remote stores



(a) Consumer welfare (b) Lorenz curve

Figure 1.19: Change in consumer welfare due to subsidies to remote stores

equivalent to an increase of 3.1% compared to the benchmark scenario. Firms benefit more from the policy than consumers, primarily because retailers in remote markets have some degree of local market power, which allows them to retain a significant portion of the change in the margin derived from the subsidies on marginal costs. Consequently, despite the modest reduction in price and the slight increase in variety, most of the subsidy is captured in the increased profit margins for the retailers.

Additionally, I calculate the policy cost as the product of the number of composite goods purchased in the subsidized stores and the subsidy granted, which is equal to 10% of the marginal costs for each particular store. The resulting cost of the policy is 307 MNOK.

Ultimately, the total welfare effect from the intervention is determined as follows:

$$\Delta W = \sum_i CV_{i(l)} + \sum_j \Delta\Pi_j - MCPF \times 0.1 \sum_{j \in \mathfrak{J}_{sub}} q_j mc_j. \tag{1.35}$$

Here, $CV_{i(l)}$ represents the compensating variation for consumers, and $\Delta\Pi_j$ captures the change in firms' profits. The last term represents the cost of the policy, calculated as the sum of 10% of variable costs across the subsidized stores $\mathfrak{J}_{sub}$ and adjusted by MCPF. Consequently, the net welfare effect accounts for 92.6 MNOK. Based on these figures, it appears that the policy is economically justified, even though the gains experienced by firms drive the majority of the total welfare increase.

Assortment discrimination contributes to welfare inequality by creating disparities in access to affordable products and a wide range of choices, disproportionately affecting consumers in remote markets. To tackle this issue, it is necessary to adopt policies that enhance assortment and minimize welfare disparities. One potential solution could be to incentivize retail chains to provide equal assortment across all their stores in a country. However, as demonstrated earlier, such an approach leads to substantial profit reductions and induces certain stores to become unprofitable, potentially exacerbating market concentration. Moreover, implementing this solution in practice poses practical challenges.

An alternative policy could be to target consumers of those areas with limited assortment. In this study, I examine a policy aimed at reducing travel costs for residents who lack a grocery store within a reasonable distance, which results in increased competition, leading to lower prices and greater variety. This policy could be implemented by improving transportation infrastructure or providing lump-sum compensations to offset travel expenses. The counterfactual analysis demonstrates that this policy has the potential to enhance competition and improve consumer welfare effectively.

Alternatively, policies can be targeted toward retailers operating in remote areas. This can involve providing cost subsidies or tax deductions to incentivize retailers in remote areas to offer more products at affordable prices. While technically, this policy may be relatively easier to implement, its effectiveness remains questionable. Although in remote markets, retailers improve assortment with the help of subsidies, local market power enables them to withhold a portion of the subsidy rather than fully pass it on to consumers.

1.8 Conclusion

In this paper, I study how multi-store firms strategically adjust product assortment in response to local competition when product-level prices are fixed. Consistent with previous literature (DellaVigna and Gentzkow, 2019; Adams and Williams, 2019; Hitsch, Hortacsu and Lin, 2019), I document that retailers do not adjust product-level prices when the competitive environment changes. Nevertheless, they adjust product selection, which could potentially serve as a powerful means to generate margins in the uniform pricing scenario.

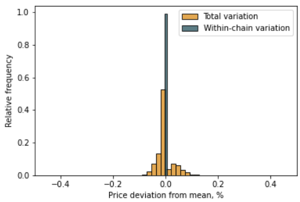
Employing a structural, spatial model of consumer and retailer behavior, I show that product selection can significantly differ across stores of the same chain. The model also allows me to attribute these changes to the local market power. This result leads to substantial assortment inequalities across the country, leading to urban residents enjoying access to more affordable food options. At the same time, consumers in remote markets have access to limited and pricier product selection.

Via counterfactual simulations, I explore the impact of adopting a uniform assortment policy. While this policy enhances consumer welfare, it would lead to substantial losses for firms. Furthermore, the policy of uniform assortment only partially addresses consumer inequality, with consumers in remote areas still incurring higher transportation costs compared to urban residents. As a result, I explore the potential impact of reducing travel costs for consumers in remote areas. The policy is relatively successful in improving competition in remote markets. The findings reveal improvement in assortment in remote areas and increased total welfare. Lastly, I examine a policy of providing subsidies to retailers in remote areas. The findings show modest improvements in assortment for consumers and an increase in total welfare. Both policies are beneficial for consumers and have a positive net welfare effect.

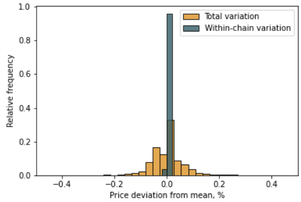
It is worth noting that the model in the paper focuses on assortment decisions and abstracts from modeling prices for individual products. Suppose market changes lead to a significant increase in market power. In that case, a firm might want to revise the entire pricing policy rather than make marginal changes in the assortment. Nonetheless, the model offers some flexibility in accommodating potential price adjustments by higher or lower optimal price points for assortment.

Another aspect that remains outside the scope of this study is the choice of formats. When entering new markets, retail groups strategically choose a store format. The choice of format implies a specific store size, prices, location, and other characteristics. For the purposes of this research, I take stores' format as a given and analyze assortment decisions conditional on the given format. While this approach allows me to examine marginal changes in the assortment, it is crucial to consider the format choice to gain a comprehensive understanding of the competitive landscape. This would allow exploring policies to stimulate more entry into remote markets that would improve competition and reduce inequality in store access.

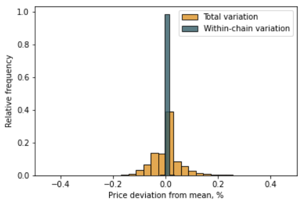
Appendix



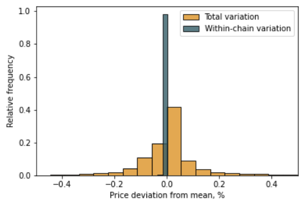
(a) Beer



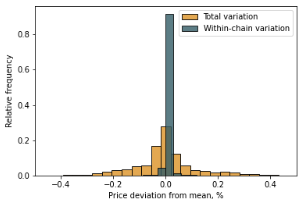
(b) Canned fish



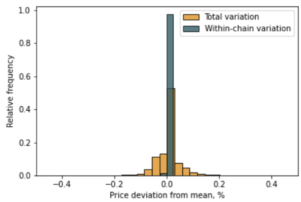
(c) Cheese



(d) Chocolate bars

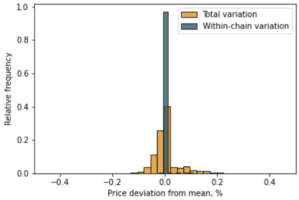


(e) Coffee

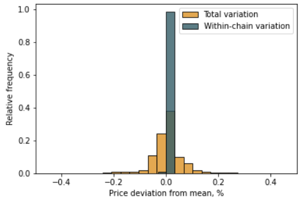


(f) Dry bread

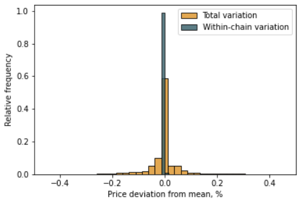
Figure A.1: Price variation within and across chains in different categories (first 6 categories)



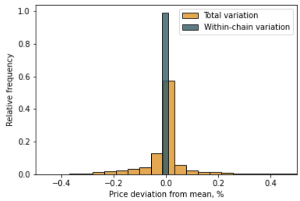
(a) Eggs



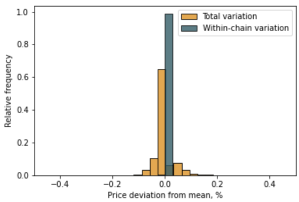
(b) Fresh bread



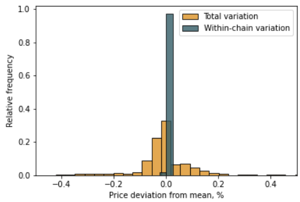
(c) Frozen fish



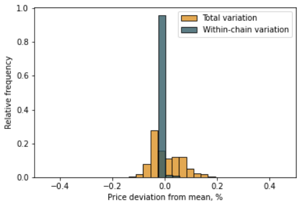
(d) Frozen pizza



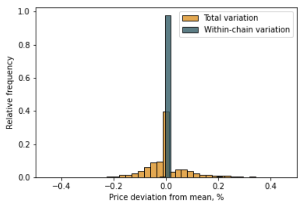
(e) Jam



(f) Juice



(g) Milk



(h) Yoghurt

Figure A.2: Price variation within and across chains in different categories (last 8 categories)

Chapter 2

Preemption in Spatial Competition: Evidence from the Retail Pharmacy Market

Abstract: We study the entry decisions of the three retail pharmacy chains in Norway over the period from 2004 to 2012. Following a deregulation of entry, the market grew rapidly, doubling the number of pharmacies. We document that repeated entry by an already present incumbent chain occurs with non-trivial frequency and set out to investigate whether preemptive motives play a key role. We propose and estimate a highly flexible spatial demand model with overlapping sets of consumers across space. While the estimates imply substantial demand heterogeneity, we reject the hypothesis that the repeated incumbent entries can be explained by market segmentation by store format differentiation. Instead, we propose that *private information* about local market conditions may play a role. Indeed, we find that an incumbent chain is significantly more likely to respond to local market heterogeneity than competing chains.

This chapter is written together with Anders Munk-Nielsen and Morten Sæthre.

2.1 Introduction

Understanding why firms decide to enter a market is a complex task that requires taking into account many factors. Specifically, it is crucial to accurately specify firms' payoffs, information, and the way they interact strategically. For instance, if an incumbent firm is observed to enter a market again, it may either be to preempt the entry of a competitor or because it has information that the market can sustain more active firms than it appears at first glance. This paper aims to empirically explore whether firms have access to such private information about local market conditions.

One important aspect to consider is that local markets can be quite heterogeneous, also with regard to unobserved characteristics. The omission of market heterogeneity can lead to bias in estimating competition, as has been shown by Orhun (2013) and Igami and Yang (2016). However, all prior empirical studies have, to our knowledge, assumed that such unobservables are common knowledge among firms. Misspecified information structure may result in bias, but we will furthermore show that it is fundamentally at odds with much of the entry decisions we observe.

To accurately measure unobserved market heterogeneity, it is important to have a demand model that is flexible enough to capture competitive overlap and demand heterogeneity at a highly localized level. In this paper, we adopt a novel approach proposed by Ellickson, Grieco and Khvastunov (2020), which allows spatially heterogeneous consumers to have location-specific choice sets. This framework differs from the traditional isolated markets approach used in previous literature (Bresnahan and Reiss, 1991; Zheng, 2016). In particular, we employ a spatial discrete choice model that explicitly accounts for the distance between consumers and firms, allowing us to measure local competitive pressure more accurately. In the model, the set of available firms and the degree of substitution depend on how consumers trade off travel distance and firm characteristics.

In this paper, we focus on the market of prescription drugs in Norway. In our analysis, we use two datasets, one containing sales of prescription pharmaceuticals in Norway from June 2004 to December 2011 and the other pertaining to consumer demographics. The pharmacy sales dataset encompasses all daily prescription transactions at the pharmacy level. Each transaction record includes information such as the total purchase amount, the number of packages, and the demographic attributes of the purchasing consumer. Notably, consumers' residential locations are not directly observed. Instead, transactions are aggregated at the pharmacy-month level within consumer segments. Consumer segments are defined based on gender and six age categories. Thus, we observe sales for a combination of the pharmacy, month, and demographic group. We supplement this with data from administrative registers on the demographic composition across finely detailed spatial subsets of Norway (called basic units). This way, we know how many packages (and at what price) were sold to females of a certain age group and education, and we also know how many of these consumers live at various locations with different distances to the pharmacy. This is how we are able to estimate demographic preference heterogeneity and travel disutility.

There are two key advantages to studying the market for prescription pharmaceuticals. First, demand is primarily driven by factors unrelated to pharmacy entry. That is, there is virtually no effect on total demand from the entry of additional pharmacies. Second, there is no scope for price competition, and all pharmacies must stock all products. While there are other sales, e.g., shampoo, that sales component is much smaller than, for example, in the US. Therefore, pharmacies can largely only compete where they are located, which is precisely what we wish to study.

Having estimated our demand model, we are ready to investigate the entry decisions by the pharmacy chains. Following deregulation just prior to the beginning of our sample period (2004), the number of pharmacies grew rapidly and effectively doubled over the period. Conveniently, the pharmacies consolidated into three competing chains virtually overnight in response to the deregulation.

We begin our entry analysis by categorizing entry events based on whether or not the nearest existing pharmacy belonged to the same chain, which we consider as an incumbent. We document that 24% of our entry events fall into this category. This number is surprisingly large given that in this market, such entry decisions result primarily in the cannibalization of existing sales within the same chain. Prior work has, therefore, often viewed such entry decisions as "preemptive".

In this paper, we propose that these entries are the result of *information asymmetry* between firms present in a local market and a new entrant. In particular, we argue that the information asymmetry stems from firms' presence in a local market, which enables them to gain insights into the local demand that may not be apparent from observed demographics. This information asymmetry can be due to foot traffic patterns in the area that might not be fully captured by observed demand characteristics. Consequently, incumbents are better informed to adapt their entry strategy.

To document unobserved market heterogeneity, we use the demand model estimates to compute the residual demand as the difference between the observed and predicted number of transactions in local markets where entries occurred. In the following analysis, we use the residual demand before entry to measure the information asymmetry between incumbents and new entrants. Our findings show that when a new pharmacy belonging to the same chain as an existing one enters a market, the residual demand is consistently higher compared to when a new chain enters a market.

To further explore these entry events, we use a linear probability model to control for differences in observable characteristics related to the events, and the result still holds. In particular, we find a significant conditional correlation between the demand residual and an indicator for the new pharmacy belonging to the same chain. This supports the information asymmetry hypothesis, which suggests that the incumbent chain reacts to the local demand residual while competing chains do not.

The closest related paper that studies unobserved market heterogeneity is Igami and Yang (2016). In the context of hamburger chains in Canada, they show that unobserved market heterogeneity leads to biased estimates of competition. However, their argument implies that *all chains* are equally likely to enter if residual demand is high. The importance of differentiation for location decisions is also demonstrated by Orhun (2013), who solves a static discrete location choice game among retail supermarkets. Orhun (2013) demonstrates that local market heterogeneity, which is assumed to be common information to players, has an important role in shaping spatial competition. We have two contributions to this literature: first and foremost, we allow information to be private. Second, we observe demand and thus do not need to infer flow profits indirectly from entry decisions.

Our approach also diverges from much of the preceding literature on retail entry, as we refrain from assuming any specific configuration of local markets. Instead, we follow Ellickson, Grieco and Khvastunov (2020) in assuming that consumers are distributed across space according to register data at a finely disaggregated level. Ellickson, Grieco and Khvastunov (2020) show that this demand model can capture rich substitution patterns that better reflect the nature of spatial competition. Consumers dislike traveling and are heterogeneous in their preferences for individual store formats. One key difference between our approach and that of Ellickson, Grieco and Khvastunov (2020) is that our disaggregated sales data allows us to estimate heterogeneity for preferences for pharmacy characteristics and travel distance across consumer groups without relying on predetermined functional forms.

We are related to a large empirical literature on entry decisions (e.g. Bresnahan and Reiss, 1991; Berry, 1992; Seim, 2006; Aguirregabiria, Mira and Roman, 2007), and in particular spatial competition; see, e.g., the recent survey by Aguirregabiria and Suzuki (2016). Specifically, the part that emphasizes the chain-affiliation aspects (Jia, 2008; Aguirregabiria and Vicentini, 2016). However, another strand has emphasized network effects induced, e.g., by logistic concerns (Holmes, 2011; Ellickson, Houghton and Timmins, 2013). Our research contributes to this strand of literature by demonstrating that information asymmetry is a significant factor in location choice for entry decisions.

The paper is organized as follows: Section 2.2 describes our data and institutional setting, and Section 2.3 provides descriptive evidence regarding entry patterns. Section 2.4 presents out model for demand and competition, and 2.5 presents estimates from the demand model. Section 2.6 presents evidence regarding incumbent entry decisions, and Section 2.7 concludes.

2.2 Data and Institutional Setting

Data

We rely on two datasets: one regarding pharmacy sales and one regarding consumers.

Our data on pharmacies has the universe of all daily prescription transactions at the pharmacy level. For each transaction, we observe the total amount purchased, the number of packages, as well as demographic characteristics of the purchasing consumer. We do not observe the residential location of the consumer, and instead combine sales at the demographic level together with location-specific population at the demographic level to estimate demographic preference heterogeneity and travel disutility.

Next, we want to aggregate transactions at the pharmacy-period level within consumer segments. To do so, we discretize consumers into groups based on their gender as well as six age categories. Table 2.1 shows summary statistics for the transactions and the aggregate demographic composition of sales.

For the same discretization of demographics, we count the number of individuals living in each *location*. Throughout the paper, location will refer to a *Basic Unit (BU)*. A BU is a zone defined by Statistics Norway, which is far smaller than a zip code area - there are 12,164 BUs across Norway. Figure 2.1 shows the BUs in the city of Bergen. For each such location, we compute the number of residing consumers of each demographic type. For the purpose of distance calculations, we will use centroids. Table 2.2 shows summary statistics for the demographic characteristics across the

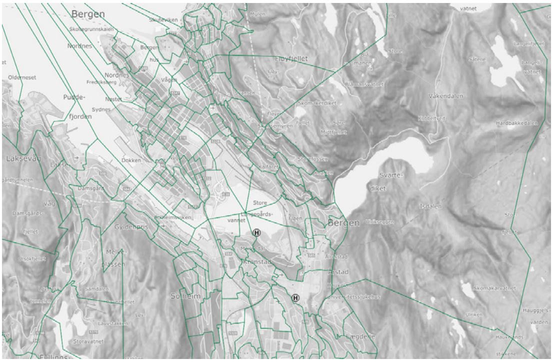


Figure 2.1: Basic Units in the City of Bergen

12,164 locations. For example, we see that there are several locations where only a single segment resides.

We likewise find the location of all pharmacies and also attribute them to the corresponding centroid. We then compute the travel distances in minutes by car from all consumer locations to pharmacy coordinates, including any tolls that would be incurred during the shortest path of travel.

Institutional Setting

Entry (de)regulation

Sale of pharmaceuticals in Norway is highly regulated and only permitted at licensed pharmacies, with the exception of a few deregulated over-the-counter drugs that can be sold in grocery stores.[1] Until 2001, pharmacies were subject to a strict licensing scheme, where only licensed pharmacists could own a pharmacy,

[1] E.g., tablets with paracetamol up to 500 mg or ibuprofen up to 200 mg with limitations on package size, and nasal sprays containing fluticasone, mometasone, triamcinolone and budesonide.

and the number and location of pharmacies was decided by the Ministry of Health Care and Services. In 2001, both ownership and establishment of pharmacies were deregulated, allowing both individuals and companies to own multiple pharmacies. The new regulation lead to the establishment of three pharmacy chains based on existing groups of pharmacies with joint purchase agreements from wholesalers. Existing pharmacies were bought up by the new chains, in addition to a noticeable and persistent increase in establishment of new pharmacies. The pharmacy chains became vertically integrated with the three large existing medical product wholesalers. The majority of pharmacies—both existing and newly established—have since belonged to one of the three chains, with a smaller number of private independent pharmacies and publicly owned pharmacies in the larger hospitals. Even though anyone can establish and own a pharmacy, each pharmacy outlet is required to have a licensed pharmacist at the location to manage and oversee the operations. Note that licensing of pharmacists is purely a matter of educational qualification, and therefore separate from the licensing of pharmacies that existed before 2001.[2]

On the wholesaler side, only full-line wholesalers are allowed to sell to pharmacies, meaning that they need to carry all prescription drugs that have marketing permission in Norway, while more specialized, medical wholesalers are barred from selling directly to pharmacies. Furthermore, wholesalers selling to pharmacies are—with limited exceptions—required to deliver drugs anywhere in Norway within 24 hours.[3] These regulatory features together with scale economies in standardized product logistics is a likely ex-

[2]For more information on regulation of pharmacies, see the Norwegian Pharmacy Act (Helse- og omsorgsdepartementet, 2000).

[3]For more information, see the Norwegian Regulation of Medical Wholesalers (Helse- og omsorgsdepartementet, 1993).

planation for the low and stable number of vertically integrated pharmaceutical chains over time.

Price regulation, reimbursement and generic substitution

Prices of prescription drugs are subject to reference price regulation, where the maximum price is set based on international averages, which is almost always binding for branded drugs, likely due to both high degree of reimbursement and low elasticity of demand for most pharmaceutical treatments.[4] Norway has a single-payer health care system, where drugs used in treatment of chronic conditions (treatments longer than 3 months) are reimbursed with a coinsurance rate of 36%, where copayments are capped at approximately 50 EUR per 3 months and total medical copayments are capped at 200 EUR per year (including copayments for doctor consultations, pharmaceuticals and laboratory services), while treatment of most contagious diseases are fully reimbursed. After generic entry, the maximum reimbursement is regulated down over time according to a common, pre-specified schedule tied to the maximum price at the time of generic entry. Pharmacies are required to have at least one generic option that is priced no higher than the maximum reimbursed price, and to suggest substitution to the cheapest generic substitute if the prescription specifies the brand name of the drug. If the customer refuses generic substitution, their reimbursement is calculated according to the maximum reimbursed price, while the remaining price is covered fully out-of-pocket.

Technically, the price regulation features a maximum price from

[4]Specifically, the maximum is set as the average of the three lowest prices from Sweden, Finland, Denmark, Germany, UK, Netherlands, Austria, Belgium and Ireland.

the wholesaler to the pharmacy and a maximum pharmacy margin, which together determine the maximum price to consumer. From the perspective of an integrated pharmacy chain, the maximum margin is only a matter of accounting, while the relevant margin is given by the difference between the consumer price and the wholesale price paid to the manufacturer, where the latter is not subject to regulation. For independent pharmacies, the maximum margin itself can play a role in determining profitability and the relative importance of prescription drug sales versus over-the-counter drugs and other products, and is likely a contributing factor to the majority of new establishments being undertaken by chains.

2.3 Descriptive Evidence

Figure 2.2 shows the number of active pharmacies over time. The figure shows a broader period around our sample period (2004-2012) to provide context. We clearly see a rapid growth following the liberalization of entry in 2001, with an almost constant growth of between 20 and 30 pharmacies per year. A constant growth rate is indicative of firms facing some form of constraints in the number of pharmacies they can open per year.

Table 2.3 presents descriptive statistics for entry events. Out of the full set of entry events, we restrict attention to a subset that are most relevant for our study, resulting in 225 events in our final dataset. Out of these, 55 (24%) are by the same chain as the nearest pharmacy. Given that there are three chains, this is a remarkably high propensity, given that such pharmacies will be cannibalizing the chain's own sales.

To better understand these 55 events, we show the empirical

Table 2.1: Summary statistics: transactions

Transactions	
Observations (million transactions)	144.6
Total revenue (billion NOK)	70.2
Average transaction (NOK)	485
No. packages	1.25
Gender composition	
Male	0.436
Female	0.564
Age composition	
0-24	0.064
25-45	0.103
46-59	0.182
60-74	0.297
75-89	0.280
90+	0.074

Note: For each transaction in our dataset, we observe the demographics of the purchasing individual, since this gets recorded as the transaction is encoded. The table shows averages computed over all transactions unweighted.

Table 2.2: Summary statistics by location (basic unit, BU)

	Mean	St.dev.	Min	Max
		Gender composition		
Male	0.510	0.064	0.000	1
Female	0.490	0.058	0.000	1
		Age composition		
0-24	0.312	0.082	0.002	1
25-45	0.277	0.090	0.004	1
46-59	0.208	0.062	0.004	1
60-74	0.143	0.073	0.001	1
75-89	0.071	0.060	0.000	1
90+	0.013	0.023	0.000	0.412
Observations (basic units)				12,164

Note: For each location (BU), we compute each statistic based on all residing individuals and take the average over all periods in our sample. The source for the demographic data is the Norwegian register data.

distribution of the density to the nearest existing pharmacy at the time of entry separately for the 55 same-chain entries and the 170 competing-chain entries in Figure 2.3. The graph clearly shows a tendency for incumbents to locate further away from the nearest existing pharmacy than a competitor would. This is consistent with the broad intuition from the classic Hotelling (1929*a*) model of competition, where one locates as close as possible to a competitor to maximize business stealing.

2.4 Model

2.4.1 Consumer Choice

Our model of consumer demand builds on Ellickson, Grieco and Khvastunov (2020). We assume that individual quantity demanded, q_{it}, does not respond to market structure. This is reasonable due

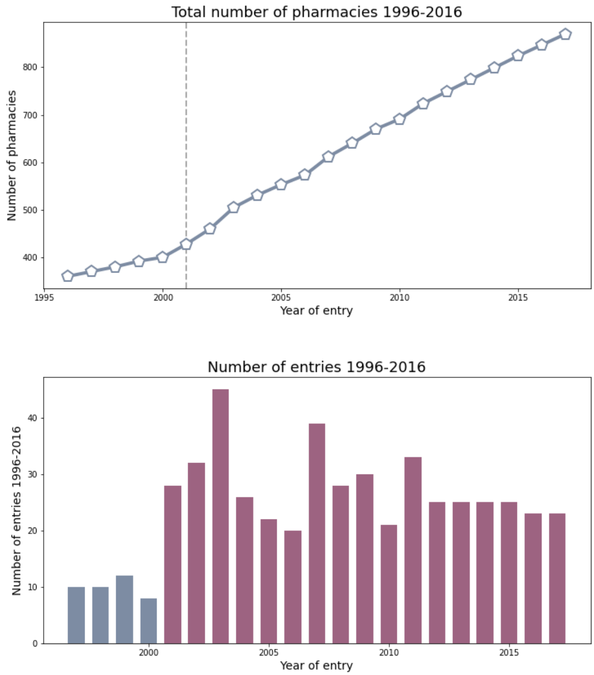


Figure 2.2: Entry over time

Note: The upper tab represents the cumulative number of pharmacies over time, while the lower tab shows the number of entries per year. The dashed line indicates the entry deregulation in 2001.

Table 2.3: Descriptive statistics for entry events

		Closest pharmacy to entrant	
	Total	Same chain	Competing chain
Distance to neighbor (driving min.)	8.0	10.5	7.2
Center	47.6%(107)	56.4%(31)	44.7%(76)
Shopping mall	32.0%(72)	29.1%(16)	32.9%(56)
Wine monopoly	7.6%(17)	3.6%(2)	8.8%(15)
Number of entries	225	55	170

Note: The numbers in parentheses are frequencies corresponding to the fractions.

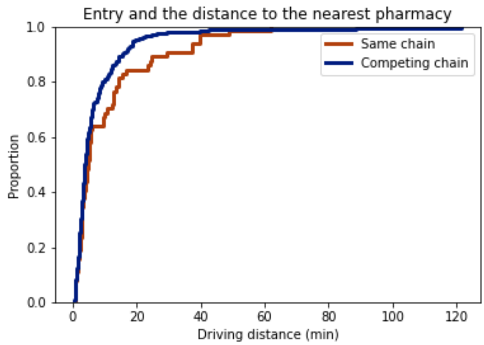


Figure 2.3: Distance to nearest existing pharmacy

Note: The graph shows the empirical CDF of the driving distance from an entering pharmacy to the nearest pharmacy operating at the time of entry. That is, an observation is an entry. There are two lines, one conditioning on the nearest pharmacy belonging to the same chain, and one where it belongs to a competing chain.

to the nature of (single-payer) insurance in Norway, and we have not found any indications that consumers forego prescribed treatments due to travel costs.[5] We therefore model consumers as choosing a pharmacy based on travel costs, characteristics of the pharmacy and characteristics of its location, and then buying an amount q_{it} that does not depend on the pharmacy itself. That is, the total quantity sold at pharmacy j in period t is

$$Q_{jt} = \sum_{i \in \mathcal{I}_t} \Pr(j|it) q_{it},$$

where $\mathcal{I}_t$ is the set of consumers and $\Pr(j|it)$ is the probability that consumer i visits pharmacy j in period t.

Consumers choose only which pharmacy to visit, taking q_{it} as

[5] Analysis available from authors upon request.

given. Consumer i residing at location $\ell \in \mathcal{L}$ and belonging to demographic group k in period t gets indirect utility

$$
\begin{aligned}
u_{i\ell ktj} &= v_{\ell ktj} + \varepsilon_{i\ell ktj}, && (2.1) \\
v_{\ell ktj} &= \gamma_0^k + \gamma_1^k d_{\ell j} + \gamma_2^k \text{toll}_{\ell j} + \mathbf{x}_{tj}' \boldsymbol{\beta}^k + \eta_{f_j}^k, && (2.2)
\end{aligned}
$$

where $d_{\ell j}$ and $\text{toll}_{\ell j}$ are driving distance and the costs of toll roads between location ℓ and pharmacy j, $\eta_{f_j}^k$ is a set of dummies for the chain affiliation pharmacy j, where the chain index is $f \in \{1, 2, 3\}$, $\varepsilon_{i\ell ktj}$ is IID Extreme Value Type I, and $\mathbf{x}_{tj}$ is a vector of pharmacy characteristics including dummies for being located in a mall, a large mall, next to a wine monopoly, and interactions of large mall and wine monopoly.[6] Given this, the probability that consumers living in location ℓ buys from pharmacy j takes the usual logit form:

$$
\Pr(j|\ell, k, t) \equiv \frac{\exp(v_{\ell ktj})}{\sum_{j' \in \mathcal{J}_t} \exp(v_{\ell ktj'})}, \qquad (2.3)
$$

where $\mathcal{J}_t$ is the set of pharmacies available in period t.

The individual quantity purchased is parameterized as

$$
q_{\ell kt} = \beta_0^k + \beta_1^k t + \sum_{m=2}^{12} \beta_m^k \mathbf{1}\{\text{month}(t) = m\} + \xi_{\ell kt}. \qquad (2.4)
$$

where ℓ denotes location of residence, and $k \in \{1, ..., K\}$ is the demographic group. This specification allows demand to vary by discrete demographic groups, k, and both to drift over time within demographic groups and to have seasonal variation, of which there

[6] In Norway, the sale of alcoholic beverages with higher alcohol volume than 4.7% is restricted to the so-called wine monopoly stores.

is a lot, e.g., due to the seasonal flu. The group unobservables, $\xi_{\ell kt}$, will play an important role later. It determines the value of being present in the local market, and we will argue that the assumption that incumbent firms have private information about $\xi_{\ell kt}$ will be crucial.

2.4.2 Role of Asymmetric Information in Entry Decisions

Pharmacy chains make entry decisions across the whole country of Norway. The set of possible locations for entry in space is the set of actual pharmacy locations that are distinct from the consumer locations. We denote locations just by j, thus omitting chain affiliation since a pharmacy and its location are indistinguishable, given that, at most, one pharmacy can be located at any given point in space.

Chains make entry decisions in order to maximize expected discounted profits, where the flow profit from a pharmacy at location j takes the form

$$\pi_{jt} = R_{jt} - C_{jt}. \tag{2.5}$$

The flow profit is a function of the total revenue R_{jt} earned at pharmacy j in period t and the total variable cost C_{jt}. Revenue depends on the composition of sales

$$R_{jt} = \sum_{\ell} \sum_{k} N_{\ell kt} q_{\ell kt} \Pr(j|\ell, k, t) p_{kt}, \tag{2.6}$$

where $\Pr(j|\ell, k, t)$ and $q_{\ell kt}$ come from Equations 2.3 and 2.4, $N_{\ell kt}$ is the number of consumers in demographic group k in period t, and p_{kt} is the earnings per transaction for a consumer of demographic group k.

The variable of interest in the entry decision is the state of local demand $\xi_{\ell t} = \sum_k N_{\ell kt}\xi_{\ell kt}$ at all consumer locations $\ell \in \mathcal{L}$. This variable is unobserved to the econometrician, and we may think of it as decomposed into two (orthogonal) parts,

$$\xi_{\ell t} = \rho\xi^C_{\ell t} + (1-\rho)\xi^I_{\ell t}, \quad \rho \in [0;1], \tag{2.7}$$

where $\xi^C_{\ell t}$ is commonly observed by all firms and $\xi^I_{\ell t}$ is privately observed by the incumbent firm. We define the incumbent chain c at location ℓ to be the firm that operates the pharmacy closest to ℓ. The parameter ρ controls the extent to which market-level excess demand is common ($\rho \to 1$) or private ($\rho \to 0$) information.

Let us now consider how we can recover the value of ρ. Intuitively, we can obtain the residuals, $\xi_{\ell t}$, using demand data, but we cannot further decompose those residuals based on demand data alone. That is, we can only hope to learn about ρ from the entry decisions that firms make.

In the ideal experiment, the initial firm network is exogenously given and orthogonal to $(\xi^C_{\ell t}, \xi^I_{\ell t})$, which are furthermore independent across locations ℓ. We can start by noting that we should observe more entries where $\xi_{\ell t}$ is large. Recovery of ρ then boils down to whether the incumbent firm is more likely to enter close to ℓ compared to a competitor, all else equal.

Our framework nests that of Igami and Yang (2016), who implicitly assume that the market-level unobservable is common information, $\rho = 1$. While this will indeed explain why an incumbent chain chooses to enter again in a nearby location, a competing chain will also know this information and should have an even stronger incentive to enter. This is because the incumbent chain will, to a large extent, be cannibalizing its own sales.

2.5 Demand

2.5.1 Econometric Methodology

Given that we do not observe the locations of the consumers for all transactions, our data leaves us unable to estimate by maximum likelihood. Instead, we predict purchases from consumers at all locations (BUs) and match them to the sales to each demographic at all pharmacies. Estimation is then conducted using the method of simulated moments.

To be precise, we observe Q_{jkt}^{obs}, the total sales to demographic group k at pharmacy j in period t. The corresponding predicted quantity from our model is obtained as

$$Q_{jkt}(\theta) = \sum_{\ell \in \mathcal{L}} N_{\ell kt} q_{\ell kt}(\theta) \Pr(j|\ell, k, t; \theta), \tag{2.8}$$

where $q_{\ell kt}$ is expected individual demand for consumers in demographic group k residing at location ℓ defined in Equation 2.4. This equation also illustrates the problem: we are predicting demand at the pharmacy-level (j), but we observe consumers at the location-level (ℓ). For simplicity, we assume that demand has fully died out after 1 hour of driving, so we impose a probability of zero mechanically thereafter.

Hence, our estimator is

$$\hat{\theta} = \arg\min_{\theta} \sum_{j \in \mathcal{J}_t} \sum_t \sum_k \left[Q_{jkt}^{\text{obs}} - Q_{jkt}(\theta)\right]^2. \tag{2.9}$$

That is, we minimize the squared residuals of observed and predicted pharmacy-level transactions for each demographic segment in all periods.

2.5.2 Results

In this section, we present and discuss the estimates of the parameters in our demand model. There are two sets of parameter estimates: those affecting pharmacy choice and those affecting quantity purchased. Table 2.4 presents estimates regarding the parameters affecting the pharmacy choice, i.e. equation (2.2). We start with the common patterns across all demographic groups and before looking the aspects that are heterogenous across consumers.

With the exception of one of the 12 demographic groups, all consumers dislike distance and travel costs. Consumers all prefer larger pharmacies and all but one group dislike pharmacy locations to be in either the center or in a small mall. Regarding location in a large mall or close to a liquor store ("wine monopoly"), we see substantial heterogeneity: younger households prefer pharmacies located in a large mall, while older households dislike it. Similarly, most households, particularly females, are attracted to pharmacies located near a wine monopoly, although there are interesting differences depending on whether it is in a mall or not: again, the elderly appear to be discouraged by mall locations, whereas it has the opposite effect for younger households.

The second part of parameters indexing consumer demand is those that affect the quantity, i.e., those in equation (2.4). Those results are presented in Tables 2.5 and 2.6. The results are as one might expect: demand is higher during the winter months, e.g., due to the seasonal flu, and increases with the age of the consumer group.

2.6 Investigating Entry Decisions

In this section, we investigate entry decisions in lieu of our demand model. Specifically, we want to focus on the events where an incumbent firm in a location enters with another pharmacy – such entry events are the ones that can potentially be viewed as *preemptive*. To do so, we restrict attention to the set of such entry events that are the cleanest examples: monopoly to duopoly transitions. That is, locations (BUs) where a single pharmacy is present prior and then another pharmacy enters with the same chain affiliation as the incumbent.

Figure 2.4 shows an event study of residual demand for the incumbent pharmacy around entry, where we have separated entry events depending on whether the incumbent and entering pharmacy belong to the same chain (*present in market*) or not (*not present in market*). Residual demand is the difference between predicted and observed demand measured in thousands of transactions aggregated over all demographic groups per month. The residual demand for incumbent firms is consistently higher than for new entrants and the difference can be up to 900 transactions. This difference is significant considering that the average pharmacy typically has around 2.6 thousand transactions per month, with an average transaction amount of roughly 480 NOK ($\sim$ 48 EUR).

If taken at face value, this result is consistent with the hypothesis that incumbent firms choose to enter a market if their private information tells them that local demand is high. However, the graph merely presents raw averages, so there may be confounders between entry events in markets where the incumbent is already present versus ones where it is not.

To control for differences, we instead pursue a linear regression

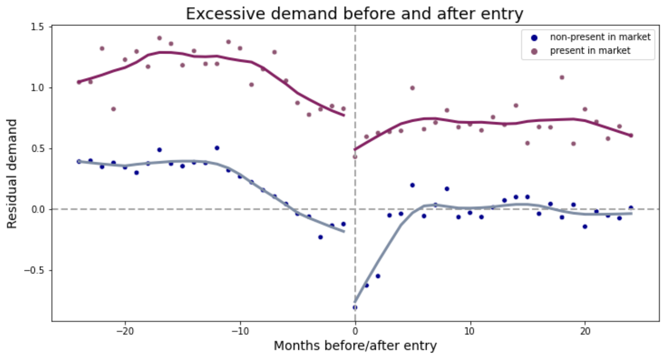


Figure 2.4: Residual demand before and after entry event

specification:

$$\mathbf{1}\{\text{same chain}\}_e = \alpha\hat{\xi}_e + x_e\kappa + \mu_e,$$

where e denotes entry events, and $\hat{\xi}_e$ is the predicted demand residual averaged over 3, 6, or 12 months prior to the entry (in separate columns), and where x_e is a vector of characteristics of the entry event, including chain dummies, characteristics of the store, and the market size.

The results are shown in Table 2.7. We find statistically significant estimates of the coefficient on residual demand, α. Furthermore, the estimates are remarkably similar whether we average over 3, 6, or 12 months prior to entry. This is consistent with the graphical evidence in Figure 2.4, where residual demand was fairly stable in many months before entry.

One important alternative hypothesis that might explain what we see is network effects. As suggested by e.g. Holmes (2011); Ellickson, Houghton and Timmins (2013), there can be various reasons for economies of density such as logistic networks. We run the risk of attributing both network effects and private information shocks to only the latter, just as the prior literature has attributed both to network effects. However, given the complexity in solving for the single agent logistic problem of rolling out stores, we leave the integration of both strands for future work.

2.7 Conclusion

In this paper, we studied the entry decisions of retail pharmacy chains in Norway following a deregulation of entry. We documented that incumbent pharmacies enter with non-trivial frequency. Therefore, we set out to investigate whether such entry decisions should best be viewed as preemptive.

In this paper, we have formulated and estimated a rich model of demand in the retail pharmacy market. Our model features rich heterogeneity across consumer segments and, thus, scope for differentiation across retail store formats. Our first hypothesis was thus that incumbent entry events were simply market segmentation due to differentiation in store format. However, we found evidence against this hypothesis.

Next, we proposed that unobservable market-level heterogeneity in demand might play a role. This has been emphasized previously by, e.g., Igami and Yang (2016); Orhun (2013). Next, we proposed that such unobservables might be either private or common information to the players of the game. We then presented evidence that such unobservables are likely private information to a larger extent. This is because large demand residuals are more likely to attract pharmacies with nearby chains compared to competing chains.

In conclusion, we find evidence for preemptive motives based on private information about local market attractiveness.

Table 2.4: Demand Estimates Part I: Pharmacy Choice

	Female					
	F0-24	F25-45	F46-59	F60-74	F75-89	F90+
Distance	-0.352***	-0.287***	-0.280***	-0.311***	-0.722***	-1.208
	(0.003)	(0.003)	(0.002)	(0.004)	(0.011)	(2.266)
Travel cost	-0.025***	-0.018***	-0.015***	-0.018***	-0.004*	0.824
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(1.582)
Distance x	0.000***	0.000***	0.000***	-0.000***	0.000***	-0.341
pop.density	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.634)
Pharmacy size, m^2	0.001***	0.001***	0.001***	0.001***	0.003***	2.091
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(3.885)
Center	0.008	-0.077***	-0.072***	-0.080***	0.189***	-0.965
	(0.009)	(0.007)	(0.007)	(0.008)	(0.021)	(10.02)
Mall	-0.269***	-0.269***	-0.164***	-0.186***	-0.426***	0.140
	(0.013)	(0.011)	(0.009)	(0.010)	(0.022)	(19.65)
Large mall	0.563***	0.407***	0.220***	0.078***	-0.485***	0.018
	(0.009)	(0.007)	(0.008)	(0.009)	(0.023)	(8.932)
Wine monopoly	0.094***	-0.084***	0.193***	0.193***	0.442***	-0.077
	(0.027)	(0.023)	(0.021)	(0.025)	(0.064)	(19.35)
Wine monopoly x	0.554***	0.319***	0.291***	0.253***	-0.473***	-1.046
large mall	(0.007)	(0.006)	(0.006)	(0.007)	(0.017)	(9.024)
	Male					
	M0-24	M25-45	M46-59	M60-74	M75-89	M90+
Distance	-0.033***	-0.39***	0.300	-0.309***	-0.449***	-11.410
	(0.001)	(0.004)	(0.257)	(0.004)	(0.01)	(10.85)
Travel cost	0.002***	-0.057***	-1.588***	-0.024***	-0.065***	-35.73
	(0.000)	(0.002)	(0.544)	(0.001)	(0.004)	(34.04)
Distance x	0.000***	0.000***	-0.971***	-0.000***	-0.000***	-0.581
pop.density	(0.000)	(0.000)	(0.335)	(0.000)	(0.000)	(0.550)
Pharmacy size, m^2	0.001***	0.001***	0.216***	0.001***	0.002***	1.317
	(0.000)	(0.000)	(0.074)	(0.000)	(0.000)	(1.246)
Center	-0.169***	-0.076***	0.006	-0.132***	-0.076***	-3.366
	(0.006)	(0.009)	(0.879)	(0.008)	(0.015)	(4.445)
Mall	-0.127***	-0.366***	0.059	-0.233***	-0.312***	2.712
	(0.010)	(0.013)	(2.497)	(0.010)	(0.017)	(8.180)
Large mall	0.282***	0.231***	-0.091	-0.080***	-0.362***	3.244
	(0.008)	(0.009)	(1.162)	(0.009)	(0.017)	(6.830)
Wine monopoly	0.114***	-0.217***	-0.078	0.084***	0.239***	-2.560
	(0.021)	(0.029)	(1.586)	(0.026)	(0.047)	(105.5)
Wine monopoly x	0.400***	0.155***	-0.153	0.138***	-0.201***	-5.944
large mall	(0.006)	(0.008)	(0.608)	(0.007)	(0.012)	(6.717)
Observations (pharmacy-months)						55,285

Note: Standard errors are in parentheses. Significance levels are *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$. The data contains 724 pharmaices and predicted demand comes from 31,569,626 pharmacy-month-location tuples. There are 29 parameters per demographic group.

Table 2.5: Demand Estimates part II: Quantity (Female)

	F0-24	F25-45	F46-59	F60-74	F75-89	F90+
Const	0.629***	1.419***	2.976***	7.231***	9.257***	19.31***
	(0.006)	(0.011)	(0.024)	(0.056)	(0.178)	(0.563)
Time trend	0.556***	0.515***	1.837***	2.081***	9.712***	-10.09***
	(0.006)	(0.01)	(0.022)	(0.051)	(0.17)	(0.509)
January	-0.243***	-0.665***	-1.418***	-2.678***	-3.267***	-0.834***
	(0.008)	(0.014)	(0.032)	(0.073)	(0.241)	(0.725)
February	-0.274***	-0.632***	-1.174***	-2.099***	-2.57***	-0.389
	(0.008)	(0.014)	(0.032)	(0.073)	(0.241)	(0.725)
March	-0.165***	-0.357***	-0.478***	-0.549***	-0.403*	1.768**
	(0.008)	(0.014)	(0.032)	(0.073)	(0.241)	(0.725)
April	-0.047***	-0.096***	-0.407***	-1.04***	-1.77***	-0.442
	(0.008)	(0.014)	(0.032)	(0.073)	(0.241)	(0.725)
May	0.031***	0.042**	-0.179***	-0.458***	-0.773**	1.041
	(0.008)	(0.014)	(0.032)	(0.073)	(0.241)	(0.725)
July	-0.239***	-0.321***	-0.689***	-1.194***	-1.395***	1.052
	(0.008)	(0.014)	(0.031)	(0.071)	(0.23)	(0.706)
August	-0.263***	-0.437***	-0.68***	-0.679***	-0.625**	2.155***
	(0.008)	(0.014)	(0.031)	(0.071)	(0.231)	(0.706)
September	-0.237	-0.407	-0.49	-0.302	-0.296	2.147
	(0.008)	(0.014)	(0.031)	(0.071)	(0.231)	(0.706)
October	-0.249***	-0.403***	-0.479***	-0.357***	-0.439*	1.989**
	(0.008)	(0.014)	(0.031)	(0.071)	(0.231)	(0.706)
November	-0.212***	-0.372***	-0.383***	-0.12	-0.051	2.341***
	(0.008)	(0.014)	(0.031)	(0.071)	(0.232)	(0.707)
December	-0.145***	-0.043**	0.704***	2.22***	1.734***	2.546***
	(0.008)	(0.014)	(0.031)	(0.071)	(0.233)	(0.708)

Note: Standard errors are in parentheses. Significance levels are * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$. Base categories are June and Apotek 1. Number of pharmacies - 724. Number of pharmacy-month pairs - 55285. Number of pharmacy-month-BUs - 31569626. Number of parameters per group - 29.

Table 2.6: Demand Estimates part II: Quantity (Male)

	M0-24	M25-45	M46-59	M60-74	M75-89	M90+
Const	0.748***	0.968***	1.658***	6.804***	13.27***	56.61***
	(0.006)	(0.008)	(0.024)	(0.055)	(0.164)	(0.903)
Time trend	0.321***	0.529***	1.183***	2.183***	5.209***	-37.62***
	(0.005)	(0.008)	(0.022)	(0.05)	(0.154)	(0.806)
January	-0.39***	-0.457***	-0.699***	-2.588***	-4.665***	-8.356***
	(0.007)	(0.011)	(0.032)	(0.071)	(0.219)	(1.148)
February	-0.374***	-0.452***	-0.601***	-1.995***	-3.502***	-7.151***
	(0.007)	(0.011)	(0.032)	(0.071)	(0.219)	(1.148)
March	-0.236***	-0.265***	-0.206***	-0.479***	-0.619**	-2.232*
	(0.007)	(0.011)	(0.032)	(0.071)	(0.219)	(1.149)
April	-0.064***	-0.092***	-0.219***	-1.005***	-1.917***	-5.519***
	(0.007)	(0.011)	(0.032)	(0.071)	(0.22)	(1.149)
May	0.031***	0.029**	-0.094**	-0.483***	-0.751	-2.058
	(0.007)	(0.011)	(0.033)	(0.071)	(0.22)	(1.149)
July	-0.286***	-0.207***	-0.331***	-1.009***	-1.498***	-2.698**
	(0.007)	(0.011)	(0.031)	(0.069)	(0.211)	(1.121)
August	-0.392***	-0.33***	-0.329***	-0.573***	-0.457**	-0.626
	(0.007)	(0.011)	(0.031)	(0.069)	(0.211)	(1.122)
September	-0.35***	-0.325	-0.218	-0.189	-0.093	0.359
	(0.007)	(0.011)	(0.031)	(0.07)	(0.212)	(1.122)
October	-0.346***	-0.313***	-0.212***	-0.239***	-0.165	0.685
	(0.007)	(0.011)	(0.031)	(0.07)	(0.212)	(1.124)
November	-0.316***	-0.3***	-0.172***	-0.071	0.251	1.623
	(0.007)	(0.011)	(0.031)	(0.07)	(0.212)	(1.125)
December	-0.224***	-0.107***	0.454***	2.657***	4.456***	5.38***
	(0.007)	(0.011)	(0.031)	(0.07)	(0.213)	(1.127)

Note: Standard errors are in parentheses. Significance levels are * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$. Base categories are June and Apotek 1. Number of pharmacies - 724. Number of pharmacy-month pairs - 55285. Number of pharmacy-month-BUs - 31569626. Number of parameters per group - 29.

Table 2.7: Preliminary Regressions

	Same/competing chain entry			
	I	II	III	IV
Const	-0.12	-0.01	-0.06	-0.17
	(0.22)	(0.21)	(0.21)	(0.22)
Apotek 1	0.50**	0.50***	0.54***	0.49**
	(0.19)	(0.18)	(0.17)	(0.19)
Boots	0.12	0.04	0.09	0.08
	(0.19)	(0.18)	(0.18)	(0.19)
Vitus	0.15	0.10	0.15	0.15
	(0.20)	(0.19)	(0.18)	(0.20)
Ditt apotek	0.15	0.07	0.06	0.09
	(0.20)	(0.19)	(0.19)	(0.20)
Center	0.12	0.11	0.12	0.15
	(0.14)	(0.13)	(0.13)	(0.15)
Small mall	0.33	0.41*	0.60**	0.72**
	(0.24)	(0.22)	(0.26)	(0.27)
Wine mon. in small mall	-0.25	-0.19	-0.20	-0.28
	(0.25)	(0.22)	(0.22)	(0.27)
Large mall	0.15	0.08	-0.03	0.06
	(0.15)	(0.14)	(0.14)	(0.15)
Wine mon. in large mall	0.04	-0.00	-0.02	0.05
	(0.16)	(0.15)	(0.15)	(0.16)
Market size	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Residuals prior 3 months		0.13***		
		(0.03)		
Residuals prior 6 months			0.12***	
			(0.04)	
Residuals prior 12 months				0.13**
				(0.06)
Observations (entries)	53	53	53	53

Note: Observation is an entry case of monopoly-duopoly transition

Chapter 3

A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers

Abstract: Our novel approach to modeling monopolistic competition with heterogeneous firms and consumers involves spatial product differentiation. Space can be interpreted either as a geographical space or as a space of characteristics of a differentiated good. In addition to price setting, each firm also chooses its optimal location in this space. We formulate conditions for positive sorting: more productive firms serve larger market segments and face tougher competition; and for the existence and uniqueness of the equilibrium. To quantify the role of the sorting mechanism, we calibrate the model using cross-sectional haircut market data and perform counterfactual analysis. We find that inequality in the distribution of the gains among consumers caused by positive market shocks can be substantial: the gains of consumers from more populated locations are 3-4 times higher.

This chapter is written together with Sergey Kokovin, Shamil Sharapudinov, Alexander Tarasov, and Philip Ushchev.

3.1 Introduction

Ever since Dixit and Stiglitz (1977), monopolistic competition has been a workhorse model in international trade, economic geography, growth, and macroeconomics. A large literature on monopolistic competition[1] demonstrates the important role of firm heterogeneity in determining general-equilibrium outcomes and in explaining a broad array of empirically observed phenomena (Melitz, 2003; Chaney, 2008; Zhelobodko et al., 2012; Mrazova and Neary, 2017; Dhingra and Morrow, 2019; Matsuyama and Ushchev, 2022). At the same time, little attention has been paid to the role of consumer heterogeneity and the interplay between heterogeneous demand and heterogeneous supply under monopolistic competition (which can be, for instance, crucial for policy analysis). We seek to narrow this gap in the literature and to make one more step towards understanding the implications of this two-sided heterogeneity in a free entry equilibrium framework.

In this paper, we develop a novel theory of monopolistic competition with bilateral heterogeneity: (i)*horizontal heterogeneity* of consumers in their spatial locations (where the space can be interpreted as either a geographical space or a product space); (ii)*vertical heterogeneity* of firms in productivities. The distribution of consumers in space is one-dimensional, symmetric, and unimodal, with a compact support. In the geographical interpretation, these assumptions capture the idea of a "monocentric city", in which population density is higher towards the city center. In the product-space interpretation, in which the horizontal heterogeneity across consumers becomes taste heterogeneity, these assumptions capture the idea of "popularity": the product type located at the origin is

[1]See Thisse and Ushchev (2018) for a recent survey.

the most popular among consumers, while the endpoint locations are the least popular. In modeling firm behavior, our major departure from traditional Melitz-type models of monopolistic competition with variable elasticity of substitution is that, apart from setting the profit-maximizing price, each active firm chooses its location in the product space.[2] This new dimension of firm behavior can be considered as either a geographical location choice or a product niche choice, i.e., which group of consumers (defined by their common tastes) to serve.[3]

Each firm's location choice entails the following trade-off. On the one hand, a more popular niche results in a higher demand for the firm's product and, thereby, in a potentially higher profit. On the other hand, assume that all active firms choose to serve the most popular niche. Then, the local competitive pressure there becomes so high that incentives arise to switch to less popular but less competitive niches. To sum up, each firm compromises between access to a *larger local market* and *softer local competition*. Or, as in our epigraph, a firm (a fox) wishes to "hunt" for numerous consumers (chickens) but tries to avoid fierce competitors

[2]Recent work on monopolistic competition with variable elasticity of substitution (see, for instance, Behrens and Murata 2007) has pointed out that not only this model is tractable but also flexible and capable of explaining a broad array of empirically observed phenomena, e.g. variable markups (Bellone et al. 2014) and incomplete pass-through (De Loecker et al. 2016).

[3]Our paper is obviously not the only one that considers a monopolistically competitive setup, in which a product has more than one "dimension". An additional dimension is often associated with the product's quality/appeal. In particular, the growing literature extends a monopolistic competition framework allowing firms to choose both price and quality of their varieties (see e.g. Feenstra and Romalis 2014; Kugler and Verhoogen 2012). Although our paper abstracts from many specific issues discussed in these studies, it complements this literature by providing a fairly general yet parsimonious model and discussing general conditions under which assortative matching between firm productivities and product characteristics – whether it is a product niche or some other product/consumer-specific attribute – can occur.

(hunters). Such a setup provides new insights on the equilibrium outcomes of monopolistic competition models (for instance, the distribution of firm sales, prices, markups, etc.), which standard representative-consumer-based models fail to deliver. Moreover, it enables us to explore the interaction between two very different aspects of product differentiation: (i) the *hedonic*aspect (see Rosen 1974) and (ii) the *market power* aspect.

We then ask what patterns of equilibria may arise in this new setting. As the baseline model, we consider the case with *fully localized competition* in which firms serve only those consumers for whom their products are the most preferred ones. Although this simplification assumes away direct spatial competition among firms, there is still *indirect* spatial competition channeled through the general equilibrium mechanism. Moreover, it is in line with recent evidence that households tend to concentrate their spending on a few preferred products that vary across households (see, for instance, Neiman and Vavra 2019). In our analysis, we do not impose any parametric restrictions on the functional forms of consumer utility or population density. We find that if the price elasticity of demand is decreasing with consumption[4] (the Marshall's Second Law of Demand), then (i) the equilibrium always exists, and (ii) all equilibria exhibit positive assortative matching - more productive firms choose larger local markets. If, in addition, the population density is log-concave, then the equilibrium is always unique. Note that the matching between firms and market niches explored in the present paper has important implications for the distribution of a firm's sales, prices, and markups and may result in a deeper understanding of data: in particular, a firm may be

[4]This case is often viewed as the most relevant one in monopolistic competition with variable elasticity of substitution. See, e.g., Zhelobodko et al. (2012), Dhingra and Morrow (2019).

smaller than another, not only because it has higher costs of production, but also because it is forced to take a narrower market niche.

Another implication of our theory is that markups can vary non-monotonically across the space. As a result, the relationship between firms' markups and productivities can be non-monotonic as well. Specifically, we prove that, under some non-restrictive conditions, the markups are highest in the most populated locations (where the most productive firms are located) and in the least populated ones (where the least productive firms are located). This result on markups differs from that in models of "spaceless" monopolistic competition (see, for instance, Zhelobodko et al. 2012), where firms' markups increase with their productivity. Our non-monotonicity result is driven by the interplay of two forces: firm heterogeneity and consumer heterogeneity. If firms were homogeneous, then the markup distribution would follow the spatial distribution of local competitive toughness. Since less popular niches exhibit lower competitive pressure, markups there are higher. In other words, to compensate for lower demand in more "remote" locations, homogeneous firms would charge higher prices there. However, because firms are actually heterogeneous, positive assortative matching drives less productive firms further away from denser locations. Since less productive firms charge, ceteris paribus, lower markups, positive assortative matching creates another component in the markup distribution, which decreases with the distance from the densely populated but extremely competitive niche – the origin. As a result, the markup distribution appears to be non-monotonic over the space. This pattern of the markup behavior is consistent with empirical findings in Díez, Fan and Villegas-Sánchez (2021), which document a U-shaped relationship between firm size and markups employing a firm-level dataset on private and listed firms from 20 countries. Moreover, in the data we use to calibrate the model, the relationship between markups and productivities seems to be slightly non-monotonic as well.

Next, we calibrate the model to assess the quantitative distributional consequences of different shocks on consumer welfare. In particular, we use cross-sectional data on the haircut market in Bergen, Norway. The city has a distinct central area with the highest population density, which declines as we move further from the city center. The haircut market closely corresponds to the assumptions made in the monopolistic competition framework (see also Asplund and Nocke 2006, which employs the data on the haircut market in Sweden). Moreover, the dataset we use provides a number of variables we need to calibrate the model. Specifically, in addition to the distribution of population in the city, we observe locations, turnovers, and profits of hairdressers in the sample. The latter allows us to back out the distribution of firm productivity without relying on the structure of the theoretical model.

We find that the model performs quite well in fitting the relationships between firms' prices/markups and productivities in the data (these two moments in the data are not directly targeted by the calibration procedure), capturing, in particular, the potential non-monotonic pattern of the markups. We then perform two counterfactual experiments: a 20% proportional increase in the population density and setting the fixed cost of production to zero (a policy aimed to facilitate entry into the market and/or to reduce exit). In both experiments, we observe that more firms enter the market, increasing the level of competition in each city location. This, in turn, changes the matching pattern: firms relocate to less populated locations, and the range of served locations expands.

We also find that consumers gain from these changes in the parameters. However, the gains are not equally distributed across consumers. Our quantitative analysis shows that consumers located closer to the city center gain 3-4 times more than those in more remote locations. This difference in the gains seems substantial and emphasizes the quantitative importance of the sorting mechanism explored in the paper. Interestingly, Bau (2019) documents that a rise in the competition level between schools in Pakistan raises the level of inequality in learning test scores benefiting strong students relatively more compared to poorly performing students. Though this empirical fact is related to a different story, it resembles our quantitative results.

Note also that in our theoretical framework, a proportional rise in the population density can be interpreted as the effects of frictionless trade with a similar country.[5] As we find, such a change increases the range of served niches/locations in the equilibrium. This finding is in line with patterns in the trade data. In particular, Fieler and Harrison (2019) find that one of the implications of tariff reductions on manufacturing in China in 1998-2007 was the introduction of new products. Also, our theory is potentially in line with findings in Holmes and Stevens (2014), which show that in the US, smaller firms are less affected by competition with China as they produce custom or specialty goods. As foreign exporting firms are typically more productive, in our framework, they choose more populated niches with a weaker impact on firms located in less populated niches (that can be interpreted as custom or specialty product types).

[5]Non-uniform gains from trade are explored in a number of studies (see, for instance, Nigai (2016), who assumes away the standard assumption about a representative consumer). In these papers, consumers are typically different in terms of their income.

Finally, we relax the assumption of fully localized competition and consider a more general case in which firms have a non-zero-measure range of service. By doing so, we allow consumers to purchase product types different from their most preferred ones. This comes at a cost: given other things equal, the utility derived from consuming product types different from the most preferred one is lower and negatively related to the distance between the product types (as in the Hotelling model). In other words, besides monopolistic competition, we consider direct spatial competition among firms (Hotelling 1929*b*; Kaldor 1935; Lancaster 1966; Beckmann 1972; Rosen 1974; Salop 1979). Although a complete analytical characterization of equilibria is a prohibitively complex task in this case, we are able to describe some properties of the equilibrium (provided that it exists). We find that more productive firms charge lower prices and produce larger volumes. More importantly, we show that if the firm's profit function (as a function of the firm's productivity, location, and price) is supermodular in location and price, then each equilibrium displays positive assortative matching.

3.2 Literature Review

Our paper contributes to at least three important strands of literature. First, it adds to the literature that analyzes markets with spatially distributed consumers (see, e.g., Lancaster, 1966; Salop, 1979; Chen and Riordan, 2007; Vogel, 2008). Regarding this literature, it is important to stress fundamental differences between our framework and standard spatial competition approach. Indeed, although the space is described as a one-dimensional interval, which is akin to Hotelling (1929*b*), we assume that consumers (i) buy in volume, and (ii) exhibit love for variety. This leads to a very different demand structure compared to Hotelling-type setups. Another distinctive feature of our approach is that monopolistically competitive firms make decisions on entry, price, and location. To the best of our knowledge, no existing market competition model captures a similarly rich pattern of firm behavior. Our setup allows studying the interactions between two types of heterogeneity: on the firm side and on the consumer side; which have been considered separately, but not together, in the spatial competition literature. In particular, Vogel (2008) considers a model of spatial competition, where heterogeneous firms strategically choose locations and prices. However, since Vogel (2008) assumes a uniform distribution of consumers, more productive firms end up facing less elastic residual demand curves. In our model, the pattern of demand elasticities firms face in equilibrium is bell-shaped w.r.t. firm's productivity (see Proposition 4 and the corresponding discussion in Subsection 2.5). Loertscher and Muehlheusser (2011) consider a sequential location game among homogeneous firms in a space with unevenly distributed consumers. These authors show that locations with a higher population density attract more firms. However, since there is no price competition in the model and firms differ only with respect to when they can enter, their model does not allow comparisons of more productive firms versus less productive firm behavior or study the equilibrium markup patterns. Goryunov, Kokovin and Tabuchi (2022) consider a monopolistic competition framework with spatially distributed consumers. However, in contrast to the present paper, this work focuses on the case of homogeneous firms and uniformly distributed consumers. As a result, it does not examine the sorting of firms across product niches. Another paper related to ours is Ushchev and Zenou

(2018), which develops a model of price competition in product-variety networks. Both consumers and suppliers of a differentiated product are embedded into a network that captures proximity between product varieties: two varieties are linked to each other if they are close substitutes; otherwise, no link exists. Each consumer's location is her most preferred variety, while her willingness to pay for other varieties decays exponentially with their geodesic distance (induced by the network) from her most preferred variety. Like in most of the network literature, the network structure of the economy is assumed *fixed*. Therefore, Ushchev and Zenou (2018) abstracts from niche choices of firms and spatial sorting.

Second, our paper is related to the literature on spatial selection/sorting of heterogeneous firms. One of the most related papers is Nocke (2006) who considers sorting of heterogeneous firms across imperfectly competitive markets of different sizes. He finds a similar outcome - more productive firms choose to locate in larger markets. However, our paper differs in at least two aspects. We tackle sorting between firms and product niches in a continuous fashion, somewhat similar to continuous economic geography in Allen and Arkolakis (2014). More importantly, Nocke (2006) mainly focuses on sorting per se, while we consider a free entry equilibrium framework with monopolistic competition analyzing its existence and uniqueness and exploring its implications for markups and consumer welfare. Among other studies, Okubo, Picard and Thisse (2010) explores how trade liberalization affects sorting across locations in a two-country model with linear demand. Behrens, Duranton and Robert-Nicoud (2014) construct a model of selection of talented individuals across ex-ante homogeneous cities.[6] Gaubert (2018) develops a quantitative model of sorting of heterogeneous firms across cities where a firm's choice depends on local input prices and agglomeration externalities. Faber and Fally (2020) document that more productive firms endogenously sort into serving the taste of richer households, implying asymmetric effects on household price indices. Our paper complements this strand of the literature by focusing in more detail on the selection of firms across product niches in a quite general setup with continuous space. Carballo, Ottaviano and Martincus (2018) empirically study self-selecting of firms into specific foreign market niches, but their approach to modeling product space is very different from ours. There is some similarity of our approach with Eckel and Neary (2010) who develop a model of flexible manufacturing with the core competence of every firm.[7] However, the sorting of firms is not addressed in this paper. Finally, the present paper also complements the literature on the role of consumer heterogeneity in monopolistic competition and its implications for the distribution of the gains from trade.[8] Focusing on horizontal consumer heterogeneity that assumes away income effects, our paper provides a new rationale for the unequal distribution of the gains from trade. Another related paper is Sharapudinov (2022),

[6]See also Behrens and Robert-Nicoud (2015) for a survey.

[7]In Eckel and Neary (2010), each firm chooses the product line to produce based on the market conditions and competition with other firms. In our paper, each firm produces just one product but decides about its location in the product/geographical space.

[8]Among this large literature, Fajgelbaum, Grossman and Helpman (2011) and Tarasov (2012) develop models of international trade with income heterogeneity and non-homothetic preferences. Osharin et al. (2014) consider a model of monopolistic competition where the elasticity of substitution between any pair of varieties is consumer-specific. Nigai (2016) considers a quantitative trade model with heterogeneous (in income and preferences) consumers and shows that the assumption of a representative consumer may overestimate (underestimate) the welfare gains from the trade of the poor (rich).

which explores the implications of costly international trade between countries in a general equilibrium setup with matching between heterogeneous firms and various product markets/niches.

The rest of the paper is organized as follows. In Section 3.3 we develop a baseline model of fully localized spatial monopolistic competition with an unspecified functional form of consumer demand. In Section 3.4, we calibrate our baseline model using detailed cross-sectional data on the haircut market in Bergen, Norway, and study the distributional consequences of various shocks on consumer welfare. In Section 3.5, we discuss an extension of our model to the case when firms compete not just within but also across niches. Section 3.6 concludes.

3.3 Baseline Model

In this section, we develop a model of a closed economy, which blends the features of monopolistic competition à la Melitz (2003) with the characteristics approach to product differentiation developed by Lancaster (1966). This model allows us to study the role of interactions between two very different facets of product differentiation: (i) the *hedonic* aspect: the price of a certain type of product depends on its type-specific characteristics (possibly including the geographical location where it is supplied) (Rosen 1974); and (ii) the *market-power* aspect: because varieties are differentiated, pricing above marginal cost need not result in losing all the customers. In the model, the demand for a certain type of product is not only affected by its price, but also by the "location" of the product in the space of product characteristics. As a result, each firm chooses both price and location. In this context, a firm's location choice means targeting a certain market segment (taking into account its size and the level of competition).

Product Space and Demand

Spatial structure. The space X, which can be interpreted either as a geographical space or a product space, is one-dimensional and represented as a real line: $X \equiv \mathbb{R}$.[9] Let $l(x) \geq 0$ be the population density at location $x \in X$, and denote by $L \equiv \int_X l(x)dx$ the total population in the economy. We assume that the population density is continuously differentiable (except, possibly, at the origin), symmetric w.r.t. the origin, decreasing with the distance from the origin, and has compact support $[-S, S]$, where $S > 0$. In the geographical interpretation, this means that we are considering a spatial structure similar to a "monocentric city" with a negative density gradient. In the product-space interpretation, this means that product types are ordered by "popularity" in the descending order: product type $x \in X$ is preferred by more consumers than product type $y \in X$ if and only if $|x| < |y|$. In this context, we refer to $l(\cdot)$ as the *spatial distribution of consumer tastes*, which we use interchangeably with "population density" in what follows. We do so both for brevity and for the sake of exploiting the intuitive appeal of Hotelling's spatial metaphor.

In our baseline model, each consumer located at x values only varieties supplied at location x. This is the case of *fully localized* competition: varieties compete for consumer's attention *within but not across locations*. The reason for introducing this assumption

[9] In the product-space interpretation, each point in space corresponds to a certain type of product, so that consumer's location $x \in X$ represents her most preferred product type. This bears some resemblance with the *ideal variety* concept introduced by Hotelling (1929*b*). We refrain from using the term "ideal variety" to avoid confusion: in our model, a variety is something different from a product type, as each type of product available on the market is represented by a continuum of varieties.

is that price competition among firms can be described as an aggregative game (Anderson, Erkal and Piccinin, 2020), which makes the analysis of firm behavior and equilibrium characterization relatively simple. In Section 3.5, we discuss the consequences of relaxing this assumption.

The utility function of a consumer located at x is given by

$$\mathcal{U}_x = V\left(\int_{\omega\in\Omega_x} u(q(\omega,x))\mathrm{d}\omega\right) + q_0. \tag{3.1}$$

where Ω_x is the set of varieties of type x, $q(\omega,x)$ is the individual consumption volume of a specific variety $\omega \in \Omega_x$ by a consumer located at x, and q_0 is the consumption of the outside good produced in a perfectly competitive market under constant returns to scale, which we choose to be the numeraire. The function $V : \mathbb{R}_+ \to \mathbb{R}$ is an upper-tier utility function, which captures the substitutability between the differentiated good and the outside good, while $u : \mathbb{R}_+ \to \mathbb{R}$ is a lower-tier utility function, which captures the substitutability between varieties of the differentiated good. We make the following assumptions:

Assumption 1. The upper-tier utility $V(\cdot)$ is sufficiently differentiable, satisfies $V'(\cdot) > 0$ and $V''(\cdot) < 0$, and has a finite choke price: $V'(0) < \infty$.

Assumption 2. The lower-tier utility $u(\cdot)$ is sufficiently differentiable, and satisfies the conditions $u'(\cdot) > 0$, $u''(\cdot) < 0$, $u(0) = 0$, and $u'(0) < \infty$.[10]

A consumer located at $x \in X$ seeks to maximize her utility (*3.1*)

[10]The last condition in Assumption 2, $u'(0) < \infty$, is equivalent to say that the individual demand schedule generated by the lower-tier utility $u(\cdot)$ has a finite choke price.

subject to the budget constraint given by

$$\int_{\omega \in \Omega_x} p(\omega)\, q(\omega, x)\, \mathrm{d}\omega + q_0 \leq I,$$

where $p(\omega)$ is the market price for variety $\omega \in \Omega_x$, while I is consumer's income. Assuming that I is sufficiently high, the consumer's utility maximization problem can be restated as follows:

$$\max_{q(\cdot,x)} \left[V\left(\int_{\omega \in \Omega_x} u(q(\omega, x)) \mathrm{d}\omega \right) - \int_{\omega \in \Omega_x} p(\omega) q(\omega, x) \mathrm{d}\omega \right]. \quad (3.2)$$

The individual inverse demand for each variety $\omega \in \Omega_x$ follows from the consumer's FOC:

$$p(\omega) = \frac{u'(q(\omega, x))}{\lambda(x)}, \quad (3.3)$$

where $\lambda(x)$ is a product-type specific demand shifter defined by

$$\lambda(x) \equiv \frac{1}{V'\left(\int_{\omega \in \Omega_x} u(q(\omega, x)) \mathrm{d}\omega \right)}. \quad (3.4)$$

The local aggregator $\lambda(x)$ can be viewed as a measure of *local competitive toughness* associated with the market segment $x \in X$: a higher $\lambda(x)$ means a downward shift of the demand schedule for each particular variety $\omega \in \Omega_x$.

Solving (3.3) for $q(\omega, x)$, we obtain the individual Marshallian demand of an x-type consumer — i.e. a consumer whose preferred product type is x — for variety ω:

$$q(\omega, x) = D(\lambda(x) p(\omega)), \quad (3.5)$$

where $D(\cdot)$ is the downward-sloping individual demand schedule defined by

$$D(z) \equiv \begin{cases} u'^{-1}(z), & \text{if } z < u'(0), \\ 0, & \text{otherwise}, \end{cases} \tag{3.6}$$

for all $z > 0$.

Since location x hosts $l(x)$ identical consumers, (3.5) implies that the market demand $Q(p(\omega), x)$ for variety $\omega \in \Omega_x$ is given by

$$Q(p(\omega), x) \equiv q(\omega, x)\, l(x) = D(\lambda(x)p(\omega))\, l(x). \tag{3.7}$$

As can be seen from equation (3.7), the market demand at x is affected by two demand shifters: the population density $l(x)$, which plays the role of a vertical shifter, and the local toughness of competition $\lambda(x)$, which plays the role of a horizontal shifter.

Firms

The supply side in the model follows Melitz (2003). Each firm is single-product, i.e. it can produce, at most, one variety. The only factor of production is labor, one unit of which is inelastically supplied by each individual.

The timing of the game among firms is as follows. First, to enter the market, firms pay a sunk entry cost equal to $f_e > 0$ units of labor and draw their marginal cost $c > 0$ from an absolutely continuous univariate distribution described by a differentiable cdf $G : [c_{\min}, \infty) \to [0, 1]$, or, alternatively, by a pdf $g(\cdot)$ defined by $g(c) \equiv G'(c)$ for any $c > c_{\min}$. Here $c_{\min} \geq 0$ is the marginal cost of the most efficient firm.[11] In what follows, we call a firm whose draw is c, a *c-type* black firm. Second, based on their draws

[11] We assume that, for the case when $c_{min} = 0$, the aggregates in the model are well defined.

of c, firms decide whether to stay in business or exit by assessing their operating profits and comparing them with the fixed production cost equal to $f > 0$ units of labor. Third, the active firms (i.e. those who decided to stay in business) choose their profit-maximizing locations, taking the pattern $\lambda(\cdot)$ of local competitive toughness as given. Fourth and last, the active firms choose their profit-maximizing prices. It is worth noting that, as there are no strategic interactions among firms in the model, the corresponding first-order conditions are the same as in the case when firms choose price and location simultaneously (due to the envelope theorem).

Using equation (3.7) for the market demand, we obtain firm ω's profit function:

$$\Pi(p, x; c(\omega)) \equiv (p - c(\omega))Q(p, x) = (p - c(\omega)) D(\lambda(x)p) l(x),$$

where $c(\omega)$ is the marginal cost of the firm, while p and x are, respectively, price and location choices. Up to a zero-measure subset of firms, pricing and location decisions of any two firms, ω and ω', of the same type, i.e., such that $c(\omega) = c(\omega')$, will be identical. Hence, it is legitimate to re-index firms so that they are indexed by their type c. As a result, it suffices to consider c-type firm's operating profit:

$$\Pi(p, x; c) \equiv (p - c) D(\lambda(x)p) l(x), \tag{3.8}$$

Note that since $l(x)$ has the property of mirror symmetry w.r.t. the origin, firms are indifferent between locating at x and locating at $-x$ for every $x > 0$. Hence, it is natural to focus on equilibrium configurations where both the firm's location pattern and spatial pattern $\lambda(x)$ of competitive pressure are also mirror-symmetric w.r.t. zero. Therefore, without loss of generality, we only consider locations $x \geq 0$ from now on. In other words, we assume that the space X is represented by $[0, S]$ interval.

Let $p(c)$ and $x(c)$ be, respectively, c-type firm's profit-maximizing price and location choice:

$$(p(c), x(c)) \equiv \arg\max_{(p,x)} \{ \Pi(p, x; c) \mid p \geq c,\, x \geq 0 \},$$

and let $\pi(c)$ stand for the c-type firm's maximum profit:

$$\pi(c) \equiv \Pi(p(c), x(c); c).$$

Using (3.8) and the envelope theorem, we get:

$$\pi'(c) = -D\left(\lambda(x(c))p(c)\right) l(x(c)) < 0,$$

hence, more productive firms earn higher profits. A c-type firm chooses to produce if and only if $\pi(c) \geq f$. If, in addition, we can guarantee that $\pi(c_{\min}) > f > \pi(\infty)$, then the equation $\pi(c) = f$ has the unique solution $\bar{c} > c_{\min}$. Following the literature, we call $\bar{c}$ the cutoff cost. In other words, $\bar{c}$ is the marginal cost of the least productive active firm, which is indifferent between producing and non-producing.

Sorting between Firms and Locations

In this section, we show that, under a quite general assumption about the lower-tier utility $u(\cdot)$, firms that choose internal locations, $S > x(c) > 0$, are completely sorted across the locations: less productive firms choose to locate further from zero. In other words, $x\left(c\right)$ is increasing in c.

For each active firm type $c \in [c_{\min}, \bar{c}]$, the profit-maximizing price and location choices $(p(c), x(c))$ solve the firm's FOCs, $\Pi_p = \Pi_x = 0$. The FOC w.r.t. price, $\Pi_p = 0$, can be written as follows:

$$\frac{p-c}{p} = \frac{1}{\mathcal{E}_D\left(\lambda(x)p\right)}, \tag{3.9}$$

where $\mathcal{E}_D(\cdot)$ is the price elasticity of demand,

$$\mathcal{E}_D(z) \equiv -\frac{zD'(z)}{D(z)}.$$

Equation (3.9) is the standard monopoly pricing condition. Solving (3.9) w.r.t. p, we obtain the relationship between the price and the firm's location, which we define as $p(x, c)$. Given this relationship, the firm's profit-maximizing location choice is obtained by solving the FOC w.r.t. location, $\Pi_x = 0$, which implies[12]

$$\frac{l(x)}{l'(x)} \cdot \frac{\lambda'(x)}{\lambda(x)} = \frac{1}{\mathcal{E}_D\left(\lambda(x)p(x,c)\right)}. \tag{3.10}$$

Combining (3.9) and (3.10), we derive a neat expression for the markup $\mathcal{M}(x, c)$:

$$\mathcal{M}(x,c) \equiv \frac{p(x,c)-c}{p(x,c)} = \frac{\lambda'(x)}{\lambda(x)} \cdot \frac{l(x)}{l'(x)}. \tag{3.11}$$

The expression for markups given by (3.11) implies the following lemma.

Lemma 1. *If $l(x)$ is strictly decreasing w.r.t. x over $(0, S)$, then in equilibrium $\lambda(x)$ is strictly decreasing over (a, b), where $(a, b) \subseteq (0, S)$ is any interval such that Ω_x is non-empty for every $x \in (a, b)$.*

Proof. If Ω_x is not empty for any $x \in (a, b)$ in equilibrium, then any point x on (a, b) is an optimal location for some firms that stay

[12]In what follows, we assume that $\lambda(x)$ is differentiable.

in the market. The markups set by these firms are strictly positive (since there is a fixed cost of production). From (3.11), positive markups imply that $\lambda'(x) < 0$ on (a, b) (as $l'(x) < 0$ on (a, b)). $\square$

The result of the lemma can be explained by a simple trade-off. Choosing an optimal location, firms face a trade-off between the size of the location and the level of competition there. Decreasing $l(x)$ means that, all else equal, the further is firm's location from zero, the lower is the demand for its product. Hence, if firms find it profitable to locate further from zero, lower demand must be compensated by a lower level of competition at this location, which in turn means lower $\lambda(x)$. The expression in (3.11) also implies that depending on the behavior of the fraction $\lambda'(x)l(x)/(\lambda(x)l'(x))$ (which is, in fact, the ratio of the elasticities of the population and competition measures), markups can, in general, grow or decline with a rise in the distance from the zero location.

Next, we explore how a firm's location choice depends on its type, i.e., the marginal cost of production. It turns out that the necessary and sufficient conditions for spatial equilibria to exhibit positive (or negative) spatial sorting of firms can be expressed in terms of the demand schedule properties. More precisely, the following proposition holds.

Proposition 1. *Assume that $l(x)$ is strictly decreasing in x for all $x \in (0, S)$. If, in addition, $\mathcal{E}_D(\cdot)$ is strictly increasing (decreasing), then, in equilibrium, for all c such that $S > x(c) > 0$, we have: $dx(c)/dc > 0$ (< 0).*

Proof. The proof is based on the log-supermodularity property of the operating profit function. Specifically, we have

$$\log \Pi(p, x, c) = \log(p - c) + \log l(x) + \log D(\lambda(x)p).$$

Thus,

$$\frac{\partial^2 \log \Pi}{\partial p \partial c} = \frac{1}{(p-c)^2} > 0,$$

$$\frac{\partial^2 \log \Pi}{\partial x \partial c} = 0,$$

$$\frac{\partial^2 \log \Pi}{\partial p \partial x} = -\lambda'(x) \frac{d\mathcal{E}_D (\lambda p)}{d\lambda p} > 0 \Longleftrightarrow \frac{d\mathcal{E}_D (\lambda p)}{d\lambda p} > 0,$$

since $-\lambda'(x) > 0$. The above log-supermodularity properties of the profit function result in the statements of the proposition. □

One can readily verify that linear demand has an increasing demand elasticity. Most specifications that are well established in the literature[13] also satisfy this property. It is worth noting that CES demand has a constant elasticity of demand. In particular, the variable profit of a firm can be written as follows:

$$\Pi(c, p(x,c), x) = (p(x,c) - c) l(x) D (\lambda(x) p(x,c)) = \\ = \frac{(\sigma-1)^{\sigma-1}}{(\sigma)^{\sigma}} c^{1-\sigma} \frac{l(x)}{(\lambda(x))^{\sigma}}.$$

Such a profit function implies that, given $\lambda(x)$, all firms (irrespective of their marginal cost) choose the location(s) where $l(x)/(\lambda(x))^{\sigma}$ achieves its maximum on $[0, S]$. This outcome may result in multiple equilibria. Indeed, if there exists an equilibrium with a certain schedule of $\lambda(\cdot)$, then any reallocation of firms across the locations that keeps $\lambda(x)$ the same is also an equilibrium (see more on the equilibrium concept in the model in the next section).

[13]Other examples include the CARA demand system (Behrens and Murata 2007) and Stone-Geary demand system (Simonovska 2015). See Zhelobodko et al. (2012) and Arkolakis et al. (2018) for more examples.

Note also that the presence of the numeraire good assumes away income effects on consumption, firms' prices and locations, etc. If the income effects were allowed, then the choice of firm's location would be affected not only by the distribution of location size $l(x)$, but also by the distribution of income among consumers. In this case, different scenarios are possible. For instance, if consumers in more distant and, therefore, less populated locations have also lower income, then we would expect the same assortative matching between firms and consumers as stated in Proposition 1. In other cases, the outcome is ambiguous in general.

Equilibrium

In this section, we describe the free entry equilibrium in our baseline model. We assume that $l(S)$ is sufficiently low. This assumption together with the presence of the fixed cost of production implies that the location of the firm with marginal cost $\bar{c}$, $x(\bar{c})$, always belongs to $[0, S)$. That is, there are some locations (close to S) that are not served by firms (consumers there purchase only the numeraire). This case is of particular interest as it implies one more endogenous margin of production - the set of niches served by firms in the market.

We showed that when the demand elasticity is strictly increasing (see Proposition 1), firms are positively sorted on $(0, S) : dx(c)/dc > 0$. This implies that the most productive firms choose zero as the optimal location: $x(c_{\min}) = 0$. The mass of firms at location $x \geq 0$ is then given by

$$\mu(x) = M_e \, g\left(c(x)\right) \, c'(x),$$

where M_e is the mass of entrants into the economy and $c(x)$ is the inverse function of $x(c)$ and represents the productivity of firms located at x.

Then **an equilibrium** can be described as a following bundle $\left(M_e, \bar{c}, \{\lambda(x),\ p(x,c), x(c)\}_{x\in\Omega, c\in[c_{\min},\bar{c}]}\right)$, such that the ensuing conditions hold:

C1 The measure of competition intensity satisfies:

$$\lambda(x) = \frac{1}{V'\left(\mu(x)u(q(x))\right)}, \tag{3.12}$$

where $q(x) = D\left(\lambda(x)p(x, c(x))\right)$ is the per capita consumption of one variety produced by a firm located at x. As there are no firms located at $x > x(\bar{c}) \equiv \bar{x}$, $\lambda(x) = 1/V'(0)$ for all $x \in (\bar{x},\ S]$. To hold the continuity of the problem, the value of $\lambda(x)$ defined in (3.12) at the rightmost location $\bar{x}$ must be equal to $1/V'(0)$. Equivalently, $c'(\bar{x})$ must be equal to zero.

C2 The schedule of prices, $p(x,c)$, solves with respect to p

$$\frac{p-c}{p} = \frac{1}{\mathcal{E}_D\left(\lambda(x)p\right)}. \tag{3.13}$$

C3 The profit-maximizing location $x(c)$ of a c-type firm solves with respect to x

$$\frac{p(x,c)-c}{p(x,c)} = \frac{\lambda'(x)}{\lambda(x)}\frac{l(x)}{l'(x)}, \tag{3.14}$$

with $x(c_{\min}) = 0$.

C4 The cutoff $\bar{c}$ is determined by the zero-profit condition:

$$\Pi\left(\bar{c}, p(\bar{c}), x(\bar{c})\right) = f. \tag{3.15}$$

C5 The mass of entrants is determined by the free entry condition:

$$\int_{c_{\min}}^{\bar{c}} \left(\Pi\left(c, p(c), x(c)\right) - f\right) \cdot g\left(c\right) dc = f_e. \tag{3.16}$$

Next, we explore the existence and uniqueness of the equilibrium defined above. Note that the above definition of equilibrium implies that the spatial pattern $\{c(x), \lambda(x)\}_{x \in [0, \overline{x}]}$ is described by the following system of differential equations

$$\frac{\mathrm{d}\lambda}{\mathrm{d}x} = -a(x)\lambda\mathcal{M}(x,c),$$

$$\frac{\mathrm{d}c}{\mathrm{d}x} = \frac{1}{M_e}\frac{\left(V'\right)^{-1}(1/\lambda)}{g(c)\,u(q(x))},$$

where $a(x) \equiv -l'(x)/l(x) > 0$ is the rate at which population decreases with the distance $|x|$ from the origin. It is straightforward to show (see Section 3.3) that $\mathcal{M}(x,c)$ and $q(x)$ are functions of $\lambda(x)c$. Thus, the system can be rewritten as follows:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}x} = -a(x)\lambda\mathcal{M}(\lambda c), \tag{3.17}$$

$$\frac{\mathrm{d}c}{\mathrm{d}x} = \frac{1}{M_e}\frac{\left(V'\right)^{-1}(1/\lambda)}{g(c)\,u(q(\lambda c))}. \tag{3.18}$$

Hence, the existence of the equilibrium is, in fact, determined by the existence of the solution of the above system with the following boundary conditions: $c\left(0\right) = c_{\min}$ and $\lambda\left(\overline{x}\right) = 1/V'\left(0\right) \equiv \lambda_{\min}$. In particular, the following proposition holds.

Proposition 2. *If $l\left(S\right)$ is sufficiently low and $l\left(0\right)$ is sufficiently high, then there exists an equilibrium in the model described by the conditions in C1-C5.*

Proof. In the Appendix. □

Sufficiently low $l(S)$ implies that $\bar{x} < S$, while sufficiently high $l(0)$ is necessary to guarantee the positive mass of entrants, M_e, into the market. In the Appendix, we formulate the exact conditions on $l(S)$ and $l(0)$ in terms of the primitives in the model. We also show that, under quite a general condition on $l(x)$, the equilibrium is unique. Specifically, the following proposition holds.

Proposition 3. *Assume that, in addition to the conditions in Proposition 2, $a'(x) \geq 0$. Then, the equilibrium is unique.*

Proof. In the Appendix. □

Notice that $a'(x) \geq 0$ if and only if $l'(x)^2 - l''(x)l(x) \geq 0$.[14] Note that the condition is sufficient meaning that the equilibrium can be unique even when $a'(x) < 0$ for some x.

Distribution of Markups

In this section, we explore how firm markups depend on firm locations and marginal costs of production. To do so, we first express the firm's markups in terms of quantities sold. Specifically, the firm's profit maximization problem can be reformulated in the following way. Given the inverse demand function, a firm maximizes its profit with respect to its location and the quantity per consumer sold at this location, q. Taking into account (3.3), the inverse demand function is given by

$$p(q, x) = \frac{u'(q)}{\lambda(x)}.$$

Hence, a firm's variable profit function can be written as follows:

[14]We need this condition on $l(x)$ to guarantee the uniqueness of the cutoff $\bar{c}$, which is not straightforward in our framework.

$$\Pi\left(c,q,x\right)=\left(\frac{u'\left(q\right)}{\lambda(x)}-c\right)ql(x).$$

This implies that given firm's location x, the quantity per consumer supplied by the firm solves

$$\frac{\partial\Pi\left(c,q,x\right)}{\partial q}=0\Leftrightarrow u'\left(q\right)+q\,u''\left(q\right)=\lambda(x)c. \tag{3.19}$$

Let us define the solution of the above expression as $q(x,c)$: a quantity per consumer sold at x by a firm with cost c. Note that $q(x,c)$ is completely determined by $\lambda(x)c$ and is a decreasing function of $\lambda(x)c$.

Given $q(x,c)$, the firm then chooses its optimal location (in the case, when the optimal location is internal: $x\in(0,\,S)$) by solving:

$$\begin{aligned}\frac{\partial\Pi(q,x,c)}{\partial x}=0\Leftrightarrow\frac{\lambda'(x)}{\lambda(x)}\frac{l(x)}{l'(x)}&=1-\frac{\lambda(x)c}{u'\left(q(x,c)\right)}=\\&=-\frac{q(x,c)\,u''\left(q(x,c)\right)}{u'\left(q(x,c)\right)}.\end{aligned}$$

The latter implies that a firm's markup, $\mathcal{M}(x,c)$, is equal to $\mathcal{E}_{u'}\left(q(x,c)\right)$. Since, $q(x,c)$ is a function of $\lambda(x)c$, $\mathcal{M}(x,c)$ is a function of $\lambda(x)c$. Moreover, if $\mathcal{E}_D$ is increasing in price, $\mathcal{E}_{u'}$ is increasing in quantity. This, in turn, implies that $\mathcal{M}(x,c)$ is a decreasing function of $\lambda(x)c$.

In equilibrium, less productive firms choose locations that are further from zero: $c(x)$ is increasing in x for all $x>0$. At the same time, $\lambda(x)$ is decreasing in x. As a result, $\lambda(x)c(x)$ and, therefore, the markup function can be non-monotonic in x. In fact, the behavior of the markup function in the equilibrium is determined by the interplay of two forces: firm heterogeneity and consumer heterogeneity. In particular, when firms are homogeneous in terms of their productivity and consumers have different locations, the behavior of the markup function is solely determined by $\lambda(x)$, which decreases in x. This implies that the markup function is increasing in x: firms located further from zero set higher markups. Indeed, to compensate for lower demand in more "remote" locations, homogeneous firms charge higher prices there. When firms are heterogeneous, less productive firms choose more remote locations to avoid tougher competition in denser locations. Since less productive firms charge lower markups, the presence of firm heterogeneity adds a decreasing trend in the behavior of the markup function. As a result, the markup function can be non-monotonic. In particular, we can prove the following proposition.

Proposition 4. *1) The markup function $\mathcal{M}(\lambda(x)c(x))$ always locally increases w.r.t. x around $x = \bar{x}$. 2) If $\mid l'(0) \mid < \infty$ and $c_{\min}$ is sufficiently close to zero, then the markup function $\mathcal{M}(\lambda(x)c(x))$ locally decreases w.r.t. x around $x = 0$. 3) Finally, if, in addition, $g'(c) \geq 0$ and $(l'(x)/l(x))'_x \leq 0$, then the markup function, $\mathcal{M}(\lambda(x)c(x))$, has a U-shape on $[0, \bar{x}]$.*

Proof. In the Appendix. □

The first two statements in the proposition mean that the markup function is decreasing around zero (under some restrictions on the parameters) and increasing around $\bar{x}$. The intuition behind that is as follows. Other things equal, lower $c_{\min}$ implies a higher level of firm heterogeneity in the neighborhood of 0 in the equilibrium. When this level is high enough (which is specified in the Appendix), we have the decreasing markup function in the neighborhood of 0, as within the markup shifter $\lambda(x)c(x)$ the second multiplier $c(x)$ changes faster than the other one. In the neighborhood of $\bar{x}$, $c'(x)$ is close to zero, implying a low level of firm heterogeneity there. As a result, the markup function is increasing. Finally, under some additional assumptions on $g(c)$ and $l(x)$, the markup function is globally U-shaped. Note that the assumption on $g(c)$ seems to be natural: it is more likely to get a bad productivity draw than a good one. For instance, a Pareto distribution satisfies this property.

An important implication of the above findings is that, due to the positive sorting in the equilibrium, the relationship between a firm's marginal costs and markups has a U-shape as well. In other words, in the equilibrium, the most and least productive firms set the highest markups, while in traditional models of monopolistic competition with firm heterogeneity, the highest markups are set by the most productive firms only – the relationship between a firm's marginal costs and markups is negative.

Another implication of Proposition 4 is that the demand elasticity $1/\mathcal{M}(\lambda(x)c(x))$ is bell-shaped w.r.t. x. Combining this with our perfect sorting result, we infer that the demand elasticities faced by firms in equilibrium are bell-shaped w.r.t. productivity. This result contrasts with Vogel (2008), who finds that more productive firms end up facing less elastic demands.

Comparative Static: A Proportional Rise in the Population Density

In this section, we analyze the implications of a proportional change in $l(x)$ in all locations: $l^{new}(x) = (1 + \Delta)\, l^{old}(x)$; that can be interpreted as the comparison of equilibrium outcomes between cities with different population sizes or the outcome of free trade between symmetric countries. Without loss of generality, we assume $\Delta > 0$ meaning that the population density uniformly rises.

To explore the effects of the change in $l(x)$, we distinguish between the short-run and long-run effects. This also simplifies understanding of the intuition behind it. By the short-run effects, we mean the implications of the change in $l(x)$ when the mass of entrants, M_e, does not react to changes in $l(x)$. The following lemma holds.

Lemma 2. *Under fixed M_e, a proportional rise in $l(x)$ increases the cutoffs $\bar{x}$ and $\bar{c}$. Given this change in $l(x)$, the values of the functions $\lambda(x)$ and $c(x)$ rise in all locations (only $c(0) = c_{\min}$ does not change).*

Proof. In the Appendix. □

The intuition of the findings above is as follows. All else equal, a rise in the population size implies higher firm profits. As a result, some inefficient firms that did not produce before find it profitable to produce now under a higher level of the population size: $\bar{c}$ rises. Similarly, as some product niches that were not attractive to firms before now become larger and start generating positive profits, $\bar{x}$ rises. Finally, a rise in the number of firms in the neighborhood of $\bar{x}$ leads to a higher level of competition in this region (increasing $\lambda(x)$). As a result, tougher competition forces firms to relocate closer to the origin, implying that $c(x)$ rises in all locations except for $x = 0$.

To analyze the long-run effects, one needs to take into account the corresponding change in M_e and its effects on the equilibrium outcomes. We expect that a uniform rise in the population density leads to a higher value of M_e. Though this outcome is very intuitive (and confirmed by our numerical simulations), under the presence of sorting between firms and product niches, we cannot provide strict proof for this statement. Nevertheless, in the below considerations, we assume that M_e increases. In the proof of the uniqueness of the equilibrium (see Step 4 in the Appendix), we show that a rise in M_e implies that $\lambda(x)$ increases at all locations. Combining this with the results in Lemma 2, we can formulate the following lemma.

Lemma 3. *Given a proportional rise in $l(x)$, if the number of entrants in the equilibrium,M_e, increases under this change in $l(x)$, then the function $\lambda(x)$ shifts upwards implying that the cutoff $\bar{x}$ increases.*

The above lemma implies that a uniform rise in the population size makes some firms choose product niches that were not served before. This is because the short-run and long-run forces work in the same direction with respect to $\lambda(x)$ and $\bar{x}$. In the long run, new entrants induce tougher competition at each location. As a result, less productive firms are forced to move to less populated niches to avoid competition, which in turn increases $\bar{x}$.

Regarding $c(x)$ and $\bar{c}$, the short-run and long-run effects seem to be different. On the one hand, a uniform rise in $l(x)$ shifts $c(x)$ upwards and increases $\bar{c}$ (as stated in Lemma 2 and discussed after). On the other hand, in the long run, there are new entrants that force less productive firms to choose less populated niches and least productive firms to exit: $c(x)$ shifts downwards, and $\bar{c}$ decreases. It appears that it is very complicated to show which effect is stronger in our model. However, we run numerous simulations, and in all of them, the long-run effect is stronger, meaning that a uniform rise in the population density shifts $c(x)$ downwards and decreases the productivity cutoff $\bar{c}$. The latter outcome is in line with results in standard models of monopolistic competition with variable markups: a rise in the market size makes the least productive firms leave the market.

3.4 Calibration

In this section, we calibrate the model to explore the distributional consequences of different shocks on consumer welfare. In doing so, we use cross-sectional data on the haircut market in Bergen, Norway. Bergen is the second-largest city in Norway, with population of around 236000 as of 2021. The city has a distinct central area with the highest population density there, which then declines as we move further from the city center. This is consistent with the assumption about the population density in our model.[15]

We use data on the regular haircut sector for two reasons. First, the haircut industry seems to satisfy, with a reasonable degree of precision, the assumptions we make in the theory part.[16] In our sample, each hairdresser is too small to strategically manipulate the market environment, which makes the monopolistic competition framework an obvious modeling choice.[17] Also, hairdressers are present in most parts of Bergen, which is in accordance with

[15]Bergen is also the most homogeneous in income among large cities in Norway (the Gini coefficient is around 25.9 according to the Statistics Norway).

[16]It is worth noting that Asplund and Nocke (2006) also employ the data on the haircut market, but in Sweden, motivating this by that such a market closely corresponds to the assumptions related to the monopolistic competition framework.

[17]In Norway, there is only one hairdresser chain, Cutters, that runs multiple hairdressers. Specifically, in Bergen, there are 12 Cutters hairdressers. These hairdressers have been excluded from the sample for the following reasons. Their multi-store nature allows us to observe the revenue and profit information only at the chain level and not at the level of each hairdresser, which in turn prevents us from using them in the analysis. Moreover, their "format" differs from the one that regular hairdressers in our sample have. In particular, they offer a drop-in concept of a quick haircut. They are also usually located in large shopping malls, attracting consumers that come in a mall to shop for other goods and services rather than to have a haircut. With the exception of Cutters, other hairdressers in the sample are small (compared to the whole industry) single-product firms.

the assumption of a continuous distribution of firms. Furthermore, we limit our analysis to regular hairdressers that typically offer traditional haircuts homogeneous in quality, which is in line with our focus on horizontal product differentiation. The absence of significant quality differences also suggests that consumers have a haircut in their neighborhood rather than in a more distant hairdresser, which substantiates our assumption of fully localized competition. Moreover, a haircut is tied to the location of a hairdresser and cannot be "delivered", which is in accordance with the absence of shipping costs in our model. Finally, the regular haircut market is rather free from the income effects, which are not present in our model.

Second, the data set we consider provides several important variables we need to calibrate the model. In particular, in addition to the distribution of population in the city, we observe locations, turnovers, and profits of hairdressers in the sample. The latter allows us to calibrate the distribution of firm productivity employing just the data without relying on the structure of the theoretical model (see details in Subsection 3.2.2).

Data Description

The data that we use for calibrating the model comes from three sources: Geodata, Business Compensation Scheme, and manually collected data on regular haircut prices. We now describe each data source in more detail. The primary data source is the database provided by Geodata (2020)[18], the primary Norwegian provider of spatial data. The database ("Bedriftsregister") contains information for the period 2015-2020 on all businesses registered in Nor-

[18] Geodata. 2020. Bedriftsregister database.

way, including location, turnover, profit, and some store characteristics. We then use the Standard Industrial Classification (SIC 2007) to select hairdressers. Specifically, we consider all firms that fall into the ”96.020 Hairdressing and other beauty treatment” code. Further, we keep only firms specializing in haircuts rather than in beard grooming, nail care, or other beauty treatments, using the information on the corresponding websites or Facebook pages. As a result, our final sample for the city of Bergen contains 116 hairdressers, for which we observe yearly data on revenues. Data on profits are available only for 86 firms. We replace the missing data on profits by employing a standard imputation procedure.[19] To calibrate the model, we employ revenues and profits for 2019.

The other important data source became available due to the BRC (2020)[20] – a part of the measures introduced by the Norwegian government to support firms facing significant losses due to the Covid-19 crisis. The scheme was introduced in March 2020 and lasted until October 2021. It allocated grants to firms that were subject to a decrease in their turnover of at least 20 percent in March 2020. Since all hairdressers had to be closed due to safety measures, all of them were eligible to apply for this support. Specifically, the Business Compensation Scheme allows us to get some measure of the fixed costs of production associated with the haircut market, which is then used in our calibration procedure. More specifically, firms that applied for the support had to specify their turnover and fixed costs in March 2020 and in the corresponding period one year ago (March 2019). The fixed costs are defined as the costs that cannot be reduced in the short term, together with

[19]In the procedure, the conditional expectation is based on a linear regression with firms’ revenue, distance to the city center, and their interaction.

[20]Brønnøysund Register Centre (BRC). Business Compensation Scheme. Accessed August 18, 2021. https://www.kompensasjonsordning.no/index_en.html.

the firm's activity level. In particular, these costs include the cost of leasing of commercial premises, lighting and heating, rental of machinery, costs for electronic communication, and various financial fees related to accounting, audit, and insurance. In our analysis, we use data on the fixed costs for March 2019 to avoid the effects of the Covid-19 crisis. Note that a strict verification process, which each application had undergone before receiving the support, guarantees the reliability of this data source. To match the firm-level data from the Business Compensation Scheme with the Geodata database, we use an organization number as a unique identifier of a firm.

We also have data on regular female and male haircut prices collected manually, using the information on hairdressers' websites, Facebook groups, or by asking hairdressers directly by phone.[21] We checked the accuracy of the data by physically visiting some of the hairdressers. To construct the price data for 2019, we use the general inflation rate in Norway, which is relatively modest (about 3.9%), and assume that inflation increased prices proportionally among firms. The descriptive statistics for our main variables are presented in Table 3.1.

The demographics data is taken from publicly available databases managed by Statistics Norway and Geonorge (2020)[22] (a public initiative for managing spatial data). To calibrate the distribution of population, we use the division of Norway into the smallest geo-

[21] Kokovin S., Ozhegova A., Sharapudinov S., Tarasov A., and Ushchev P (2023) "Data and code for: A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. http://doi.org/10.3886/E186482V1.

[22] Geonorge. 2020. Befolkning på grunnkretsnivå, accessed August 20, 2021. https://kartkatalog.geonorge.no/metadata/befolkning-paa-grunnkretsniv/7eb907de-fdaa-4442-a8eb-e4bd06da9ca8.

Table 3.1: Descriptive statistics for hair salons

Variable	Mean	SD	Min	Median	Max
Turnover, thous. NOK	2998	3025	195	2180	19538
Profit, thous. NOK	196	210	2	151	1270
Fixed costs, thous. NOK	557	526	118	385	2771
Price for a male haircut, NOK	536	113	250	490	760
Price for a female haircut, NOK	730	143	250	765	1099
Distance to the city center, km	4.1	3.9	0.02	2.6	11.7

graphical unit - *Basic unit* (BU). A BU is a zone defined by Statistics Norway; it is similar to the census blocks used in the US. In Bergen, there are 361 BUs, with the median area equal to 0.28 squared km. Figure 1 in the Appendix shows the BUs in the city of Bergen. For each BU, we count the number of people residing there. We use the Euclidean distance between the city center and the centroid of the corresponding BU for the distance between a certain BU and the city center. The descriptive statistics for basic units are presented in Table 3.2.

Table 3.2: Descriptive statistics for basic units in Bergen

Variable	Mean	SD	Min	Median	Max
Area, sq.km	1.58	4.23	0.01	0.29	52.7
Population	675	522	3	493	4108
Pop. density, people per sq.km	3276	4134	7.9	2100	24526
Distance to the city center, km	4.8	3.5	0	4.5	16.2
Number of firms	0.32	0.96	0	0	8

Calibration Strategy

In this subsection, we describe our calibration procedure.

Population Distribution

We assume that the distribution of consumers in our space is represented by $l(x) = A(1 - (x/S)^{\gamma})$, where $A > 0$ and $\gamma > 0$ are parameters that capture the total population size and the curvature of the distribution. To calibrate the parameters in the distribution function, we employ the distribution of population in Bergen across BUs normalized by the BU areas. This distribution as a function of the BU distance from the city center is presented in Figure 2 in the Appendix.

To calibrate γ, we note that γ is the elasticity of $1 - l(x)/l(0)$ with respect to x. Using the empirical counterpart of $1 - l(x)/l(0)$, where x is the BU distance from the city center and $l(0)$ is the maximum population size across all BUs, we run the corresponding OLS regression and find that the estimate of γ is significant and equal to 0.18. We set S to 16.2 - the distance from the city center to the most remote BU. Finally, we set A to be equal to the maximum value of normalized density in a BU, which is 24526. As a result, our distribution of the population takes the following form: $l(x) = 24526(1 - (x/16.2)^{0.18})$.

Productivity Distribution

To calibrate the distribution of firm productivity, we construct its empirical counterpart employing the data on firm turnovers, profits, fixed costs of production and prices. It is worth noting that, in our sample, hairdressers' revenues and operating profits aggregated at the BU level and normalized by its area are decreasing as functions of the distance of the corresponding BU to the city center: the total revenues and profits are lower in more remote basic units, which is consistent with our theory. The relationship between the hairdresser's fixed costs of production and its distance to the city center appears to be not significant, suggesting that hairdressers' fixed costs of production are barely affected by hairdresser's remoteness.

As a proxy for the price of a haircut, we use the price of a regular male haircut. This is done for at least two reasons. First, a male haircut is a more standardized product than a female one. Second, as we consider regular hairdressers that do not offer nail care or other beauty treatments, it is more likely that the role of male haircuts in determining revenues and profits prevails over that of female ones. Moreover, in our data, prices for male and female haircuts are highly positively correlated (0.79). In fact, our calibration procedure shows that employing prices for female haircuts instead of male ones does not substantially affect the quantitative predictions of the model (see below). The left panel in Figure 3.1 shows the BU-level distribution of the weighted (by revenues) average prices. As can be seen, the slope is positive, which is consistent with the theory but not significant. The right panel in the figure represents the relationship between the prices of male haircuts and distance without averaging prices at the BU level: each dot in the picture represents the price level of a certain hairdresser. As can be inferred, the slope is again positive and significant at the 5% significance level.

The data on revenues, profits, prices, and fixed costs allow us to calculate the marginal costs of production of each hairdresser in the sample.[23] Figure 3.2 depicts the relationship between the hair-

[23]To compute the marginal costs of a firm, we first derive the quantity as the revenue of this firm divided by the price. Then, we find the total variable costs by subtracting the profit and the fixed costs from the revenue. Assuming that marginal costs are constant, we calculate the marginal costs by dividing the total variable costs by the quantity. The markup is then the ratio between the difference in the price and marginal costs and the price. Note that for two hairdressers, we derive

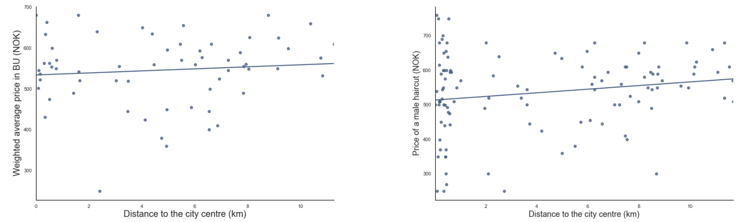


Figure 3.1: Haircut prices in Bergen

Note: The left panel: each dot represents one basic unit of Bergen. The number of dots (which is 62) corresponds to the number of basic units with at least one hair salon. The estimate of the slope parameter is 2.22 with no significance. The right panel: each dot represents the price of a haircut from a certain hairdresser. The estimate of the slope parameter is 5.25 at the 5% level of significance.

dressers' marginal costs and their remoteness from the city center. As can be seen, less productive hairdressers tend to be located further from the city center. This is in line with our theory when we assume the increase in price demand elasticity. Figure 3.3 presents the distribution of the markups across space. One can see that the further a hairdresser is located from the city center, the lower the markup it charges. Recall that, according to our theoretical results, the markup schedule can have a U-shape: the markup function is first decreasing and then increasing in distance. In Figure 3.3, we do not observe such a pattern. However, the relationship between markups and marginal costs (see Figure 3.5) is "closer" to being non-monotonic: markups are first decreasing in marginal costs and then seem to be slightly increasing.

We assume that marginal costs are drawn from a Weibull dis-

negative marginal costs. These observations have been dropped, when calibrating the productivity distribution.

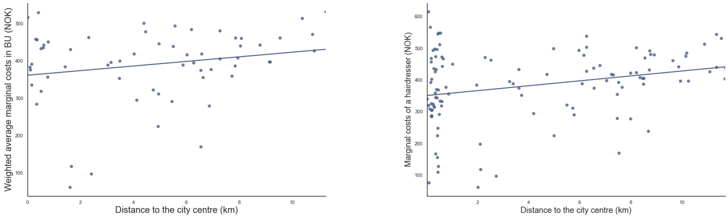


Figure 3.2: Marginal costs

Note: The left panel: each dot represents one basic unit of Bergen. The number of dots (which is 62) corresponds to the number of basic units with at least one hair salon. The estimate of the slope parameter is 6.15 with 10% level significance. The right panel: each dot represents the marginal cost of a certain hairdresser. The estimate of the slope parameter is 7.74 with 5% level significance.

tribution over $[c_{min}, \infty)$:

$$G(c) = 1 - \frac{e^{-\left(\frac{c}{\alpha}\right)^k}}{e^{-\left(\frac{c_{\min}}{\alpha}\right)^k}}, \qquad g(c) = \frac{e^{-\left(\frac{c}{\alpha}\right)^k}}{e^{-\left(\frac{c_{\min}}{\alpha}\right)^k}} \frac{k}{\alpha} \left(\frac{c}{\alpha}\right)^{k-1},$$

where k is the shape and α is the scale parameter. The choice of the Weibull distribution is mainly determined by that the empirical density function is not monotone (see Figure 3.4), which is captured by the Weibull functional form. We set c_{min} to 0.062 - the minimum marginal cost of production in the data measured in thousands of krones. To calibrate the shape and scale parameters, we employ the maximum likelihood (ML) procedure using the empirical distribution of marginal costs. Note that in the data, we observe the conditional distribution of the marginal cost of production as firms with $c > \bar{c}$ exit the market. Therefore, to calibrate k and α, we fit $g(c)/G(\bar{c})$ to the data, where $\bar{c}$ is the marginal cost of the least productive firm in the market, which is 0.615. The ML procedure results in k being equal to 3.8 and α equal to 0.43. Thus,

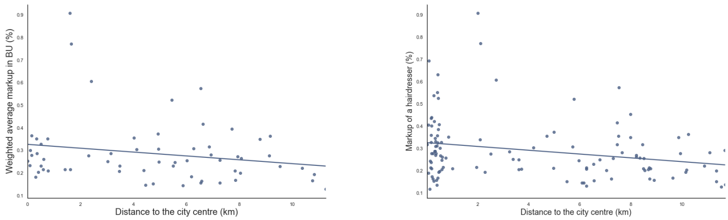


Figure 3.3: Markups

Note: The left panel: each dot represents one basic unit of Bergen. The number of dots (which is 62) corresponds to the number of basic units with at least one hair salon. The estimate of the slope parameter is -0.01 with 10% level significance. The right panel: each dot represents the markup of a certain hairdresser. The estimate of the slope parameter is -0.01 with 5% level significance.

the calibrated distribution function is

$$G(c) = 1 - \frac{e^{-\left(\frac{c}{0.43}\right)^{3.8}}}{e^{-\left(\frac{0.062}{0.43}\right)^{3.8}}}.$$

Figure 3.4 presents the fit of the ex-post density function to its empirical counterpart. As can be seen, the Weibull distribution fits the empirical density function for the marginal costs of production quite well.

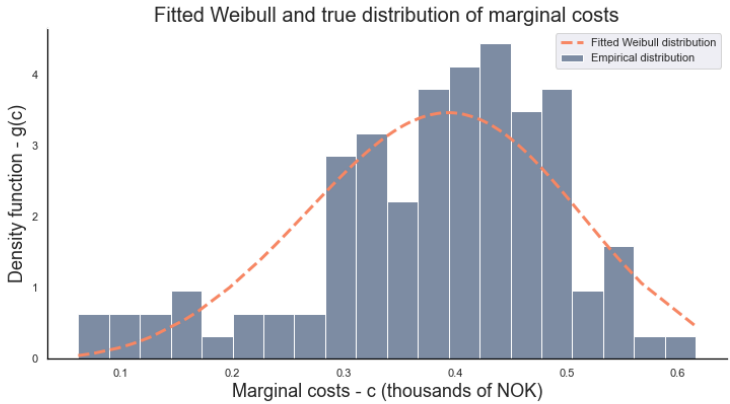


Figure 3.4: Empirical and calibrated distribution of marginal costs
Note: Marginal costs are calculated using prices of male haircuts. One observation is one hairdresser.

Note that if we use prices for female haircuts to calibrate $G(c)$, then the calibrated value of the shape parameter barely changes (it increases from 3.8 to 4.0). The value of the scale parameter changes more substantially (it rises from 0.43 to 0.57). However, the latter change is "quantitatively compensated" by the changes in the calibrated values of the other parameters (see the next subsection).

The Other Parameters

For the upper-tier utility function, we assume $V(x) = \ln(1 + x)$. For the lower-tier utility we choose the quadratic function: $u(q) = q - aq^2/2$; where parameter a is calibrated to match the location cutoff $\bar{x}$, which is 11.7. For the fixed costs of production f, we take the average fixed costs across all hairdressers in the data, which is 553.948 thousand krones. Finally, to calibrate the entry costs f_e, we match the productivity cutoff $\bar{c}$, which is 0.615. Table 3 summarizes our calibration strategy.

Table 3.3: Calibration Strategy

Function	Parameterization	Values from the data	Fitted moment and value
$V(x)$:	$\ln(1+x)$		
$u(q)$:	$q - \frac{a}{2}q^2$		$\bar{x} = 11.7$, $a = 0.107$
$l(x)$:	$A\left(1 - \left(\frac{x}{S}\right)^{\gamma}\right)$	$A = 24526$, $\gamma = 0.18$, $S = 16.2$	
$g(c)$:	$\frac{e^{-\left(\frac{c}{\alpha}\right)^k}}{e^{-\left(\frac{c_{\min}}{\alpha}\right)^k}} \frac{k}{\alpha} \left(\frac{c}{\alpha}\right)^{k-1}$	$k = 3.8$, $c_{\min} = 0.062$, $\alpha = 0.43$	
f_e:			$\bar{c} = 0.615$, $f_e = 2018.4$
f :		553.948	

3.4.1 Results and Counterfactual Analysis

Our calibration strategy results in f_e and a being equal to 2018.4 and 0.107, respectively. To assess how well the model fits the data, we present two figures. Figure 3.5 depicts the relationship between marginal costs and markups in the data and the one generated by the calibrated model. As can be seen, the markup function generated by the model fits the empirical relationship quite well. The model implies, on average, slightly lower markups for more productive firms and higher markups for less productive ones. The average markup generated by the model is 0.31, while the average markup in the data is 0.29. The model also generates non-monotonicity of markups, which does not contradict the pattern in the data. Figure 3.6 stands for the relationship between marginal costs and prices. Again, it can be seen that the model performs well in fitting this relationship. The prices generated by the model are, on average, slightly higher for less productive firms than those in the data and lower for firms with the lowest marginal costs. We also compare the average revenues in the data and those generated by the model. In the data, the revenues per firm are around 3000 thous. krones, the model predicts the average revenues to be around 7159 thous. krones. The fit is not perfect, but taking into account that, when calibrating the model, we do not target the moment related to revenues at all, the difference is not that substantial.

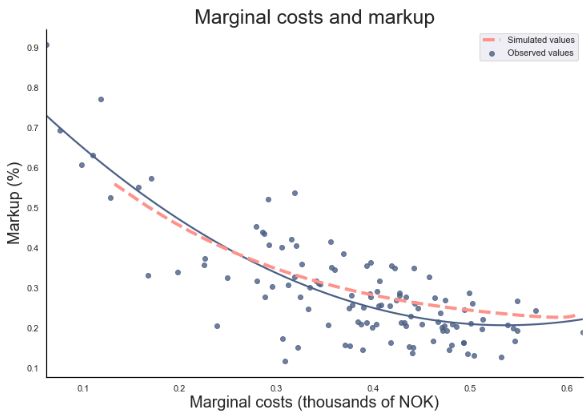


Figure 3.5: Simulated and observed marginal costs and markups

Note: Each dot represents one hairdresser.

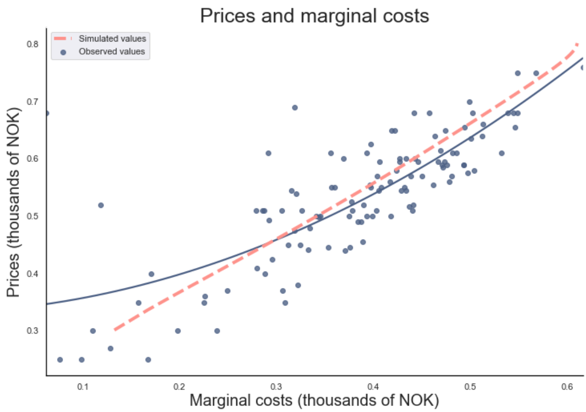


Figure 3.6: Simulated and observed marginal costs and prices

Note: Each dot represents one hairdresser.

Next, we perform two counterfactual experiments. First, we consider a 20% proportional increase in the population density: that is, we increase A by 20%. Second, we eliminate the fixed cost of production by setting f to zero. The latter counterfactual can be interpreted as a policy aimed to facilitate entry into the market and/or to reduce exit. In our experiments, we are mainly interested in the distribution of welfare changes across consumers.

Our quantitative analysis shows that, as discussed in Section 2.6, a proportional increase in the population density, $l(x)$, leads to a higher level of competition in each location resulting in an upward shift of $\lambda(x)$. Tougher competition in the market, in turn, implies tougher selection with a lower cutoff $\bar{c}$ in the new equilibrium. At the same time, the location cutoff $\bar{x}$ goes up, as a higher level of population makes more remote locations attractive for firms. Our experiment shows that the matching function $c(x)$

shifts downward: in the new equilibrium, each location (except the most populated one) is served by more productive firms. We also observe a decrease in the price levels in all locations in the city. However, the impact on firms' markups is non-monotonic. We find that, in the most populated locations, the markups decrease, but the least populated locations experience an increase in markups. The reason behind this outcome is that the sorting effect on markups for these locations is positive and strong enough to compensate for the downward pressure of higher competition on markups.

Finally, we explore the changes in consumer welfare in the economy. Note that the quasi-linear structure of consumer preferences implies that welfare changes can be interpreted as equivalent changes in money income. Figure 3.7 reports the distribution of welfare gains across consumers caused by a 20% proportional rise in the population density. As can be inferred, consumers located closer to the city center gain relatively more than those more "remote" consumers. In particular, the gains around the center are about 33 NOK, while consumers located around the original location cutoff $\bar{x}$ (which is 11.7) gain 3-4 times less, about 8-10 NOK. The relative difference in the gains is quite substantial and, thereby, emphasizes the quantitative importance of the sorting mechanism explored in the paper.

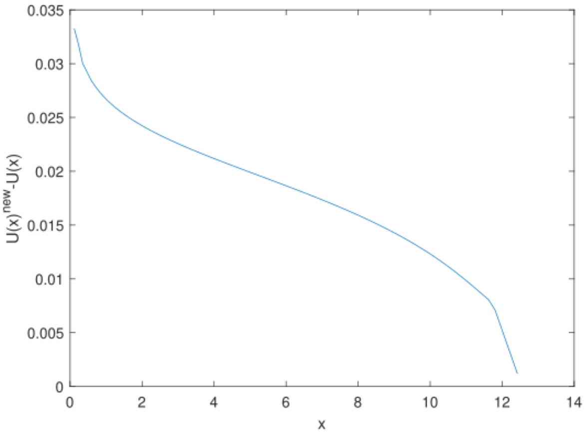


Figure 3.7: Welfare gains: A rise in the population density

Note: Welfare gains for each location are calculated as follows: $\mathcal{U}_x^{new} - \mathcal{U}_x$, where $\mathcal{U}_x$ denotes welfare level under the baseline parameterization, while $\mathcal{U}_x^{new}$ is the counterfactual welfare level under a 20% increase in population density.

In our second counterfactual experiment, we set the fixed costs of production f to zero. Our quantitative analysis shows that, in this case, the level of competition in each location rises: $\lambda(x)$ shifts upwards. At the same time, since we reduce the fixed costs of production, the selection into the market is less tough, implying a higher cutoff $\bar{c}$ and a higher location cutoff $\bar{x}$. Moreover, each location, except for the most populated one, is served by more productive firms in the new equilibrium. This shift in the matching function $c(x)$, together with a higher level of competition, yields a downward shift in prices across the city. Our analysis also shows that, in this experiment, the competitive pressure is strong enough to outweigh the sorting effect and eventually generates a downward shift in markups across the whole city.

As for welfare gains, the more remote locations gain less than those closer to the city center. The pattern of the distribution of the gains across locations is very similar to that derived in the case of a proportion rise in the population density. We find that the gains around the city center constitute about 45 NOK, while consumers located around the original location cutoff $\bar{x}$ (which is 11.7) gain about 15 NOK. It is worth noting that the relative difference in the gains between the central and most remote locations seems to be stable across our experiments - the gains around the city center are 3-4 times higher than those at the "peripheral" locations.[24]

3.5 Extension: Competition across Locations

In this section, we discuss what happens if we relax the assumption of fully localized competition. More precisely, assume that consumers value varieties supplied at locations other than their place of residence and that the appeal of a product type y to a x-type consumer decays with the distance $|x - y|$ between x and y. Under these circumstances, the utility function of a consumer located at $x \in X$ is given by

$$\mathcal{U}_x = V\left(\int_X k_\tau(x,y)\int_{\Omega_y} u(q(\omega,x))\mathrm{d}\omega\mathrm{d}y\right) + q_0, \qquad (3.20)$$

where Ω_y is the set of varieties of niche $y \in X$, $k_\tau(x,y)$ is a spatial discount factor, $q(\omega, x)$ is the individual consumption of variety

[24] In the earlier version of the paper (see Sergey G Kokovin, Shamil Sharapudinov, Alexander Tarasov and Philip Ushchev 2020), we numerically explore how the parameters characterizing the distribution of consumer tastes, A and γ, and the shape parameter of the firm productivity distribution affects the implications of a uniform increase in population.

$\omega \in \Omega_y$ by a consumer located at x (where y may differ from x), while $V(\cdot)$, $u(\cdot)$ and q_0 have the same meaning as in (3.1).

This way of modeling preferences is akin to the model proposed by Ushchev and Zenou (2018), where a consumer's willingness to pay for a variety decreases with the geodesic distance from a consumer to a firm in a product-variety network. However, unlike these authors, we do not assume specific functional forms for preferences and the distance decay patterns. We only impose Assumptions 1-2 from Section 3.3 on $V(\cdot)$ and $u(\cdot)$, respectively. In addition to these, we impose the following assumption about the spatial discount factor $k_\tau(x,y)$:

Assumption 3. The kernel $k_\tau : X \times X \to \mathbb{R}_+$ representing the spatial discount factor in (3.20) has the following structure:

$$k_\tau(x,y) = \tau\psi(\tau|x-y|), \tag{3.21}$$

where $\tau > 0$ is a "transport cost" parameter which captures the decay rate of utility with distance from the most preferred product type, while $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is the distance decay function, which (i) decreases with distance: $\psi'(\cdot) < 0$, and (ii) sums up to one: $2\int_{\mathbb{R}_+} \psi(z)\mathrm{d}z = 1$.

In other words, the family $\{k_\tau\}_{0<\tau<\infty}$ of decay kernels constitutes a *standard mollifier* (see, e.g., Evans 2010, p. 713). To give a few examples, the distance decay function $\psi(\cdot)$ maybe (i) negative exponential: $\psi(z) \equiv \exp\{-z\}$; (ii) Gaussian: $\psi(z) \equiv (2\pi)^{-1/2}\exp\{-z^2/2\}$.

Assumption 3 implies that, when $\tau \to \infty$, we obtain our baseline model (Section 3.3) as the limit case. Indeed, since the distance decay kernel $k_\tau(x,\cdot)$ is a standard mollifier, it converges (weakly) to the *Dirac's delta* with support $\{x\}$.[25] As a result, when $\tau \to \infty$, (3.20) becomes (3.1).

A consumer located at $x \in X$ seeks to maximize her utility (3.20) subject to the budget constraint, which is now given by

$$\int_X \int_{\omega \in \Omega_y} p(\omega)\, q(\omega, x) \mathrm{d}!\mathrm{d}y + q_0 \leq I, \tag{3.22}$$

where $p(\omega)$ is the market price for variety ω of the y-type product, while I is consumer's income. The consumer's utility maximization problem can be restated as follows:

$$\begin{aligned} \max_{q(\cdot)} \Bigg[V \left(\int_X k_\tau(x,y) \int_{\Omega_y} u(q(\omega,x)) \mathrm{d}\omega \mathrm{d}y \right) - \\ - \int_X \int_{\omega \in \Omega_y} p(\omega) q(\omega, x) \mathrm{d}\omega \mathrm{d}y \Bigg]. \end{aligned} \tag{3.23}$$

The individual demand $q(\omega, x)$ is the solution to the consumer's FOC, which now takes the form

$$\frac{p(\omega)}{k_\tau(x,y)} = \frac{u'(q(\omega,x))}{\lambda(x)}, \tag{3.24}$$

where y is the product niche variety ω belongs to ($\omega \in \Omega_y$), while $\lambda(x)$ is the local competitive toughness, which now takes the form

$$\lambda(x) \equiv \frac{1}{V'\left(\int_X k_\tau(x,y) \int_{\Omega_y} u\left(q\left(\omega, x\right)\right) \mathrm{d}\omega \mathrm{d}y\right)}. \tag{3.25}$$

[25]More precisely, we have: $m_\tau \rightharpoonup \delta_x$ as $\tau \to \infty$ were m_τ is the linear functional defined by $m_\tau(\varphi) \equiv \int_X k_\tau(x,y)\varphi(y)\mathrm{d}y$ for any function φ which is continuous over X, while $\rightharpoonup$ stands for convergence w.r.t. the weak topology. The Dirac's delta δ_x concentrated at $x \in X$ is a linear functional defined as follows: $\delta_x(\varphi) \equiv \varphi(x)$ for any function φ which is continuous over X. See Evans (2010) for details.

Solving (3.24) for $q\,(\omega, x)$, we obtain the individual Marshallian demand of an x-type consumer — i.e. a consumer whose preferred product type is x — for variety ω:

$$q\,(\omega, x) = D\left(\lambda(x)\frac{p(\omega)}{k_\tau(x,y)}\right), \tag{3.26}$$

where $D(\cdot)$ is the downward-sloping demand schedule defined by (3.6). To obtain the market demand $Q(\omega, x)$ for variety $\omega \in \Omega_x$, we integrate (3.26) across the product space X with respect to the population density:

$$Q(\omega, x) = \int_X D\left(\lambda(y)\frac{p(\omega)}{k_\tau(x,y)}\right) l(y)\mathrm{d}y. \tag{3.27}$$

Equation (3.27) implies that the shape of the market demand is affected by: (i) the exogenous spatial distribution $l(\cdot)$ of consumers; (ii) the endogenous spatial distribution $\lambda(\cdot)$ of local competitive toughness; and (iii) the spatial discount factor.

Using the market demands (3.27), we obtain the profit of a c-type firm as a function of price and location choices:

$$\Pi(c, p,\, x) \equiv (p - c)\int_X D\left(\frac{\lambda(y)p}{k_\tau(x,y)}\right) l(y)\mathrm{d}y. \tag{3.28}$$

As in Section 3.3, we use the following notation:

$$(p(c), x(c)) \equiv \arg\max_{(p,x)} \Pi(c, p,\, x).$$

We also denote by $Q(c)$ the c-type firm's profit-maximizing production scale:

$$Q(c) \equiv \frac{\Pi(c, p(c),\, x(c))}{p(c) - c}.$$

Proposition 5. *(i) More productive firms produce at larger scales and charge lower prices:*

$$\frac{dp(c)}{dc} > 0, \qquad \frac{dQ(c)}{dc} < 0. \tag{3.29}$$

(ii) More productive firms choose more competitive locations on $[0, S)$ *if and only if the profit is supermodular along the price-location curve:*

$$\Pi_{px}(c, p(c), x(c)) > 0. \tag{3.30}$$

Proof. In the Appendix. □

When $\tau \to \infty$, we fall back to the baseline model of fully localized competition (Section 3.3). Providing full analytical characterization of equilibria and a clear-cut comparative statics for the case when $\tau < \infty$ is problematic. The issue with that case when $\tau < \infty$ is that the supermodularity condition in Proposition 5 cannot be expressed in terms of the primitives of the model, as it is imposed on the reduced form of the profit function. The lack of tractability of the case when $\tau < \infty$ stems from the fact that, as firms compete both within and across locations, the price competition among firms cannot be described as an aggregative game even locally (i.e., within the same location), since the whole schedule $\lambda(\cdot)$ of competitive toughness matters for the individual pricing behavior of each firm. One can clearly see that from the expression (3.28).

3.6 Conclusion

This paper develops a monopolistic competition model that features matching between heterogeneous firms and product niches. Specifically, we formulate a sufficient condition for positive sorting between firms and product niches: more productive firms choose more populated product niches. This outcome provides new insights into the equilibrium distribution of firm sales, prices, and markups that are now explained not only by the comparative costs of these firms but also by the distribution and size of available market niches. Moreover, the positive sorting of firms in the product space implies a new channel through which market shocks can affect the distribution of welfare across consumers. To quantify the role of the sorting mechanism, we calibrate the model using cross-sectional data on the haircut market in Bergen, Norway , and perform counterfactual analysis. We find that the unequal distribution of the gains among consumers caused by positive market shocks can be quite substantial: the gains of consumers located in more populated niches are 3-4 times higher than those of more remote consumers. It is worth noting that the baseline model considered in the paper assumes away the direct spatial competition among firms. As mentioned, the analysis of this more general case is rather complicated. However, this research direction seems to be rich in its theoretical and quantitative implications. Another interesting research direction is related to the behavior of multi-product firms within the considered framework with consumer heterogeneity. We leave these questions for further research.

Appendix

The Proof of Proposition 2

We proceed in four steps.

Step 1. We start with a series of definitions. First, we define the following function:

$$\pi(\lambda c) \equiv \max_{z \geq 0}[(u'(z) - \lambda c)z].$$

In fact, this is the rescaled profit of a c-type firm under local competitive toughness λ. We define

$$x_{\max} \equiv l^{-1}\left(\frac{\lambda_{\min} f}{\pi(\lambda_{\min} c_{\min})}\right). \tag{A.1}$$

We assume that $x_{\max} < S \Longleftrightarrow l(S) < \lambda_{\min} f / \pi(\lambda_{\min} c_{\min})$ (that is, $l(S)$ is sufficiently low). We also define

$$c_{max} \equiv \frac{1}{\lambda_{\min}} \pi^{-1}\left(\frac{\lambda_{\min} f}{l(0)}\right). \tag{A.2}$$

We assume that $c_{\max} > c_{\min} \Longleftrightarrow l(0) > \lambda_{\min} f / \pi(\lambda_{\min} c_{\min})$ (that is, $l(0)$ is sufficiently high). Note that, if the latter condition fails to hold, there clearly exists no equilibrium. Indeed, in this case, the most productive firm would not break at $x = 0$, even if the competitive toughness λ is at its minimum possible level: $\lambda = \lambda_{\min} > 0$. Therefore, $l(0) > \lambda_{\min} f / \pi(\lambda_{\min} c_{\min})$ is an absolutely necessary condition for the set of active firms to be non-empty.

Next, we define the *cutoff curve* $C \subset \mathbb{R}^2_+$ as follows:

$$C \equiv \left\{ (x, c) \in \mathbb{R}^2_+ : \begin{array}{l} l(x)\pi(\lambda_{\min} c) = \lambda_{\min} f, \\ 0 \leq x \leq x_{\max},\ c_{\min} \leq c \leq c_{\max} \end{array} \right\}.$$

Clearly, C is the set of all a priory feasible solutions $(\overline{x}, \overline{c})$ of the zero-profit condition. Geometrically, C is a downward sloping curve on the (x, c)-plane connecting the points $(0, c_{\max})$ and $(x_{\max}, c_{\min})$, where $x_{\max}$ and $c_{\max}$ are defined, respectively, by

(A.1) and (A.2). Note that, from the definition of $c_{\max}$, it follows that $\lambda_{\min} c_{\max} < u'(0)$ (since $\pi(\lambda_{\min} c_{\max}) = \lambda_{\min} f / l(0) > 0$).

Since $x_{\max} < S$, the population decay rate $a(x) \equiv -l'(x)/l(x)$ is a bounded continuous function over $[0, x_{\max}]$.[26] Therefore, using the Weierstrass theorem, we can define:

$$A \equiv \max_{0 \le x \le x_{\max}} a(x) < \infty. \tag{A.3}$$

Step 2. Consider any $\overline{x} \in (0, x_{\max}]$. Because the cutoff curve C is downward sloping, there exists a unique $\overline{c} \in [c_{\min}, c_{\max})$ such that $(\overline{x}, \overline{c}) \in C$. By Picard's theorem (see, e.g., Pontryagin 1962), there exists $\varepsilon > 0$ such that, for any $x \in (\overline{x} - \varepsilon, \overline{x}]$, there exists a unique solution $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ to (3.14) – (3.15) satisfying the boundary conditions: $\lambda_{\overline{x}}(\overline{x}) = \lambda_{\min}$, $c_{\overline{x}}(\overline{x}) = \overline{c}$. Picard's theorem applies here, since the right-hand sides of (3.14) – (3.15) are well-defined and continuously differentiable and, thereby, locally Lipshitz in (λ, c) in the vicinity of $(\lambda_{\min}, \overline{c})$. In particular, the denominator of the right-hand side of (3.15) never equals zero. Indeed, because $(\overline{x}, \overline{c}) \in C$, we have: $\lambda_{\min} \overline{c} < \lambda_{\min} c_{\max} < u'(0)$ (see Step 1).

Next, we show that the above local solution $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ can be extended backwards either on $[x_0, \overline{x}]$, where $x_0 \in [0, \overline{x})$ and $c_{\overline{x}}(x_0) = c_{\min}$, or on $[0, \overline{x}]$. In intuitive geometric terms, it means the following: the solution $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ can be extended backwards either until it hits the plane $\{(x, \lambda, c) \in \mathbb{R}^3 \, : \, x = 0\}$ or up to the plane$\{(x, \lambda, c) \in \mathbb{R}^3 \, : \, c = c_{\min}\}$. Note that the case when $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ hits the intersection line of these two planes, i.e. the

[26] Observe that $a(x)$ need not be bounded and continuous over the whole range $[0, S]$. To see this, set $S = 1$ and consider a linear symmetric population density: $l(x) = 1 - |x|$ for $x \in (-S, S)$. Then, we have $a(x) = 1/(1 - x)$, which is clearly unbounded over $(0, 1)$.

straight line $\{(x, \lambda, c) \in \mathbb{R}^3 : x = 0, c = c_{\min}\}$, is not ruled out.

Assume the opposite: $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ can be **only** extended backwards on $(x_0, \overline{x}]$, where $x_0 \in (0, \overline{x})$ and $\lim_{x \downarrow x_0} c_{\overline{x}}(x) > c_{\min}$. By the continuation theorem for ODE solutions (Pontryagin 1962), this may only hold true in two cases:

Case 1: an "explosion in finite time" occurs, i.e.

$$\limsup_{x \downarrow x_0} \|(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))\| = \infty, \tag{A.4}$$

where $|| \cdot ||$ stands for the standard Euclidean norm in $\mathbb{R}^2$.

Case 2: the right-hand side of the system (3.14)–(3.15) is not well defined at (x_0, λ, c), where $(\lambda, c) = \lim_{x \downarrow x_0} (\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$.

Let us first explore the possibility of Case 1. One can show that $\lambda_{\overline{x}}(x)$ is bounded on $(x_0, \overline{x}]$. Indeed, we have on $(x_0, \overline{x}]$ (recall that $\mathcal{M}(\lambda c)$ is decreasing in λc, as the price elasticity of demand is increasing)

$$0 > \frac{\mathrm{d}\lambda_{\overline{x}}(x)}{\mathrm{d}x} > -A\mathcal{M}(\lambda_{\min} c_{\min}) \lambda_{\overline{x}}(x).$$

This implies that $\mathrm{d}\ln \lambda_{\overline{x}}(x)/\mathrm{d}x$ is uniformly bounded from above in the absolute value, which in turn means that $\lambda_{\overline{x}}(x)$ is bounded from above on $(x_0, \overline{x}]$. Clearly, $c_{\overline{x}}(x)$ is also bounded, as it increases in x and satisfies:

$$0 \leq c_{\min} < \lim_{x \downarrow x_0} c_{\overline{x}}(x) \leq c_{\overline{x}}(x) \leq c_{\overline{x}}(\overline{x}) = \overline{c} < \infty,$$

for all $x \in (x_0, \overline{x}]$. As a result, (A.4) cannot hold, meaning that Case 1 is not possible.

Let us now explore the possibility of Case 2. When $u'(0) = \infty$, this clearly cannot be the case, as the right-hand side of (3.14)–

(3.15) is well defined for all $c > c_{\min}$, for all $\lambda > \lambda_{\min}$, and for all $x \geq 0$. Thus, it remains to explore the case when $u'(0) < \infty$. In this case, the ODE system (3.14)–(3.15) is not well defined, when $\lim_{x \downarrow x_0} \lambda_{\overline{x}}(x) c_{\overline{x}}(x) = u'(0)$ (in this case, the denominator of the right-hand side in (3.15) is equal to zero). Assume that this is the case. Then, $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))_{x \in (x_0, \overline{x}]}$ and $\lambda c = u'(0)$ define each a curve in the (λ, c)-plane. Note that $u'(0) > \lambda_{\overline{x}}(x) c_{\overline{x}}(x)$ for any $x \in (x_0, \overline{x}]$, otherwise $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ could not be extended backwards on $(x_0, \overline{x}]$. Hence, the curve $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))_{x \in (x_0, \overline{x}]}$ lies strictly below the curve $\lambda c = u'(0)$ in the (λ, c)-plane and intersects it at $\left(\lim_{x \downarrow x_0} \lambda_{\overline{x}}(x), \lim_{x \downarrow x_0} c_{\overline{x}}(x)\right)$ (the limits exist, as $\lambda_{\overline{x}}(x)$ and $c_{\overline{x}}(x)$ are monotone and bounded). This in turn implies that

$$\lim_{x \downarrow x_0} \left| \frac{\mathrm{d} c_{\overline{x}}(x) / \mathrm{d} x}{\mathrm{d} \lambda_{\overline{x}}(x) / \mathrm{d} x} \right| \leq \frac{u'(0)}{\lim_{x \downarrow x_0} \lambda_{\overline{x}}^2(x)}. \tag{A.5}$$

However, using (3.14)–(3.15), we have:

$$0 > \lim_{x \downarrow x_0} \frac{\mathrm{d} \lambda_{\overline{x}}(x)}{\mathrm{d} x} > -\infty, \qquad \lim_{x \downarrow x_0} \frac{\mathrm{d} c_{\overline{x}}(x)}{\mathrm{d} x} = +\infty,$$

which contradicts the inequality (A.5) when $u'(0) < \infty$. That is, Case 2 is not possible as well. Hence, we observe a contradiction to that $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ can be only extended backwards on $(x_0, \overline{x}]$, where $x_0 \in (0, \overline{x})$ and $\lim_{x \downarrow x_0} c_{\overline{x}}(x) > c_{\min}$.

As a result, the solution $(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))$ can be extended backwards either up to the plane $\{(x, \lambda, c) \in \mathbb{R}^3 : x = 0\}$ or up to the plane$\{(x, \lambda, c) \in \mathbb{R}^3 : c = c_{\min}\}$, or both options hold simultaneously.

Step 3. We now construct an equilibrium without taking into account free entry into the market: i.e., we assume that M_e is given. To do this, we define the following function over $[0, x_{\max}]$:

$$\varphi(\overline{x}) = \begin{cases} c_{\overline{x}}(0) - c_{\min}, & \text{if } (\lambda_{\overline{x}}(x), c_{\overline{x}}(x)) \text{ up to } \{x = 0\}, \\ -c_{\overline{x}}^{-1}(c_{\min}), & \text{if } (\lambda_{\overline{x}}(x), c_{\overline{x}}(x)) \text{ up to } \{c = c_{\min}\}. \end{cases} \tag{A.6}$$

By continuity of solutions to ODE w.r.t. initial values (Pontryagin 1962), $\varphi(\overline{x})$ is a continuous function of $\overline{x}$. Furthermore, it is readily verified that the following inequalities hold:

$$\varphi(0) = c_{\max} - c_{\min} > 0, \qquad \varphi(x_{\max}) = -x_{\max} < 0.$$

Hence, by the intermediate value theorem, there exists $\overline{x}^* \in (0, x_{\max})$, such that $\varphi(\overline{x}^*) = 0$. Setting $(\lambda^*(x), c^*(x)) \equiv (\lambda_{\overline{x}^*}(x), c_{\overline{x}^*}(x))$ and $\overline{c}^* \equiv c_{\overline{x}^*}(\overline{x}^*)$, derive a candidate equilibrium:

$$\left\{\overline{x}^*, \overline{c}^*, (\lambda^*(x), c^*(x))_{x \in [0, \overline{x}^*]}\right\}. \tag{A.7}$$

We now verify that the candidate equilibrium (A.7) is indeed an equilibrium when M_e is given. That $(\lambda^*(x), c^*(x))$ is a solution to (3.14) – (3.15) follows by construction. The equality $\varphi(\overline{x}^*) = 0$ means that $(\lambda^*(x), c^*(x))$ can be extended simultaneously up to both planes: $\{x = 0\}$ and $\{c = c_{\min}\}$. This, in turn, is equivalent to $c^*(0) = c_{\min}$, i.e. $(\lambda^*(x), c^*(x))$ satisfies one of the boundary conditions. The other boundary condition, $\lambda^*(\overline{x}^*) = \lambda_{\min}$, is satisfied by construction. Finally, $(\overline{x}^*, \overline{c}^*) \in C$ means that $(\overline{x}^*, \overline{c}^*)$ satisfy the zero-profit condition (3.12).

Step 4. So far, we have been proceeding as if M_e were a constant. However, M_e is endogenous, and is determined by the free entry condition given by:

$$\Pi_e(M_e) \equiv \int\limits_{c_{\min}}^{\overline{c}^*(M_e)} \left[\frac{l(x^*(c,M_e))}{\lambda^*(c,M_e)}\pi\left(\lambda^*(c,M_e)c\right) - f\right] g(c)\mathrm{d}c = f_e, \tag{A.8}$$

where $\lambda^*(c, M_e)$ is a decreasing function parametrically described by the downwards-sloping curve $(\lambda^*(x, M_e),\, c^*(x, M_e))|_{x\in[0,\overline{x}^*]}$, while $x^*(\cdot, M_e)$ is the inverse to $c^*(\cdot, M_e)$. We assume that $l(0)$ is such that

$$f_e < \int\limits_{c_{\min}}^{c_{\max}} \left[\frac{l(0)}{\lambda_{\min}}\pi\left(\lambda_{\min}c\right) - f\right] g(c)\mathrm{d}c. \tag{A.9}$$

Further, we show that this condition is sufficient for equation (A.8) to have a solution $M_e^* > 0$.

First, we show that $\Pi_e(\infty) = 0$. Observe that, when $M_e \to \infty$, equation (3.15) implies that $\mathrm{d}c^*/\mathrm{d}x$ becomes uniformly small. Taking into account that $c^*(0) = c_{\min}$, we have that

$$\lim_{M_e\to\infty} \overline{c}^*(M_e) = c_{\min}, \qquad \lim_{M_e\to\infty} \overline{x}^*(M_e) = x_{\max}.$$

It is straightforward to see that the above implies that $\Pi_e(\infty) = 0$.

Next, we consider $\Pi_e(0)$. Observe that, when $M_e \to 0$, equation (3.15) implies that $\mathrm{d}c^*/\mathrm{d}x$ becomes uniformly large or, equivalently, $\mathrm{d}x^*/\mathrm{d}c$ becomes uniformly small. This implies that

$$\lim_{M_e\to 0} \overline{x}^*(M_e) = 0, \qquad \lim_{M_e\to 0} \overline{c}^*(M_e) = c_{\max}.$$

Hence,

$$\Pi_e(0) = \int\limits_{c_{\min}}^{c_{\max}} \left[\frac{l(0)}{\lambda_{\min}}\pi\left(\lambda_{\min}c\right) - f\right] g(c)\mathrm{d}c.$$

According to our assumption, $\Pi_e(0) > f_e > 0 = \Pi_e(\infty)$. This means that equation (A.8) has a solution $M_e^* > 0$. This completes the proof.

The Proof of Proposition 3

We proceed in four steps. Until Step 4, we ignore the free-entry condition and treat the mass $M_e > 0$ of entrants as exogenous. At Step 4, we take (A.8) into account and show that it uniquely determines M_e.

Step 1. Assume there are at least two equilibrium outcomes corresponding to the same value of M_e:

$$\left\{\overline{x}^*, \overline{c}^*, (\lambda^*(x), c^*(x))_{x\in[0,\overline{x}^*]}\right\}$$

and

$$\left\{\overline{x}^{**}, \overline{c}^{**}, (\lambda^{**}(x), c^{**}(x))_{x\in[0,\overline{x}^{**}]}\right\}.$$

Note that $\overline{x}^* \neq \overline{x}^{**}$. Indeed, if $\overline{x}^* = \overline{x}^{**}$, then $\overline{c}^* = \overline{c}^{**}$ (since the cutoff curve C is downward-sloping). Hence, $(\lambda^*(x), c^*(x))$ and $(\lambda^{**}(x), c^{**}(x))$ are solutions to the same system of ODE satisfying the same boundary conditions. By Picard's theorem, this implies that $(\lambda^*(x), c^*(x)) = (\lambda^{**}(x), c^{**}(x))$ pointwise.

Let us assume without loss of generality that $\overline{x}^* < \overline{x}^{**}$. Because $(\overline{x}^*, \overline{c}^*) \in C$ and $(\overline{x}^{**}, \overline{c}^{**}) \in C$, $\overline{x}^* < \overline{x}^{**}$ implies that $\overline{c}^* > \overline{c}^{**}$. Since $\left\{\overline{x}^{**}, \overline{c}^{**}, (\lambda^{**}(x), c^{**}(x))_{x\in[0,\overline{x}^{**}]}\right\}$ is an equilibrium for given M_e, we have that $c^{**}(0) = c_{\min}$. Furthermore, $\left(c^{**}\right)'_x(x) > 0$. Combining this with $\overline{x}^* < \overline{x}^{**}$, we derive the following inequalities:

$$c^{**}(\overline{x}^{**} - \overline{x}^*) > c^{**}(0) = c_{\min} = c^*(0) = c^*(\overline{x}^* - \overline{x}^*). \qquad \text{(A.10)}$$

For each $z \in [0, \overline{x}^*]$, define $\Delta(z)$ as follows:

$$\Delta(z) \equiv c^{**}(\overline{x}^{**} - z) - c^{*}(\overline{x}^{*} - z). \tag{A.11}$$

As has been shown, $\Delta(\overline{x}^{*}) > 0$. Taking into account that $\overline{c}^{*} > \overline{c}^{**}$, $\Delta(0) < 0$. By the intermediate value theorem, there exists $\xi \in (0, \overline{x}^{*})$, such that $\Delta(\xi) = 0$. Let ξ_0 be the smallest of such ξs. Clearly, we have: $c^{**}(\overline{x}^{**} - \xi_0) = c^{*}(\overline{x}^{*} - \xi_0)$ and $c^{**}(\overline{x}^{**} - z) < c^{*}(\overline{x}^{*} - z)$ for all $z < \xi_0$.

Step 2. Next, we show that

$$\lambda^{**}(\overline{x}^{**} - \xi_0) > \lambda^{*}(\overline{x}^{*} - \xi_0). \tag{A.12}$$

Using (3.14) yields (recall that $\lambda^{**}(\overline{x}^{**}) = \lambda_{\min} = \lambda^{*}(\overline{x}^{*})$)

$$\left.(\lambda^{**}(\overline{x}^{**} - z))'_z\right|_{z=0} = a\left(\overline{x}^{**}\right)\lambda_{\min}\mathcal{M}\left(\lambda_{\min}\overline{c}^{**}\right) >$$
$$> a\left(\overline{x}^{*}\right)\lambda_{\min}\mathcal{M}\left(\lambda_{\min}\overline{c}^{*}\right) = \left.(\lambda^{*}(\overline{x}^{*} - z))'_z\right|_{z=0},$$

which holds true because $a'(x) \geq 0$, $\overline{c}^{*} > \overline{c}^{**}$, and the markup function $\mathcal{M}(\cdot)$ is strictly decreasing. Furthermore, we have:

$$\left.(\lambda^{**}(\overline{x}^{**} - z))'_z\right|_{z=0} > \left.(\lambda^{*}(\overline{x}^{*} - z))'_z\right|_{z=0} > 0.$$

Thus, $\lambda^{**}\left(\overline{x}^{**} - z\right) > \lambda^{*}\left(\overline{x}^{*} - z\right)$ holds true for sufficiently small values of z.

Assume that there is some $\xi_1 \in (0, \xi_0)$, such that $\lambda^{**}(\overline{x}^{**} - \xi_1) = \lambda^{*}(\overline{x}^{*} - \xi_1)$, while $\lambda^{**}(\overline{x}^{**} - z) > \lambda^{*}(\overline{x}^{*} - z)$ for all $z < \xi_1$. Denote $\lambda_1 \equiv \lambda^{*}(\overline{x}^{*} - \xi_1)$. Differentiating the log of the ratio $\lambda^{**}(\overline{x}^{**} - z)/\lambda^{*}(\overline{x}^{*} - z)$ w.r.t. z at $z = \xi_1$ yields (recall that, from the previous step, $c^{**}(\overline{x}^{**} - z) < c^{*}(\overline{x}^{*} - z)$ for all $z < \xi_0$):

$$\left[\ln\left(\frac{\lambda^{**}(\overline{x}^{**}-z)}{\lambda^{*}(\overline{x}^{*}-z)}\right)\right]'_z\Bigg|_{z=\xi_1} =$$
$$= a\left(\overline{x}^{**}-\xi_1\right)\mathcal{M}\left(\lambda_1 c^{**}\left(\overline{x}^{**}-\xi_1\right)\right) -$$
$$- a\left(\overline{x}^{*}-\xi_1\right)\mathcal{M}\left(\lambda_1 c^{*}\left(\overline{x}^{*}-\xi_1\right)\right) > 0.$$

By continuity, $\left[\ln\left(\frac{\lambda^{**}(\overline{x}^{**}-z)}{\lambda^{*}(\overline{x}^{*}-z)}\right)\right]'_z > 0$ must hold for any $z \in (\xi_1 - \varepsilon, \xi_1)$, where $\varepsilon > 0$ is sufficiently small. Hence, the ratio $\lambda^{**}(\overline{x}^{**}-z)/\lambda^{*}(\overline{x}^{*}-z)$ increases over $(\xi_1 - \varepsilon, \xi_1)$ and strictly exceeds 1 at $z = \xi_1 - \varepsilon$. Thus, $\lambda^{**}(\overline{x}^{**}-\xi_1)/\lambda^{*}(\overline{x}^{*}-\xi_1)$ also strictly exceeds 1, i.e. $\lambda^{**}(\overline{x}^{**}-\xi_1) > \lambda^{*}(\overline{x}^{*}-\xi_1)$. Based on that, we conclude that ξ_1 does not exist. This proves (A.12).

Step 3. Differentiating the function $\Delta(z)$ defined by (A.11) at $z = \xi_0$, we obtain:

$$\Delta'_z(\xi_0) = -\frac{1}{M_e g\left(c_0^*\right)}\left[\frac{\left(V'\right)^{-1}\left(1/\lambda_0^{**}\right)}{u\left(q\left(\lambda_0^{**}c_0^*\right)\right)} - \frac{\left(V'\right)^{-1}\left(1/\lambda_0^{*}\right)}{u\left(q\left(\lambda_0^{*}c_0^*\right)\right)}\right] < 0. \tag{A.13}$$

where $c_0^* \equiv c^*(\overline{x}^* - \xi_0) = c^{**}(\overline{x}^{**} - \xi_0)$, $\lambda_0^* \equiv \lambda^*(\overline{x}^* - \xi_0)$, and $\lambda_0^{**} \equiv \lambda^{**}(\overline{x}^{**} - \xi_0)$. The inequality (A.13) holds true because, by (A.12), we have $\lambda_0^{**} > \lambda_0^*$, while the function $\left(V'\right)^{-1}\left(1/\lambda\right)/u\left(q\left(\lambda c\right)\right)$ increases in λ for any given $c > c_{\min}$. However, by definition of ξ_0, $\Delta(z)$ must change sign from negative to positive at $z = \xi_0$. Hence, it must be true that $\Delta'_z(\xi_0) \geq 0$. This contradicts (A.13) and implies that, for any fixed value of M_e, there is a unique equilibrium outcome corresponding to this value of M_e.

Step 4. To finish the proof of uniqueness, it remains to show that $\mathrm{d}\Pi_e(M_e)/\mathrm{d}M_e < 0$ for any $M_e > 0$. Let us define

$$\mathfrak{N}(c, M_e) \equiv \frac{l(x^*(c, M_e))}{\lambda^*(c, M_e)} \pi\left(\lambda^*(c, M_e)c\right).$$

Then, we have:

$$\frac{\mathrm{d}\Pi_e(M_e)}{\mathrm{d}M_e} = \int_{c_{\min}}^{\bar{c}^*(M_e)} \frac{\partial \mathfrak{N}(c, M_e)}{\partial M_e} g(c)\mathrm{d}c + \left[\mathfrak{N}(\bar{c}^*(M_e), M_e) - f\right] \frac{\mathrm{d}\bar{c}^*(M_e)}{\mathrm{d}M_e},$$

where the last term equals zero due to the cutoff condition. Hence,

$$\frac{\mathrm{d}\Pi_e(M_e)}{\mathrm{d}M_e} = \int_{c_{\min}}^{\bar{c}^*(M_e)} \frac{\partial \mathfrak{N}(c, M_e)}{\partial M_e} \mathrm{d}G(c).$$

Thus, a sufficient condition for $\mathrm{d}\Pi_e(M_e)/\mathrm{d}M_e < 0$ for any $M_e > 0$ is given by

$$\frac{\partial \mathfrak{N}(c, M_e)}{\partial M_e} < 0 \text{ for any } M_e > 0 \text{ and any } c \in \left[c_{\min}, \bar{c}^*(M_e)\right].$$

It is straightforward to see that, due to the envelope theorem, the latter is hold when

$$\frac{\partial \lambda^*(x, M_e)}{\partial M_e} > 0 \text{ for any } M_e > 0 \text{ and any } x \in \left[0, \bar{x}^*(M_e)\right].$$

In fact, it is sufficient to show that

$$\frac{\partial \lambda^*(x, M_e)}{\partial M_e} \geq 0 \text{ for any } M_e > 0 \text{ and any } x \in \left[0, \bar{x}^*(M_e)\right]$$

and $\partial\lambda^*(x, M_e)/\partial M_e > 0$ on some non-zero measure subset of $[0, \overline{x}^*(M_e)]$. The rest of the proof amounts to establishing the latter statement.

Assume that, on the contrary, for some $M_e > 0$, there exists a compact interval $[x_1, x_2] \subseteq [0, \overline{x}^*(M_e)]$, such that $\partial\lambda^*(x, M_e)/\partial M_e \leq 0$ for all $x \in [x_1, x_2]$. Without loss of generality, let us also assume that $[x_1, x_2]$ cannot be extended further without violating the condition $\partial\lambda^*(x, M_e)/\partial M_e \leq 0$ (otherwise, we can replace it with a larger one). We will therefore refer to $[x_1, x_2]$ as a *non-extendable* interval. We consider several possible cases.

Case 1: Assume that $x_1 = 0$. In this case, we have: $c^*(x_1, M_e) = c_{\min}$, hence $\partial c^*(x_1, M_e)/\partial M_e = 0$. Recall that

$$\frac{\mathrm{d}c}{\mathrm{d}x} = \frac{1}{M_e} \frac{(V')^{-1}(1/\lambda)}{g(c)\, u(q_x)}.$$

Since $\partial\lambda^*(x_1, M_e)/\partial M_e \leq 0$, $\partial c^*(x_1, M_e)/\partial M_e = 0$, and M_e rises, $\partial\,(c^*)'_x\,(x_1, M_e)/\partial M_e < 0$ (the right-hand side of the above equation decreases at $x_1 = 0$ with a rise in M_e). Note that $\partial c^*(x_1, M_e)/\partial M_e = 0$ and $\partial\,(c^*)'_x\,(x_1, M_e)/\partial M_e < 0$ imply that $\partial c^*(x, M_e)/\partial M_e < 0$ in some right neighborhood of $x_1 = 0$.

Case 2: Assume that $x_2 = \overline{x}^*(M_e)$. We have $\lambda^*(\overline{x}^*(M_e), M_e) = \lambda_{\min}$. This implies that

$$\frac{\partial\lambda^*(\overline{x}^*(M_e), M_e)}{\partial x} \frac{d\overline{x}^*(M_e)}{dM_e} + \frac{\partial\lambda^*(\overline{x}^*(M_e), M_e)}{\partial M_e} = 0.$$

The second term in the left-hand side of the above equation is non-positive (as assumed). Recall that $\lambda^*(x, M_e)$ is strictly decreasing in x. As a result, $d\overline{x}^*(M_e)/dM_e \leq 0$. Combining this with the fact $(\overline{x}^*(M_e), \overline{c}^*(M_e)) \in C$, where C is the downward sloping cutoff curve, we get: $\mathrm{d}\bar{c}^*(M_e)/\mathrm{d}M_e \geq 0$. That is,

$$\frac{\partial c^*\left(\overline{x}^*(M_e), M_e\right)}{\partial x}\frac{\mathrm{d}\overline{x}^*(M_e)}{\mathrm{d}M_e} + \frac{\partial c^*\left(\overline{x}^*(M_e), M_e\right)}{\partial M_e} \geq 0,$$

where the first term is non-positive because, as shown above, $\mathrm{d}\overline{x}^*(M_e)/\mathrm{d}M_e \leq 0$, while $\partial c^*\left(\overline{x}^*(M_e), M_e\right)/\partial x > 0$. Hence, the second term, $\partial c^*\left(\overline{x}^*(M_e), M_e\right)/\partial M_e$, must be non-negative. If $\partial c^*\left(\overline{x}^*(M_e), M_e\right)/\partial M_e = 0$, then one can show that $\partial\left(c^*\right)'_x\left(\overline{x}^*(M_e), M_e\right)/\partial M_e < 0$. Here, we use again the fact that

$$\frac{\mathrm{d}c}{\mathrm{d}x} = \frac{1}{M_e}\frac{\left(V'\right)^{-1}(1/\lambda)}{g(c)\,u(q_x)}.$$

This in turn implies that $\partial c^*\left(\overline{x}^*(M_e), M_e\right)/\partial M_e > 0$ in some left neighborhood of $x_2 = \overline{x}^*(M_e)$.

Case 3: Assume that $0 < x_1 < x_2 < \overline{x}^*(M_e)$. Because $[x_1, x_2]$ is non-extendable, there exists a small open left half-neighborhood $\mathcal{N}_1$ of x_1, and a small right half-neighborhood $\mathcal{N}_2$ of x_2, such that $\partial\lambda^*(x, M_e)/\partial M_e > 0$ for all $x \in \mathcal{N} \equiv \mathcal{N}_1 \cup \mathcal{N}_2$. Hence, for a c-type firm where $c = c^*(x, M_e)$ with $x \in [x_1, x_2]$, relocating marginally beyond $[x_1, x_2]$ in response to a marginal increase in M_e is not profit-maximizing behavior. Indeed, that $\partial\lambda^*(x, M_e)/\partial M_e \leq 0$ over $[x_1, x_2]$ means that the profit function increases uniformly over $[x_1, x_2]$, while $\partial\lambda^*(x, M_e)/\partial M_e > 0$ for all $x \in \mathcal{N}$ means that relocating from $[x_1, x_2]$ into $\mathcal{N}$ would lead to a reduction of maximum feasible profit.[27] This immediately imply that

[27] One may wonder why no firm would relocate from $[x_1, x_2]$ to somewhere beyond $\mathcal{N}$ in response to a marginal increase of M_e. This would mean, for at least some firm type c, that the firm's profit-maximizing location choice $x^*(c, M_e)$ has a discontinuity in M_e. However, by the maximum theorem (Sundaram 1996), $x^*(c, M_e)$ must be upper-hemicontinuous in M_e. Furthermore, by strict quasi-concavity of the profit function, $x^*(c, M_e)$ is single-valued. For single-valued mappings, upper-hemicontinuity implies continuity. Hence, $x^*(c, M_e)$ cannot exhibit

$$\frac{\partial c^*(x_1, M_e)}{\partial M_e} \le 0, \qquad \frac{\partial c^*(x_2, M_e)}{\partial M_e} \ge 0.$$

Moreover, for $j = 1,2$ we have (the proof is the same as in the previous cases)

$$\frac{\partial c^*(x_j, M_e)}{\partial M_e} = 0 \Rightarrow \frac{\partial \left(c^*\right)'_x (x_j, M_e)}{\partial M_e} < 0.$$

The findings in the above cases allow us to formulate the following important result. *There exists a location x_4 in an arbitrary small right half-neighborhood of x_1, such that $\partial c^*(x_4, M_e)/\partial M_e < 0$. Similarly, there exists a location x_5 in an arbitrary small left half-neighborhood of x_2, such that $\partial c^*(x_5, M_e)/\partial M_e > 0$.*

By the intermediate value theorem, there must exist a location $x_3 \in (x_4, x_5) \subset [x_1, x_2]$ such that

$$\frac{\partial c^*(x_3, M_e)}{\partial M_e} = 0, \; \frac{\partial \left(c^*\right)'_x (x_3, M_e)}{\partial M_e} \ge 0.$$

The non-negative sign of the derivative follows from the fact that $c^*(x, M_e)$ is increasing in x. This in turn implies that the derivative of

$$\frac{1}{M_e} \frac{\left(V'\right)^{-1} (1/\lambda^*(x_3, M_e))}{g(c^*(x_3, M_e))\, u(q(\lambda^*(x_3, M_e)c^*(x_3, M_e)))}$$

with respect to M_e is non-negative. That is, the derivative of

$$\frac{\left(V'\right)^{-1} (1/\lambda^*(x_3, M_e))}{g(c^*(x_3, M_e))u(q(\lambda^*(x_3, M_e)c^*(x_3, M_e)))}$$

with respect to M_e is strictly positive. This means that $\partial\lambda^* \left(x_3, M_e\right) / \partial M_e > 0$ (recall that $\partial c^*(x_3, M_e)/\partial M_e = 0$). However, since $x_3 \in [x_1, x_2]$, it

---

discontinuities.

must be that $\partial\lambda^*(x_3, M_e)/\partial M_e \leq 0$, which is a contradiction. This completes the proof of uniqueness of the equilibrium.

The proof of Proposition 4

To prove the proposition, we use the equilibrium conditions for $\lambda'(x)$ and $c'(x)$. Specifically, from (3.11) and (3.9),

$$\lambda'(x) = \frac{l'(x)\lambda(x)}{l(x)} \frac{p(x, c(x)) - c(x)}{p(x, c(x))},$$

$$M_e g\left(c(x)\right) c'(x) u\left(q(x, c(x))\right) = \left(V'\right)^{-1} \left(1/\lambda(x)\right) \iff$$

$$\iff c'(x) = \frac{\left(V'\right)^{-1} \left(1/\lambda(x)\right)}{M_e g\left(c(x)\right) u\left(q(x, c(x))\right)}.$$

Hence,

$$(\lambda(x)c(x))'_x = c(x)\lambda'(x) + \lambda(x)c'(x) =$$

$$= \frac{\lambda(x)}{g\left(c(x)\right)} \left[c(x) g\left(c(x)\right) \frac{l'(x)}{l(x)} \frac{p(x, c(x)) - c(x)}{p(x, c(x))} + \frac{\left(V'\right)^{-1} \left(1/\lambda(x)\right)}{M_e u\left(q(x, c(x))\right)} \right].$$

Consider,

$$(\lambda(x)c(x))'_{x=0} =$$

$$= \frac{\lambda(0)}{g\left(c_{\min}\right)} \left(c_{\min}\, g\left(c_{\min}\right) \frac{l'(0)}{l(0)} \frac{p(0, c_{\min}) - c_{\min}}{p(0, c_{\min})} + \frac{\left(V'\right)^{-1} \left(1/\lambda(0)\right)}{M_e u\left(q(0, c_{\min})\right)} \right).$$

Since $g(c)$ is a density function, $lim_{c_{\min}\to 0} c_{\min}\, g\left(c_{\min}\right) = 0$. Hence, if $\mid l'(0) \mid < \infty$, then for sufficiently low $c_{\min}$,

$$c_{\min}\, g\left(c_{\min}\right) \frac{l'(0)}{l(0)} \frac{p(0, c_{\min}) - c_{\min}}{p(0, c_{\min})} + \frac{\left(V'\right)^{-1} \left(1/\lambda(0)\right)}{M_e u\left(q(0, c_{\min})\right)} > 0.$$

Similarly,

$$(\lambda(x)c(x))'_{x=\bar{x}} = \frac{\lambda(\bar{x})}{g(\bar{c})}\left(\bar{c}\,g(\bar{c})\,\frac{l'(\bar{x})}{l(\bar{x})}\frac{p(\bar{x},\bar{c})-\bar{c}}{p(\bar{x},\bar{c})} + \frac{(V')^{-1}(1/\lambda(\bar{x}))}{M_e u\,(q(\bar{x},\bar{c}))}\right).$$

Note that, as there is the fixed cost of production f, $p(\bar{x},\bar{c}) > \bar{c}$. Moreover, $\lambda(\bar{x}) = 1/V'(0)$ in the equilibrium, implying that $(V')^{-1}(1/\lambda(\bar{x})) = 0$ (this also means that $c'(\bar{x}) = 0$). As a result, since $l'(\bar{x}) < 0$,

$$\bar{c}\,g(\bar{c})\,\frac{l'(\bar{x})}{l(\bar{x})}\frac{p(\bar{x},\bar{c})-\bar{c}}{p(\bar{x},\bar{c})} + \frac{(V')^{-1}(1/\lambda(\bar{x}))}{M_e u\,(q(\bar{x},\bar{c}))} < 0.$$

To prove the third statement of the proposition, we rewrite $(\lambda(x)c(x))'_x$ in the following way:

$$(\lambda(x)c(x))'_x = \\ = \frac{\lambda(x)}{g(c(x))}\left(\frac{l'(x)}{l(x)}c(x)\,g(c(x))\,\mathcal{M}(\lambda(x)c(x)) + \frac{(V')^{-1}(1/\lambda(x))}{M_e u\,(q(\lambda(x)c(x)))}\right),$$

where $\mathcal{M}(.)$ is the markup function. Let us denote $\tilde{x}\in(0,\bar{x})$ as an interior extremum of $\lambda(x)c(x)$: $(\lambda(\tilde{x})c(\tilde{x}))'_x = 0$. We know that $(\lambda(x)c(x))'_{x=0} > 0$ and $(\lambda(x)c(x))'_{x=\bar{x}} < 0$. Hence, $\lambda(x)c(x)$ has at least one interior local maximizer.

Next, we show that, for any $\tilde{x}$, $(\lambda(\tilde{x})c(\tilde{x}))''_{xx} < 0$. We have

$$(\lambda(\tilde{x})c(\tilde{x}))''_{xx} = \\ = \left(\frac{\lambda(\tilde{x})}{g(c(\tilde{x}))}\right)'\left(\frac{l'(\tilde{x})}{l(\tilde{x})}c(\tilde{x})\,g(c(\tilde{x}))\,\mathcal{M}(\lambda(\tilde{x})c(\tilde{x})) + \frac{(V')^{-1}(1/\lambda(\tilde{x}))}{M_e u\,(q(\lambda(\tilde{x})c(\tilde{x})))}\right) \\ + \frac{\lambda(\tilde{x})}{g(c(\tilde{x}))}\left(\frac{l'(\tilde{x})}{l(\tilde{x})}c(\tilde{x})\,g(c(\tilde{x}))\,\mathcal{M}(\lambda(\tilde{x})c(\tilde{x})) + \frac{(V')^{-1}(1/\lambda(\tilde{x}))}{M_e u\,(q(\lambda(\tilde{x})c(\tilde{x})))}\right)'_x.$$

Note that the first term in the right hand side of the above formula is equal to zero. Thus, we have (recall that $(\lambda(\tilde{x})c(\tilde{x}))'_x = 0$)

$$
\begin{aligned}
&(\lambda(\tilde{x})c(\tilde{x}))''_{xx} = \\
&= \frac{\lambda(\tilde{x})}{g\left(c\left(\tilde{x}\right)\right)}\left(\frac{l'(\tilde{x})}{l(\tilde{x})}c\left(\tilde{x}\right)g\left(c\left(\tilde{x}\right)\right)\mathcal{M}(\lambda(\tilde{x})c(\tilde{x})) + \frac{\left(V'\right)^{-1}\left(1/\lambda(\tilde{x})\right)}{M_e u\left(q(\lambda(\tilde{x})c(\tilde{x}))\right)}\right)'_x \\
&= \frac{\lambda(\tilde{x})}{g\left(c\left(\tilde{x}\right)\right)}\left(\left(\frac{l'(\tilde{x})}{l(\tilde{x})}c\left(\tilde{x}\right)g\left(c\left(\tilde{x}\right)\right)\right)'_x\mathcal{M}(\lambda(\tilde{x})c(\tilde{x})) + \frac{\left(\left(V'\right)^{-1}\left(1/\lambda(\tilde{x})\right)\right)'_x}{M_e u\left(q(\lambda(\tilde{x})c(\tilde{x}))\right)}\right).
\end{aligned}
$$

We have

$$
\begin{aligned}
&\left(\frac{l'(x)}{l(x)}c(x)g\left(c(x)\right)\right)'_x = \\
&= \frac{l'(x)}{l(x)}\left(c(x)g\left(c(x)\right)\right)'_x + c(x)g\left(c(x)\right)\left(\frac{l'(x)}{l(x)}\right)'_x < 0,
\end{aligned}
$$

since $c'(x) > 0$, $g'(c) \geq 0$, and $(l'(x)/l(x))'_x \leq 0$. At the same time, $\left(V'\right)^{-1}\left(1/\lambda(x)\right)$ is decreasing in x as $V''(\cdot) < 0$ and $\lambda'(x) < 0$. Hence, $(\lambda(\tilde{x})c(\tilde{x}))''_{xx} < 0$.

We now finish the proof of part (iii) of Proposition 3. As derived above, $\lambda(x)c(x)$ has no interior local minimum over $(0, \overline{x})$ and at least one interior local maximizer. Assume that $\lambda(x)c(x)$ has at least two distinct local maximizers. Then, there must be a local minimizer in between, which contradicts our above finding. We conclude that $\lambda(x)c(x)$ is bell-shaped in x, while the markup function $\mathcal{M}(\lambda(x)c(x))$ is U-shaped in x. This completes the proof.

The proof of Lemma 2

Note that in this proof it is important that $\partial\lambda(x, M_e, \delta)/\partial\delta$ and $\partial c(x, M_e, \delta)/\partial\delta$ are analytic in x over $(0, \overline{x})$, meaning that they can be represented by convergent power series (this is the case, when, for instance, the primitives in the model are analytic):

$$\frac{\partial\lambda(x, M_e, \delta)}{\partial\delta} = \sum_{k=0}^{\infty} a_k(M_e, \delta)x^k, \qquad \frac{\partial c(x, M_e, \delta)}{\partial\delta} = \sum_{k=0}^{\infty} b_k(M_e, \delta)x^k.$$

This makes the case when $\partial\lambda(x, M_e, \delta)/\partial\delta = 0$ and $\partial(\lambda)'_x(x, M_e, \delta)/\partial\delta = 0$ at some x impossible. Why? If this is the case, then $\partial c(x, M_e, \delta)/\partial\delta = 0$ and $\partial(c)'_x(x, M_e, \delta)/\partial\delta = 0$ as well implying that the derivatives of all orders of $\partial\lambda(x, M_e, \delta)/\partial\delta$ w.r.t. x at this point equal to zero. An analytic function with this property must be identically zero (Courant and John 2012, p. 545). This in turn means that $\lambda(x)$ does not change on the whole interval $[0, \overline{x}]$ when δ changes, which is impossible. For the same reason, it is not possible that $\partial c(x, M_e, \delta)/\partial\delta = 0$ and $\partial(c)'_x(x, M_e, \delta)/\partial\delta = 0$ at some x.

To simplify the exposition of the proof, we divide it into several parts.

Part 1

In this part, we prove that $\partial\overline{x}(M_e, \delta)/\partial\delta > 0$. Assume, on the contrary, that $\partial\overline{x}(M_e, \delta)/\partial\delta \leq 0$. Then, because an increase in δ leads to an upward shift of the cutoff curve C, it must be that $\partial\overline{c}(M_e, \delta)/\partial\delta > 0$. Note also that if $\partial\overline{x}(M_e, \delta)/\partial\delta < 0$, then (by continuity) $\lambda(x, M_e, \delta)$ must decrease w.r.t. δ in some neighborhood of $\overline{x}$ (as $\lambda(x, M_e, \delta)$ is decreasing in x). If $\overline{x}$ does not change with the change in δ, one can derive from (3.14) that $\partial\left(-(\lambda)'_x(\overline{x}, M_e, \delta)\right)\partial\delta <$

0. This is because $\partial\bar{c}(M_e,\delta)/\partial\delta > 0$ and $\lambda(\bar{x},M_e,\delta) = \lambda_{\min}$. This in turn also means that $\partial\lambda(x,M_e,\delta)/\partial\delta < 0$ in some neighborhood of $\bar{x}$. That is, if $\partial\bar{x}(M_e,\delta)/\partial\delta \leq 0$, $\lambda(x,M_e,\delta)$ must decrease w.r.t. δ over some interval $(x_1,\bar{x})$. Two cases may arise.

Case 1: $x_1 = 0$. In this case, $\partial\lambda(0,M_e,\delta)/\partial\delta < 0$. Then, taking into account the boundary condition $c(0,M_e,\delta) = c_{\min}$, it is straightforward to see from the equilibrium condition in (3.15) that $\partial(c)'_x(0,M_e,\delta)/\partial\delta < 0$. This in turn implies that $\partial c(x,M_e,\delta)/\partial\delta < 0$ in the vicinity of $x = 0$ (since $c(0,M_e,\delta) = c_{\min}$ is not affected by δ). As a result, we have the following situation: given the rise in δ, $c(x)$ falls in the neighborhood of zero and rises in the neighborhood of $\bar{x}$ as $\partial\bar{c}(M_e,\delta)/\partial\delta > 0$. This implies that there exists $x_2 \in (0,\bar{x})$ such that $\partial c(x_2,M_e,\delta)/\partial\delta = 0$ - the value of $c(x)$ at x_2 is not affected by the rise in δ. Moreover, $\partial(c)'_x(x_2,M_e,\delta)/\partial\delta > 0$ (as $c(x)$ falls around zero). This in turn means (here we use the equilibrium condition in (3.15)) that $\partial\lambda(x_2,M_e,\delta)/\partial\delta > 0$ which contradicts the assumption that $\partial\lambda(x,M_e,\delta)/\partial\delta < 0$ for all $x > 0$. *Note that we will use this particular way of deriving the contradiction throughout the whole proof of the lemma.*

Case 2: $x_1 > 0$. In this case, it must be true that $\partial\lambda(x_1,M_e,\delta)/\partial\delta = 0$. Moreover, the absolute value of the slope of $\lambda(x)$ at this point increases: $\partial\left(-(\lambda)'_x(x_1,M_e,\delta)\right)/\partial\delta > 0$, as $\partial\lambda(x,M_e,\delta)/\partial\delta < 0$ on $(x_1,\bar{x})$. In this case, from the equilibrium condition in (3.14) we derive that $\partial c(x_1,M_e,\delta)/\partial\delta < 0$. Now, we use the same argument as in the previous case. There exists $x_3 \in (x_1,\bar{x})$ such that $\partial c(x_3,M_e,\delta)/\partial\delta = 0$ and $\partial(c)'_x(x_3,M_e,\delta)/\partial\delta > 0$. This in turn implies that $\partial\lambda(x_3,M_e,\delta)/\partial\delta > 0$ which contradicts the assumption that $\partial\lambda(x,M_e,\delta)/\partial\delta < 0$ for all $x > x_1$.

Thus, we show that $\partial\bar{x}(M_e,\delta)/\partial\delta > 0$.

Part 2

Next, we show that $\partial\lambda(x, M_e, \delta)/\partial\delta > 0$ for all x. Assume that, on the contrary, there exists a non-extendable interval $(x_4, x_5) \subset [0, \overline{x}]$ such that $\partial\lambda(x, M_e, \delta)/\partial\delta \leq 0$ on this interval. Note that since $\overline{x}$ rises (implying that $\partial\lambda(x, M_e, \delta)/\partial\delta > 0$ in some neighborhood of $\overline{x}$), $x_5 < \overline{x}$. Consider again two cases.

Case 1: $x_4 > 0$. In this case, because (x_4, x_5) is a non-extendable interval where $\partial\lambda(x, M_e, \delta)/\partial\delta < 0$, it must be that:

$$\frac{\partial\lambda(x_4, M_e, \delta)}{\partial\delta} = 0 = \frac{\partial\lambda(x_5, M_e, \delta)}{\partial\delta}.$$

Moreover,

$$\frac{\partial\left(-\left(\lambda\right)'_x\left(x_4, M_e, \delta\right)\right)}{\partial\delta} > 0 > \frac{\partial\left(-\left(\lambda\right)'_x\left(x_5, M_e, \delta\right)\right)}{\partial\delta}.$$

In this case, (3.14) implies that

$$\frac{\partial c(x_4, M_e, \delta)}{\partial\delta} < 0 < \frac{\partial c(x_5, M_e, \delta)}{\partial\delta}.$$

Hence, there exists $x_6 \in (x_4, x_5)$, such that

$$\frac{\partial c(x_6, M_e, \delta)}{\partial\delta} = 0, \qquad \frac{\partial\left(c\right)'_x\left(x_6, M_e, \delta\right)}{\partial\delta} > 0.$$

This means that $\partial\lambda(x_6, M_e, \delta)/\partial\delta > 0$, which contradicts the assumption that $\partial\lambda(x, M_e, \delta)/\partial\delta \leq 0$ for all $x \in (x_4, x_5)$.

Case 2: $x_4 = 0$. In this case, it can potentially be that $\partial\lambda(0, M_e, \delta)/\partial\delta = 0$ or $\partial\lambda(0, M_e, \delta)/\partial\delta < 0$. Note that if $\partial\lambda(0, M_e, \delta)/\partial\delta = 0$, then $\partial(\lambda)'_x(x, M_e, \delta)/\partial\delta = 0$ (as $\partial c(0, M_e, \delta)/\partial\delta = 0$). As discussed at the beginning of the proof, this case is impossible. If $\partial\lambda(0, M_e, \delta)/\partial\delta < 0$, then from (3.15), $\partial(c)'_x(0, M_e, \delta)/\partial\delta < 0$, meaning that in some neighborhood of zero $c(x)$ falls with the rise in δ. Then, we use again the logic from the previous case and, thereby, derive the contradiction.

Part 3

The next step is to show that $\partial c(x, M_e, \delta)/\partial\delta > 0$ for all $x \in (0, \overline{x}]$. Assume that, on the contrary, that there exists a non-extendable interval $(x_7, x_8) \subset [0, \overline{x}]$, such that $\partial c(x, M_e, \delta)/\partial\delta \leq 0$ on this interval. If $x_7 = 0$, then $\partial(c)'_x(0, M_e, \delta)/\partial\delta \leq 0$ and $\partial c(0, M_e, \delta)/\partial\delta = 0$. In this case, $\partial\lambda(0, M_e, \delta)/\partial\delta \leq 0$ which contradicts our previous results. If $x_7 > 0$, then again $\partial c(x_7, M_e, \delta)/\partial\delta = 0$ and $\partial\left(c\right)'_x(x_7, M_e, \delta)/\partial\delta < 0$ (recall that $\partial\left(c\right)'_x(x_7, M_e, \delta)/\partial\delta$ cannot be equal to zero). That is, we derive the contradiction: $\partial\lambda(x_7, M_e, \delta)/\partial\delta < 0$.

Finally, since $\partial c(x, M_e, \delta)/\partial\delta > 0$, $\partial\overline{x}(M_e, \delta)/\partial\delta > 0$, and $(c)'_x > 0$, $\partial\overline{c}(M_e, \delta)/\partial\delta > 0$.

The proof of Proposition 5

(i) Totally differentiating both sides of the FOCs, $\Pi_p = 0$ and $\Pi_x = 0$, w.r.t. c yields

$$\begin{pmatrix} dp(c)/dc \\ dx(c)/dc \end{pmatrix} = -\begin{pmatrix} \Pi_{pp} & \Pi_{px} \\ \Pi_{px} & \Pi_{xx} \end{pmatrix}^{-1} \begin{pmatrix} \Pi_{cp} \\ \Pi_{cx} \end{pmatrix}, \tag{A.14}$$

where the right-hand side is evaluated at $(p, x) = (p(c),\, x(c))$. As implied by the FOCs and the definition of the profit function, we have: $\Pi_{cp} = -Q_p > 0$, $\Pi_{cx} = -Q_x = \frac{\Pi_x}{p-c} = 0$. Plugging these expressions for Π_{cp} and Π_{cx} back to (A.14) yields

$$\begin{pmatrix} dp(c)/dc \\ dx(c)/dc \end{pmatrix} = \frac{1}{\Pi_{pp}\Pi_{xx} - \Pi_{px}^2} \begin{pmatrix} \Pi_{xx}Q_p \\ -\Pi_{px}Q_p \end{pmatrix}. \tag{A.15}$$

Using (A.15) and the chain rule, and taking into account that $Q_x = 0$, we obtain:

$$\frac{dp(c)}{dc} = \frac{\Pi_{xx}}{\Pi_{pp}\Pi_{xx} - \Pi_{px}^2} Q_p > 0,$$

$$\frac{d}{dc} Q(p(c), x(c)) = \frac{\Pi_{xx}}{\Pi_{pp}\Pi_{xx} - \Pi_{px}^2} Q_p^2 < 0,$$

where both inequalities hold due to the SOC. This proves the inequalities in (3.29).

(ii) The equivalence of the inequality in (3.30) to $\mathrm{d}x(c)/\mathrm{d}c > 0$ follows immediately from (A.15) and the SOC.

Figures

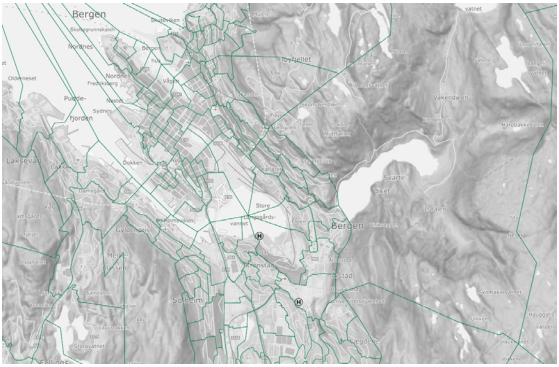


Figure A.1: Basic Units in the City of Bergen
Note: Source: `https://kart.ssb.no/`

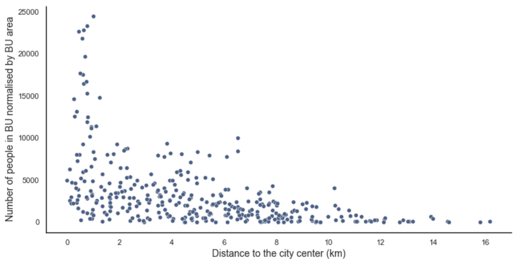


Figure A.2: Distribution of population in Bergen
Note: Each dot in the figure represents the number of people living in a certain basic unit of Bergen divided by the basic unit area.

Bibliography

Adams, Brian, and Kevin R Williams. 2019. "Zone pricing in retail oligopoly." *American Economic Journal: Microeconomics*, 11(1): 124–56.

Aguirregabiria, Victor, and Gustavo Vicentini. 2016. "Dynamic Spatial Competition Between Multi-Store Retailers." *The Journal of Industrial Economics*, 64(4): 710–754.

Aguirregabiria, Victor, and Junichi Suzuki. 2016. "Empirical games of market entry and spatial competition in retail industries." In *Handbook on the Economics of Retailing and Distribution*. Edward Elgar Publishing.

Aguirregabiria, Victor, Pedro Mira, and Hernan Roman. 2007. "An Estimate Dynamic Model of Entry, Exit, and Growth in Oligopoly Retail Markets." *American Economic Review*, 97(2): 449–454.

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.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *The Quarterly Journal of Economics*, 129(3): 1085–1140.

Anderson, Simon P, Andre De Palma, and Jacques-Francois Thisse. 1992. *Discrete choice theory of product differentiation.*

Anderson, Simon P, Nisvan Erkal, and Daniel Piccinin. 2020. "Aggregative games and oligopoly theory: short-run and long-run analysis." *The RAND Journal of Economics*, 51(2): 470–495.

Argentesi, Elena, Paolo Buccirossi, Roberto Cervone, Tomaso Duso, and Alessia Marrazzo. 2021. "The effect of mergers on variety in grocery retailing." *International Journal of Industrial Organization*, 79: 102789.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare. 2018. "The Elusive Pro-Competitive Effects of Trade." *The Review of Economic Studies*, 86(1): 46–80.

Asplund, Marcus, and Volker Nocke. 2006. "Firm Turnover in Imperfectly Competitive Markets." *The Review of Economic Studies*, 73(2): 295–327.

Atal, Juan Pablo, Jose Ignacio Cuesta, and Morten Sæthre. 2022. "Quality Regulation and Competition: Evidence from Pharmaceutical Markets." National Bureau of Economic Research.

Ater, Itai, and Adi Shany. 2021. "Exercising Market Power Without Using Prices: Service Time in Online Grocery." Maurice Falk Institute for Economic Research in Israel Discussion Paper 5.

Bau, Natalie. 2019. "Estimating an Equilibrium Model of Horizontal Competition in Education." CEPR Discussion Papers.

Beckmann, Martin J. 1972. "Spatial Cournot Oligopoly." *Papers in Regional Science*, 28(1): 37–48.

Behrens, Kristian, and Frederic Robert-Nicoud. 2015. "Agglomeration Theory with Heterogeneous Agents." *Handbook of Regional and Urban Economics*, 5: 171 – 245.

Behrens, Kristian, and Yasusada Murata. 2007. "General Equilibrium Models of Monopolistic Competition: A New Approach." *Journal of Economic Theory*, 136(1): 776–787.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy*, 122(3): 507–553.

Bellone, Flora, Patrick Musso, Lionel Nesta, and Frederic Warzynski. 2014. "International Trade and Firm-Level Markups when Location and Quality Matter." *Journal of Economic Geography*, 16(1): 67–91.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, Steven T. 1992. "Estimation of a Model of Entry in the Airline Industry." *Econometrica: Journal of the Econometric Society*, 889–917.

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262.

Berry, Steven T, and Joel Waldfogel. 2001. "Do mergers increase product variety? Evidence from radio broadcasting." *The Quarterly Journal of Economics*, 116(3): 1009–1025.

Bjornerstedt, Jonas, and Frank Verboven. 2016. "Does merger simulation work? Evidence from the Swedish analgesics market." *American Economic Journal: Applied Economics*, 8(3): 125–64.

Bresnahan, Timothy F. 1987. "Competition and collusion in the American automobile industry: The 1955 price war." *The Journal of Industrial Economics*, 457–482.

Bresnahan, Timothy F, and Peter C Reiss. 1991. "Entry and competition in concentrated markets." *Journal of political economy*, 99(5): 977–1009.

Carballo, Jeronimo, Gianmarco I.P. Ottaviano, and Christian Volpe Martincus. 2018. "The Buyer Margins of Firms' Exports." *Journal of International Economics*, 112: 33 – 49.

Chamberlin, Edward. 1933. *The Theory of Monopolistic Competition.* Cambridge, MA:Harvard University Press.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review*, 98(4): 1707–21.

Chen, Yongmin, and Michael H Riordan. 2007. "Price and Variety in the Spokes Model." *The Economic Journal*, 117(522): 897–921.

Courant, Richard, and Fritz John. 2012. *Introduction to Calculus and Analysis I.* Springer Science & Business Media.

Crawford, Gregory S, Oleksandr Shcherbakov, and Matthew Shum. 2019. "Quality overprovision in cable television markets." *American Economic Review*, 109(3): 956–95.

Davis, Peter. 2006. "Spatial competition in retail markets: movie theaters." *The RAND Journal of Economics*, 37(4): 964–982.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform pricing in us retail chains." *The Quarterly Journal of Economics*, 134(4): 2011–2084.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform." *Econometrica*, 84(2): 445–510.

Dhingra, Swati, and John Morrow. 2019. "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity." *Journal of Political Economy*, 127(1): 196–232.

Díez, Federico J, Jiayue Fan, and Carolina Villegas-Sánchez. 2021. "Global declining competition?" *Journal of International Economics*, 132: 103492.

Dixit, Avinash K, and Joseph E Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, 67(3): 297–308.

Domenchich, T, and D McFadden. 1975. "A Theory of individual travel demand." *Urban Travel Demand*, 33–46.

Duarte, Marco, Lorenzo Magnolfi, and Camilla Roncoroni. 2020. "The competitive conduct of consumer cooperatives." Working paper.

Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do prices and attributes explain international differences in food purchases?" *American Economic Review*, 104(3): 832–67.

Eaton, B Curtis, and Richard G Lipsey. 1989. "Product differentiation." *Handbook of industrial organization*, 1: 723–768.

Eckel, Carsten, and J Peter Neary. 2010. "Multi-Product Firms and Flexible Manufacturing in the Global Economy." *The Review of Economic Studies*, 77(1): 188–217.

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

Ellickson, Paul B, Paul LE Grieco, and Oleksii Khvastunov. 2020. "Measuring competition in spatial retail." *The RAND Journal of Economics*, 51(1): 189–232.

Ellickson, Paul B, Stephanie Houghton, and Christopher Timmins. 2013. "Estimating network economies in retail chains: a revealed preference approach." *The RAND Journal of Economics*, 44(2): 169–193.

Evans, Lawrence C. 2010. *Partial Differential Equations.* Graduate Studies in Mathematics, Vol. 19. American Mathematical Society, Providence, Rhode Island.

Faber, Benjamin, and Thibault Fally. 2020. "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data." *The Review of Economic Studies (forthcoming).*

Fajgelbaum, Pablo, Gene M Grossman, and Elhanan Helpman. 2011. "Income Distribution, Product Quality, and International Trade." *Journal of Political Economy*, 119(4): 721–765.

Fan, Ying, and Chenyu Yang. 2020. "Competition, product proliferation, and welfare: A study of the US smartphone Market." *American Economic Journal: Microeconomics*, 12(2): 99–134.

Feenstra, Robert C, and John Romalis. 2014. "International Prices and Endogenous Quality." *The Quarterly Journal of Economics*, 129(2): 477–527.

Fieler, Ana Cecília, and Ann Harrison. 2019. "Escaping Import Competition in China." Working Paper.

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.

Gandhi, Amit, and Jean-François Houde. 2019. "Measuring substitution patterns in differentiated-products industries." National Bureau of Economic Research Working Paper w26375.

Gaubert, Cecile. 2018. "Firm Sorting and Agglomeration." *American Economic Review*, 108(11): 3117–53.

Goryunov, Maxim, Sergey Kokovin, and Takatoshi Tabuchi. 2022. "Continuous spatial monopolistic competition: matching goods with consumers." *Economic Theory*, 74(3): 793–832.

Handbury, Jessie. 2019. "Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities." National Bureau of Economic Research.

Handbury, Jessie, and David E Weinstein. 2015. "Goods prices and availability in cities." *The Review of Economic Studies*, 82(1): 258–296.

Hanemann, W Michael. 1984. "Discrete/continuous models of consumer demand." *Econometrica: Journal of the Econometric Society*, 541–561.

Helse- og omsorgsdepartementet. 1993. "Forskrift om grossistvirksomhet med legemidler." *https://lovdata.no/dokument/SF/forskrift/1993-12-21-1219*.

Helse- og omsorgsdepartementet. 2000. "Lov om apotek (apotekloven)." *https://lovdata.no/dokument/NL/lov/2000-06-02-39*.

Hitsch, Günter J, Ali Hortacsu, and Xiliang Lin. 2019. "Prices and promotions in us retail markets: Evidence from big data." National Bureau of Economic Research.

Holmes, Thomas J. 2011. "The diffusion of Wal-Mart and economies of density." *Econometrica*, 79(1): 253–302.

Holmes, Thomas J, and John J Stevens. 2014. "An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade." *Journal of Political Economy*, 122(2): 369–421.

Hotelling, Harold. 1929*a*. "Stability in Competition." *The Economic Journal*, 39(153): 41–57.

Hotelling, Harold. 1929*b*. "Stability in Competition." *The Economic Journal*, 39(153): 41–57.

Houde, Jean-François. 2012. "Spatial differentiation and vertical mergers in retail markets for gasoline." *American Economic Review*, 102(5): 2147–2182.

Igami, Mitsuru, and Nathan Yang. 2013. "Cannibalization and Preemptive Entry of Multi-Product Firms." *Rand Journal of Economics*, 36: 908–929.

Igami, Mitsuru, and Nathan Yang. 2016. "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada." *Quantitative Economics*, 7(2): 483–521.

Jia, Panle. 2008. "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry." *Econometrica*, 76(6): 1263–1316.

Judd, Kenneth L. 1985. "Credible spatial preemption." *The RAND Journal of Economics*, 153–166.

Kaldor, Nicholas. 1935. "Market Imperfection and Excess Capacity." *Economica*, 2(5): 33–50.

Kim, Hyunchul, and Jungwon Yeo. 2021. "The Effect of Product Variety in Multiproduct Retail Pricing: The Case of Supermarkets." *Available at SSRN 3794048*.

Kokovin, Sergey G, Shamil Sharapudinov, Alexander Tarasov, and Philip Ushchev. 2020. "A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers." *CESifo Working Paper No. 8082*.

Kugler, Maurice, and Eric Verhoogen. 2012. "Prices, Plant Size, and Product Quality." *The Review of Economic Studies*, 79(1): 307–339.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy*, 74(2): 132–157.

Loertscher, Simon, and Gerd Muehlheusser. 2011. "Sequential location games." *The RAND Journal of Economics*, 42(4): 639–663.

MacDonald, James M, and Paul E Nelson Jr. 1991. "Do the poor still pay more? Food price variations in large metropolitan areas." *Journal of Urban Economics*, 30(3): 344–359.

Mankiw, N Gregory, and Michael D Whinston. 1986. "Free entry and social inefficiency." *The RAND Journal of Economics*, 48–58.

Matsa, David A. 2011. "Competition and product quality in the supermarket industry." *The Quarterly Journal of Economics*, 126(3): 1539–1591.

Matsuyama, Kiminori, and Philip Ushchev. 2022. "Selection and Sorting of Heterogeneous Firms through Competitive Pressures." *CEPR Discussion Paper No. DP17092*.

Mazzeo, Michael J, Katja Seim, and Mauricio Varela. 2018. "The welfare consequences of mergers with endogenous product choice." *The Journal of Industrial Economics*, 66(4): 980–1016.

McFadden, Daniel. 1974. "The measurement of urban travel demand." *Journal of public economics*, 3(4): 303–328.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Merker, Tyra. 2022. "Explaining Uniform Pricing in Retail Markets – The Role of Consumer Self-Sorting."

Mrazova, Monika, and J. Peter Neary. 2017. "Not So Demanding: Demand Structure and Firm Behavior." *American Economic Review*, 107(12): 3835–74.

Neiman, Brent, and Joseph S Vavra. 2019. "The Rise of Niche Consumption." National Bureau of Economic Research Working Paper 26134.

Nigai, Sergey. 2016. "On Measuring the Welfare Gains from Trade Under Consumer Heterogeneity." *The Economic Journal*, 126(593): 1193–1237.

Nocke, Volker. 2006. "A Gap for Me: Entrepreneurs and Entry." *Journal of the European Economic Association*, 4(5): 929–956.

Okubo, Toshihiro, Pierre M Picard, and Jacques-François Thisse. 2010. "The Spatial Selection of Heterogeneous Firms." *Journal of International Economics*, 82(2): 230–237.

Orhun, A Yeşim. 2013. "Spatial differentiation in the supermarket industry: The role of common information." *Quantitative Marketing and Economics*, 11(1): 3–37.

Osharin, Alexander, Jacques-François Thisse, Philip Ushchev, and Valery Verbus. 2014. "Monopolistic competition and income dispersion." *Economics Letters*, 122(2): 348 – 352.

Pontryagin, Lev S. 1962. *Ordinary Differential Equations.* Elsevier, Amsterdam.

Quan, Thomas W, and Kevin R Williams. 2018. "Product variety, across-market demand heterogeneity, and the value of online retail." *The RAND Journal of Economics*, 49(4): 877–913.

Richards, Timothy J, Lauren Chenarides, and Metin Çakir. 2022. "Dollar store entry."

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1): 34–55.

Salop, Steven C. 1979. "Monopolistic Competition with Outside Goods." *Bell Journal of Economics*, 10(1): 141–156.

Schmalensee, Richard. 1978. "Entry deterrence in the ready-to-eat breakfast cereal industry." *The Bell Journal of Economics*, 305–327.

Seim, Katja. 2006. "An empirical model of firm entry with endogenous product-type choices." *The RAND Journal of Economics*, 37(3): 619–640.

Shaked, Avner, and John Sutton. 1982. "Relaxing price competition through product differentiation." *The review of economic studies*, 49(1): 3–13.

Shaked, Avner, and John Sutton. 1987. "Product differentiation and industrial structure." *The Journal of Industrial Economics*, 131–146.

Sharapudinov, Shamil. 2022. "Trade in Niches, Variable Markups, and Firm Sorting in General Equilibrium." *Available at SSRN: https://ssrn.com/abstract=4010994 (January 16, 2022)*.

Simonovska, Ina. 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." *The Review of Economic Studies*, 82(4): 1612–1656.

Spence, A Michael. 1975. "Monopoly, quality, and regulation." *The Bell Journal of Economics*, 417–429.

Spence, Michael. 1976*a*. "Product differentiation and welfare." *The American Economic Review*, 66(2): 407–414.

Spence, Michael. 1976*b*. "Product selection, fixed costs, and monopolistic competition." *The Review of economic studies*, 43(2): 217–235.

Sundaram, Rangarajan K. 1996. *A First Course in Optimization Theory.* Cambridge University Press.

Sutton, John. 1986. "Vertical product differentiation: Some basic themes." *The American Economic Review*, 76(2): 393–398.

Sweeting, Andrew. 2010. "The effects of mergers on product positioning: evidence from the music radio industry." *The RAND Journal of Economics*, 41(2): 372–397.

Tarasov, Alexander. 2012. "Trade Liberalization and Welfare Inequality: A Demand-Based Approach." *The Scandinavian Journal of Economics*, 114(4): 1296–1317.

Thisse, Jacques-Francois, and Philip Ushchev. 2018. "Monopolistic Competition without Apology." *Handbook of Game Theory and Industrial Organization, Volume I*, 1: 93.

Ushchev, Philip, and Yves Zenou. 2018. "Price Competition in Product Variety Networks." *Games and Economic Behavior*, 110: 226–247.

Utredninger, Norges Offentlige. 1997. "Nytte-kostnadsanalyser. Prinsipper for lønnsomhetsvurderinger i offentlig sektor."

Vogel, Jonathan. 2008. "Spatial Competition with Heterogeneous Firms." *Journal of Political Economy*, 116(3): 423–466.

Waterson, Michael. 1989. "Models of product differentiation." *Bulletin of Economic Research*, 41(1): 1–27.

Yonezawa, Koichi, and Timothy J Richards. 2016. "Competitive package size decisions." *Journal of Retailing*, 92(4): 445–469.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-Francois Thisse. 2012. "Monopolistic Competition: Beyond the Constant Elasticity of Substitution." *Econometrica*, 80(6): 2765–2784.

Zheng, Fanyin. 2016. "Spatial competition and preemptive entry in the discount retail industry." Columbia Business School Research Paper 16-37.