CESifo WORKING PAPERS

10263
2023

February 2023

A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers

Sergey Kokovin, Alina Ozhegova, Shamil Sharapudinov, Alexander Tarasov, Philip Ushchev

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10263*

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.

JEL-Codes: F100, L110, L130.

Keywords: firm heterogeneity, geographical space, product space, positive sorting, product niches.

Sergey Kokovin
HSE University, Moscow / Russia
skokovin@hse.ru

Alina Ozhegova
Norwegian School of Economics
Department of Economics, Bergen / Norway
alina.ozhegova@nhh.no

Shamil Sharapudinov
Pennsylvania State University, Department of Economics, University Park / PA / USA
sbs6259@psu.edu

Alexander Tarasov
HSE University, Moscow / Russia
atarasov@hse.ru

Philip Ushchev
HSE University, Moscow / Russia
fuschev@hse.ru

January 2023
This paper was previously circulated under the title “Heterogeneous Consumers Matching Heterogeneous Firms in Monopolistic Competition” as well as CESifo Working Paper No. 8082. We are grateful to K. Behrens, S. Demidova, R. Edwards, H. Egger, D. Gomtsyan, E. Kazakova, K. Matsuyama, Y. Murata, M. Parenti, A. Rodríguez-Clare, and conference and seminar participants at the HSE (St. Petersburg), XVIII April Conference on Economic and Social Development, the Oligo Workshop 2017, the 2017 Asian Meeting of the Econometric Society, the 2nd World Congress of Comparative Economics, the 6th “IOSE” conference, the 2017 PSE Summer School on Trade, EEA 2017, EARIE 2017, and International research seminar "International Trade and Regional Economics" (Moscow) for useful comments and suggestions. Our special thanks to M. Goryunov, N. Schutz, and J.-F. Thisse for fruitful conversations during the early stages of this project. Kokovin, Ushchev and Tarasov gratefully acknowledge support from the Basic Research Program of the National Research University Higher School of Economics (HSE University). All remaining errors are ours.

... The Fox seemed perplexed, and very curious. ... "Are there hunters on that planet?" "No." "Ah, that is interesting! Are there chickens?" "No." "Nothing is perfect," sighed the fox.

"Le Petit Prince". Antoine de Saint-Exupéry.

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). 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 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 as a product niche choice, i.e., which group of consumers (defined by their common tastes) to serve.[3]

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

[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 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

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 (hunters). Such a setup provides new insights on 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's 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 firm's sales, prices, and markups and may result in a deeper understanding of data: in particular, a firm may be 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.

productivities and product characteristics – whether it be a product niche or some other product/consumer-specific attribute – can occur.

[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).

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 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 et al. (2021), who 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, who employ 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), who 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).

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 purchasing 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; 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 firm's productivity, location, and price) is supermodular in location and price, then each equilibrium displays positive assortative matching.

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 Lancaster 1966, Salop 1979; Chen and Riordan 2007, and Vogel 2008 among others). 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), 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

[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.

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 studying the equilibrium markup patterns. Goryunov et al. (2022) consider a monopolistic competition framework with spatially distributed consumers. However, in contrast to the present paper, this work focuses on the case with homogeneous firms and uniformly distributed consumers. As a result, it does not examine sorting of firms across product niches. Another paper related to ours is Ushchev and Zenou (2018), who develop a model of price competition in product-variety networks. Both consumers and suppliers of a differentiated product are embedded into a network which 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) abstract 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 size. 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 explore its implications for markups and consumer welfare. Among other studies, Okubo et al. (2010) explores how trade liberalization affects sorting across location in a two-country model with linear demand. Behrens et al. (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 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 et al. (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 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), who 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 2 we develop a baseline model of fully localized spatial monopolistic competition with unspecified functional form of consumer demand. In Section 3, 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 4, we discuss an extension of our model to the case when firms compete not just within but also across the niches. Section 5 concludes.

2 The 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

[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 et al. (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 trade of the poor (rich).

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).

2.1 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 is that price competition among firms can be described as an aggregative game (Anderson et al., 2020), which makes the analysis of firm behavior and equilibrium characterization relatively simple. In Section 4, we discuss the consequences of relaxing this assumption.

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

$$\mathcal{U}_x = V\left(\int_{\omega \in \Omega_x} u(q(\omega, x))\mathrm{d}\omega\right) + q_0. \tag{2.1}$$

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 numéraire. The function $V : \mathbb{R}_+ \to \mathbb{R}$ is an upper-tier utility function, which captures the

[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). 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.

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 (*2.1*) subject to the budget constraint given by

$$\int_{\omega \in \Omega_x} p(\omega)\, q(\omega, x)\, \mathrm{d}\omega + q_0 \le I, \tag{2.2}$$

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 and, therefore, allows for positive consumption of the numeraire, and using the standard monotonicity argument, the consumer's utility maximization problem can be restated as follows:

$$\max_{q(\cdot,x)} \left[V\left(\int_{\omega \in \Omega_x} u(q(\omega, x))\mathrm{d}\omega \right) - \int_{\omega \in \Omega_x} p(\omega) q(\omega, x)\mathrm{d}\omega \right]. \tag{2.3}$$

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

$$p(\omega) = \frac{u'(q(\omega, x))}{\lambda(x)}, \tag{2.4}$$

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

$$\lambda(x) \equiv \frac{1}{V'\left(\int_{\omega \in \Omega_x} u(q(\omega, x))\mathrm{d}\omega \right)}. \tag{2.5}$$

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 (2.4) 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:

$$q(\omega, x) = D(\lambda(x) p(\omega)), \tag{2.6}$$

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

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

$$D(z) \equiv \begin{cases} u'^{-1}(z), & \text{if } z < u'(0), \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

for all $z > 0$.

Since location x hosts $l(x)$ identical consumers, (2.6) implies that the market demand $Q(p(\omega), x)$ for variety $\omega \in \Omega_x$ is given by

$$Q(p(\omega), x) \equiv q\,(\omega, x)\, l(x) = D\,(\lambda(x)p(\omega))\, l(x). \tag{2.8}$$

As can be seen from equation (2.8), 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.

2.2 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* firm. Second, based on their draws 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. Forth, 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 (2.8) 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

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

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:

$$\Pi(p, x; c) \equiv (p - c)\, D\left(\lambda(x)p\right) l(x), \tag{2.9}$$

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 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:

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

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

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

Using (2.9) 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.

2.3 Sorting between firms and locations

In this section, we show that, under 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:

$$\frac{p - c}{p} = \frac{1}{\mathcal{E}_D\left(\lambda(x)p\right)}, \tag{2.10}$$

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

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

Equation (2.10) is the standard monopoly pricing condition. Solving (2.10) w.r.t. p, we obtain the relationship between the price and 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]

$$\frac{l(x)}{l'(x)} \cdot \frac{\lambda'(x)}{\lambda(x)} = \frac{1}{\mathcal{E}_D(\lambda(x)p(x,c))}. \tag{2.11}$$

Combining (2.10) and (2.11), we derive a neat expression for the markup $\mathcal{M}(x,c)$:

$$\mathcal{M}(x,c) \equiv \frac{p(x,c)-c}{p(x,c)} = \frac{\lambda'(x)}{\lambda(x)} \cdot \frac{l(x)}{l'(x)}. \tag{2.12}$$

The expression for markups given by (2.12) 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 in the market. The markups set by these firms are strictly positive (since there is the fixed cost of production). From (2.12), positive markups imply that $\lambda'(x) < 0$ on (a,b) (as $l'(x) < 0$ on (a,b)). □

The result in 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 (2.12) 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 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.

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

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 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).

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.

[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.

2.4 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 imply 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 a 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.

An equilibrium is then a 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 following conditions hold:

C1 The measure of competition intensity satisfies:

$$\lambda(x) = \frac{1}{V'\left(\mu(x)u(q(x))\right)}, \tag{2.13}$$

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 $x > x(\bar{c}) \equiv \bar{x}$, $\lambda(x) = 1/V'\left(0\right)$ for all $x \in (\bar{x},\ S]$. To hold the continuity of the problem, the value of $\lambda(x)$ defined in (2.13) at the rightmost location $\bar{x}$ must be equal to $1/V'\left(0\right)$. Equivalently, $c'(\bar{x})$ must be equal to zero.

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

$$\frac{p-c}{p} = \frac{1}{\mathcal{E}_D\left(\lambda(x)p\right)}. \tag{2.14}$$

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

$$\frac{p(x,c)-c}{p(x,c)} = \frac{\lambda'(x)}{\lambda(x)}\frac{l(x)}{l'(x)}, \tag{2.15}$$

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

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

$$\Pi\left(\bar{c}, p(\bar{c}), x(\bar{c})\right) = f. \tag{2.16}$$

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

$$\int_{c_{\min}}^{\bar{c}} \left(\Pi\left(c, p(c), x(c)\right) - f\right) \cdot g\left(c\right) dc = f_e. \tag{2.17}$$

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 2.5) that $\mathcal{M}(x,c)$ and $q(x)$ are functions of $\lambda(x)c$. Thus, the system can be rewritten as follows:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}x} = -a(x)\lambda\mathcal{M}(\lambda c), \tag{2.18}$$

$$\frac{\mathrm{d}c}{\mathrm{d}x} = \frac{1}{M_e}\frac{\left(V'\right)^{-1}(1/\lambda)}{g(c)\, u(q(\lambda c))}. \tag{2.19}$$

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 $l\left(S\right)$ implies that $\bar{x} < S$, while sufficiently high $l\left(0\right)$ 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.

2.5 The 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 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 an account (2.4), 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:

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

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

$$\frac{\partial \Pi\,(c, q, x)}{\partial q} = 0 \Leftrightarrow u'\,(q) + q\, u''\,(q) = \lambda(x)c. \tag{2.20}$$

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:

$$\frac{\partial \Pi\,(q, x, c)}{\partial x} = 0 \Leftrightarrow \frac{\lambda'(x)}{\lambda(x)} \frac{l(x)}{l'(x)} = 1 - \frac{\lambda(x)c}{u'\,(q(x, c))} = -\frac{q(x, c)\, u''\,(q(x, c))}{u'\,(q(x, c))}.$$

The latter implies that a firm's markup, $\mathcal{M}(x, c)$, is equal to $\mathcal{E}_{u'}\,(q(x, c))$. 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

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

and consumers have different locations, the behavior of the markup function is solely determined by $\lambda(x)$, which is decreasing in x. This implies that the markup function is increasing in x: firms located further from zero set higher markups. Indeed, to compensate 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 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 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.

2.6 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. 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's 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 a 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 least productive firms leave the market.

3 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 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 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

[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.

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, 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).

3.1 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, the primary Norwegian provider of spatial data. The database ("Bedriftsregister") contains information for the period 2015-2020 on all businesses registered in Norway, 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.[18] To calibrate the model, we employ revenues and profits for 2019.

The other important data source became available due to the Business Compensation Scheme – 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, which applied for the support, had to specify their turnover and fixed costs in March 2020 and in the corresponding

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

period one year ago (March 2019). The fixed costs are defined as the costs that cannot be reduced in the short term together with 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 been 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. 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 1.

Table 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

The demographics data is taken from publicly available databases managed by Statistics Norway and Geonorge (a public initiative for managing spatial data). To calibrate the distribution of population, we use the division of Norway into the smallest geographical 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 8 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 2.

3.2 Calibration strategy

In this subsection, we describe our calibration procedure.

Table 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

3.2.1 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 which capture, respectively, 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 9 in the Appendix.

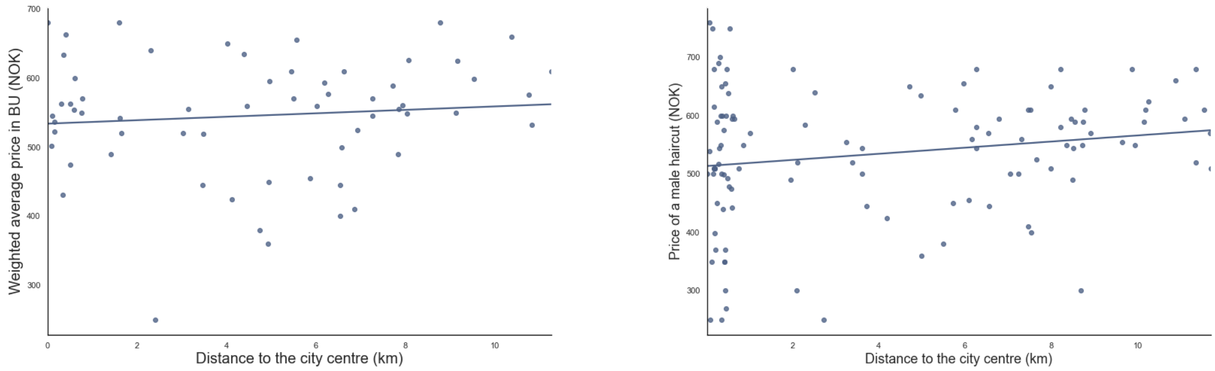
To calibrate γ, we note that γ is the elasticity of $1 - l(x)/l(0)$ with respect to x. Using the empirical counterpart of $1 - l(x)/l(0)$, where x is the BU distance from the city center and $l(0)$ is the maximum population size across all BUs, we run the corresponding OLS regression and find that the 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: $l(x) = 24526(1 - (x/16.2)^{0.18})$.

3.2.2 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 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.

Figure 1: Haircut prices in Bergen



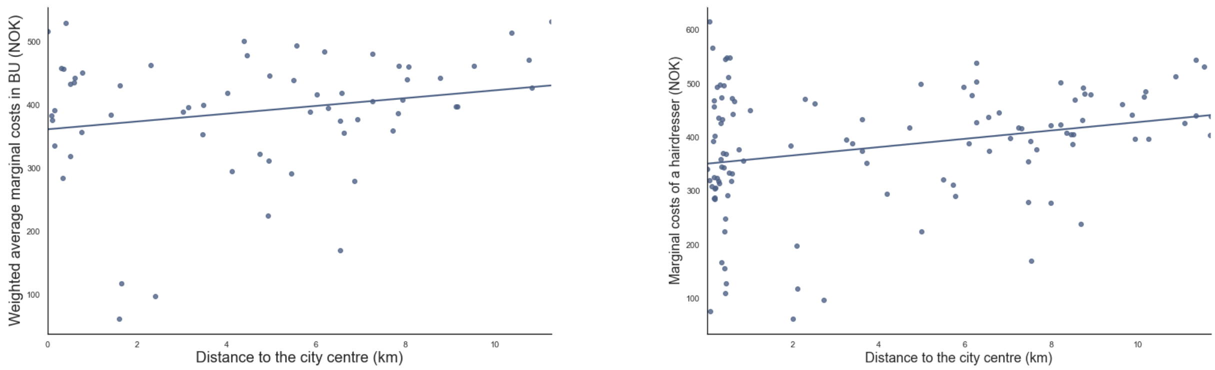
Note: The left panel: each dot represents one basic unit of Bergen. The number of dots (which is 62) corresponds to the number of basic units with at least one hair salon. The estimate of the slope parameter is 2.22 with no significance. The right panel: each dot represents the price of a certain hairdresser.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

The data on revenues, profits, prices, and fixed costs allow us to calculate the marginal costs of production of each hairdresser in the sample.[19] Figure 2 depicts the relationship between the hairdressers' marginal costs and its remoteness from the city center. As can be seen, less productive hairdressers tend to locate further from the city center. This is in line with our theory when we assume the increasing in price demand elasticity. Figure 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 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, we do not observe such a pattern. However, the relationship between markups and marginal costs (see Figure 5) is "closer" to being non-monotonic: markups are first decreasing in marginal costs

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

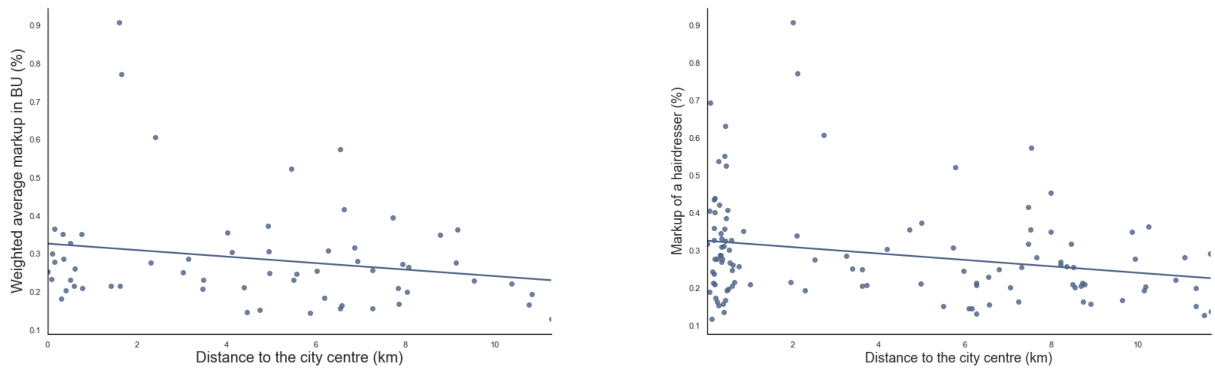
and then seem to be slightly increasing.

Figure 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 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.

Figure 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 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.

For the theoretical distribution of firm productivity, we assume the Weibull functional form on $[c_{min}, \infty)$:

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

with the density function

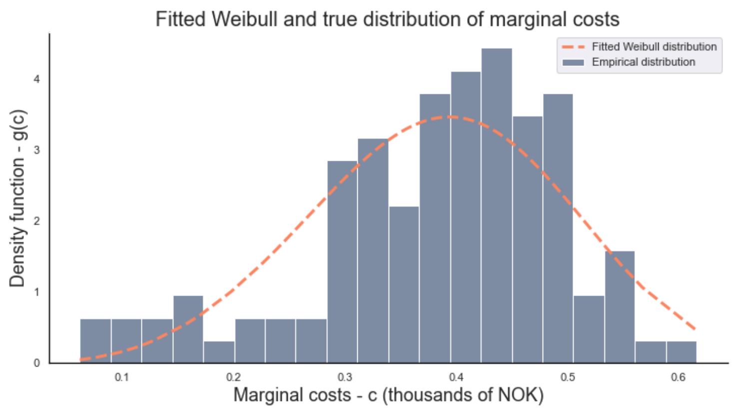
$$g(c) = \frac{e^{-\left(\frac{c}{\alpha}\right)^k}}{e^{-\left(\frac{c_{\text{min}}}{\alpha}\right)^k}} \frac{k}{\alpha} \left(\frac{c}{\alpha}\right)^{k-1}$$

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 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 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, 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 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 4: Empirical and calibrated distribution of marginal costs



Note: Marginal costs area 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).

3.2.3 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.

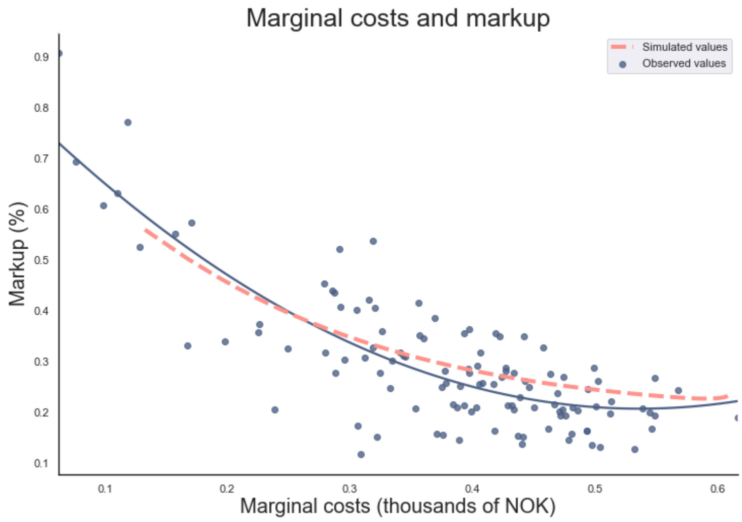
Table 3: Calibration Strategy

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

3.3 Results and counterfactual analysis

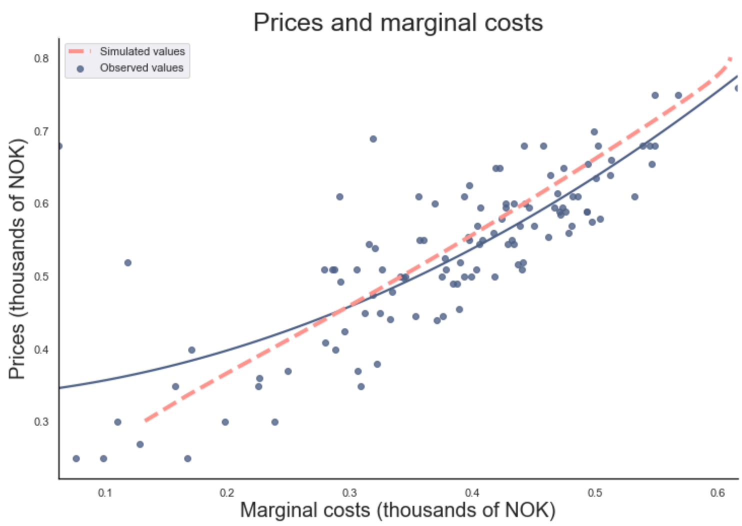
Our calibration strategy results in f_e and a being equal to 2018.4 and 0.107, respectively. To assess how well the model fits the data, we present two figures. Figure 5 depicts the relationship between marginal costs and markups in the data and the one generated by the calibrated model. As can be seen, 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 is consistent with the data. Figure 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 being 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 5: Simulated and observed marginal costs and markups



Note: Each dot represents one hairdresser. Both parameters of the fitted parabola are 1% significant.

Figure 6: Simulated and observed marginal costs and prices



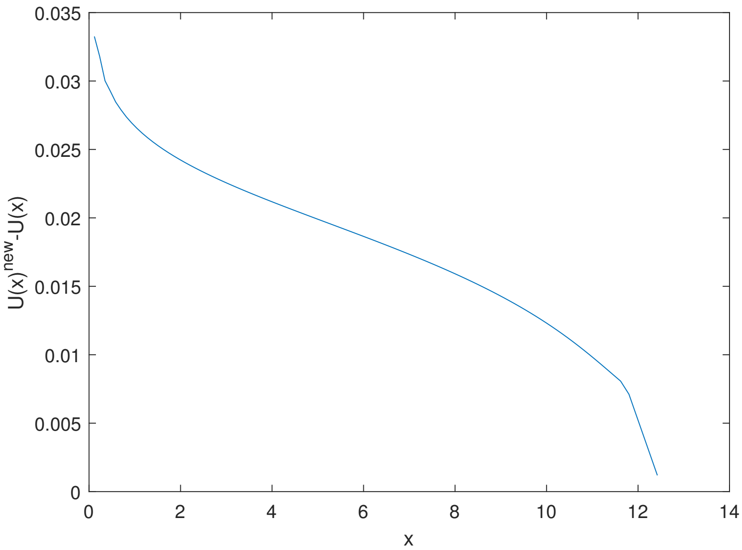
Note: Each dot represents one hairdresser.

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 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 for these locations on markups is positive and strong enough to compensate 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 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 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 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 yield 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, as in the previous experiment, the more remote locations gain less than those closer to the city center. In fact, 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 have 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.[20]

4 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

$$\mathcal{U}_x = V\left(\int_X k_\tau(x,y)\int_{\Omega_y} u(q(\omega,x))\mathrm{d}\omega\mathrm{d}y\right) + q_0, \tag{4.1}$$

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 $\omega \in \Omega_y$ 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 (2.1).

This way of modeling preferences is akin to the model proposed by Ushchev and Zenou (2018), where 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 2 on $V(\cdot)$ and $u(\cdot)$, respectively. In addition to these, we impose the following assumption about the spatial discount factor $k_\tau(x,y)$:

[20] In the earlier version of the paper (see Kokovin, Sharapudinov, Tarasov, and 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 affect the implications of a uniform increase in population.

Assumption 3. The kernel $k_\tau : X \times X \to \mathbb{R}_+$ representing the spatial discount factor in (4.1) has the following structure:

$$k_\tau(x, y) = \tau\psi(\tau|x - y|), \tag{4.2}$$

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 $\psi(\cdot)$ may be (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 2) 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 $\{x\}$.[21] As a result, when $\tau \to \infty$, (4.1) becomes (2.1).

A consumer located at $x \in X$ seeks to maximize her utility (4.1) subject to the budget constraint, which is now given by

$$\int_X \int_{\omega\in\Omega_y} p(\omega)\, q(\omega, x)\mathrm{d}\omega\mathrm{d}y + q_0 \leq I, \tag{4.3}$$

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:

$$\max_{q(\cdot)} \left[V\left(\int_X k_\tau(x, y) \int_{\Omega_y} u(q(\omega, x))\mathrm{d}\omega\mathrm{d}y \right) - \int_X \int_{\omega\in\Omega_y} p(\omega)q(\omega, x)\mathrm{d}\omega\mathrm{d}y \right]. \tag{4.4}$$

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

$$\frac{p(\omega)}{k_\tau(x, y)} = \frac{u'(q(\omega, x))}{\lambda(x)}, \tag{4.5}$$

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

$$\lambda(x) \equiv \frac{1}{V'\left(\int_X k_\tau(x, y) \int_{\Omega_y} u\left(q\left(\omega, x\right)\right)\mathrm{d}\omega\mathrm{d}y\right)}. \tag{4.6}$$

Solving (4.5) for $q(\omega, x)$, we obtain the individual Marshallian demand of an x-type consumer —

[21] 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.

i.e. a consumer whose preferred product type is x — for variety ω:

$$q\left(\omega, x\right) = D\left(\lambda(x)\frac{p(\omega)}{k_\tau(x,y)}\right), \tag{4.7}$$

where $D(\cdot)$ is the downward-sloping demand schedule defined by (2.7). To obtain the market demand $Q(\omega, x)$ for variety $\omega \in \Omega_x$, we integrate (4.7) across the product space X with respect to the population density:

$$Q(\omega, x) = \int_X D\left(\lambda(y)\frac{p(\omega)}{k_\tau(x,y)}\right) l(y)\mathrm{d}y. \tag{4.8}$$

Equation (4.8) 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 (4.8), we obtain the profit of a c-type firm as a function of price and location choices:

$$\Pi(c, p,\, x) \equiv (p-c)\int_X D\left(\frac{\lambda(y)p}{k_\tau(x,y)}\right) l(y)\mathrm{d}y. \tag{4.9}$$

As in Section 2, 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:

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

We have the following result.

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

$$\frac{dp(c)}{dc} > 0, \qquad \frac{dQ(c)}{dc} < 0. \tag{4.10}$$

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

$$\Pi_{px}(c, p(c),\, x(c)) > 0. \tag{4.11}$$

Proof. In the Appendix. □

Recall that, in the baseline model of fully localized competition (Section 2), when $\tau \to \infty$, the Marshall's Second Law of Demand appears to be sufficient for perfect sorting among firms located on $(0, S)$ in the equilibrium. 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 structure of the expression (4.9) for the profit.

5 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; while less productive firms move to smaller niches to avoid competition with the leaders. This outcome provides new insights on the equilibrium distribution of firm sales, prices, and markups that are now explained not only by 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. This channel is absent in standard spaceless models of monopolistic competition. The framework we develop seems to be quite rich in implications. 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 multiproduct firms within the considered framework with consumer heterogeneity. We leave these questions for further research.

References

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

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

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2018). The Elusive Pro-Competitive Effects of Trade. *The Review of Economic Studies 86*(1), 46–80.

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

Bau, N. (2019). Estimating an Equilibrium Model of Horizontal Competition in Education. Technical report, CEPR Discussion Papers.

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

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy 122*(3), 507–553.

Behrens, K. and Y. Murata (2007). General Equilibrium Models of Monopolistic Competition: A New Approach. *Journal of Economic Theory 136*(1), 776–787.

Behrens, K. and F. Robert-Nicoud (2015). Agglomeration Theory with Heterogeneous Agents. *Handbook of Regional and Urban Economics 5*, 171 – 245.

Bellone, F., P. Musso, L. Nesta, and F. Warzynski (2014). International Trade and Firm-Level Markups when Location and Quality Matter. *Journal of Economic Geography 16*(1), 67–91.

Carballo, J., G. I. Ottaviano, and C. V. Martincus (2018). The Buyer Margins of Firms' Exports. *Journal of International Economics 112*, 33 – 49.

Chaney, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review 98*(4), 1707–21.

Chen, Y. and M. H. Riordan (2007). Price and Variety in the Spokes Model. *The Economic Journal 117*(522), 897–921.

Courant, R. and F. John (2012). *Introduction to Calculus and Analysis I.* Springer Science & Business Media.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, Markups, and Trade Reform. *Econometrica 84*(2), 445–510.

Dhingra, S. and J. Morrow (2019). Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. *Journal of Political Economy 127*(1), 196–232.

Díez, F. J., J. Fan, and C. Villegas-Sánchez (2021). Global declining competition? *Journal of International Economics 132*, 103492.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review 67*(3), 297–308.

Eckel, C. and J. P. Neary (2010). Multi-Product Firms and Flexible Manufacturing in the Global Economy. *The Review of Economic Studies 77*(1), 188–217.

Evans, L. C. (2010). *Partial Differential Equations*. Graduate Studies in Mathematics, Vol. 19. American Mathematical Society, Providence, Rhode Island.

Faber, B. and T. Fally (2020). Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data. *The Review of Economic Studies (forthcoming)*.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). Income Distribution, Product Quality, and International Trade. *Journal of Political Economy 119*(4), 721–765.

Feenstra, R. C. and J. Romalis (2014). International Prices and Endogenous Quality. *The Quarterly Journal of Economics 129*(2), 477–527.

Fieler, A. C. and A. Harrison (2019). Escaping Import Competition in China. Working paper.

Gaubert, C. (2018). Firm Sorting and Agglomeration. *American Economic Review 108*(11), 3117–53.

Goryunov, M., S. Kokovin, and T. Tabuchi (2022). Continuous spatial monopolistic competition: matching goods with consumers. *Economic Theory 74*(3), 793–832.

Holmes, T. J. and J. 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, H. (1929). Stability in Competition. *The Economic Journal 39*(153), 41–57.

Kaldor, N. (1935). Market Imperfection and Excess Capacity. *Economica 2*(5), 33–50.

Kokovin, S. G., S. Sharapudinov, A. Tarasov, and P. Ushchev (2020). A Theory of Monopolistic Competition with Horizontally Heterogeneous Consumers. *CESifo Working Paper No. 8082*.

Kugler, M. and E. Verhoogen (2012). Prices, Plant Size, and Product Quality. *The Review of Economic Studies 79*(1), 307–339.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy 74*(2), 132–157.

Loertscher, S. and G. Muehlheusser (2011). Sequential location games. *The RAND Journal of Economics 42*(4), 639–663.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Mrazova, M. and J. P. Neary (2017). Not So Demanding: Demand Structure and Firm Behavior. *American Economic Review 107*(12), 3835–74.

Neiman, B. and J. S. Vavra (2019). The Rise of Niche Consumption. Working Paper 26134, National Bureau of Economic Research.

Nigai, S. (2016). On Measuring the Welfare Gains from Trade Under Consumer Heterogeneity. *The Economic Journal 126*(593), 1193–1237.

Nocke, V. (2006). A Gap for Me: Entrepreneurs and Entry. *Journal of the European Economic Association 4*(5), 929–956.

Okubo, T., P. M. Picard, and J.-F. Thisse (2010). The Spatial Selection of Heterogeneous Firms. *Journal of International Economics 82*(2), 230–237.

Osharin, A., J.-F. Thisse, P. Ushchev, and V. Verbus (2014). Monopolistic competition and income dispersion. *Economics Letters 122*(2), 348 – 352.

Pontryagin, L. S. (1962). *Ordinary Differential Equations*. Elsevier, Amsterdam.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy 82*(1), 34–55.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *Bell Journal of Economics 10*(1), 141–156.

Sharapudinov, S. (2022). Trade in Niches, Variable Markups, and Firm Sorting in General Equilibrium. *Available at SSRN: https://ssrn.com/abstract=4010994 (January 16, 2022).*

Simonovska, I. (2015). Income Differences and Prices of Tradables: Insights from an Online Retailer. *The Review of Economic Studies 82*(4), 1612–1656.

Sundaram, R. K. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

Tarasov, A. (2012). Trade Liberalization and Welfare Inequality: A Demand-Based Approach. *The Scandinavian Journal of Economics 114*(4), 1296–1317.

Thisse, J.-F. and P. Ushchev (2018). Monopolistic Competition without Apology. *Handbook of Game Theory and Industrial Organization, Volume I 1*, 93.

Ushchev, P. and Y. Zenou (2018). Price Competition in Product Variety Networks. *Games and Economic Behavior 110*, 226–247.

Vogel, J. (2008). Spatial Competition with Heterogeneous Firms. *Journal of Political Economy 116*(3), 423–466.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012). Monopolistic Competition: Beyond the Constant Elasticity of Substitution. *Econometrica 80*(6), 2765–2784.

6 Appendix

In this Appendix, we provide the proofs of some lemmas and propositions as well as some figures.

Proofs of some Lemmas and Propositions

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

$$x_{\max} \equiv l^{-1}\left(\frac{\lambda_{\min} f}{\pi(\lambda_{\min} c_{\min})}\right). \tag{6.1}$$

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

$$c_{max} \equiv \frac{1}{\lambda_{\min}}\pi^{-1}\left(\frac{\lambda_{\min} f}{l(0)}\right). \tag{6.2}$$

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:

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

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 (6.1) and (6.2). 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 $[0, x_{\max}]$.[22] Therefore, using the Weierstrass theorem, we can define:

$$A \equiv \max_{0 \le x \le x_{\max}} a(x) < \infty. \tag{6.3}$$

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 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.

$$\limsup_{x \downarrow x_0} \|(\lambda_{\overline{x}}(x), c_{\overline{x}}(x))\| = \infty, \tag{6.4}$$

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)

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

[22] 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 $(0, 1)$.

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, (6.4) 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 $(\lim_{x\downarrow x_0}\lambda_{\overline{x}}(x), \lim_{x\downarrow x_0} c_{\overline{x}}(x))$ (the limits exist, as $\lambda_{\overline{x}}(x)$ and $c_{\overline{x}}(x)$ are monotone and bounded). This in turn implies that

$$\lim_{x\downarrow x_0}\left|\frac{\mathrm{d}c_{\overline{x}}(x)/\mathrm{d}x}{\mathrm{d}\lambda_{\overline{x}}(x)/\mathrm{d}x}\right| \leq \frac{u'(0)}{\lim_{x\downarrow x_0}\lambda^2_{\overline{x}}(x)}. \tag{6.5}$$

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 (6.5) 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 $[0, x_{\max}]$:

$$\varphi(\overline{x}) = \begin{cases} c_{\overline{x}}(0) - c_{\min}, & \text{if } (\lambda_{\overline{x}}(x), c_{\overline{x}}(x)) \text{ can be extended up to } \{x = 0\}, \\ -c_{\overline{x}}^{-1}(c_{\min}), & \text{if } (\lambda_{\overline{x}}(x), c_{\overline{x}}(x)) \text{ can be extended up to } \{c = c_{\min}\}. \end{cases} \tag{6.6}$$

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:

$$\left\{ \overline{x}^*, \overline{c}^*, (\lambda^*(x), c^*(x))_{x \in [0, \overline{x}^*]} \right\}. \tag{6.7}$$

We now verify that the candidate equilibrium (6.7) 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:

$$\Pi_e(M_e) \equiv \int_{c_{\min}}^{\overline{c}^*(M_e)} \left[\frac{l(x^*(c, M_e))}{\lambda^*(c, M_e)} \pi\left(\lambda^*(c, M_e)c\right) - f \right] g(c) \mathrm{d}c = f_e, \tag{6.8}$$

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

$$f_e < \int_{c_{\min}}^{c_{\max}} \left[\frac{l(0)}{\lambda_{\min}} \pi\left(\lambda_{\min} c\right) - f \right] g(c) \mathrm{d}c. \tag{6.9}$$

Further, we show that this condition is sufficient for equation (6.8) 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 (6.8) 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 (6.8) 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:

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

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

$$c^{**}(\overline{x}^{**} - \overline{x}^*) > c^{**}(0) = c_{\min} = c^*(0) = c^*(\overline{x}^* - \overline{x}^*). \tag{6.10}$$

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

$$\Delta(z) \equiv c^{**}(\overline{x}^{**} - z) - c^*(\overline{x}^* - z). \tag{6.11}$$

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

$$\lambda^{**}(\overline{x}^{**} - \xi_0) > \lambda^*(\overline{x}^* - \xi_0). \tag{6.12}$$

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

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

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:

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

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 $z < \xi_0$):

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

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 (6.12).

Step 3. Differentiating the function $\Delta(z)$ defined by (6.11) at $z = \xi_0$, we obtain:

$$\Delta'_z(\xi_0) = -\frac{1}{M_e g\left(c_0^*\right)}\left[\frac{\left(V'\right)^{-1}\left(1/\lambda_0^{**}\right)}{u\left(q\left(\lambda_0^{**}c_0^*\right)\right)} - \frac{\left(V'\right)^{-1}\left(1/\lambda_0^{*}\right)}{u\left(q\left(\lambda_0^{*}c_0^*\right)\right)}\right] < 0. \quad (6.13)$$

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 (6.13) holds true because, by (6.12), we have $\lambda_0^{**} > \lambda_0^*$, while the function $\left(V'\right)^{-1}(1/\lambda)/u\left(q\left(\lambda c\right)\right)$ 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 (6.13) 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:

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

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

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

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

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

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

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

In fact, it is sufficient to show that

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

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, $\partial \left(c^*\right)'_x (x_1, M_e)/\partial M_e < 0$ (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 $\partial \left(c^*\right)'_x (x_1, M_e)/\partial M_e < 0$ 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, $d\overline{x}^*(M_e)/dM_e \le 0$. 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: $\mathrm{d}\overline{c}^*(M_e)/\mathrm{d}M_e \ge 0$. That is,

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

where the first term is non-positive because, as shown above, $\mathrm{d}\overline{x}^*(M_e)/\mathrm{d}M_e \le 0$, 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 $\partial\lambda^*\left(x, M_e\right)/\partial M_e > 0$ for all $x \in \mathcal{N} \equiv \mathcal{N}_1 \cup \mathcal{N}_2$. Hence, for a c-type firm where $c = c^*\left(x, M_e\right)$ 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 $\partial\lambda^*\left(x, M_e\right)/\partial M_e \le 0$ over $[x_1, x_2]$ means that the profit function increases uniformly over $[x_1, x_2]$, while $\partial\lambda^*\left(x, M_e\right)/\partial M_e > 0$ 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.[23] This immediately imply that

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

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

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

[23] 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 discontinuities.

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

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

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

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

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

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

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

Hence,

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

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

Consider,

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

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

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

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

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

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

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

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

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

We have

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

since $c'(x) > 0$, $g'(c) \geq 0$, and $(l'(x)/l(x))'_x \leq 0$. At the same time, $(V')^{-1}(1/\lambda(x))$ 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 $\partial\overline{c}(M_e, \delta)/\partial\delta > 0$ and $\lambda(\overline{x}, M_e, \delta) = \lambda_{\min}$. This in turn also means that $\partial\lambda(x, M_e, \delta)/\partial\delta < 0$ in some neighborhood of $\overline{x}$. That is, if $\partial\overline{x}(M_e, \delta)/\partial\delta \leq 0$, $\lambda(x, M_e, \delta)$ must decrease w.r.t. δ over some interval $(x_1, \overline{x})$. 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 $\overline{x}$ as $\partial\overline{c}(M_e, \delta)/\partial\delta > 0$. This implies that there exists $x_2 \in (0, \overline{x})$ 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 $(x_1, \overline{x})$. 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 $x_3 \in (x_1, \overline{x})$ 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 $\partial\overline{x}(M_e, \delta)/\partial\delta > 0$.

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,

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

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

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

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 $\partial\,(c)'_x\,(x_7, M_e, \delta)/\partial\delta < 0$ (recall that $\partial\,(c)'_x\,(x_7, M_e, \delta)/\partial\delta$ 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

$$\begin{pmatrix} dp(c)/dc \\ dx(c)/dc \end{pmatrix} = -\begin{pmatrix} \Pi_{pp} & \Pi_{px} \\ \Pi_{px} & \Pi_{xx} \end{pmatrix}^{-1} \begin{pmatrix} \Pi_{cp} \\ \Pi_{cx} \end{pmatrix}, \tag{6.14}$$

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 (6.14) yields

$$\begin{pmatrix} dp(c)/dc \\ dx(c)/dc \end{pmatrix} = \frac{1}{\Pi_{pp}\Pi_{xx} - \Pi_{px}^2} \begin{pmatrix} \Pi_{xx}Q_p \\ -\Pi_{px}Q_p \end{pmatrix}. \tag{6.15}$$

Using (6.15) 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 (30).

(ii) The equivalence of the inequality in (31) to $\mathrm{d}x(c)/\mathrm{d}c > 0$ follows immediately from (6.15) and the SOC.

Some Figures

Figure 8: Basic Units in the City of Bergen

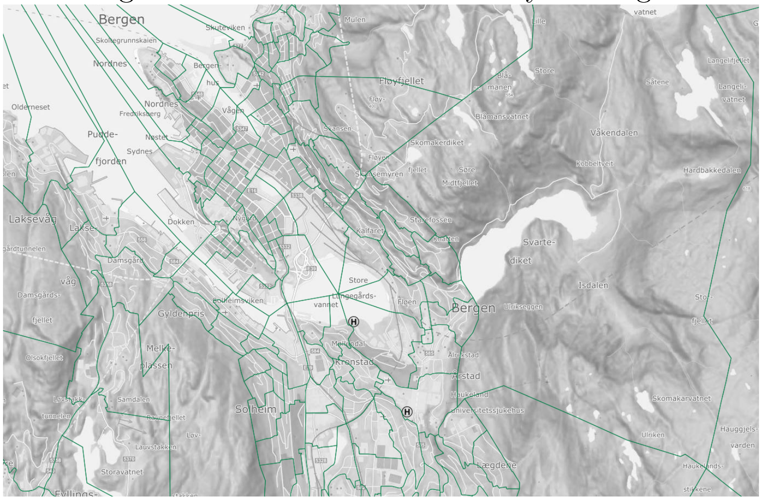
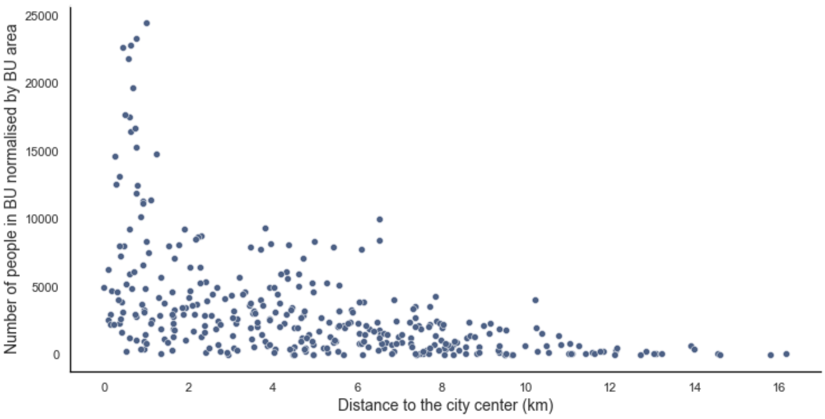


Figure 9: Distribution of population in Bergen



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.