



Master Thesis

Jonas K. Sekamane

Competitive location behaviour with foresight

Supervisor: Johan Lagerlöf

ECTS points: 30

Date of submission: 15/06/2016

Keystrokes: Approx. 185.493



Competitive location behaviour with foresight

Jonas K. Sekamane

Abstract

This paper considers a competitive location model, where firms compete with one another through their location decision in a two-dimensional market space. The paper studies how firms locate and the corresponding competitive environment. In particular the paper investigates when firms agglomerate at the centre or seek out niche segments of the market. And whether one firm captures a predominant share of customers, when firms compete through location, or if the market is evenly partitioned among firms. Both a symmetric unimodal and an asymmetric and bimodal distribution of consumers are examined. An agent-based modelling approach is used and the paper proposes a new decision rule with foresight – as an alternative to the current heuristic decision rules, where firms are oblivious to the simultaneous action and future response from competing firms. The new decision rule uses *inductive reasoning*, i.e. the firm holds several hypotheses on how competing firms chooses to locate, and use the most probable hypothesis to predict the future location of competing firms. In addition the firm gradually discards poorly performing hypotheses and forms new hypotheses, thus the hypotheses of the firms are endogenous to the model. Firms acting with foresight enable strategic considerations.

1	Introduction	1
2	Literature review	5
2.1	Multiple firms and two-dimensional space	6
2.2	Location problems	8
2.3	Voronoi diagram	9
2.4	Hinterlands and vertical product differentiation	11
2.5	Foresight	12
2.6	Agent-based modelling	15
3	Model	19
3.1	Consumers	20
3.1.1	Consumer distribution	21
3.2	Firm behaviour	26
3.2.1	Sticker-rule	26
3.2.2	Aggregator-rule	27
3.2.3	Hunter-rule	28
3.2.4	Deliberate reaction to competitors	28
3.2.5	Maxcov-rule	30
3.3	Summary variables	32
4	Methodology	33
4.1	Model parameterisation	35
4.2	Markov chain	36
4.2.1	Stochastic process	36
4.2.2	Time-homogenous Markov chain	37
4.2.3	Estimating output variables	40
4.2.4	Determining burn in	43
4.2.5	Oscillation in all-maxcov model	44
5	Analysis	46
5.1	Baseline model and decision rules	47
5.2	Models with an asymmetric and bimodal distribution	52
5.3	Decision rules with foresight	58
5.3.1	Maxcov-inductor rule	59
5.3.2	Maxcov-inductor-GA rule	62
5.4	Models with foresight	63
5.4.1	Maxcov-inductor model	64
5.4.2	Maxcov-inductor-GA model	70
6	Conclusion	77
Appendix A	Limitations on the effective number of firms (ENP)	80
Appendix B	Model overview	80
Appendix C	Maxcovrnd decision rule	82
Appendix D	Further details on condition/forecast rules	83
Appendix D.1	Genetic algorithm	85
Appendix E	Mean representation in maxcov-inductor model	87
References		88

1. Introduction

In the field of economics, competition is often synonymous with price competition. However businesses compete with one another on a multitude of other factors. Retailers compete for customers when they choose their location. Manufacturers compete when they choose product characteristics such as size, colour, shape, style, flavour, etc.

The range of available products is a response to the diversity of consumer preferences. Specific products and specific locations are selected to appeal to different segments of the market. Consumers have idiosyncratic preferences over their preferred variant. Some prefer a sweeter cider to a more sour cider, and vice versa. Similarly consumers favour shops located in their respective neighbourhood over shops located in other neighbourhoods. This inherent heterogeneity among individuals creates a collective demand for product diversity and a demand across all neighbourhoods. Consequently consumers are willing to pay more for variants that better mirror their preferences. With these premiums firms have the possibility to maintain a price level above marginal costs and thus the premiums are a source of market power.

The market is unlikely to sustain a large number of products or retailers – even with broad diversity in consumer preferences – because of *increasing returns to scale* in production, distribution, R&D, etc. Each firm must capture a sufficiently large share of the market to cover its fixed costs. This in turn creates a natural barrier that prohibits each individual consumer from simply producing his or her preferred product. Instead the market will consist of few firms that compete with one another for customers. With few firms in the market it is hard to defend the assumption of *competitive price-taking* firms, or similar assumptions that treat firms as identical passive entities. Instead competition and interaction among firms is strategic. When a profit-maximising firm evaluates different actions – such as its location or product characteristics – it should take account of the potential demand associated with each action. The firm plays an active role in shaping its outcome and can appeal to different segments through its choices. Additionally firms with foresight should consider the action and response from competing firms. If all competitors are contemplating opening a shop on Main Street, then the retailer might secure a larger profit by locating off Main Street despite the lower concentration of consumers there. It is not enough that the action of the firm is profitable given the current state. The firms compete for the same consumers, therefore the long-term profitability is contingent on the firm's ability to predict the future action of competing firms and respond accordingly. This may include, but is not limited to, finding a niche segment of the market that is sufficiently large to sustain the firm, or engaging in direct competition while maintaining the flexibility to adjust to changes.

The natural setting when considering competitive location or competitive product differ-

entiation is a market with few firms where firms act strategically and with foresight. The choice of each firm reflects the diversity of consumer preferences and so the aggregated distribution of consumer preferences is likely to be of great importance. Moreover the aggregated distribution of each consumer's preferred product or location does not necessarily follow a convenient uniform, symmetric or even unimodal distribution. The distribution may well be asymmetric and multi-modal.

This type of competition has received far less attention than price competition, and consequently many important questions have yet to be answered. Most notably what happens when multiple firms compete in a multidimensional space. For instance several retailers locating in a two-dimensional geographic space, or multiple cider producers considering product differentiation along several dimensions such as sweet-sour, sparkling-still, clear-cloudy etc. Every day business owners make decisions in settings such as these. And yet the literature on economics lack a solid theoretical understanding of how firms make these decisions and how it affects the overall market. In light of this it seems evident that product differentiation is a crucial factor missing in our understanding of how modern market economies operate.

This paper will be purely theoretical. It does not aim to provide business owners with best practices on how to locate nor how to differentiate itself from competitors. The aim is to understand the decision process of firms. This requires developing a theoretical model capable of analysing the decisions, yet versatile enough to handle the settings that firms often find themselves competing in (i.e. few firms acting strategic with foresight in a market not necessarily restricted to a single dimension and where the distribution of consumer preferences might be asymmetric and multi-modal). Developing a theoretical model opens the future field of study to more practical applications and analyses. One such analysis might concern the above mentioned market power that arises from the consumers' willingness to pay a premium for product variants reflective of their preferences. If this severely distorts the market, then antitrust and competition laws should take account of it. Another analysis might concern the additional source of market failure that can arise from the selection process. When firms compete through their choice of product characteristics or location, it may lead to an under- or oversupply. But unique to models, where the product selection is endogenous, is also the possibility that competition leads to the wrong products or locations. For instance competition among retailers may lead all retailers to locate in the city centre, even if it is socially optimal for retailers to disperse evenly across the neighbourhoods of the city. If this is the case, then the government or local zoning laws should adjust incentives such that firms choose efficient locations and the efficient set of products from the spectrum of possible variants.

Models studying this type of competition are collectively known as *competitive location*

models. The pioneering model is Hotelling’s line market, where two firms compete for consumers that are evenly distributed along one dimension (Hotelling, 1929). This model has been extended in numerous ways, such as the number of firms, the distribution of consumers, and the number of dimensions. The traditional approach is to solve for the equilibria of the model analytically. The conclusions then follow from analysing these equilibrium settings. The limited number of new insights discovered the last 15 years, suggests that this approach might be reaching its limits. Biscaia & Mota (2013, p. 869) write that “*it would seem that most of the important features that justify the spatial behaviour of firms have already been explored*”. In addition it has been proven that some location models – despite being fairly simple in their setup – belong to a set of problems suspected to be infeasible to solve. An alternative approach to these types of models, recently advanced in the political science literature, rely on *agent-based modelling*. The location problem is set up as a computer model in which the agents act simultaneously and according to a given set of decision rules. The model is analysed by running the simulation and observing how the interaction of agents play out. The agent-based model allows researchers to study settings that are beyond the reach of the analytical approach, such as multiple firms in a two-dimensional space. However in the current models agents act using simple rules of thumb or heuristics without any type of foresight. That is, each agent is oblivious to the future behaviour of the other agents and does not take their behaviour into account when deciding its own location. Thus the agent-based models developed so far lack the strategic considerations fundamental to earlier competitive location models.

This paper uses the agent-based model laid out by Laver & Sergenti (2011). We construct a new decision rule with foresight. With this rule the firm forecasts the future location of competing firms and uses the predicted locations when it decides where to locate. The forecast method used in this paper is a modified version of the one advanced by the *Santa Fe Artificial Stock Market* model (Arthur, 2014, chapter 11). With this method firms use *inductive reasoning*, i.e. they hold several hypotheses on how a competing firm chooses to locate, and use the most probable hypothesis given the current state to predict the future location of the competing firm. The firm gradually discard poorly performing hypotheses and form new hypotheses. The model is dynamic and firms act simultaneously. The results of the model are derived by observing how the interaction of firms play out.

Motivating this paper is not the desire to discover new equilibrium solutions to the competitive location model. Nor is the objective to check existing equilibrium conditions under different assumptions. As the literature review will show this is the path many have gone before and yet main key questions regarding the competitive location behaviour of firms remain unanswered. Rather than continuing along this inefficacious cycle, the paper attempts to con-

struct a model from a set of reasonable assumptions, to see if this path might provide the literature with new answers. The agent-based modelling approach provides this opportunity, and recent attempts using this approach show promise.

This paper disregards price competition and focuses solely on the type of competition that takes place through product differentiation or location differentiation. While disregarding price competition leaves much to be desired in terms of developing a unified and fundamental theoretical model, it reduces the complexity of the model significantly. We leave it up to future research to appropriately incorporate price competition and determine its effect on the results. The paper assumes that all firms charge the same fixed price. A situation like this might arise because of government regulation, or if firms previously committed themselves to a fixed price. The situation may also arise if firms are subject to a *retail price maintenance* agreement, where prices are fixed earlier in the distribution channel, e.g. by manufacturers or wholesalers. For the sake of consistency and to avoid ambiguity all the models are formulated in terms of geographic location. However the results also apply to product differentiation and in the context of political parties competing for voters through their choice of policy positions.

The paper is structured as follows. The literature review covers previous results from competitive location models such as Hotelling's line market, its extensions, and the agent-based model. The section also examines research on related concepts – such as Voronoi diagrams, distance metrics and implementation of learning in economic models – which are used in later sections. The final model is quite complex due to its many interconnected components. To aid the reader the paper gradually adds components to the model and introduces decision rules, starting with the most simple. The model and basic decision rules without foresight are introduced in the *model* section. The *methodology* section presents the approach used to solve the model and estimate the output values of the model. In addition the section explains the various ways used to set the parameter values and initiate the model, and the methods used to determine when the underlying process has *burnt in*, such that the results obtained are independent of the initial position of firms. The *analysis* section first presents the results from the models where firms use the basic decision rules in a market where the aggregated consumer distribution is symmetric and unimodal. Then the section presents the results from the models with an asymmetric and bimodal distribution. This allows the reader to grasp the fundamental parts of the model, before being presented with the more elaborate decision rules with foresight. The section contains a detailed introduction to the decision rules where firms use *inductive reasoning*, and finally presents the results from models with foresight.

2. Literature review

The literature on spatial competition began with [Hotelling's \(1929\)](#) paper. Hotelling proposes a line market where consumers are uniformly distributed along the line and firms produce identical goods. Consumers inelastically demand one unit from the firm with the lowest price. The price is composed of the *mill price* and the *transportation cost*. Each firm charges a *mill price*, i.e. the price for one unit when purchased at the doorsteps of the firm. And customers bear a cost by going to the firm to shop. This *transport cost* is linear in the distance between the customer and the firm. Two firms decide where to locate and which *mill price* to charge. Firms first simultaneously choose their location, and thereafter simultaneously decide on their prices. Two opposing forces come into play; If the firm locates close to its competitor, it will attract more consumers. Meanwhile locating closer also lead to more fierce price competition from its competitor.

If *mill prices* are fixed and equal then there is no price competition and both firms locate at the centre of the market and split the market equally. The agglomeration of firms has been coined the *principle of minimal differentiation*¹. The average distance travelled by consumers is one-fourth the length of the line. This is socially suboptimal. In the socially optimal location, the market is split in two equal halves, with a firm located at the centre in each of these halves. This reduces the average distance travelled by consumers to one-eighth of the length of the line.

An error lead Hotelling to conclude that the model with price competition would also exhibit the *principle of minimal differentiation*. But when two firms are located close to each other near the centre, price competition drives down prices to a level where it is no longer optimal for firms to locate in the middle half of the line. The paper by [d'Aspremont, Gabszewicz & Thisse \(1979\)](#) shows this and that no equilibrium exists in the model laid out by Hotelling. Instead they modify *transportation cost* to be quadratic in the distance between the customer and the firm, rather than linear. With this modification a unique equilibrium exist, but where firms locate at either ends of the line, i.e. *maximal differentiation*. This shows how the *principle of minimal differentiation* breaks down when price competition is introduced. Unfortunately it is just one of many examples of how sensitive the conclusions from the Hotelling model are to changes in the setup.

Below I will review some of the literature on the Hotelling model and more generally spatial competition models. I will present the reader with papers that relax the assumptions and generalise the Hotelling model. [Biscaia & Mota \(2013\)](#) find 352 journal articles published since 1979 that reference 'Hotelling' or 'spatial competition'. The sheer volume of papers on

¹The *principle of minimal differentiation* was coined by Boulding in 1955 ([Biscaia & Mota, 2013](#)).

the topic makes it a colossal task to provide an exhaustive review. This literature review is not exhaustive, but selective and focused on the papers deemed relevant for this particular paper. The following set of papers provide an exhaustive review of the topic; The early developments and historical development of the spatial competition literature is reviewed by [Eiselt & Marianov \(2011, chapter 1\)](#) and [Smith, Laporte & Harper \(2009\)](#). [Biscaia & Mota \(2013\)](#) contain a bibliometric analysis of the literature. [Eiselt \(2011\)](#) provide a reissue of Hotelling’s original paper with comments and references to later contributions. [Graitson \(1982\)](#); [Brenner \(2001\)](#); [Eiselt, Laporte & Thisse \(1993\)](#) provide literature reviews focused on the Hotelling model. While [Kilkenny & Thisse \(1999\)](#) and [ReVelle & Eiselt \(2005\)](#) give a more general review of spatial competition models. [Daskin \(2008\)](#) provide a review of discrete spatial competition models. Sequential location models are reviewed by [Eiselt & Laporte \(1997\)](#) and [Kress & Pesch \(2012\)](#). And the review by [Plastria \(2001\)](#) focuses on the optimisation approaches to competitive location models.

2.1. Multiple firms and two-dimensional space

[Eaton & Lipsey \(1975\)](#) considers both the line market and two-dimensional space. They assume that firms charge a fixed price, and they investigate the equilibrium with N number of firms in the market. The number of firms is exogenously determined.

[Eaton & Lipsey \(1975\)](#) prove analytically that no equilibrium solution exists when three firms have similar prices. In the line market the two firms at each perimeter will move towards the centre of the line, reducing the market of the third firm. Eventually this forces the firm at the centre to leapfrog to the outside. This process continues with the two outer firms again moving towards the centre of the line. This has been referred to as the *dancing equilibria* although it does not fulfil the requirements of an equilibrium ([Eiselt, 2011](#)). In the line market with four firms, two firms will locate at one-fourth the length, and two firms will locate a three-fourths the length of the line. Similarly in the case of five firms, two firms are located at one-fifth the length, two at four-fifths the length, and the fifth firm locates at the centre. The equilibrium ceases to be unique for six or more firms, instead it is a range of equilibria. If there is an odd number of firms, then one firm will be located at the centre of the line. The firms on the perimeters are paired. While the remaining interior firms may be paired, or may be separated by some distance. However no firm has a total market area that is less than the largest *half-market* of any other firm. The *half-market* refers to the two market areas on either side of the firm. [Eaton & Lipsey \(1975\)](#) conjecture that the *principle of minimum differentiation* should perhaps be replaced by the *principle of local clustering*, since pairs of firms choose to locate close to one another. [Eaton & Lipsey \(1975\)](#) also consider multimodal and asymmetric distributions of consumers along the line. Here they prove that an equilibrium requires that the number of firms in the market does not exceed twice the

number of peaks in the consumer density function, i.e. if the distribution of consumers is bimodal (two peaks), then no equilibrium exists with four or more firms in the market. Osborne and Pitchik (1986) investigate mixed strategy equilibrium in the line market with a non-uniform distribution of consumers (Eiselt, 2011). They prove that no equilibrium exists with more than four firms, if the distribution is strictly concave or strictly convex. If the distribution is symmetric around the centre of the line then an equilibrium exists with three or more firms, where the mixed strategy is also symmetric. In the equilibrium with three firms, one firm locates at the centre and the two other firms use mixing strategies.

The results above from Eaton & Lipsey (1975) rely on the assumption that firms behave without foresight – referred to as *zero conjectural variation*. Eaton & Lipsey (1975) drop this assumption to analyse how firms with foresight choose to locate. To do this they assume that firms adopt a *minimax strategy* where firms anticipate the entry of a new firm into the market. This leads firms to minimise the size of their largest *half-market*. When a firm locates at the centre of its own market, rather than at the perimeters, it minimises its exposure to new firms entering the market. With this assumption firms never locate in pairs as described above. Instead, for any number of firms, there exists a unique equilibrium where firms are evenly spaced along the line. These equilibria are socially optimal as they minimise the average distance to consumers. The *principle of local clustering* no longer holds when firms use a *minimax strategy*. Further note that the *minimax strategy* does not consider whether or not entry is a rational strategy for new firms, and thus it leads to an inadequate equilibrium concept.

Eaton & Lipsey (1975, p. 40) note that it is not obvious how three or more firms will choose to locate in two-dimensional space, and write that “*using conventional analytical techniques the problem is very complex, perhaps intractable*”. Instead they use simulation techniques and once again assume that firms behave without foresight. In their simulation consumers are uniformly distributed on a disc, rather than a square plane. With any number of firms between 3 and 17, they show that a local equilibrium exists for the firm, where all firms are evenly spaced along a circle with the same centre as the disc, and with a radius that is less than the radius of the disc. However this is not a global equilibrium, and the circular configuration immediately breaks up, since one firm will choose to pair up with one of its neighbours rather than remain fixed on the circle. The same is true with less than 9 firms in the market, where one firm is located at the centre, and the remaining firms are evenly spaced in a circular configuration. With 9 or more firms in the market, the firm at the centre is in a global equilibrium, but the firms on the circle are only in a local equilibrium. In addition they show that no equilibrium exists for firms located in a hexagonal pattern, square pattern or rectangular pattern. The firms at the perimeter of the market will pair up

with neighbours, breaking these patterns. In sum they suspect that no equilibrium solution exists for three or more firms in two-dimensional space. Shaked (1975) proved their suspicion in the case with three firms. But it remains an open question when there is more than 3 firms in the market.

The general result – from the 85 years of research since Hotelling published his paper – is that conclusions and equilibrium are non-robust across alternative model specifications. Researchers have generalised the model, but most results are falsified when assumptions are modified. And even minor modification may remove any equilibrium constellations.

2.2. Location problems

The Hotelling model belongs to a wider set of *competitive location models*, which again falls into the more general realm of *location problems*. The difficulty of finding optimal solutions is not restricted to the Hotelling model. It shows up in many other location problems – even in problems without any kind of competition among the firms or agents. This subsection takes a more general look at location problems and optimal solutions.

The location problem aimed at finding the location of the firm that minimises the sum of distances to all its consumers is called the *minisum problem*. The minisum problem on the continuous plane is known as the *Weber problem* (Eiselt & Marianov, 2011, chapter 1). And the *multistore Weber problem* is when the firm has multiple facilities and the aim is to find the locations on the plane that minimise the sum of distances from each facility to the consumers (Eiselt & Marianov, 2011, chapter 15). The *multistore Weber problem* is difficult since it consists of two subproblems – an allocation problem and a location problem – that have to be solved simultaneously. It has been proven that the computational difficulty in solving the *multistore Weber problem* is *NP-hard*², and thus heuristic approaches are required for large problems (Eiselt & Marianov, 2011, p. 336). Heuristic methods find good solutions, but do not guarantee that the solution is optimal.

A relative to the *minisum problem* is the *minimax problem*. In this problem the objective is to minimise the longest distance between the firm and its consumers (similar to a *minimax strategy*). On the continuous plane this problem is known as the *continuous 1-centre*. And more generally, when the firm has multiple facilities, the *continuous p-centre*. This problem minimises the longest distance between p facilities and the consumers (Eiselt & Marianov, 2011, chapter 4). These problems resemble *set covering problems*. More specifically, when

²In the field of computational complexity theory a distinction is made between problems that can be solved in *polynomial time*, and those problems that cannot. A problem is said to be “easy” and “feasible” when it can be solved in polynomial time. A solution to an *NP-hard* problem may not be verifiable in polynomial time. *NP-hard* problems are at least as difficult as *NP problems*, and a solution to an *NP problem* can be verified in polynomial time. (Ahuja, Magnanti & Orlin, 1988, Appendix B).

there is positive demand everywhere in the space, then the *continuous p-centre* is equivalent to the problem of covering an area with p circles with the minimum possible radius, also known as *circle covering*. Masuyama, Ibaraki & Hasegawa (1981) show that the problem is NP-Complete³.

Thus finding the optimal solution is difficult even when disregarding the competitive aspects, where firms through their location decision compete with one another for customers.

2.3. Voronoi diagram

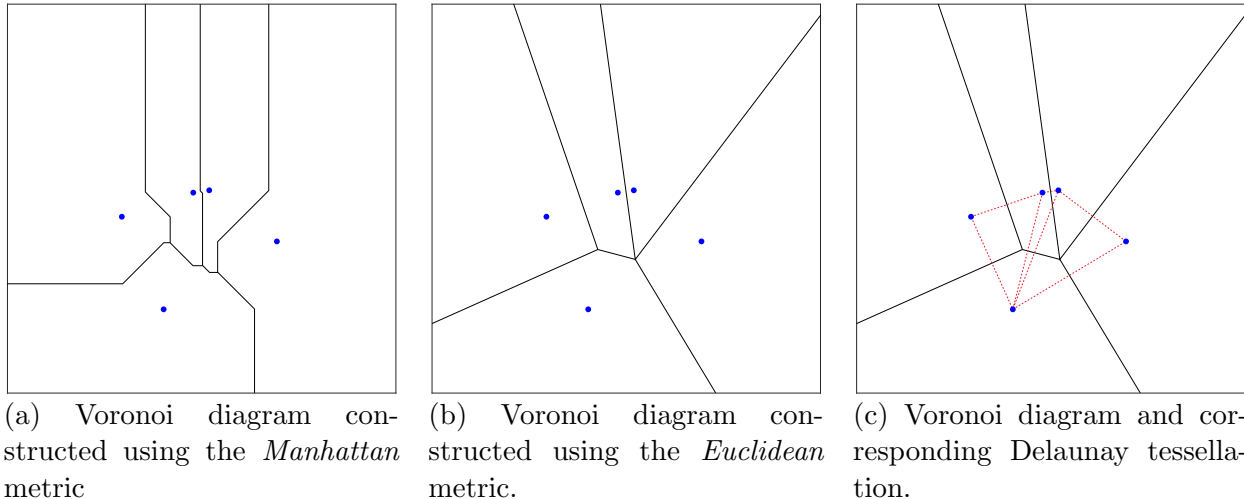
The *Voronoi diagram* is a geometric construction used to characterise allocation and studied for more than 175 years in different fields (Eiselt & Marianov, 2011, chapter 19). This subsection reviews the properties of Voronoi diagrams and related concepts that spring from the research on Voronoi diagrams. A Voronoi diagram consists of several Voronoi sets. Each Voronoi set has an associated *seed*, and each set contains all the points that are closer to the associated *seed* than any of the other *seeds*. We can construct a Voronoi diagram from the location of firms, if consumers shop at the nearest firm (which is the case in the fixed-price two-dimensional Hotelling model). The *seed* is the location of the firm and the Voronoi set corresponds to the market of the firm. The Voronoi set contains all of the customers that are located closer to the firm than any of the other firms. See example of Voronoi diagram with five seeds in figure 1a and 1b. There are several different ways of measuring distance. The most commonly used are the *Euclidean* distance metric and the *Manhattan* metric⁴. The choice of distance metric influences the structure of the Voronoi diagram. And methods for constructing Voronoi diagrams which rely on one distance metric, do not easily generalise to other distance metrics – in particular methods that construct Voronoi diagrams incrementally. This limitation will come into play later in the paper, when we discuss the optimal location of a new firm locating in a market with existing competing firms.

The average point or average location of a Voronoi set is referred to as the *centroid*. This point minimises the sum of distances to all points within the Voronoi set. When a firm locates at the *centroid* of its market, it constitutes a local socially optimal location. The *centroidal Voronoi tessellation* (CVT) is a special case of the Voronoi diagram where each *seed* is located at the centroid of its respective set. This is a global socially optimal solution, since it minimises the average distance between all consumers and firms (Laver & Sergenti,

³An NP-complete problem is both an *NP problem* and *NP-hard problem*. That is the solution can be verified in polynomial time, but no efficient method exists that can find the solution to these problems.

⁴The *Manhattan* metric measures the distance between point A and B as: $d_{AB}^{(1)} = |a_1 - b_1| + |a_2 - b_2|$, while the *Euclidean* metric is $d_{AB}^{(2)} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$. More generally the distance is measured as $d_{AB}^{(p)} = \left(|a_1 - b_1|^p + |a_2 - b_2|^p\right)^{\frac{1}{p}}$ with the value of the parameter p chosen for the purpose at hand.

Figure 1: Example of Voronoi diagram with five seeds (firms). The solid lines are the boundaries of the Voronoi set (boundaries of market).



2011). Any *centroidal Voronoi tessellations* is optimal, but not necessarily unique. Given for instance 2 firms in a square space there are several CVTs; the firms could locate left-right splitting the market in half vertically, the firms could locate up-down such that they split the market in half horizontally, and so on. In each case the average distance to firms is minimised and the location of firms constitutes a social optimum. The *Lloyd's algorithm* generates a *centroidal Voronoi tessellation* from any generic Voronoi diagram. At every iteration, the algorithm moves the *seeds* to the centroid of their current Voronoi set. Using this iterative process any Voronoi diagram converges to a *centroidal Voronoi tessellation*. One of the decision rules used in this paper, is in fact an implementation of the *Lloyd's algorithm*. And so when we compare the results from this baseline decision rule to the other decision rules, we are comparing with the social optimum location of firms.

The dual of the Voronoi diagram is the *Delaunay tessellation* (Eiselt & Marianov, 2011, chapter 19). The *Delaunay tessellation* links *seeds* with adjacent Voronoi sets and the length of each link is the distance between the pair of *seeds*, see figure 1c. Reformulated in terms of firms, the Delaunay tessellation is a network that links firms with a common market boundary (i.e. direct competitors). In figure 1c the firm located furthest left and furthest right are not linked since they do not share a market boundary. The two firms are not in direct competition with one another over any of the customers. However both firms are linked to the firm at the bottom.

The *Voronoi game* originates from the study of Voronoi diagrams. This competitive location model shares features with the Hotelling model. In the Voronoi game two firms place n number of facilities each and consumers shop at the nearest facility. The game is

sequential so firms take turns and place one facility each turn. The game ends after $2n$ turns when both firms have placed all their facilities. The payoff of each firm is its share of the market. The winner is the firm with a majority of consumers. In the continuous line market the follower has a winning strategy that captures $\frac{1}{2+\epsilon}$ of the market, however the leader has a strategy that can make ϵ arbitrarily small (Ahn, Cheng, Cheong, Golin & Van Oostrum, 2001). The two-dimensional Voronoi game is significantly more difficult to solve. In a one-round version of the Voronoi game – where each firm places all its n facilities in its first turn – the follower still has a winning strategy, but the leader cannot make ϵ arbitrarily small, if consumers are distributed over a square area (Eiselt & Marianov, 2011, chapter 9). On the other hand, if consumers are distributed over a rectangular area that is sufficiently oblong, then the leader has a winning strategy (Fekete & Meijer, 2005). Fekete & Meijer (2005) furthermore show that the follower’s problem is NP-hard in the two-dimensional one-round game.

2.4. *Hinterlands and vertical product differentiation*

The choice of space influences results. As seen above when moving from one-dimensional to two-dimensional space it changes the equilibrium, and sometimes removes the equilibrium altogether. Similarly the shape of the space (e.g. disc, square, rectangular) and whether the space is bounded or unbounded affects the results. Certain combinations of shape and boundary creates *hinterlands*, i.e. remote regions or sparsely populated regions. Competing firms will find it less attractive to locate in these regions. And as a result the papers that consider spaces without naturally occurring hinterlands – such as the one-dimensional circular market – find that firms locate more evenly across the entire space (Eaton & Lipsey, 1975; Salop, 1979).

Competitive location models are often described in terms of geographic location. However these models are also applicable when studying product differentiation, e.g. manufactures competing with one another when choosing product characteristics such as the size, colour, and shape of their respective products. When the model is formulated in terms of geographic location, then the ideal point of the consumers is his or her location. Every consumer would ideally want a firm to locate at his or her doorsteps, so the consumer avoids transportation costs altogether. In terms of product differentiation the ideal point is the consumer’s preferred product characteristics. The transportation cost is replaced by the degree of disutility from consuming a product that differs from ones preferred product. The ideal points are distributed over the space, and so the consumers have heterogeneous preferences regarding products or firms. The models discussed so far strictly relate to *horizontal product differentiation*. Thus even when all firms charge the same price, firms are always able to achieve positive demand, through differentiation. There are niches that guarantee the firm a positive

demand. This is not the case with *vertical product differentiation*, such as product quality. With *vertical product differentiation* the preferences of consumers are homogenous, i.e. all consumers prefer high quality products to low quality products. If all firms charge the same price, then the firm with the highest product quality will capture the entire market. No niches exists. The two-dimensional Hotelling model assume *horizontal production differentiation* along both dimensions. An alternative model is the Launhardt model (Ferreira & Thisse, 1996). The model includes both *horizontal* and *vertical product differentiation*. It can be seen as an extension of the Hotelling model in that firms choose location, but the firms have different delivery prices⁵. The delivery prices is an easy way to incorporate *vertical product differentiation* into the model. The delivery price is part of the *transportation cost* paid by consumers. Consumers still shop at firm with the lowest total price (*mill price* and *transportation costs*), thus all consumers prefer low delivery prices over high delivery prices. The Launhardt model is in a sense two dimensional. The dimension along which firms locate assumes *horizontal product differentiation*, while the delivery price dimension assumes *vertical product differentiation*.

2.5. Foresight

Prescott & Visscher (1977) criticise the Hotelling model for its assumption of costless relocation. The Hotelling model implicitly assumes that firms can always relocated at a future point in time at low cost, and therefore firms – when they choose their location – do not need to take account of future entry into the market. Instead they propose a sequential location model, where it is prohibitively expensive to relocate, and where firms therefore use foresight when choosing their location once-and-for-all. Given rational expectations, firms can deduce the optimal solution by reasoning backwards from the final stage of the problem. Thereby determining the sequence of optimal actions for every firm in each stage, i.e. each firm chooses the profit maximising location subject to the location of firms already in the market, and subject to the rules that future rational entrants will use. Let N denote the exogenously determined number of firms in the market. The model is solved using backwards induction, i.e. the last firm entering the market will choose the location that maximises its profit given the position of all other firms already located in the market. The second last firm (firm number $N - 1$) uses the position of all $N - 2$ firms already located and the best response of the last entering firm when choosing its location. Firm number $N - 2$ use the position of the first $N - 3$ firms and the best response of firm number $N - 1$ and N , and so forth. This continues all the way to the first firm. This process of reasoning from a set of

⁵Although the Launhardt model can be seen as an extension of the Hotelling, the paper by Launhardt was published 44 years before Hotelling published his paper.

premises, until one reaches the logical conclusion is known as *deductive reasoning*.

Arthur (2014, chapter 11) is highly sceptical of this approach. Noting that there may be simple cases where every firm can figure out what to do, and where this approach produces the correct equilibrium predictions. But in more complicated scenarios or large problems with many firms the backwards induction approach is likely to break down. The approach requires sequential location and assumes *common knowledge* and *common knowledge of rationality*. That is each firm knows the preferences of all other firms, and each firm knows that all other firms knows the preferences of all firms, in an infinite regress. In addition each firm knows that all other firms behave rationally, and each firm knows that all other firms knows that all firms will behave rational, also in an infinite regress. To solve the model it is common to assume that firms are homogenous, both in terms of expectations and available information. Furthermore the optimal location must be unique. If any firm is uncertain about the preference or rationality of another firm, then the *pure-strategy* equilibrium solution breaks down. In essence when a firm has to choose its own location, the location of the other firms is unknown. The firm can use the predicted location of the other firms, but the location outcome that each firm is trying to predict, depends on predictions that the firm and other firms form. “Predictions are forming a world those predictions are trying to forecast” (Arthur, 2014, p. 175). This self-referential loop leads to logical indeterminacy, and thus without some coordination device or without assuming homogenous firms, the maximisation problem is ill defined and cannot be solved deductively. No logical conclusion exists rendering *deductive reasoning* void. Nonetheless firms still make decisions even when faced with such ill defined problem. One possibility is that firms use heuristics or rules of thumb. Arthur (2014, chapter 11) discusses these issues in terms of a competitive location model, but does not propose a new location model that incorporates or overcomes these limitations. Instead Arthur (2014, chapter 11) looks at a model on asset pricing with many similarities, but which is easier to solve since each investors only need to predict a single outcome, namely the stock price. In the location model each firm needs to predict $N - 1$ outcomes, i.e. the location of all other firms.

In the *Santa Fe Artificial Stock Market* discussed by Arthur (2014, chapter 3, chapter 11) investors are heterogeneous both in terms of information and the expectation models they use to predict the stock price⁶. Investors hold several hypotheses on the expected stock price, all of which are based on the recent history of the stock price. They act on the hypothesis that

⁶More precisely, investors try and predict a linear combination of stock price and dividend. Dividend follows a stochastic autoregressive process, $AR(1)$, which is unknown to investors. And the stock price is given by the market clearing price after investors have chosen between the risky stock and a risk-free bond with a fixed rate r . Investors have a constant absolute risk averse utility function.

suit the current state of the market and that has provided the most reliable predictions in the past. In each period, when the actual stock price is revealed, investors update the accuracy of their hypothesis. They continuously discard poorly performing hypotheses and form new hypotheses. [Arthur \(2014\)](#) refers to this as *inductive reasoning*, i.e. deriving principles or conclusions from observation and viewing conclusions as *probable*, rather than *certain*. Investors act on the hypothesis viewed most probable to hold true, and update hypothesis in response to new information or observations that falsifies currently held hypothesis. In their decision process investors use *inductive reasoning* as opposed to *deductive reasoning*.

More specifically, the Santa Fe Artificial Stock Market model is an agent-based model, where the investors are agents. Investors hold M different hypotheses, each with a corresponding linear forecasting model. Each hypothesis is a *condition/forecast* rule, such that if the current state of the market matches the conditions in the hypotheses, then the hypotheses is said to be *active*. Of all the active hypothesis, the investor then uses the linear forecasting model from the hypotheses that has the best accuracy. Thus investors are able to recognise patterns in state of the market, and act using their most reliable hypotheses. The state of the market is summarised by a 12-bit array. Each bit corresponds to a scenario and takes the value 0 or 1 indicating whether the scenario is true or false. They use scenarios such as the “the price has risen in the last 3 periods” or “the current price is not larger than 16 times the dividend divided by fixed rate r ”. The combination of these 12-bits yield more than 4,000 different states of the market.

Learning becomes a necessity once fully rational agents are replaced by agents with *bounded rationality*. *Bounded rational* agents do not act irrational, but act as best according to their knowledge and skills, and learning gives agents the potential to gradually expand their knowledge and discover new skills. Two types of learning exists in economics ([Ehrentreich, 2008](#), chapter 4). With *rational-based learning* agents act optimally given their available information and when new evidence arrives they update their prior opinion in a rational fashion. Bayesian learning is one example of *rational-based learning* where agents hold several hypotheses. The probability of a given event occurring is conditional on the associated hypothesis. When agents observe a given event, they increase the probability of the hypothesis for which the event is most likely. In this way agents rationally update their prior opinion. The other type of learning is *biologically inspired learning*. This type of learning mimics a natural evolutionary selection processes, and is highly effective in terms of creating well-adapted “species” or strategies in an ever-changing environment. Rather than updating existing hypothesis, agents develop new hypotheses by mutating and recombining existing hypothesis. These processes are known as *mutation* and *crossover* respectively. Mutation creates new hypotheses based on the existing hypotheses through experimentation and un-

intended mistakes. While crossover creates new hypotheses by combining randomly selected bits and pieces of existing hypothesis. To increase the performance both *mutation* and *crossover* are often coupled with *selection*, such that new hypothesis are based on the best performing existing hypothesis. Genetic algorithms are used to model *biologically inspired learning*. Arthur (2014, chapter 3) uses a genetic algorithm to create new hypothesis through *mutation* and *crossover*.

Arthur (2014, chapter 3) finds that the model is able to capture both the *rational expectations regime* and what they call the *complex regime*. In the former the model reproduces the equilibrium found analytically by assuming rational expectations and homogenous agents. That is the equilibrium where the stock price is the fundamental value, i.e. the sum of all future dividends in present value terms. There is no herd effects, technical trading remains unprofitable, and trading volume is low. The model produces the *rational expectation regime* when investors discover new hypotheses at a slow rate. While the model produces the *complex regime* when using a faster exploration rate of new hypotheses. In the *complex regime* it is profitable for investors to engage in technical trading, since the herd effect produces bubbles and crashes in the stock price, and trading volume is high.

This paper incorporates *inductive reasoning* into the competitive location model, using a modified version of the above described model. Section 5.3 provides further details and the modification need to incorporate *inductive reasoning* into a location model.

2.6. Agent-based modelling

Agent-based models are build from the bottom up, in the sense that the model consist of many autonomous and interacting agents. Each agent has strictly local motives and through the interaction of all the agents emerges the macro-behaviour of the system (Ehrentreich, 2008, chapter 2). Agents are often heterogeneous and thus the macro-behaviour is hard to predict from the local motives of agents alone. Agent-based models are typically solved on a computer by setting up the decision rules of agents and observing the system as the interaction of agents play out. Although researchers use computers to run agent-based models this is not the same as solving traditional models through numerical simulation. Instead one should view the introduction of agent-based models into economics as a response to “*the inappropriateness of traditional methods*”, in particular the assumptions of representative agents and rational expectations (Ehrentreich, 2008, chapter 2, p. 7). Models assuming representative agents are incapable of generating emergent phenomena, such as the *complex regime* described above where one observes herd effects and technical trading. While fully rational agents with perfect information are deprived of free choice – the only available choice is the rational course of action. While this limits the number of possible outcomes of the model making it easier to solve, it does so at the expense of what might have been more appropriate

and realistic assumptions. Both assumptions are widely used in order to solve traditional models, while neither are necessary in agent-based models. Agent-based models are thus particularly well suited when studying problems with heterogeneous agents or bounded rationality. Furthermore agent-based models are a generalisation of equilibrium economics, as seen above with the Santa Fe Artificial Stock Market model, where standard equilibrium behaviour is a special case among a much broader set of complex behaviours. Agent-based models can provide researchers with answers to problems that would be intractable using conventional analytical techniques. However as noted by [Laver & Sergenti \(2011, p. viii\)](#) “*rigorously designed and analysed computational experiments, however exhaustive, will never be substitutes for flawless and elegant classical formal proofs*”.

[Laver & Sergenti \(2011\)](#) construct an agent-based model. Their model investigates the competitive location behaviour of political parties competing for voters on a two dimensional policy space. Political parties choose a policy position and the voters vote for the party whose policy is closest to the voter’s ideal policy. Their model shares many similarities with the two-dimensional Hotelling model with multiple firms. However to study multiparty competition they assume that the political parties use heuristics or rules of thumb to determine their location at each iteration. They run the model for several hundred iterations. They have five types of political parties with the following decision rules. *Sticker*: a party that sticks to its ideological ideal policy point, no matter what. *Aggregator*: A party that constantly moves towards the centre of its current voter base. *Predator*: A party that moves towards its most successful competitor, trying to replicate its success. *Hunter*: A party that continues to move in the same direction, if the previous move proved fruitful, and otherwise heads in the opposite direction. *Explorer*: A party that surveys several directions and then move along the most lucrative direction. The last two decision rules attempts to maximise the share of votes directly. While the other decisions rules may maximise the vote share of the political party, but this would happen indirectly. Furthermore note that none of these decision rules considers the simultaneous move of competing parties. They all implicitly assume that the other parties stick to their current policy position. Hence there is no foresight and thus the decision rules are stripped of any and all strategic considerations.

[Laver & Sergenti \(2011\)](#) considers a continuous and unbounded space. They obtain this space by assuming that voters’ ideal policy points are drawn randomly from two *bivariate normal distributions*. By varying the relative size and means of the *bivariate distributions* they are able to analyse the location behaviour of political parties when consumers are distributed symmetric around a single peak, and when the distribution of consumers is asymmetric and has two peaks. They can account for the effect of both a symmetric unimodal distribution and an asymmetric bimodal distribution of voters. To simplify their analysis they rotate the

space such that the mean ideal points of the two distributions only differ along one dimension. They call this the *main axis of policy disagreement* between their two subpopulations of voters.

In the baseline models they investigate, the number of political parties is exogenously given and all parties use the same decision rule. They use these baseline models to analyse the location behaviour in each of the decision rules, and to grasp the differences between decision rules. The baseline model where all parties use the *aggregator*-rule is an implementation of the *Lloyd's algorithm*, and thus the policy position of parties converge to the socially optimal location, where the average distance to the ideal point of all voters is minimised. In the model where all parties use the *hunter*-rule – and thus continuously try to increase their vote share – they find that parties choose policy positions that are closer to the mean ideal point of all voters than the socially optimal positions. The *hunter*-parties do not agglomerate at the peaks of the voter's ideal points, but choose a position that lies at a distance to the peaks. The party achieves short-run gains by locating close or at the peak of voters' ideal points, however the response by competing parties quickly erodes these gains, leading to the long-run policy positions that lie at a distance to the peaks. Their extended model includes entry and exit of political parties and political parties with different decision rules compete with one another for voters. The decision rule of each new political party is randomly drawn with equal probability. The entry of new political parties resembles a citizen-candidate model. Citizens or voters that feel increasingly underrepresented by the current political parties form new parties. Political parties exit when they repeatedly fall below a given threshold share of voters. They compare the competitive advantages of the decision rules by measuring the average vote share and the longevity of each decision rule. *Hunter*-parties capture the largest share of voters and live longest when pitted against *sticker*-parties and *aggregator*-parties, and when the speed at which the *hunter*-parties moves sufficiently balances the exploration of new position and exploration of its current policy position. However the *hunter*-rule is not robust as both the *explore*-rule and *predator*-rule outperforms it, when all five decision rules are pitted against one another.

Fowler & Laver (2008) use a similar model, but run a tournament in which decision rules submitted by other researchers are pitted against one another. Researchers knew that Fowler & Laver (2008) pre-entered the *sticker*, *aggregator*, *hunter* and *predator* rules. Furthermore they assume that elections are held every twentieth iteration. The nineteen intervening periods are campaign periods. Parties can freely explore different policy positions in campaign periods. And only the policy positions and votes captured in an election period influence the exit and entry decision of parties. Decision rules were prohibited from using information on the distribution of voters ideal points. The decision rules submitted fall into the following

categories. *Tweaks of pre-entered rules*: Decision rules based on the four pre-entered decision rules, with slight modification. *Centre-seekers*: Decision rules that attempt to discover the peak or centre of voters' ideal points, and position the party at the centre. *Inter-electoral explorers*: Decision rules that take advantage of the campaign periods to try out different policy positions, and settling at the most prosperous of these positions. *Parasites*: Decision rules that try and replicate the success of other political parties, for instance by choosing a policy position close to largest political party. *Satisfiers/survivors*: Decision rules that are content with a given share of the votes, or rules that only move once their vote share approaches the threshold level.

Fowler & Laver (2008) find that *centre-seeker* rules generally underperform and argue (similarly to the *hunter* rule above) that the short-run gains by locating at the centre of voters' ideal points critically endangers the long-run survival of the party, because of the response from other political parties. *Parasite* rules are highly disruptive to the political party that "hosts" the *parasite*, but rarely outperform the other rules, since they typically end up capturing half of the *host's* market. With the *inter-electoral explorers* and *satisfier/survivor* rules the party faces the above mentioned trade-off between exploration of new positions and the exploration of its current policy position. The tournament winner was a *satisfying parasite* rule named *KQ-Strat*, submitted by Kevin Quinn. Political parties using this rule jittered slightly and randomly when above the threshold level, and when below the threshold level moved towards a randomly selected competing party that was above the threshold level. A large number of the submitted rules performed worse than the pre-entered *sticker* rule, and only 7 out of the 25 submitted rules performed better than the pre-entered *hunter* rule. None of the decision rules use foresight *per se*. They all implicitly assume that competing parties stay fixed, exempt perhaps the *fool-proof* rule by Jim Adams which moves the policy of the party one-fifth the average policy shift of all competing parties. While the rationale for the rule is not included in Fowler & Laver (2008), one could argue that parties with this rule anticipate a backlash to the policy shift of all competing firms, and thus moving one-fifth the average shift is a reasonable action given this expectation⁷. Nonetheless the *fool-proof* rule performs worse than the *sticker* rule.

Recap

This literature review has gone through the different location behaviours that firms use in competitive location models without prices. In the analytical models, the simplest behaviour

⁷The one-fifth move seems to have originated in a paper by Adams & Somer-Topcu (2009). The paper is an empirical investigation of the policy response of parties to policy shifts among competing political parties. They find a statistical significant shift between 0.149 and 0.196 in response to the average policy shift of competing parties.

is firms locating without foresight (*zero conjectural variation*). With the *minimax strategy* firms foresee the entry of new firms (or the relocation of existing firms), in a manner that cause maximum damage to the market of the firm. The firm therefore locates to minimise this damage. However the *minimax strategy* does not consider whether or not entry is a rational strategy for new firms, and thus it leads to an inadequate equilibrium concept. The rational expectations approach does consider this, but requires a sequential setup and quite restrictive assumptions. The agent-based model on the stock market provides a path where these assumptions are relaxed while maintaining the simultaneous setup. Namely agents use *inductive reasoning* to make predictions and update their hypotheses in response to new information. The agent-based model from the literature on political science show how to go about analysing the two-dimensional Hotelling firm with multiple firms, and thus picks up where the analytical approach falls short of finding equilibria. In the agent-based models numerous decision rules have been constructed and battle tested in tournaments. However no one has so far considered decision rules with foresight. This paper will build upon this model by Laver & Sergenti (2011) but introduce agents that use *inductive reasoning* as a way to reintroduce foresight into the competitive location model. The literature review has also repeatedly highlighted the difficulties in optimal solutions to location models, even those without any competitive aspects.

3. Model

To analyse the location behaviour of firms we construct an agent-based model akin to Laver & Sergenti (2011)⁸. We reformulate the model in terms of firms competing for customers, when they choose their geographic location. It is simply shorter to write ‘firms’ and ‘relocate’ rather than ‘political parties’ and ‘shift policy position’ and it reads more naturally. The results of the model are still valid in a political context, as well as in the context where manufacturers compete for customers through product differentiation when choosing product characteristic along two dimensions.

The decision rule of firms is exogenously determined. In each iteration the decision rule of the firm determines its location. All firms choose their location simultaneously. The model is executed with several iteration and analyses how the interaction of the firms affect the overall location of firms. The model assumes that all firms charge the same price, and thus

⁸All models in this paper have been programmed from the bottom up. The code is available at <https://github.com/jsekamane/Positioning/tree/master/Models/ABM>. All the models are build in MATLAB (R2015a) and run on a MacBook Pro with a 2.6 GHz quad-core Intel Core i7 processor. In addition several of the models use MATLAB’s Parallel Computing Toolbox with 4 local workers taking advantage of the quad-core processor structure to execute several runs in parallel.

this paper abstracts away price competition, and only considers location differentiation. Price competition greatly increases the complexity of the model, and is therefore beyond the scope of this paper. Furthermore firms can and do change prices more frequently than they change location. Thus price competition and location competition take place on two different time scales – a fast and a slow. It remains an open question how to appropriately incorporate the different time scales into an unified iterative model.

3.1. Consumers

The scope of this paper is limited to one side of the market – the firms and their decision-making behaviour. For this reason simplifying assumptions are used for the other side of the market – the consumers and the purchasing behaviour of consumers. For one thing consumers are static. They do not shift position, they maintain preferences, they are obliques to trends, and are in no way influenced by changes in the market. Secondly consumers behave non-strategically and demand inelastically. They always purchase a single unit from the closest firm, regardless of the distance to that firm. And finally, apart from the location, all firms are identical in the eyes of the consumer. There is no brand recognition and no loyalty – the customer purchase from the closest firm regardless of previous purchasing history.

Every consumer has an ideal point in the market. This is the point where the consumer wants a firm to locate. All other points are second to the ideal point. And the further away from the ideal point that the firm locates, the less attractive it becomes. Consumers have *single peaked preferences*. Essentially this paper reduces consumers to being merely their ideal points.

Viewed in terms of geography the ideal point would be the consumer’s physical location and the distance is the bee-line distance to firms. Alternatively the model can be viewed in terms of product differentiation. Then the ideal point of the consumer is his or her desired product characteristic along the two dimensions. We assume *horizontal product differentiation* along both dimension, i.e. product characteristic are non-ordinal. In addition we need to assume that product characteristics are continuous and infinitely divisible.

The utility function for each consumer is a function of the distance from the consumer’s ideal point and to the position of the firm. The utility function is equivalent to a quadratic loss function. The *Euclidean* distance metric is used. The utility of customer i is the negative squared distance between customer i and the closest firm j :

$$U_i(j) = -d(i, j)^2 \tag{1}$$

3.1.1. Consumer distribution

The simplification of consumer's behaviour allows focused attention on the behaviour of firms. As discussed in the literature review the distribution of consumers can fundamentally change the behaviour of firms. We would like to capture and analyse these changes – in particular the effects from asymmetric and multimodal distributions. We follow the method used by [Laver & Sergenti \(2011\)](#) and assume two consumer subpopulations in our market. When the two subpopulations have the same mean ideal point then the aggregated distribution of consumers will be symmetric and unimodal. While subpopulations with vastly different mean ideal points and different sizes will lead to an aggregated distribution of consumers that is asymmetric and bimodal. Thus this method of using two subpopulations is able to capture several different types of aggregate distributions using only two parameters; the polarisation of subpopulation ideal points, and the relative size of the two subpopulations. The distribution of consumer ideal points is exogenous – more specifically the parameter values of the distribution are exogenously determined.

The market is two-dimensional, so we use the two-dimensional equivalent of the *normal distribution* – the *bivariate normal distribution*. Both subpopulations follow a *bivariate normal distribution*. Using different mean values in the distribution imply that the subpopulations disagree over the ideal point. Without further restrictions the subpopulations would disagree on both dimensions. To simplify the analysis, and without loss of generality, we assume that the subpopulations have a common distribution of ideal points along one dimension and only disagree over the ideal point along the other dimension. This is equivalent to rotating the market space or coordinate system until disagreement only appears along one of the dimensions, i.e. rotating the market space such that the line going through the mean ideal points of the subpopulations is parallel to an axis. One can always rotate the coordinate system without loss of information, and so this simplifies our analysis without any loss of generality. This is similar to the orthogonal transformation underlying a *principal component analysis*. We start out with two dimensions (or two components), and end up with two new dimensions (two principle components); the disagreement dimension (x-axis) and the agreement dimension (y-axis). The two subpopulations – from now on referred to as ‘left’ and ‘right’ – follow a *bivariate normal distribution* with mean $(-\mu, 0)$ and $(\mu, 0)$ respectively, and both with standard deviation $(0.5, 0.5)$ ⁹. Where the parameter μ measures the mean ideal point polarisation. The parameter measuring the relative size of the left subpopulation is

⁹That is, $(\mu_{x,l}, \mu_{y,l}) = (-\mu, 0)$ and $(\mu_{x,r}, \mu_{y,r}) = (\mu, 0)$, while the standard deviation is $(\sigma_{x,i}, \sigma_{y,i}) = (0.5, 0.5)$ for $i = \{l, r\}$. We further assume no correlation among the two dimensions. The correlation coefficient ρ is zero and thus the covariance matrix is $\begin{bmatrix} 0.5^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}$.

n_l/n_r . The aggregated consumer distribution is a *mixed bivariate normal distribution* with weights based on the relative size of the subpopulations (Weisstein, 2002a; Balakrishna & Lai, 2009):

$$f(x, y) = \frac{n_l/n_r}{1 + n_l/n_r} f^{(l)}(x, y) + \frac{1}{1 + n_l/n_r} f^{(r)}(x, y), \quad f^{(i)}(x, y) = \frac{e^{-\frac{(x-\mu_{x,i})^2}{2(\sigma_{x,i})^2} - \frac{(y-\mu_{y,i})^2}{2(\sigma_{y,i})^2}}}{2\pi\sigma_{x,i}\sigma_{y,i}} \quad (2)$$

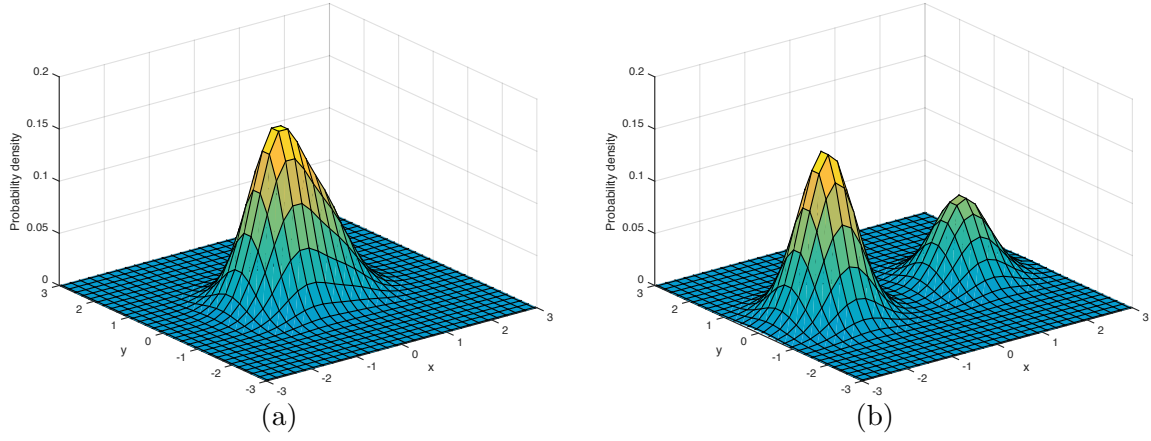
The mean ideal point on the y-axis is 0 for both subpopulations hence no disagreement along this dimension. The mean ideal point on the x-axis is $-\mu$ for the left subpopulation and μ for the right subpopulation. For $\mu > 0$ there is disagreement along the x-axis. I will analysis the polarisation parameter in the range $\mu \in [0, 1.5]$ – a range that spans both an unified and a split market. At the lower bound of the range there is no disagreement and the market consists of one unified customer base. At the upper bound there is essentially no overlap between the subpopulations and the customer base is split. The range of the relative subpopulation size is $n_l/n_r \in [1, 2]$. The subpopulations are equally large when $n_l/n_r = 1$, while $n_l/n_r = 2$ indicates that the left subpopulation is twice the size of the right subpopulation. See example of distributions in figure 2.

The default unit of measure throughout this paper will be standard deviations. This unit of measure will refer to the standard deviation measurement used in the *bivariate normal distributions*, e.g. when this paper reports that the distance between two points is 1 standard deviation, then this is equivalent to twice the standard deviation in the subpopulation distribution. The unit of measure is independent of the space and the coordinate system, making it easier to compare the results with other studies.

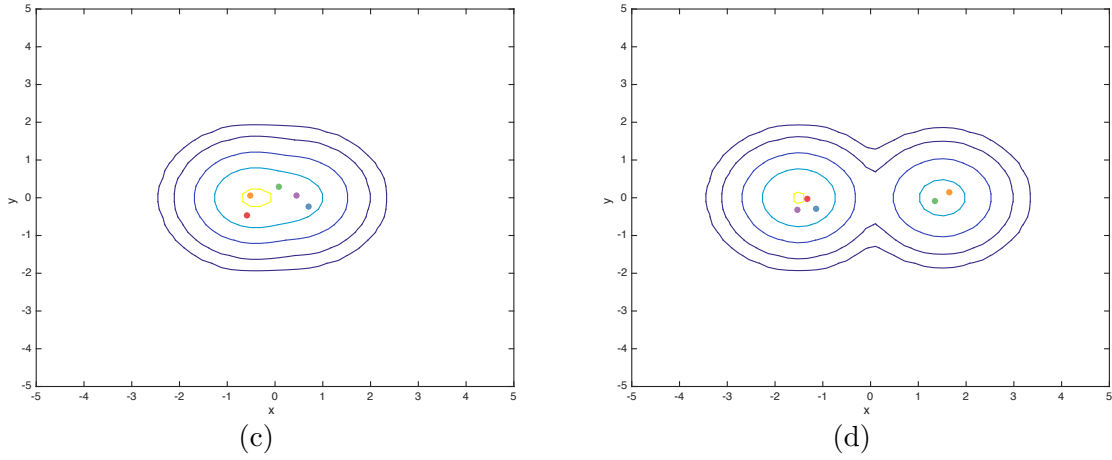
Given the aggregated consumer distribution one approach would be to randomly draw a finite number of consumers from the distribution. Market shares can then be calculated by counting the number of consumers. This paper, like Laver & Sergenti (2011), will not employ this approach. Instead the consumer distribution is carried through and market shares are calculated by aggregating probability mass. Similarly the centroid of the market of each firm is weighted by the probability mass rather than the number of consumers (centroid shown in figure 2e and 2f). Although the two approaches yield the same results on average, the latter is independent of particular draws, thus we save computational power not having to execute several repetitions to obtain average values.

We will now evaluate the consumer distribution function, with particular focus on the combinations of the parameter values that give rise to a single peak in the distribution, and the combinations which give two peaks. This information will help us interpret the results of the model later on. The consumer distribution function is a *mixed bivariate normal*

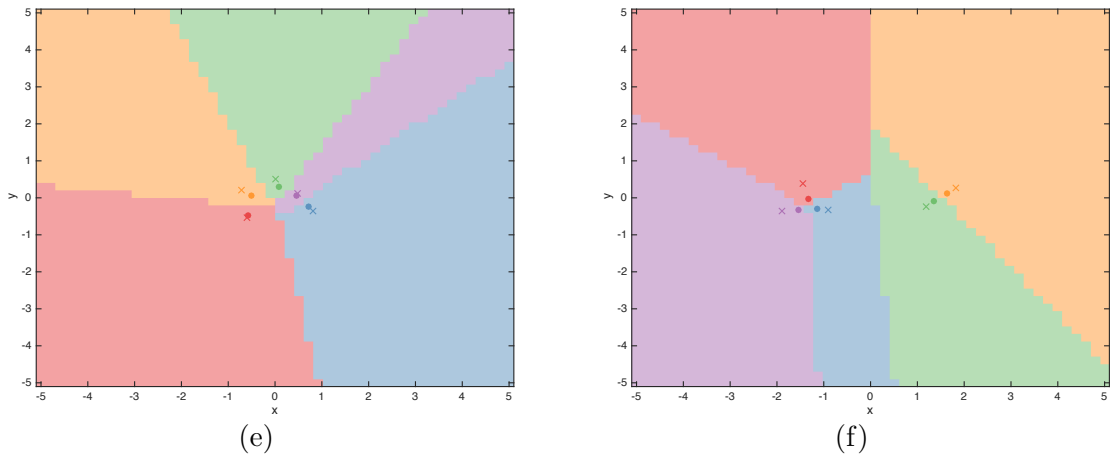
Figure 2: Example of market where the left subpopulation is twice as large as the right ($n_l/n_r = 2$) and the polarisation of the mean ideal points is respectively $\mu = 0.5$ (left) and $\mu = 1.5$ (right).



(a), (b) Distribution of consumer ideal points.



(c), (d) The contours of the distribution. The dots are an example of how five firms might locate.



(e), (f) The market area of each of the five firms. The dots are the location of the firm. The crosses are the market centroids (weighted with the probability mass within each respective market).

distribution. The distribution is symmetric along the y-axis, so the peak of unimodal, right peak of bimodal, left peak of bimodal and the saddle point of the distribution will all have coordinate zero, $y^p = y_r^p = y_l^p = y^s = 0$. We need only focus on determining the coordinates of the peaks and saddle point along the x-axis. Because of the symmetry along the y-axis and since there is no covariance between x and y in the *bivariate normal distributions*, we can determine the x-coordinates using the marginal probability function f_x . This function is equivalent to a mixed univariate normal distribution (Weisstein, 2002b,a; Balakrishna & Lai, 2009). We express the function in terms of the free parameters; polarisation of the ideal points of the subpopulation, μ , and the relative size of the subpopulation n_l/n_r :

$$f_x(x, \mu, n_l/n_r) = \frac{n_l/n_r}{1 + n_l/n_r} \frac{e^{-\frac{(x+\mu)^2}{2(\frac{1}{2})^2}}}{\sqrt{2\pi}\sqrt{\frac{1}{2}}} + \frac{1}{1 + n_l/n_r} \frac{e^{-\frac{(x-\mu)^2}{2(\frac{1}{2})^2}}}{\sqrt{2\pi}\sqrt{\frac{1}{2}}}$$

To find the extrema of the function we differentiate with respect to x :

$$\begin{aligned} \frac{\partial f_x(x, \mu, n_l/n_r)}{\partial x} &= \frac{n_l/n_r}{1 + n_l/n_r} \frac{e^{-2(x+\mu)}}{\sqrt{\pi}} (-4(x+\mu)) + \frac{1}{1 + n_l/n_r} \frac{e^{-2(x-\mu)}}{\sqrt{\pi}} (-4(x-\mu)) \\ &= \frac{-4e^{-2(x+\mu)}((x+\mu)n_l/n_r + (x-\mu)e^{8x\mu})}{(1 + n_l/n_r)\sqrt{\pi}} \end{aligned}$$

Setting the derivative equal to zero and rearranging gives:

$$\begin{aligned} \frac{\partial f_x(x, \mu, n_l/n_r)}{\partial x} = 0 &\Leftrightarrow (x+\mu)n_l/n_r + (x-\mu)e^{8x\mu} = 0 \\ &\Leftrightarrow n_l/n_r = -\frac{x-\mu}{x+\mu}e^{8x\mu} \end{aligned} \quad (3)$$

It is not possible to solve for x , so instead we plot the equation in a (μ, x) diagram to evaluate the expression, see figure 3. The curve splits into several branches revealing the *bifurcation*, i.e. for specific combinations of μ and n_l/n_r our x takes on multiple values (Strogatz, 1994). As $\mu \rightarrow 0$ implies $x \rightarrow 0^-$ when approaching from the negative side. However as $\mu \rightarrow 0$ and $x > 0$ we will observe a jump between $\mu \simeq 0.657$ and $\mu = \frac{1}{2}$. The *bifurcation* implies a non-linearity in all the models in this paper.

Unsurprisingly we see that with no polarisation, $\mu = 0$, the distribution is unimodal with a single peak at $x^p = 0$ regardless of the relative size of the subpopulation ($\forall n_l/n_r$). With a low degree of polarisation, $\mu \in]0, \frac{1}{2}]$, the distribution remains unimodal with a unique peak at the origin or in between the origin and the mean ideal point of the left-subpopulation, i.e $x^p \in [0, \bar{x}_l[$, where the mean ideal point is $\bar{x}_l = -\mu$. The relative size

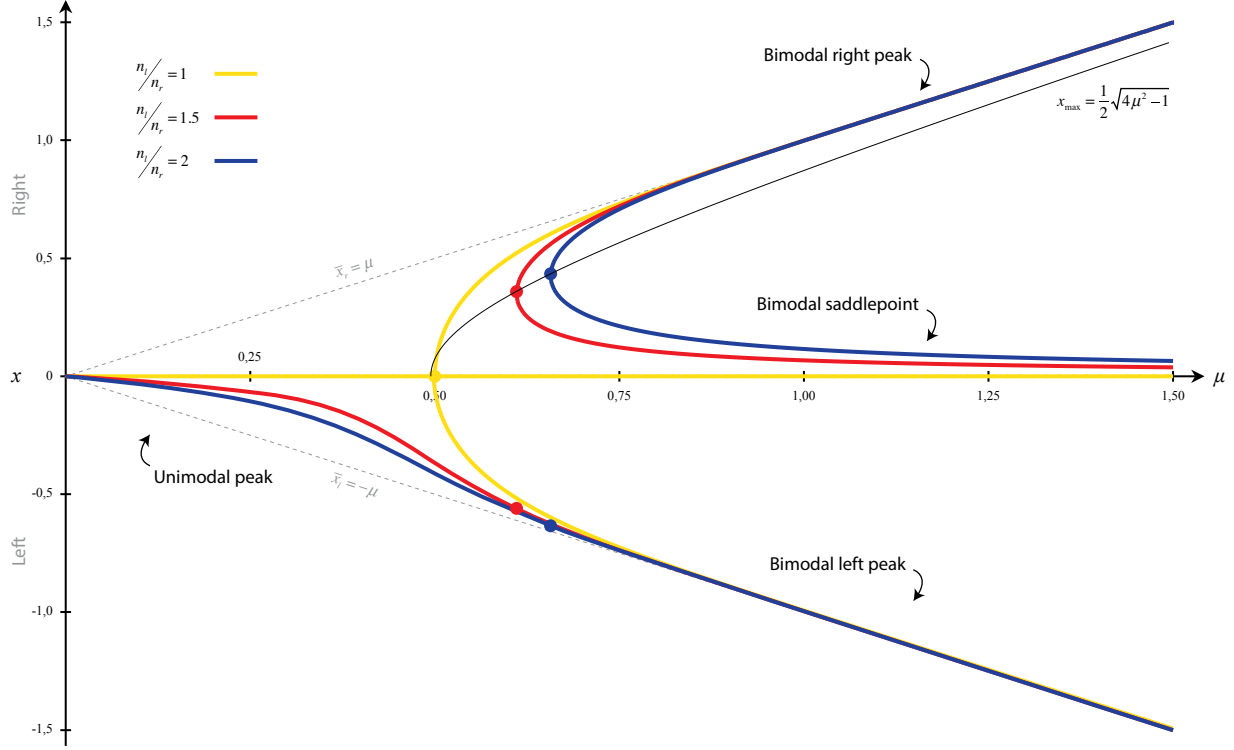


Figure 3: *Bifurcation diagram*. The curves indicate the extrema of the population density function along the x -axis. The left subpopulation is respectively twice as large (blue curve), half as large (red curve) and just as large (yellow curve) as the right subpopulation. The grey dotted lines indicates the mean ideal points of each subpopulation. The curves above the solid black curve indicates the right peak of the distribution, while curves below (and $x \geq 0$) indicate the saddle point. The dots indicate the breaking point between the unimodal distribution and the bimodal distribution at different relative sizes of the subpopulations.

of the subpopulation determines the exact value of μ , for which the distribution becomes bimodal. With subpopulations of equal size, $n_l/n_r = 1$, the distribution is bimodal for $\mu > \frac{1}{2}$. In addition the peaks of the bimodal distribution are located at equal distance to the origin, while the saddle point is located at the origin, $x^s = 0$. As the relative size of the subpopulation increases so does the value of μ , for which the distribution becomes bimodal. To determine this breaking point, we can differentiate equation 3 with respect to x ¹⁰. As the polarisation of the subpopulation, μ , increases the peaks of the distribution converges

$$10 \frac{\partial n_l/n_r}{\partial x} = -\frac{2\mu e^{8x\mu}(4\mu^2 - 4x^2 - 1)}{(\mu+x)^2} = 0 \Leftrightarrow x_{max} = \frac{1}{2}\sqrt{4\mu^2 - 1} \Leftrightarrow \mu_{max} = \frac{1}{2}\sqrt{4x^2 + 1}$$
 where x_{max} (or μ_{max}) is the breaking point between the right peak and the saddle point of the distribution. By substituting x_{max} back into equation 3 we get an expression for the breaking point in terms of the combinations of μ and n_l/n_r : $n_l/n_r = -\frac{\sqrt{4\mu^2 - 1} - 2\mu}{\sqrt{4\mu^2 - 1} + 2\mu} e^{4\sqrt{4\mu^2 - 1}\mu}$. Further note that if n_l/n_r is lower than the right-hand side of the equation, then the distribution is bimodal, and visa versa. At the upper limit of our range, when the left subpopulation is twice as large as the right ($n_l/n_r = 2$), we have that $\mu \simeq 0.657$, i.e. for any $\mu > 0.657$ the distribution is bimodal.

towards the mean ideal points of each subpopulation, $x_r^p \rightarrow \bar{x}_r = \mu$ and $x_l^p \rightarrow \bar{x}_l = -\mu$. With an unequal sized subpopulations, $n_l/n_r > 1$, the saddle point lies slightly to the right of the origin, $x^s > 0$. The asymmetric *bifurcation diagram* arises, because we choose to model the left subpopulation larger than the right, by choosing the range $n_l/n_r \in [1, 2]$.

3.2. Firm behaviour

We now turn our attention to the behaviour of firms. The paper investigates a market with anywhere from two to twelve firms, i.e. $N \in [2, 12]$. The number of firms is exogenously determined. The decision rules, the underlying rationale for each rule, and the assumptions and information necessary for each rule is presented below. Only the decision rules without foresight are presented. Once we have analysed the model with these rules and gotten a better understanding of each rule, the model is extended to include foresight and *inductive reasoning*. All firms choose their location simultaneously, and in each iteration the decision rule of the firm determines the location of the firm.

Ideally we would want each firm to choose the location that maximises its share of the market, given the location of the other firms. However all firms choose their location simultaneously, and thus when a firm has to choose its own location, the location of other firms is unknown. The firm can use the predicted location of the other firms. But the location outcome that each firm is trying to predict, depends on predictions that the firm and other firms form. This self-referential loop leads to logical indeterminacy and so the maximisation problem is ill defined. Instead of solving the maximisation problem for the optimal location, firms may use heuristics or rules of thumb, when they choose location. The literature review provided an overview of the many different decision rules previously considered. This paper will use three of the heuristic decision rules laid out by [Laver & Sergenti \(2011\)](#) as the base to which other decision rules are compared. We use the *sticker*, *aggregator* and *hunter*-rule since they represent respectively no exploration of the space, the social optimal locations, and a unceasing exploration for better locations.

3.2.1. Sticker-rule

The simplest decision rule a firm can use is the *sticker*-rule. With this decision rule the firm sticks to its initial position regardless of what happens. This could be a firm that is either unwilling to change or incapable of change. The management of the firm may have an unyielding belief in the long-run superiority of the position of the firm, discouraging it from any change, even in times of despair, i.e. a belief that the market share of the firm will always recover and excel, in and of itself. The firm might also be unable to change, due to financial constraint such as the fixed cost of relocating or investments that cannot be recuperated. And finally, given the task of predicting the future location of all other firms,

the firm might see its current position as less risky than a new location based on uncertain predictions.

3.2.2. *Aggregator-rule*

Firms using the *aggregator*-rule constantly seeks to please its own customer base. The firm does not try and predict the future. Instead it takes its current market area and locates at the centre of it. More specifically the centre or *centroid* is the mean ideal point of all customers of the firm. The centroid position takes into account the population density within the market area, drawing the firm towards the centre of mass (see figure 2).

The future location of competing firms may change and thus change the market area of the firm. And so there is no guarantee that the centroid of the current market area is identical to the centroid of the subsequent market area. Nonetheless firms using the *aggregator*-rule continually pursue the mean ideal point of their customer base. Likewise there is no guarantee that the relocation of the firm will increase its share of the market, or even maintain the current share. Market shares are in a sense a secondary priority for the firm using the *aggregator*-rule. The management of the firm may reason that they can retain and recruit new customers by pleasing their current customers¹¹.

When all firms in the market use the *aggregator*-rule then the model is an implementation of the *Lloyd's algorithm* and the location of firms will converge to the stable *Centroidal Voronoi Tessellation* (Laver & Sergenti, 2011, chapter 3, pp. 48-49). The *Centroidal Voronoi Tessellation* (CVT) is a special geometric construction where each firm is located at the centroid of its own market area. *Aggregator* firms located at the current centroid will not relocate. Hence when all firms use the *aggregator*-rule the CVT is stable – none of the firms relocate. The CVT has the useful property that the distance between all customers and their closest firm is minimised. The location of firms is socially optimal when the average distance to consumers is minimised. This will prove useful later on when comparing the social welfare of the different decision rules. The all-aggregator model will maximise our measure of social welfare.

The underlying requirement for an *aggregator* firm is that it has perfect knowledge of its current customer base. The firm knows the span of its current market area and the distribution of customers within the area, such that the firm can correctly determine the mean ideal point of its customers.

The *aggregator* firm never relocates outside its current market area. But otherwise there

¹¹Noted that customer loyalty is not incorporated in the models. Customers always choose the closest firm regardless of previous purchasing history. And this might understate the efficiency of firms using the *aggregator*-rule in markets with a high degree of customer loyalty.

is no limit as to how far an *aggregator* firm can move at each iteration. The firm nonetheless tends to move in relatively small steps, unless the market is extremely unstable.

3.2.3. Hunter-rule

The *hunter*-rule is a trial-and-error type of decision rule. The firm continues in the same direction, if it previously proved fruitful, and otherwise the firm heads in the opposite direction. At each iterations the firms move what corresponds to 0.1 standard deviation of the population distribution. If the previous move did not increase the market share, then the firm turns around and heads in a random direction drawn from the 180 degree arc now in front of it.

Firms using the *hunter*-rule never settle down. The firm endlessly hunts for higher market shares with the same speed and intensity. For a firm with this decision rule there is no comfortable threshold share of the market that suffices or slows exploration. In the trade-off between exploration and exploitation the firm always chooses the first option.

The only information used in the decision process of a *hunter*-firm is the current relative change to its market share and its current direction. No information going further into the passed is used. The firm may lack memory, or may emphasis the present to such a degree that information going further back is seen as worthless. The behaviour of a *hunter*-firm is most suited for a fast evolving market with high unpredictability.

The *hunter*-firm moves 0.1 standard deviation each iteration. This is the *speed of adaption*. It is beyond the scope of this paper to investigate the effects of changing the speed parameter. However [Laver & Sergenti \(2011, chapter 7, pp. 150-151\)](#) find evidence that a speed parameter of 0.1 standard deviation is optimal for a *hunter*-firm. They argue that this speed strikes the balance between quickly reaching better locations without overshooting these location when moving around.

3.2.4. Deliberate reaction to competitors

The above described heuristic decision rules are good first approximation on how firms might choose to locate. Especially in a simultaneous multi-agent location model where the future location of other firms is unknown. Later on we would like to investigate how foresights affects our results. But before doing so, we need to return to the deliberate process of maximising market share, i.e. before we can answer how the firm locates given the *predicted* locations of competing firms, we need a new decision rule that determines how the firm locates given *any* location of competing firms. None of the above mentioned decision rule take the location of competing firms into consideration. In this subsection we develop a decision rule that assumes that competing firms stay at their current location. Later on we

extend the decision rule so it includes location predictions and learning. In the end we will have reintroduced strategic considerations into the simultaneous location model.

For now assume that competing firms stay at their current location. We want to know where the firm should relocate to maximise its market share. When a firm relocates it gains new customers and loses others. Thus there is a trade-off between the areas of the market that it gains and the areas that it loses. To simplify matters even further we only focus on the first effect. The problem is then equivalent to a new firm entering a market – populated with competing firms – and choosing the location that will maximise its share. There is extensive research on competitive location models such as the Hotelling model and the Voronoi game, and on the geometry behind Voronoi diagrams. Yet only four papers provide methods on how to find the location that maximises the market of a firm in two-dimensional space. The methods all concern the location of newly entering firms. None consider the optimal location of an existing firm that relocates. [Cabello, Díaz-Báñez, Langerman, Seara & Ventura \(2010\)](#) use the *reverse nearest neighbour* method to find the position that maximises the number associated points or customers. However their method requires a finite number of points. Thus this method is not applicable, since we assume an infinitely continuous number of consumers by using the distribution of consumers. The remaining three papers use continuous distributions, but they assume the distribution is uniform, which also make them unsuited for our needs, since we have a *mixed bivariate normal distribution*. [Averbakh, Berman, Kalcsics & Krass \(2015\)](#) use a method that partitions the solution space into smaller regions. The partition is done in such a way that the structure of the Voronoi diagram is unaffected by how the firm locates within each region. The structure is only affected by which of the regions the firm locates in. The partition simplifies the problem to a *search problem* over all regions. Their method uses the *Manhattan* distance metric. The partitioning also work with *Euclidian* distance metric. However it might be impossible to obtain exact solutions with this distance metric, because of the complexity that the *Euclidian* norm introduces in the object function ([Averbakh et al., 2015](#), pp. 409-410). The models in this paper uses the *Euclidian* distance metric. [Cheong, Efrat & Har-Peled \(2007\)](#) note the difficulties in finding analytical solutions to the problem and develop an algorithm that approximates the maximum of the object function. Their method finds the largest circle that is contained within each existing market area and that does not contain any of the existing firms. They construct squares with lengths equal to the radius of the largest empty circles and with the same centres as the circles. These squares are partitioned into grids and they use an ϵ -*approximation* method to pick the cell in the grid that maximises the market area. As previously mentioned [Cheong et al. \(2007\)](#) assume a uniform distribution of customers. When working with non-uniform distributions the largest empty circles is a poor criteria to narrow down the search for maximum, since

there is no guarantee that a dominant part of the density mass falls within the circle. The last paper by [Dehne, Klein & Seidel \(2002\)](#) proves that if the location of neighbouring firms span a *convex hull*, then there exists a unique local maximum inside this *convex hull*. They use a *Newton method* to calculate these local maxima within each of the *Delaunay circles* (that is the smallest circles covering each triangles in the *Delaunay triangulation*). In addition they check for corner solutions, i.e. locating “on top” of existing firms. The optimal location for the new firm is then the maximum of all these local maxima. The case where the neighbouring firms do not span a *convex hull* is left open. And so they only provide a partial solution to the problem. There is nothing that prevent neighbouring firms from locating such that they do not span a *convex hull* in the models of this paper, and thus this method is not suitable either.

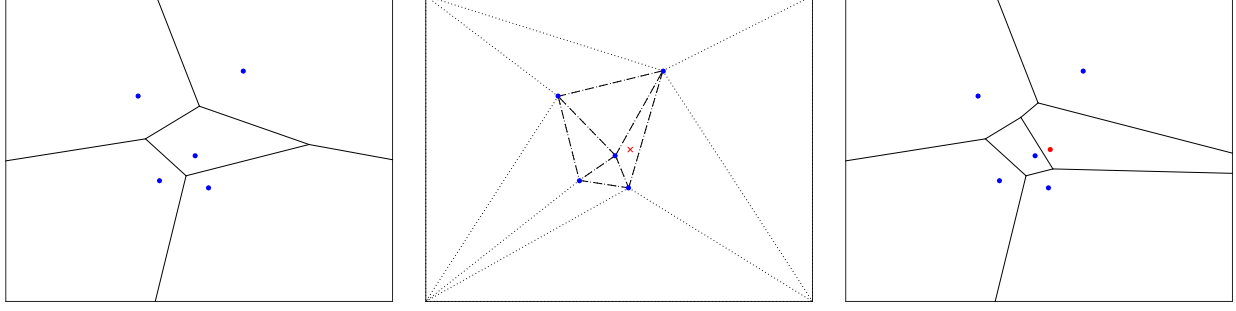
None of the currently existing methods are applicable to the models used in this paper. This is due to the non-uniform infinitely continuously distribution of consumers, where neighbouring firms may not form *convex hulls*, and because we use the *Euclidian* distance metric. Since there does not exist any methods capable of finding the optimal solution, we are forced to develop a heuristic method that roughly approximates the solution. I draw on the commonalities of these papers when constructing the decision rule that explicitly tries to find the location that maximises the market share of the firm.

3.2.5. *Maxcov-rule*

Firms using the *maxcov* decision rule aims for the location that maximises the number of customers. The firm assumes that the ideal location lies in the gaps between competing firms. The firm considers all the gaps and picks the one with the most consumers. Each gap is a triangle in the *Delaunay triangulation*. The *Delaunay triangulation* is constructed using only the location of competing firms, and the boundary points of the space, see figure 4. The latter insures that the firm also considers locations that lie outside the area spanned by competing firms ([Cheong et al., 2007](#), p. 556). The firms does not consider its own current location when determining the ideal point, i.e. the firm only considers the area it gains by relocating and not the potential loss. The Delaunay triangle with most consumers is selected¹². The

¹²The triangle with the most consumers is a proxy for the actual market share obtained by the firm once it locates within the triangle. The triangle and the actual market (obtained once the firm locates) differs in shape and size, as seen in figure 4b and 4c. The triangle with most consumers may not always be the triangle that results in the highest market share for the firm. However in 40-90% of the cases the triangle with most consumers is also the triangle that results in the highest market share. If the firm did not locate in the triangle with most consumers, but selected any one of the Delaunay triangles with equal probability, it would pick the triangle that results in the highest market share in only 4-16% of all cases. Although this approach is not perfect, it is sufficient and computationally fast. The problem is reduced to a *search problem* over all the $2 + 2(N - 1)$ Delaunay triangles, where $N - 1$ is the number of competing firms.

Figure 4: Example with five competing firms (blue markers) in a market where $\mu = 0$ and $n_l/n_r = 1$.



(a) The Voronoi diagram illustrating the market areas of competing firms.

(b) *Delaunay triangulation* of competing firms and boundary points. Red marker is weighted centroid of triangle with most customers.

(c) The Voronoi diagram once we include the ideal position of the firm (red dot).

ideal location is the weighted mean ideal point of all customers within that triangle. We use the density of consumers as weights.

Rather than move directly to the ideal position, the *maxcov*-firm, moves 0.1 standard deviations in the direction of the ideal position. There are three reasons for this gradual adjustment. First, it makes the speed at which the *maxcov*-firm moves comparable to the speed of the other decision rules. Recall that the *hunter*-firm moves 0.1 standard deviations, and that the *aggregator* never moves outside its current market area. Secondly, none of the available methods deal with the optimal relocation of the firm, so we are forced to use a method that treats the problem similar to that of a new firm entering and locating in a market already populated with existing firms. By moving incrementally the firm acknowledges its current location. Finally, the ideal position of the firm is sensitive to changes in the position of competing firms. Even minor changes may alter which of the triangles contain the most customers, and thus lead to significantly different ideal position. The gradual adjustments insures less volatile variation in the relocation pattern of the firm, which is necessary later on when other firms attempt to predict the future location of the firm.

When a *maxcov*-firm chooses its location it explicitly assumes that competing firms remain at their current position. When the firm chooses its location the location of the other firms is unknown. Not knowing how the other firms will move and without predictions it may be reasonable to use the current position of the firm as the best point of reference. In determining its location the *maxcov* firm uses information on the location of competing firms. Additionally it is assumed that the firm has perfect knowledge of the consumer distribution such that it

can determine which of the gaps contain the largest number of consumers¹³.

3.3. Summary variables

There are three main perspectives we want to analyse in the competitive location model. One perspective is the location of the firms. Do firms agglomerate or cluster at particular locations or do firms disperse across the market space? Another perspective is the competitive environment. Is competition a *winner-take-all* game in which one firm is able to capture a predominant share of customers, or are customers evenly shared among firms? The last perspective concerns *social welfare*. How does the competition among firms affect the wellbeing of customers? To investigate these perspectives we construct three summary variables. All these variables aggregate the state of the market into a single measure that is comparable across parameter settings and models.

Mean eccentricity

To summarise the position of firms we use the mean eccentricity. Eccentricity measures the distance from a firm to the mean ideal point of all customers. Eccentricity has the desired properties of a summary variable. It is a single measure as opposed to the coordinates of the firm that are two dimensional. It is a relative measure that is naturally interpretable as the distance to the population centre. And since we use the mean ideal point which changes with the parameters, mean eccentricity is comparable across different parameter settings. Had we used the origin of the coordinate system instead, this would not change with the parameters and our distance would thus depend on the specific parameter setting. In addition the origin is an arbitrary location and thus not easily interpretable. We take the average of all firms to get the mean eccentricity in each iteration.

Effective number of firms (ENP)

To summarise the competitive environment we use a relative measure of the market shares of the firms. This measure is known as the *effective number of parties* (ENP), but will in this paper be referred to as the *effective number of firms*. It measures the concentration of market shares among firms. The measure goes from 1 and up to the number of firms in the market (N). In a market with four firms where all firms have an equal share of the market the effective number of firms is four (i.e. $ENP=N$), while the effective number of firms is

¹³Alternatively – and perhaps more realistically – the firm could attempt to predict which gaps contained the largest number of customers by using only the market share of surrounding firms (Fowler & Laver, 2008, p. 75). However to avoid the forecast errors arising from such predictions this paper assumes the *maxcov*-firm has perfect knowledge of the distribution of customers.

one (ENP=1) if a single firm captures the entire market. The ENP tells us how many firms would be in the market, if they all had identical shares. The ENP is defined as:

$$ENP = \frac{\left(\sum_j^N s_j\right)^2}{\sum_j^N s_j^2} \quad (4)$$

where s_j is the share of firm j . The ENP is the inverse of the Herfindahl-Hirschman Index (Laakso & Taagepera, 1979, p. 4). ENP and the Herfindahl-Hirschman Index measure the same thing. However the ENP is the easiest to interpret across our parameter settings, since we change the number of firms in the market. Eaton & Lipsey (1975) argue that a necessary condition for equilibrium is that no firm has a total share of the market that is less than half the market of any competitor. If a firm has a lower share than this, then the firm would choose to pair with the competitor thereby archiving at least half of the competitors market. This also holds in a two-dimensional market. We can show that this condition is not fulfilled when $ENP < \frac{(1+N)^2}{3+N}$, calculations provided in [Appendix A](#).

Mean representation

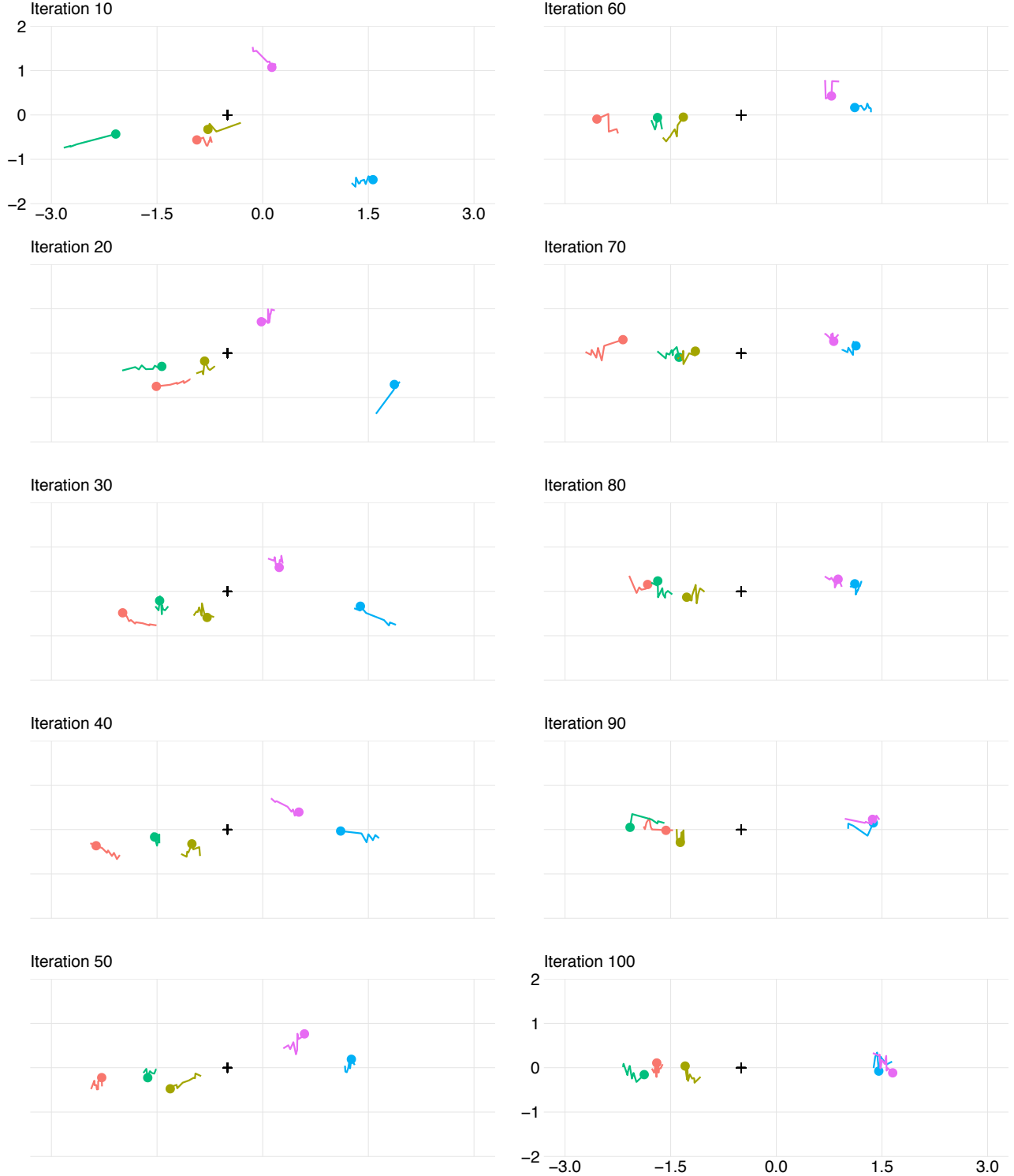
It is straightforward to create a measure that summarises the consumer welfare given our utility function (equation 1). By taking the average over all customers we get the mean utility or mean representation in every stage. This measure tells us how well the location of firms represent the ideal point of consumers. The social optimal location of firms is the one that minimises the average distance to customers. This is equivalent to maximising the mean representation.

4. Methodology

We have now gone through the basic setup of the economic model. We have discussed the economic rationale underlying consumers and the behaviour of firms. We have specified the variables of interest and how these should be interpreted. The following section gets slightly more technical as we delve into the methodology of solving the model and estimating the variables.

The decision rule of the firm determines how the firm chooses to locate in each *iteration*. We execute the model with several *iterations*, observe how the location of firms converge, and analyse how the decision rules and parameter values of the model affect the location of firms. Figure 5 provides an example of the location of firms at different iterations.

Analytical models are solved mathematically. Results, sensitivity to parameter changes and verification of the procedure follows directly from the derivation of key equations.



Whereas computer simulated models, such as agent-based models, requires some pre-planning and multiple executions in order to analyse results and insure trustworthy results. This type of pre-planning resembles what takes place in an experimental study and therefore the following outline of the procedure is also known as the *experimental design*.

4.1. Model parameterisation

To initiate the models we need to set the initial position of firms. For the results of the model to be credible, they should be independent of the initial positions. The model should produce similar end-results regardless of the initial position of firms. First of, we draw the initial positions randomly¹⁴. Secondly, we execute several *repetitions* of the model with identical parameter values, but with varying randomly selected initial positions. We compare the results across *repetitions* to insure that the location of firms converge to the same limited set of locations. This will be the case in all our models, and thus we satisfy that end-results are independent of the initial position of firms.

We want an *experimental design* such that we can evaluate how the parameters of the model affect results. So each model is executed with different combinations of parameter values and subsequently compared. To set the parameter values we employ two different methods. In the first set of models with a symmetric consumer distribution, all firms use the same decision rule, and the model only has two parameters – the decision rule and number of firms. There are 4 decision rules and the number of firms can take on 11 different values, since $N \in [2, 12]$. The parameter values are integer numbers and the entire parameter space spans 44 cells or 44 combinations, so for these models we use the simple *grid sweep* method. This method *runs* through the entire parameter space executing each combination of parameters in turn. The last set of models have two additional parameters – the polarisation of the subpopulations $\mu \in [0, 1.5]$ and the relative size of the subpopulations $n_l/n_r \in [1, 2]$. The full parameter space is huge and these parameter values are real numbers with no “natural” increments. And so there is no intrinsic reason for a grid representation to be a fair representation of the full parameter space, and even less so considering the nonlinearity of the models (see figure 3). This makes the *grid sweep* method unsuitable, instead we use the *Monte Carlo parameterisation* method. For every *run* this method draws the parameter values uniformly random. With sufficiently many *runs* the method is able to map out results spanning the entire parameter space.

To summarise we execute several *runs* to evaluate the effect of the different parameters. In each *run* we execute several *repetitions* with the same parameter values, but with varying

¹⁴The position of a firm is drawn uniformly random from a disc centred at (0,0) and with a 3 standard deviation radius.

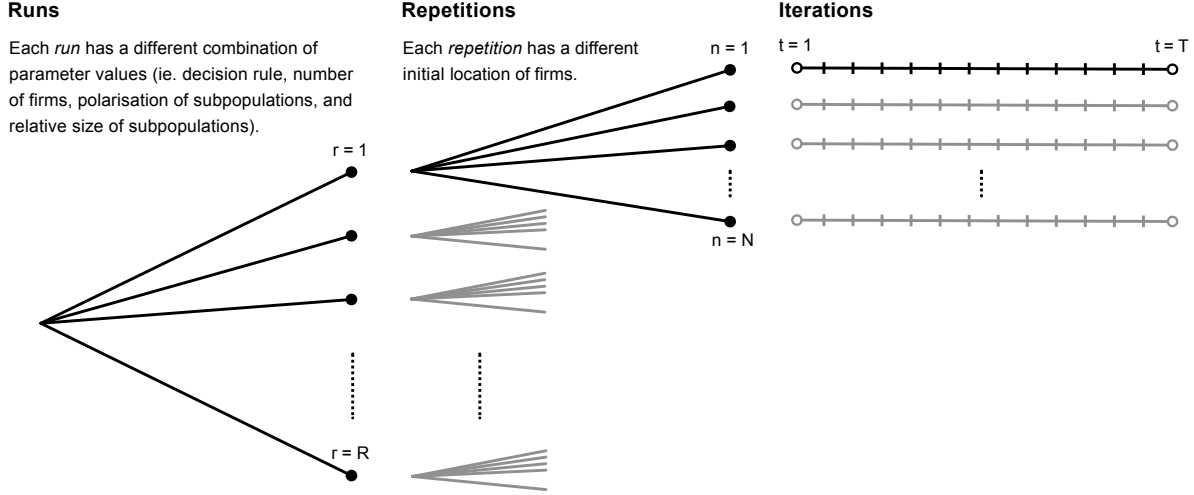


Figure 6: Outline of the experimental design.

randomly drawn initial positions. In every *repetition* the model goes through numerous *iterations* enabling the location of firms to converge. And at each *iteration* the decision rule of the firm determines how the firm locates. See outline of the experimental design in figure 6.

4.2. Markov chain

We have laid out the specifications of our model and will shortly turn to the methods used to estimate the output variables of the model. But first we take a look at the underlying process of a *run* of the model. We show that a *run* of the model constitutes a *stochastic process*. Laver & Sergenti (2011) note that most computational models can be represented by a particular *stochastic process* known as the *time-homogenous Markov chain*. We present the necessary conditions for a *time-homogenous Markov chain*. When the model satisfy these conditions we have prior knowledge of the dynamics of the process – in particular convergence and steady state. Knowing the dynamics of the process allows us to construct methods that give representative estimates of the output variables.

4.2.1. Stochastic process

The model generates a vector with the values for all output variables at each *iteration* in every *repetition*. The values of the vector could for instance be the coordinates of the firm, the market share of firms, a measure of the effective number of firms (ENP), a measure of the eccentricity of locations and a measure of how well the location of firms represents the consumers' ideal points. We follow Laver & Sergenti (2011) and use the following notation; $y_t^{(n)}$, where y is the vector of values, t is the iteration number and n the repetition number. One *repetition* of the model with ten *iterations* would produce a series of ten vectors;

$(y_1^{(1)}, y_2^{(1)}, \dots, y_{10}^{(1)})$. Because of the randomly drawn initial positions of firms, the exact values of these ten vectors depend on the *random seed* and another *repetition* with a different *random seed* returns different values of the vectors¹⁵. We can use this to our advantage, if we use a different *random seed* for each repetition. Then the vector $y_t^{(n)}$ represents a single realisation of a *random vector* Y_t , where Y_t is all possible realisations of y at iteration t . That is, $y_1^{(1)}$ is a realisation of Y_1 associated with repetition 1. Similarly the series of vectors $(y_1^{(1)}, y_2^{(1)}, \dots, y_{10}^{(1)})$ is a realisation of $(Y_1, Y_2, \dots, Y_{10})$ associated with repetition 1. The series of random vectors $(Y_1, Y_2, \dots, Y_{10})$ represents a *stochastic process*. And so, one *repetition* gives a series of vectors, e.g. $(y_1^{(1)}, y_2^{(1)}, \dots, y_{10}^{(1)})$. While a *run*, which consists of multiple *repetitions*, constitutes a *stochastic process*.

There are three factors, that distinguish different stochastic processes. One factor is the range of all the possible values that the random vector might take. This is also known as the *state space*. Another factor is the iteration *index set*. In the example above with ten iterations the index set is $\{1, 2, \dots, 10\}$. This paper will only focus on discrete-time processes. The last factor is the dependency between the random vectors, Y_t , in the process.

4.2.2. Time-homogenous Markov chain

The Markov process is a particular stochastic process, that satisfies the Markov property. The Markov property restricts the dependencies of the random vectors in the process, namely that the future state of the process may depend on the current state, but cannot depend on any of the previous states of the process. In other words the Markov process, with a vector of the state space, X_t , satisfies the Markov property if:

$$Prob[X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0] = Prob[X_{t+1} = j | X_t = i] \quad \forall t \quad (5)$$

for all states of the process i_0, \dots, i_{t-1}, i , and j . In this paper the state of the process simply summarises the location of all firms, e.g. state i refers a specific and unique location of all firms, this could be the initial position of firms. If one or more firms change location, then the process enters a different state j . If all firms returned to their respective initial positions, then the process would once again be in state i . The state space contains all possible states

¹⁵Computers are deterministic machine incapable of generating truly random numbers. Instead they use a pseudorandom number generator that approximates random numbers. The pseudorandom numbers are completely determined by the *random seed*. The *random seed* is the number used to initialise the pseudorandom number generator. Initiating the generator with the same *random seed* will produce the same sequences of random numbers. While initiating the generator with different *random seeds* produce different sequences of random numbers. Further note that the different sequences of numbers are independent and identically distributed (IID) across the *random seeds*. Although the numbers stemming from a pseudorandom number generator are not truly random, they are sufficiently random for most applications including the models in this paper.

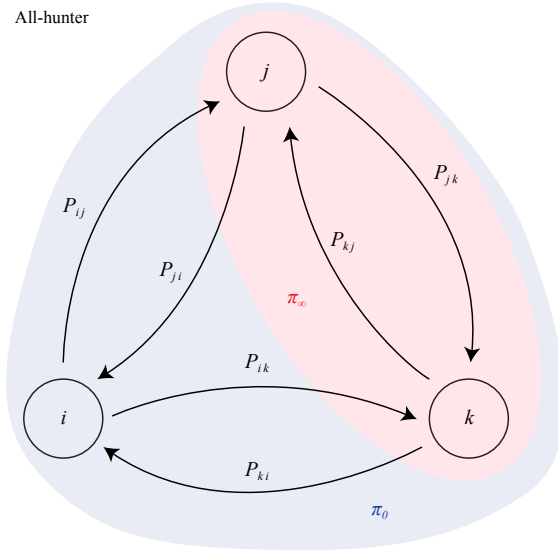
of the process. With a finite state space, we can write the *one-step transition probability* as $P_{ij}^{t,t+1} = \text{Prob}[X_{t+1} = j | X_t = i]$, which is the probability for X_{t+1} being in state j , given that X_t is in state i . A *time-homogenous Markov chain* further requires that the *transition probabilities* are independent of the iteration parameter; $P_{ij}^{t,t+1} = P_{ij}$. This is also known as a Markov chain with *stationary transition probabilities*. Only the current state affects the probability of the next state in the process and probabilities stay constant over time. Figure 7 provides stylised examples of different *time-homogenous Markov chains*. We use matrix notation to shorten the equations describing the evolution of the process¹⁶. We can represent all the stationary *one-step transition probabilities* using the *transition probability matrix*, P . The *state space distribution vector*, π_t , represents the unconditional probability distribution of the state space at time t . Each element i in the vector describes the probability that the process will be in state i at iteration t . The *state space distribution* then evolves as given by $\pi'_{t+1} = \pi'_t P$. From this equation it follows that if we know the initial state space distribution, π_0 , we can derive the state space distribution vector, π_t , using $\pi'_t = \pi'_0 P^t$.

We have argued how a *run* of our model constitutes a stochastic process and linked it to the dynamics of Markov chains. We are now at a point where we can look closer at convergence and steady state. The state space distribution vector is *stationary* when $\pi_{t+1} = \pi_t$. Because of the randomly drawn initial location of firms the initial state space distribution, π_0 , is seldom stationary and several iterations are needed. The process reaches steady state once the state space distribution becomes stationary. That is $\lim_{t \rightarrow \infty} \pi_t = \pi_\infty$, where the stationary state space distribution, π_∞ , solves $\pi_{t+1} = \pi_t$. All *time-homogenous Markov chains* converge to at least one steady state distribution. A process that converges to a unique distribution vector, π_∞ , regardless of the initial distribution vector, π_0 , is known as an *ergodic* process. There are two types of *time-homogenous Markov chains* (THMC); *stochastic THMC* that contain a random component (besides the randomly drawn initial positions), and *deterministic THMC* that do not contain any random component and where the probability that the random vector, Y_t , takes on a particular values is 1. This distinction is useful since not all THMC are *ergodic*, but as Laver & Sergenti (2011, chapter 4, p. 64) note all *stochastic THMC* with a finite state space are *ergodic*. With a random component in the process each state has strictly positive probability of being reached in a finite number of iterations, thus the process avoids “getting stuck” and eventually converges to the unique distribution vector (Laver & Sergenti, 2011, chapter 4, p. 71). Recall that a *hunter-firm* turns around and heads in a randomly selected direction. A random component such as this

¹⁶We let S denote the dimension of the state space, that is the number of possible states. The *transition probability matrix* P is a $S \times S$ sized matrix. The size of the *state space distribution vector* π_t is $S \times 1$.

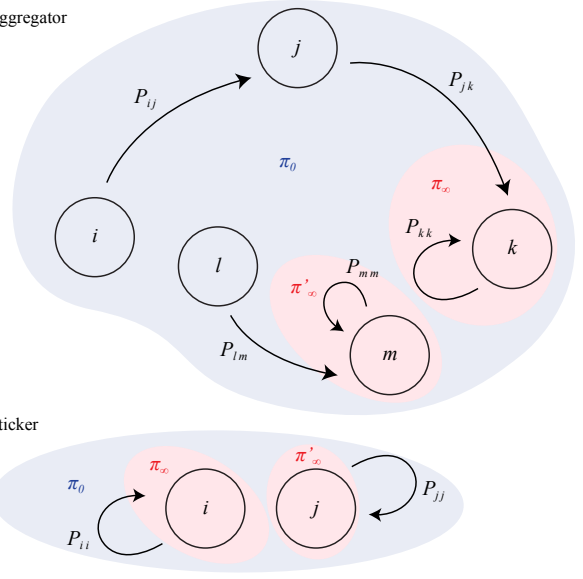
Stochastic time-homogenous Markov Chain

All-hunter



Deterministic time-homogenous Markov Chain

All-aggregator



All-sticker

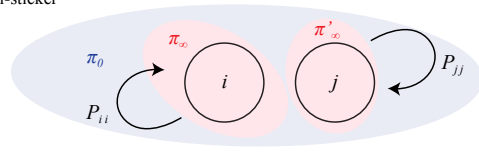


Figure 7: Stylised examples of *time-homogenous Markov chains*. Each arrow has a corresponding positive *one-step transition probability*, P_{ij} (no arrow indicates that the *one-step transition probability* is zero). The process or *run* starts with an initial state space distribution π_0 , where each element i in the vector π_0 is the probability of the process starting in state i . The process converges to the stationary state space distribution, π_∞ (states within the shaded pink area have positive probability, while states outside have zero probability). The random component in the all-hunter model insures that all *one-step transition probabilities* between different states are strictly positive. A *hunter-firm* never settles down and thus there are no self-loops (the diagonal in the *transition probability matrix*, P , consists of zeroes). However the process is *ergodic* so it converges to a unique stationary state space distribution, π_∞ , regardless of the initial state space distribution. The all-sticker and all-aggregator model are deterministic THMC, thus all *one-step transition probabilities* equal 1. Both models converge to a single state rather than a distribution of states, since there is no oscillation. However the processes are *non-ergodic*, and thus the single state depends on the specific probabilities in initial state space distribution π_0 .

insures that the process is *ergodic*. We know that this process will converge to a unique state space distribution, π_∞ , and that this distribution is independent of the initial state space distribution. There is no guarantee that a *deterministic THMC* converges to a single state, since it might oscillate between states. The deterministic process underlying the all-maxcov model – where all firms use the *maxcov* decision rule – does sometimes oscillate between several states. We will deal specifically with this special case in section 4.2.5 – all other deterministic process in this paper converge to a single state. And all the *deterministic THMC* in this paper are *non-ergodic*. The arguments for these assertions are provided in Appendix B. This means that although the process converges to a single state, this state is not unique, but depends on the initial position of firms. The method used to estimate the values of the output variables takes this into account, so our end-results are independent of the initial location of firms.

It is often possible to construct several Markov representations. When choosing the vector

of the state space, X_t , one needs to insure that the output variables, Y_t , can be derived from the vector of the state space, i.e. $Y_t = f(X_t)$. In addition the vector of the state space, X_t , has to satisfy the Markov property. In most of the models in this paper the vector of the state space only needs to contain the coordinates of the firms, since we can calculate the remaining output variables from the coordinates¹⁷. Once the vector of the state space, X_t , reaches steady state so will the output variables, Y_t .

4.2.3. Estimating output variables

We calculate a mean estimate for each of the values in the output variable, i.e. we calculate the mean estimate of the effective number of firms (ENP), the mean eccentricity and the mean representation. We want an accurate estimate of the output variable in steady state, and so none of the output variables obtained in transient states can be used to estimate the steady state. We discard all *burn-in* iterations, that is the iterations needed to reach the steady state. In the subsequent section we return to the empirical issue of determining the number of *burn-in* iterations, but for now assume that the process has burnt in.

Define $\psi_t^{(n)}$ as the expected value of the output variable Y_t at repetition n and iteration t . We know that in steady state the state space distribution vector is *stationary*. Thus in steady state the expected values of any of the output variables, $Y_t, Y_{t+1}, Y_{t+2}, \dots$, must also be stationary, $\psi_t^{(n)} = \psi_{t+1}^{(n)} = \psi_{t+2}^{(n)} = \dots$, for repetition n . Removing the time subscript we have that $\psi^{(n)}$ is the expected value of Y_t for repetition n over all iterations t in steady state. Using this we can define the process of the output variable as the sum of expected value and disturbance term,

$$Y_t = \psi^{(n)} + \varepsilon_t \quad (6)$$

The disturbance term, ε_t , is serially correlated across iterations and therefore indexed with t . We execute several repetitions not just repetition n . For each repetition we may have a different expected value, i.e. a different value of $\psi^{(n)}$. So we model $\psi^{(n)}$ as the realised value of a random variable, where $\psi^{(n)}$ is drawn from a distribution with mean ψ and standard deviations σ_ψ . Consequently ψ is the expected value of Y_t over all repetitions and all iterations in steady state.

To estimate ψ we can use the *ensemble average*. This entails executing several repetitions of a run, up until a pre-specified iteration t that is within the steady state. And then taking

¹⁷Although the coordinates take on real numbers in theory, in practice when executed on any computer there is a limit to the precision of the coordinates. Matlab stores values using up to 64-bits ([MathWorks, 2016](#)). This limited precision is enough for us to say that the coordinates are discrete (to a high level of precision), and thus the state space is finite (although quite large).

the average over all repetitions of the realised values, $y_t^{(n)}$, of the output variable, Y_t , at iteration t :

$$\text{Ensemble Average}_t \text{ of } Y_t = \sum_{n=1}^N \frac{y_t^{(n)}}{N} \quad (7)$$

where N is the total number of repetitions. The *ensemble average* converges to ψ as the total number of repetitions go to infinity:

$$\psi = \text{plim} \sum_{n=1}^N \frac{y_t^{(n)}}{N}$$

The disadvantage of using the *ensemble average* to estimate ψ is that we discard a lot of *burn-in* iterations. For instance if we determine that it takes 251 iterations for the process to burn in and we execute 100 repetitions, then the first 250 *burn-in* iterations in each repetition are of no use when estimating the output variables in steady state. In total we discard 25,000 iterations and only use information from 100 iterations (the last iteration in each repetition) in our estimate. By any stretch this is a highly ineffective use of computational resources. For this reason, when possible, we would prefer to use the *time average* to estimate ψ . This entails executing a single repetition of a *run*, up until a pre-specified number of post-burn-in iteration. And then taking the average of $y_t^{(n)}$ over all post-burn-in iterations:

$$\text{Time Average}^{(n)} \text{ of } Y_t = \sum_{t=1}^T \frac{y_t^{(n)}}{T} \quad (8)$$

where T is the total number of post-burn-in iterations and all iterations, $t \in \{1, 2, \dots, T\}$, are within the steady state. As the total number of iterations go to infinity, the *time average* converges to $\psi^{(n)}$, i.e. the expected value of Y_t for repetition n . Recall that an *ergodic* process converges to the same unique state space distribution regardless of the initial distribution. And since the only difference between different repetitions within the same run is the initial positions of firms, it implies that the limiting state space distribution is independent of the particular repetition when the process is *ergodic*. So for a *ergodic* process we have that $\psi^{(n)}$ is equal to ψ , in which case:

$$\psi = \text{plim} \sum_{t=1}^T \frac{y_t^{(n)}}{T}$$

Recall the example from above where it takes 251 iterations for the process to burn in. We further assume that the process is ergodic and execute 100 post-burn-in iterations. Using the *time average* we execute a single repetition with an overall of 351 iterations. In total we only discard 250 *burn-in* iterations, and use information from the last 100 iterations to estimate ψ . In this example we reduced the number of discarded iterations by a factor 100

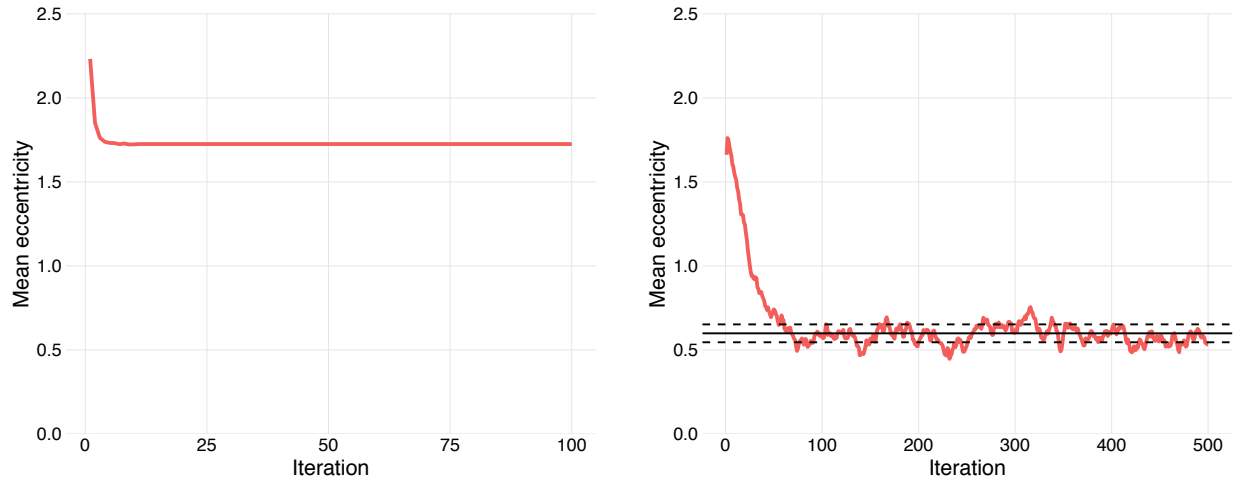
when using the *time average* instead of the *ensemble average*. However with a high degree of autocorrelation in the process it might not be possible to use the *time average* – which is what we show below. Further note that all our *deterministic time-homogenous Markov chains* are *non-ergodic*, so for these processes we have no other option, but to use the *ensemble average* to estimate ψ .

The disturbance term in equation 6 is independent and identically distributed (IID) across repetitions, because we use a different *random seed* for each repetition. As previously noted different *random seeds* give different sequences of numbers, and these sequences are IID, which in turn produce disturbance terms that at any given iteration are IID across repetitions. So when executing several repetitions the observed values at iteration t will be reasonably equally spread around the true mean, and the *ensemble average* gives a representative mean estimate of ψ . The *time average* may not give a reasonably representative estimate, because the disturbance term is serially correlated across iterations. While an ergodic process will eventually map out the entire steady state distribution vector, π_∞ , the process is slow to map out the distribution if there is a high degree of autocorrelation. A fully mapped out distribution assures that the *time average* gives a representative estimate of ψ . To check whether enough observations have been collected to map out the steady state distribution vector we run several test repetitions. For each output variable we calculate the *R-hat statistic* (Laver & Sergenti, 2011; Brooks & Gelman, 1998)¹⁸. The R-hat statistic is a relative measure of the between-repetition variance and the total within-repetition variance. And thus the measure reveals whether there is a potential to trim the state space distribution further by increasing the number of iterations. A low R-hat statistic indicates less potential to trim. In the limit the R-hat statistic tends to 1. This paper uses the typical cutoff level of 1.05. If the R-hat statistic for every output variable is lower than 1.05, then we feel confident that the steady state distribution has been mapped out and proceed to using the *time average* in the final execution of the model. If the R-hat statistic is above 1.05 we can try to increase the number of iterations and re-run the test, or we will determine that the speed of convergence is prohibitively slow and resort to using the *ensemble average* in the final execution of the model. In the following section we distinguish between *stochastic THMC* where the *time average* provides a representative estimate of ψ and those that do not, i.e. whether the R-hat statistic is below the cutoff level for all the output variables, or not.

¹⁸The *R-hat statistics* is also known as the *potential scale reduction factor*. When running our test repetitions we calculate the *R-hat statistic* using the second half of all iterations. The R-hat statistic is calculated as: $\hat{R} = \sqrt{\frac{\tilde{T}-1}{\tilde{T}}W + \frac{1}{\tilde{T}}\frac{B}{W}}$, where $B = \frac{\tilde{T}}{N-1} \sum_{n=1}^N (\bar{\gamma}_n - \bar{\gamma})^2$ and $W = \frac{1}{N(\tilde{T}-1)} \sum_{n=1}^N \sum_{t=1}^{\tilde{T}} (\gamma_{t,n} - \bar{\gamma}_n)^2$ are respectively between-repetition and within-repetition variance. The total number of repetitions is N , and $\tilde{T} = T/2$ is half of the total number of iterations, while $\gamma_{t,n}$ is the realised value in repetition n at iteration t .

4.2.4. Determining burn in

Figure 8: Example of trace plots.



(a) A single repetition from a deterministic THMC. This repetition burns in after 12 iterations, where mean eccentricity flatlines.

(b) A single test repetition from a stochastic THMC. Solid black line is the *time average* calculated using the second half of all iterations. Dotted black lines indicate \pm one standard deviation. Burns in after 61 iterations (estimate within one standard deviation of estimate).

We want an estimate of our output variables in steady state. So for each model we need to empirically determine the number of *burn-in* iterations. For a deterministic THMC this is fairly straightforward, since the process converges to a single state. The process has burnt in once the values no longer change, i.e. once the process becomes stationary. Empirically we set the *burn-in* period to the maximum number of iterations it takes before the output variables flatline, see figure 8a¹⁹. A stochastic THMC converges to a distribution of states rather than a single state (see example in figure 8b), which make it slightly more difficult to empirically determine burn in. In the processes where the *time average* provides a representative estimate of ψ we first identify the runs that require most iterations to converge – so called extreme cases. This is done by executing a few repetitions from different runs (i.e. different parameter values) of the model and then visually inspecting the trace plots of the output variables. Secondly, we execute several test repetitions with many iterations of these extreme cases. We focus on the extreme cases, since we want to find the number of *burn-in* iterations needed for all runs of the model to reach steady state. For each test repetition we

¹⁹For each repetition we find the first iteration where the value of the output variable is equal to the value at the last iteration. From all the repetitions we select the largest of these iterations. This is then the number of iterations needed for all the repetitions in the model to burn in.

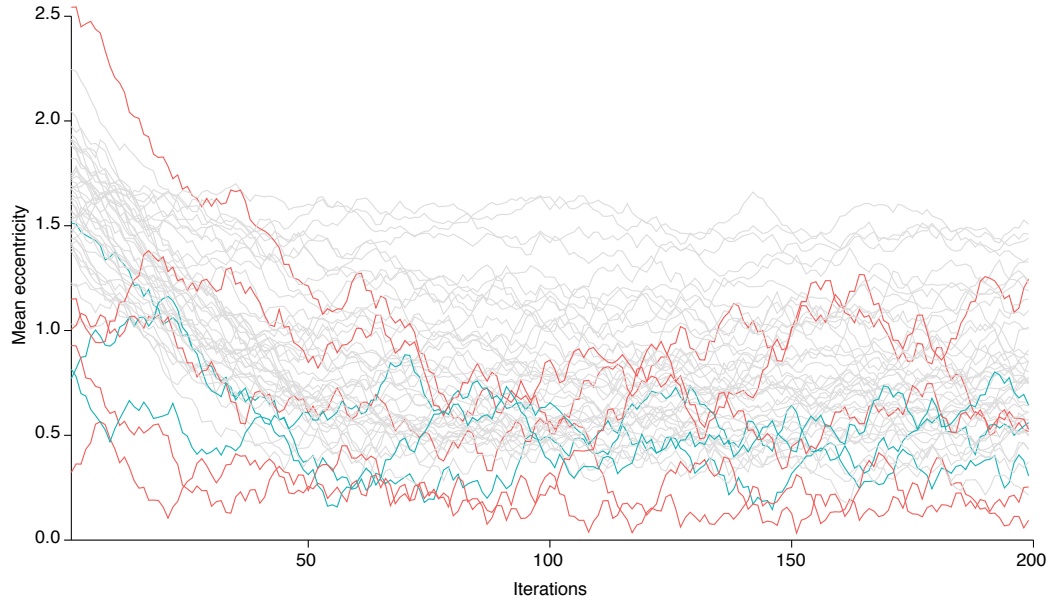
calculate the *time average* and corresponding standard deviation using the second half of all iterations²⁰. And finally, with our estimate of ψ we determine that a particular test repetition has burnt in once the output variable is within one standard deviation of the estimated ψ , see figure 8b. We set the *burn-in* period to the maximum number of iterations it takes for each of the test repetitions to burn in. In the stochastic THMC where the *time average* does not provide a representative estimate of ψ , the first two steps are unchanged; first identify runs that require most iterations to converge, and secondly execute several test repetitions. However since we cannot use the *time average* to calculate a representative estimate of ψ we must rely on a less rigorous method to determine the number of *burn-in* iterations. For every test repetition we visually inspect the trace plots of each of the output variables to determine when the test repetition appears to have reached steady state, see example in figure 9. We set the *burn-in* period to the maximum number of iterations it takes for each of the test repetitions to burn in. In the final execution of all our models we choose to err on the side of caution and set the *burn-in* period slightly higher – often rounding up to the nearest fifty, i.e. 50, 100, 150, etc.

4.2.5. Oscillation in all-maxcov model

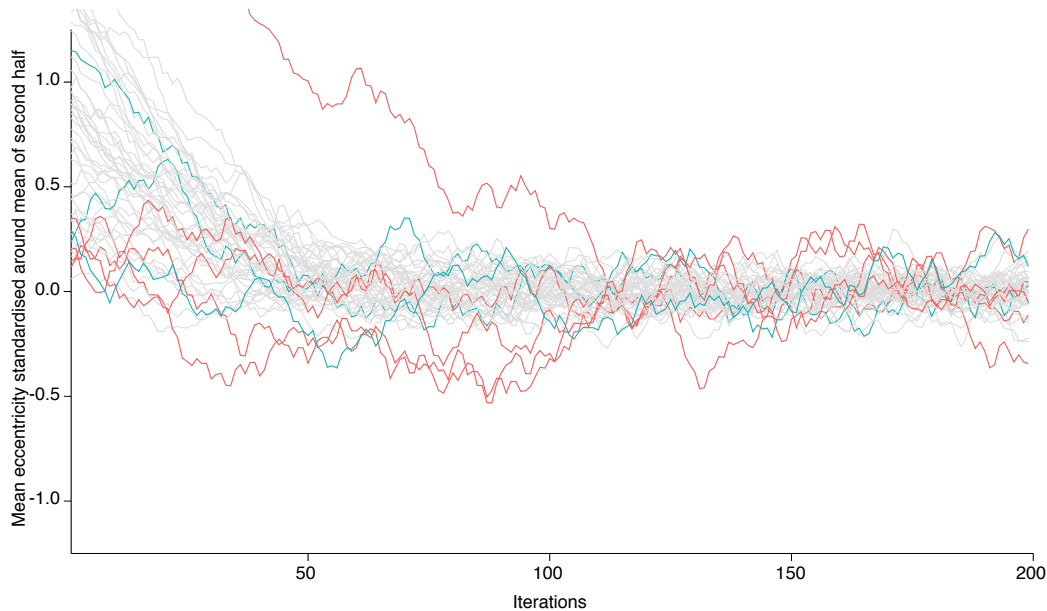
The all-maxcov model, where all firms use the *maxcov* decision rules, is a deterministic THMC. However unlike the other deterministic THMC, this process does not necessarily converge to a single state. For certain parameter values and initial locations of the firm the process oscillates between several states. If the process oscillates between many vastly different states, then calculating the ensemble-average at iteration t is unlikely to provide a representative estimate of ψ . And because the process is *non-ergodic* – the distribution vector, π_∞ depends on the initial distribution vector, π_0 – the time-average will not provide a representative estimates of ψ either. Instead we assume that the process oscillates between a few states where the location of firms only differ slightly, and then calculate the *ensemble average*. We create a new decision rule called *maxcovrnd* with one slight modification: instead of moving 0.1 standard deviation, the firm randomly draws the speed parameter for a uniformly distribution with range $[0, 0.2]$. On average the firm moves 0.1 standard deviation. But more importantly the process now contains a random component, and so it constitutes a *stochastic THMC*, where we know with certainty that the *ensemble average* provides a representative estimate of ψ . When we compare the results of the all-maxcov model to the all-maxcovrnd model we find that the results are minuscule (see Appendix C). This confirms

²⁰The second half procedure is a sufficient although not a necessary condition for the process to burn in (Laver & Sergenti, 2011, chapter 4, p. 73). We use this procedure to get a representative estimate of ψ , when we have yet to determine the actual number of *burn-in* iterations.

Figure 9: Example of multiple test repetitions. To ease the inspection of multiple test repetitions required developing a small interactive tool, see <https://github.com/jsekamane/filter-time>. The tool displays all repetition in the same trace plot, with the option to highlight a specific repetition or all repetitions with a specific combination of parameter values. As well as the option to standardise the series. These figures are derived from the tool.



(a) Multiple test repetitions from different runs (i.e. different parameter values). Red and blue highlights repetitions with respectively 2 and 3 firms.



(b) Mean eccentricity is standardised around the mean of the second half of all iterations. Standardising helps determine burn in, especially when considering different runs with different combination of parameter values. Clearly the repetitions with 2 and 3 firms require most iterations to burn in. This model appears to have reached steady state after approximately 150 iterations.

that the deterministic process does in fact oscillates between a few slightly different states. And so we feel confident that using the ensemble-average in the all-maxcov model provides representative mean estimates of our output variable. We prefer to use the *maxcov*-rule, rather than the *maxcovrnd*-rule, since we are going to extend the model with firms that try to predict the future location of the other firms. Evaluating predictions makes less sense when the location of firms contains a stochastic component.

Recap

In sum we have constructed an experimental design that allows us to evaluate the effect of parameter changes, and insures that results are independent of the initial position of firms. We have identified the appropriate methods to estimate the summary variables. The appropriate method depend on the underlying process, see outline in figure 10. Understanding the underlying process of the run provides us with prior knowledge on how we should go about solving the model. However it still requires some effort and test repetitions to determine the number of *burn-in* iterations and to determine whether the *time average* provides representative estimates. Once this has been determined the final execution of the model is straightforward.

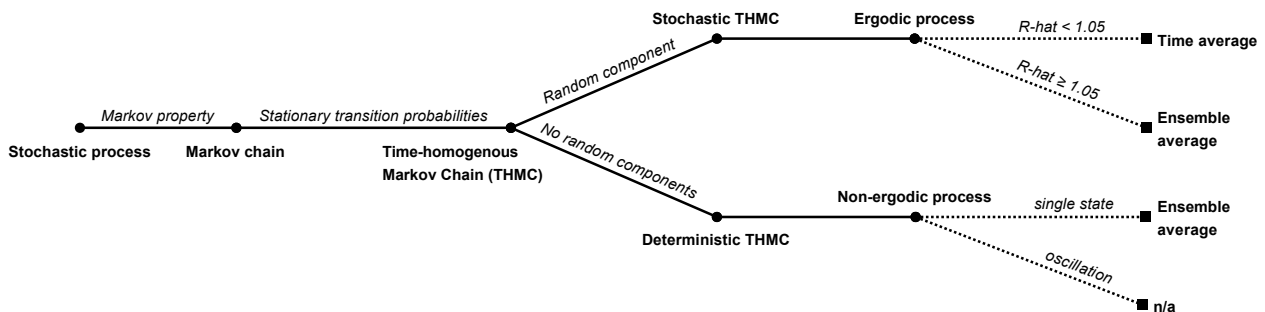


Figure 10: Outline of the various processes and their respective requirements. The respective methods used to estimate ψ are presented furthest to the right.

5. Analysis

We are now ready to investigate the first models. We start with the all-sticker, all-aggregator, all-hunter and all-maxcov models. In each of these models all the firms use the same decision rule. Unlike Fowler & Laver (2008) that run a tournament to determine the best performing decision rules, we want to analyse the underlying dynamics of each decision rule and how these different decision rules affect the location behaviour of firm. By comparing results across the four different models, we see how the location behaviour of the decision rules differ. We start with a symmetric unimodal distribution of consumers. The peak of the

consumer distribution function coincides with the mean ideal point of all consumers. There is no polarisation of the mean ideal points of the two subpopulations and the subpopulations are equally large ($\mu = 0$ and $n_l/n_r = 1$). Thus there is the only one free parameter in each model; the number of firms N . We use the grid sweep method to set this parameter value.

5.1. Baseline model and decision rules

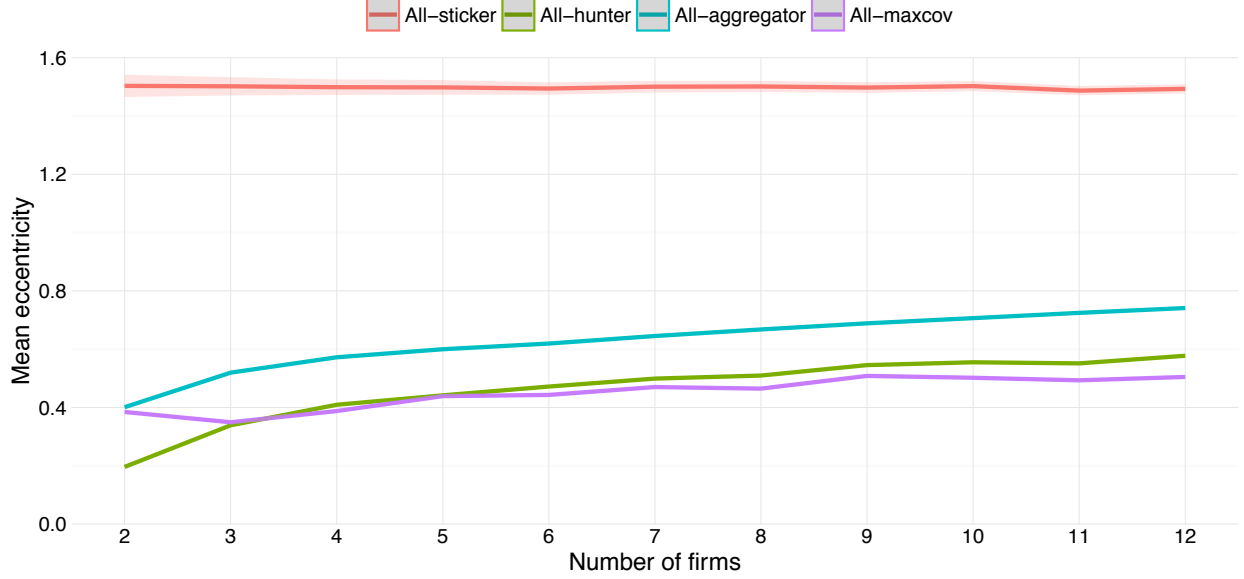


Figure 11: Mean eccentricity for respectively all-sticker, all-aggregator, all-hunter and all-maxcov model where $\mu = 0$ and $n_l/n_r = 1$.

Note: All models use grid sweep method, i.e. each model executes 11 runs (one for each value of N). The all-sticker, all-aggregator and all-maxcov model uses the *ensemble average*, and is executed with 1,000 repetitions, and respectively 1, 50, 99 burn-in iterations. The all-hunter model uses the *time average*, and is executed with 1 repetition and 1,150 iterations (burn-in after 150 iterations). The bands indicate the 95% confidence interval of our estimate.

Figure 11 plots the mean eccentricity for all four models. In the all-sticker model the firms remain at their initial position. They never relocate. As expected we see that the average distance to the population centre is 1.5 standard deviation in the all-sticker model. This result reflects how the initial positions of firms are drawn, i.e. uniformly random from a circle with radius of 3 standard deviations and centre at $(0,0)$. And thus we find that on average a *sticker* firm will be located 1.5 standard deviations away from the centre of the population – irrespective of the number of firms in the market. In the three remaining models mean eccentricity is significantly lower. The initial position of the firms in these models are drawn from exactly the same distribution, but the firms clearly moved towards the population centre.

In the all-aggregator model the firms aim to please their current customer base. The firms choose to locate 0.4-0.8 standard deviations away from the population centre with the distance increasing along with number of firms in the market. In the all-hunter model, where all firms constantly seek higher market shares, the firms choose to locate in closer proximity to the population centre than in the all-aggregator model. *Hunter* firms locate around 0.2 standard deviation closer to the centre than *aggregator* firms.

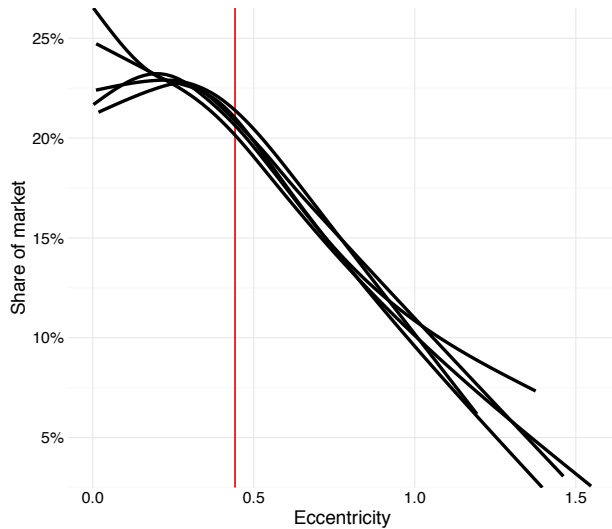


Figure 12: Eccentricity against market share for each firm. Shows the results from a single repetition with 5 *hunter* firms. At each iteration we record the eccentricity and market share. We smoothed the data. The estimated mean eccentricity for five *hunter* firms is 0.44 standard deviations (vertical red line).

Despite the fact that the population centre has the largest density of consumers of any point in the market, firms consistently locate at a distance to this point — even *hunter* firms that constantly seek larger market shares. This suggests that locating too close to the population centre is suboptimal (Laver & Sergenti, 2011, chapter 5). For a *hunter*-firm there is a short-run gain by moving closer to the high density population area. However the response from competing firms quickly erodes this gain, leading to the long-run location that lies at a distance to the population centre. Our estimate of mean eccentricity for five *hunter* firms is 0.44. Compare this to figure 12,

where we for each of the five *hunter* firms plot eccentricity against market share. When the individual firm locates 0.44 standard deviations from the population centre, its share of the market is around 21%. Locations closer to the population centre is associated with a higher share of the market for each individual firm, i.e. a short-run gain. However this gain comes at the expense of one or several of the other firms, since they simultaneously experience a decrease in their share, which forces them to respond. This response erodes the gain and the firm distances itself from the population centre. In the long-run locations close to the peak cannot be sustained, leading to the location that lies at a distance to the peak.

In the all-maxcov model the firms deliberately attempt to maximise their market share. And the firm takes the location of competing firms into account, although it assumes that competing firms maintain their current position. Firms in the all-maxcov and all-hunter model locate at similar distance to the population centre, except in the market with two firms. In this case the *maxcov* firms locate at the same distance as *aggregator* firms. Recall that a *maxcov*-firm locates in the largest Delaunay triangle. When there are two firms in

the market, then each firm chooses among four triangles. All these triangles have a common corner at the location of the competing firm²¹. When the competing firm locates reasonable close to the population centre, then only a small fraction of most densely populated area is contained within each triangle. And, since the *maxcov*-firm locates at the mean ideal point of all consumers within the triangle, this implies that centre of mass of the triangle is shifted slightly away from the population centre. This is why the estimated mean eccentricity is 0.4 standard deviations in the all-maxcov model with two firms. In the all-maxcov model there is only a slight tendency for firms to locate further away from the population centre as the number of firms increase. On average the *maxcov*-firms locate 0.4-0.5 standard deviations away from the population centre.

In the models the actual number of firms in the market is an endogenously determined parameter. Our experimental design allows us to investigate the models when there is anywhere from 2 to 12 firms in the market. To analyse the competitive environment we use a measure called *effective number of firms* (ENP). The ENP takes the relative market shares of the firms into account. The ENP tells how many firms would be in the market if they all had identical shares of the market. If the actual number of firms and the effective number of firms is the same (grey diagonal line in figure 13), then the firms in the market split the market evenly. If the effective number of firms is less than the actual number, then one or several of the firms will have a disproportionate share of the market.

Figure 13 presents the estimated effective number of firms (ENP). The effective number of firms is low in the all-sticker model, starting around 1.5 and going up to 4.5. The initial position of firms gives some firms a clear advantage over the other firms in the market. And since *sticker* firms do not relocate the uneven concentration of customers persists.

In the last three models the market is more evenly split among the firms. This is the result of firms with identical decision rules competing with one another. With four firms in the all-aggregator and all-maxcov model the market is perfectly even split among the firms. With many firms in the market it is increasingly harder to maintain the perfectly even split among the firms. And so we observe that the ENP increases slower than the number of firms in the market increases. When the ENP is below the black diagonal line, then at least one firm has a share of the market that is less than half of one competitor. This is the case in the all-hunter, all-aggregator and all-maxcov model when there is respectively 5, 7 and 10 firms

²¹The edge opposite the common corner in each triangle is the respective edges of the boundary of the space. The *maxcov* decision rule does not consider corner solutions, i.e. locating at the exact same location as competing firms, but assumes that the ideal location lies between the gaps of competing firms. When faced with only one competing firm, then there is no gap and the *maxcov* firm move towards one of four points surround the competitor. Note however that the *maxcov* firm always move towards the side of the competitor with the largest share (i.e. largest half market).

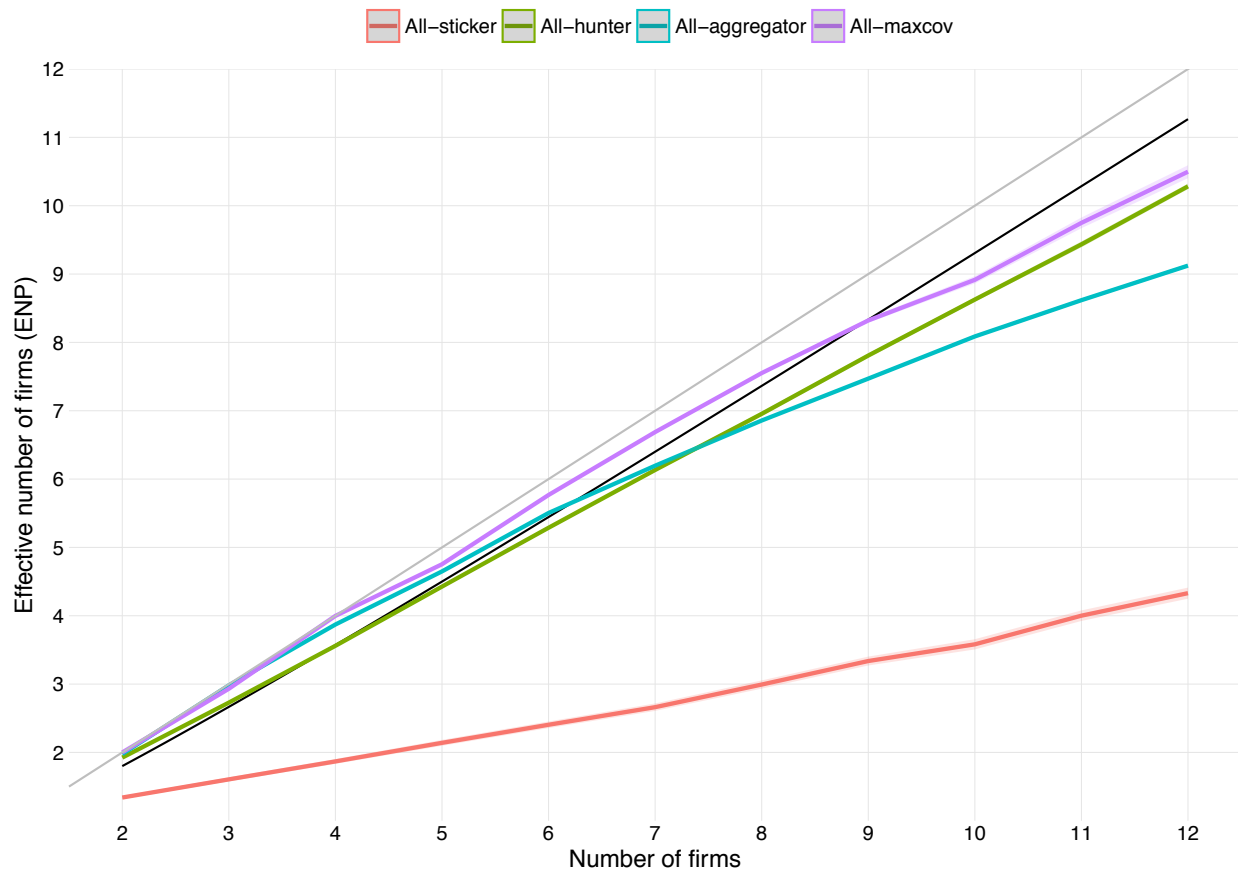


Figure 13: Effective number of firms (ENP) for respectively all-sticker, all-aggregator, all-hunter and all-maxcov model where $\mu = 0$ and $n_l/n_r = 1$. With an ENP below the black diagonal line at least one firm has a share of the market that is less than half the share of a competitor.

in the market. A firm with less than half of its competitor's share, could capture at least half of the competitor's share by pairing with the competitor. With many firms in the market the heuristic decision rules are too simplistic. They miss an important feature, namely pairing once their share falls below a certain threshold. This might also explain why the *KQ-Strat* decision rule performed well in the tournament (Fowler & Laver, 2008). Once the share of a firm using this rule falls below a certain threshold, then the firm uses a simple yet effective rule of thumb that follows directly from one of the necessary equilibrium conditions in Eaton & Lipsey (1975), namely to pair up.

In the all-aggregator model with many firms the effective number of firms is slightly lower than the all-hunter and all-maxcov model. The reason for this is that with many firms the centre of the distribution easily overcrowds. *Aggregator*-firms locate to please their current customer base. And so overcrowding leads the firms to locate on different orbits around the centre. Firms located on the inner orbits attract a larger share of the consumers than the

firms located on the orbit further away from the centre. And the unequal market share of firms reduces the *effective number of firms* (ENP).

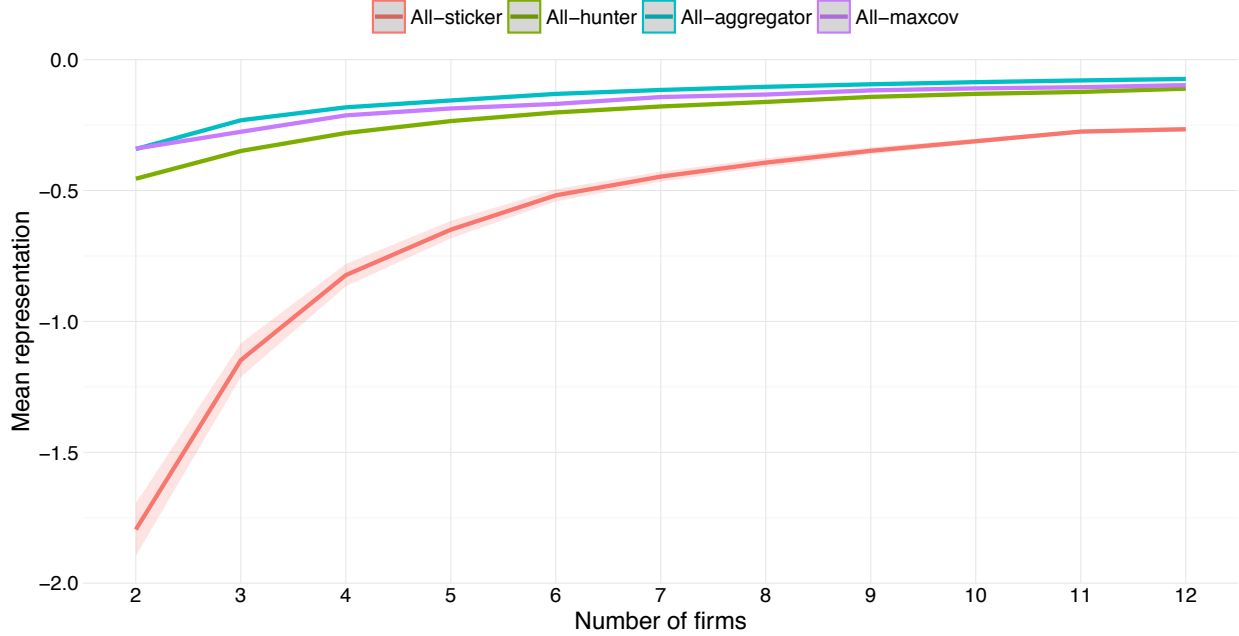


Figure 14: Mean representation for respectively all-sticker, all-aggregator, all-hunter and all-maxcov model where $\mu = 0$ and $n_l/n_r = 1$.

The mean representation measures the satisfaction of consumers, see figure 14. This social welfare measure calculates the average utility of all consumers. The utility of the consumer increases when the distance to the closest firm decreases. From earlier we know that the all-aggregator model will result in the location of firms that constitute a *Centroidal Voronoi Tessellation* (CVT). The CVT minimises the average distance to all customers and therefore maximises our social welfare measure. The all-aggregator model provide a benchmark, since the location of firms is socially optimal.

In both the all-hunter and all-maxcov model where firms aim to maximise their market share, they manage to achieve almost the same mean representation as in the all-aggregator model, where firms actively aim to please their customer base. And this despite *hunter* and *maxcov*-firms locating closer to the population centre than *aggregator*-firms. We also see that as the number of firms in the market increases the mean representation converges towards the social optimal level. Unsurprisingly the all-sticker model scores lowest on our social welfare measure.

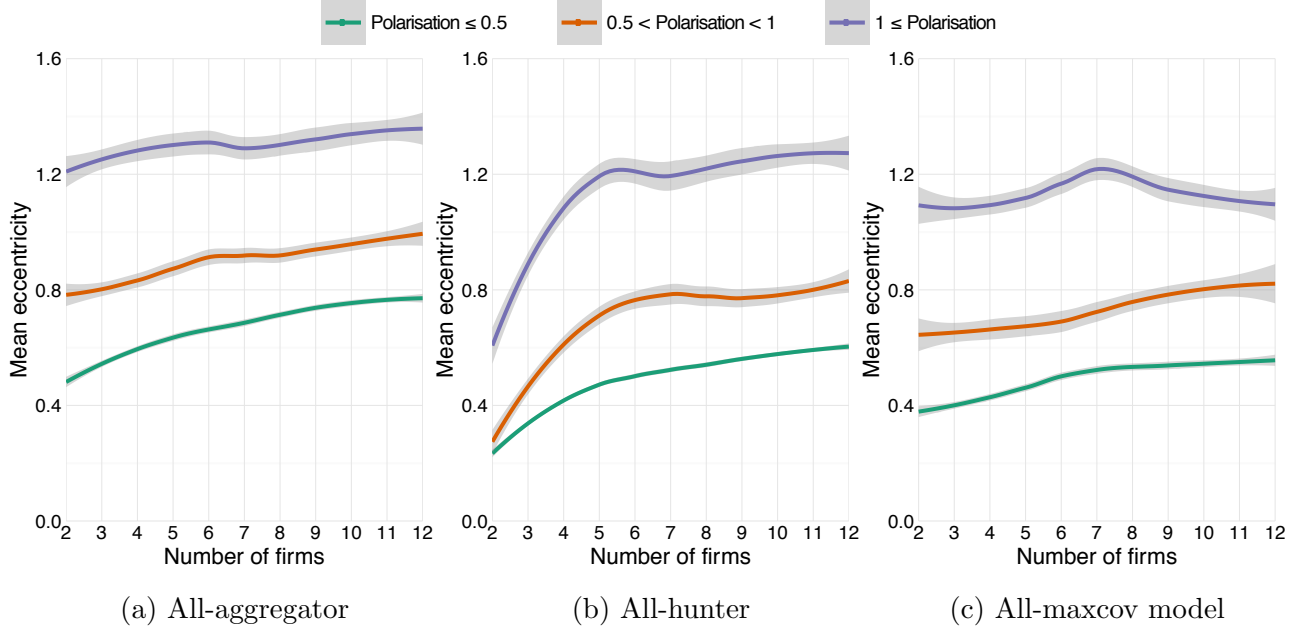
5.2. Models with an asymmetric and bimodal distribution

In the section above we got a sense of the different models and how firms react when using different decision rules. The results have so far assumed a symmetric unimodal distribution of consumers. There was no disagreement between the two subpopulations over the average ideal point along any of the two dimensions. There was a single peak in the aggregated distribution where the density of consumers was greater than any other point, and the peak corresponded to the mean ideal point of all consumers. We observed that firms chose to locate at a distance to mean ideal point of consumers. The literature review showed that different distributions of consumers have had significant impact on the results of previous competitive location models. We now want to investigate how firms locate when there is not a single peak in the distribution? How the location dynamics of firms change when the subpopulations disagree over the mean ideal point? And to what extend our results so far generalise to other distributions, such as the asymmetric and multimodal distributions?

We continue to investigate models where all firms use the same decision rule. Firms are competing against other firms using the same decision rule. Firms using the *sticker*-rule do not relocate, and thus executing the all-sticker model with an asymmetric consumer distribution would not provide further insights. Thus we disregard the *sticker* decision rule in this subsection.

In the following models there are three free parameters; The number of firms in the market N which takes integer values between 2 and 12. The polarisation of the subpopulations μ which can take any value between 0 and 1.5. And finally the relative size of the subpopulation n_l/n_r which takes any value between 1 and 2. To parametrise our models we use the Monte Carlo parameterisation method. For each *run* of the model the parameter values are uniformly random drawn from their respective ranges. We execute many *runs* of the models to map out the entire parameter space. For each *run* we obtain a mean estimate of our summary variables. Instead of presenting each of the 500 estimates, the estimates are smoothed for various degrees of polarisation. This gives three curves that capture the important patterns in the estimates. The first curve summarises the results from the *runs* where there is a low degree of polarisation ($\mu \leq 0.5$). We know from earlier, that this implies, that the distribution of consumers contain a single peak, and that the degree of asymmetry depends on the relative size of the subpopulation. The second curve is for a medium degree of polarisation ($0.5 < \mu < 1$). And the last curve is for a high degree of polarisation ($\mu \geq 1$). In this range we know that the distribution is bimodal and that the peaks of the distribution almost coincide with the mean ideal points of the two subpopulations. In addition, in this range the mean ideal point of all consumers is placed somewhere in the low density area in between the two peaks.

Figure 15: Mean eccentricity where $\mu \in [0, 1.5]$ and $n_l/n_r \in [1, 2]$.



Note: All models use Monte Carlo parameterisation method, the *ensemble average*, and is executed with 500 runs. The all-aggregator model is executed with 100 repetitions and 100 burn-in iterations. The all-hunter model with 100 repetitions and 150 burn-in iterations. The all-maxcov model with 50 repetitions and 150 burn-in iterations. For each of the three degrees of polarisation the respective estimates are smoothed using the *local polynomial regression fitting* (LOESS) method, with the bands indicating the corresponding 95% confidence interval.

Figure 15c presents the mean eccentricity, i.e the mean distance of firms to the population centre. As the polarisation of mean ideal points increases, so does the distance to the population centre. With a bimodal distribution the firms no longer locate close to the population centre. Firms using the *hunter*-rule still locate closer to the mean ideal point of consumers than the firms using the *aggregator*-rule for any degree of polarisation. However the differences in proximity between *hunter*-firms and *aggregator*-firms is much less pronounced in markets with a high degree of polarisation and many firms. Here the mean eccentricity is approximately 1.2-1.4 for both decision rules.

In the highly polarised setting we see significant differences between the market with 4 or more *hunter*-firms and then the market with 2 or 3 *hunter*-firms. For one we see that the distance to the mean ideal point of consumers increases swiftly when going from two to three and from three to four firms. While going beyond four firms in the market has negligible effect on the average distance to the population centre. This happens because firms separate when there is four or more firms in the market. In these cases we typically see that firms split into two crowds. Each crowd locates close one of the peaks of the consumer distribution function. The number of firms in each crowd is roughly proportional to the relative size of

the subpopulation. There is little competition between firms from different crowds, while the firms in each crowd compete fiercely with one another for the customers in the respective subpopulation. In the symmetric all-hunter model with four or more firms in the market, we observed that firms locate 0.4–0.6 standard deviations away from the peak of the distribution (figure 11). Similar behaviour is observed in the models with a high degree of polarisation, however here we have two peaks rather than one. And the firms tend to locate around a peak, rather than in between the two peaks. See an example of this split in a market with five firms in figure 5. Each individual *hunter* firm experiences a short-run loss when it moves away from a peak. This discourages the firm from switching between the crowds. Thus the composition of the crowd is fairly stable, and the firm sees no long-run advantage in locating between subpopulations and close to the mean ideal point of all consumers. On the other hand with two or three *hunter*-firms in the market, the firms tend to move in unity²². Two *hunter* firms tend to locate in between the two subpopulations. Thus they locate close to the average ideal point of all customers, but simultaneously they locate in an area with a low density of consumers. Later we show that this has great impact on the average utility of consumers. With two firms and little to no polarisation, all decision rules show that the firms locate within 0.6 standard deviation of the mean ideal point of all consumers. However only *hunter*-firms continue to locate like this when the polarisation among the subpopulation increases. This effect shows the interplay of decision rule and the distribution of consumers.

Laver & Sergenti (2011, chapter 5) discovered the change when going from 2-3 *hunter* firms and into the realm of four or more *hunter* firms. Yet they seem unaware of an earlier and related discovery by Eaton & Lipsey (1975). Eaton & Lipsey (1975) also analyse asymmetric distributions admittedly in a slightly different setting, namely the bounded one-dimensional space (line market). They show analytically that an equilibrium only exists if the number of firms is less or equal to twice the number of peaks in the density distribution, i.e. $N \leq 2M$ where M is the number of peaks in the consumer density function. They find that when the number of firms is exactly twice the number of peaks then the firms locate in pairs around the quantiles of the distribution. While with fewer firms (than twice the number of peaks) there is some leeway as to how firms locate. Both the pairing of firms in groups of two and the $2M$ limit stems from the single dimensionality of the space they consider. In a one-dimensional line market the only available option is which side of a given point you locate on (left or right). Using this and equilibrium conditions Eaton & Lipsey (1975) show

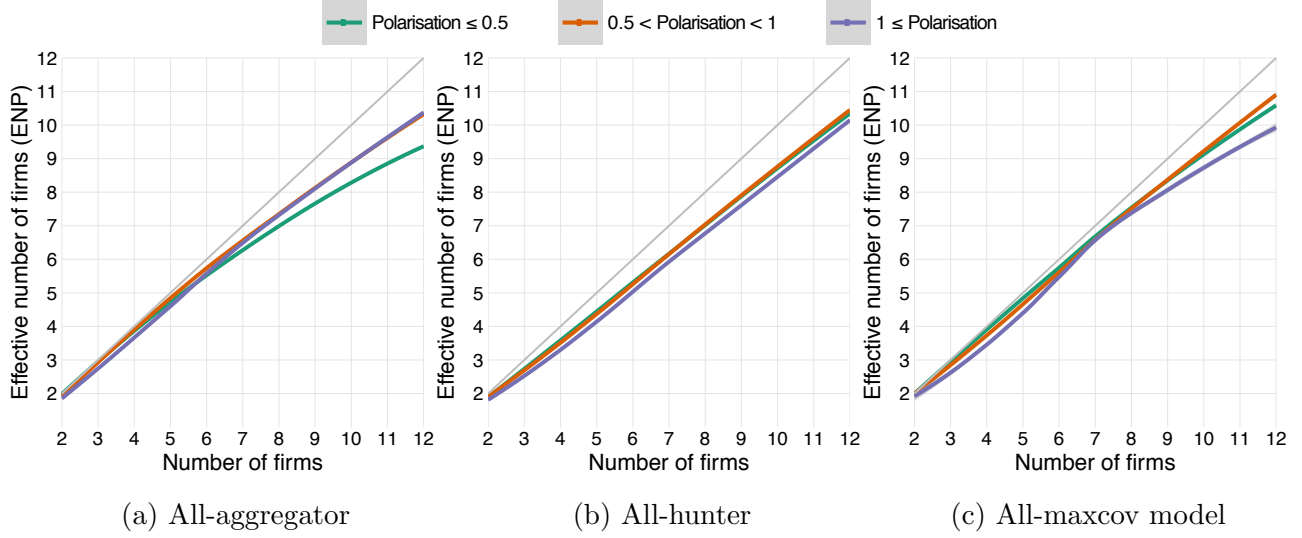
²²Because *hunter* firms cannot leapfrog to the perimeter, but move with constant speed, we don't observe the *dancing equilibrium* described Eaton & Lipsey (1975) in the market with three *hunter*-firms. Instead the three firms move in unity and the location of the unit bounces back and forth between the two peaks of the distribution.

the limit of $N \leq 2M$. In the two-dimensional market firms can locate 360 degrees around a given point, and thus firms need not pair in twos. This is what we observe in the all-hunter model where firms crowd together although not necessarily in pairs of two. Additionally in two-dimensional space the number of peaks does not create an upper limit on the number of firms in the crowd – once again because firms can locate all the way around a given point. But the number of peaks does seem to influence how firms locate. There is a single peak with a low degree of polarisation ($\mu \leq 0.5$), and two peaks with a high degree of polarisation ($\mu \geq 1$). Only when the number of firms is equal or greater than twice the number of peaks do the *hunter* firms form crowds around the peaks of the distributions. When the number of firms is less twice than the number of peaks, then there is some leeway as to how firms locate and thus we see that *hunter* firms locate between the peaks of distribution. I strongly suspect that the consumer distribution with three peaks and five or less *hunter* firms would give some leeway as to how firms locate. With four peaks, seven or less firms would have leeway as to how to locate, etc.

The symmetric distribution showed that *maxcov* and *hunter* firms locate at similar distance to the population centre, except with two firms in the market. We see similar pattern with the asymmetric distribution of customers, figure 11. The *maxcov* and *hunter* firms locate a similar distance to the population centre irrespectively of the polarisation of the subpopulations. The exception is when there are only a few firms in the market. With the *maxcov* decision rule there is no abrupt change when going from 2-3 firms and to 4 or more firms, as is the case with the *hunter* decision rule. Two *maxcov* firms will not agglomerate at the average ideal point of all customers, but instead locate at a distance to the centre. In contrast with the two *hunter* firms locate fairly close to the population centre. For any number of firms and with a high degree of polarisation the *maxcov* firms also split into two crowds that locate around each peak. However the number of firms in each crowd is not necessarily proportional to the number of consumers in each subpopulation.

The asymmetric distribution reiterates the conclusions from the symmetric distribution regarding the effective number of firms (ENP), figure 16. Both the model with *aggregator* and *hunter* firms show that the market is relatively even split among the firms in the market. When there are many firms and a low degree of polarisation then the ENP is slightly lower in the *aggregator* model than the *hunter* model. As mentioned when discussing the symmetric model, with many firms in the market the area around the single peak easily overcrowds, and *aggregator* firms locate at different orbits around the peak. This gave rise to the slightly lower ENP in the *aggregator* model, since firms on inner orbit capture a larger share than those on the outer orbit. When the degree of polarisation increases the density of consumers is dispersed across a larger area, which leads to less overcrowding. Consequently *aggregator*

Figure 16: Effective number of firms (ENP) where $\mu \in [0, 1.5]$ and $n_l/n_r \in [1, 2]$.



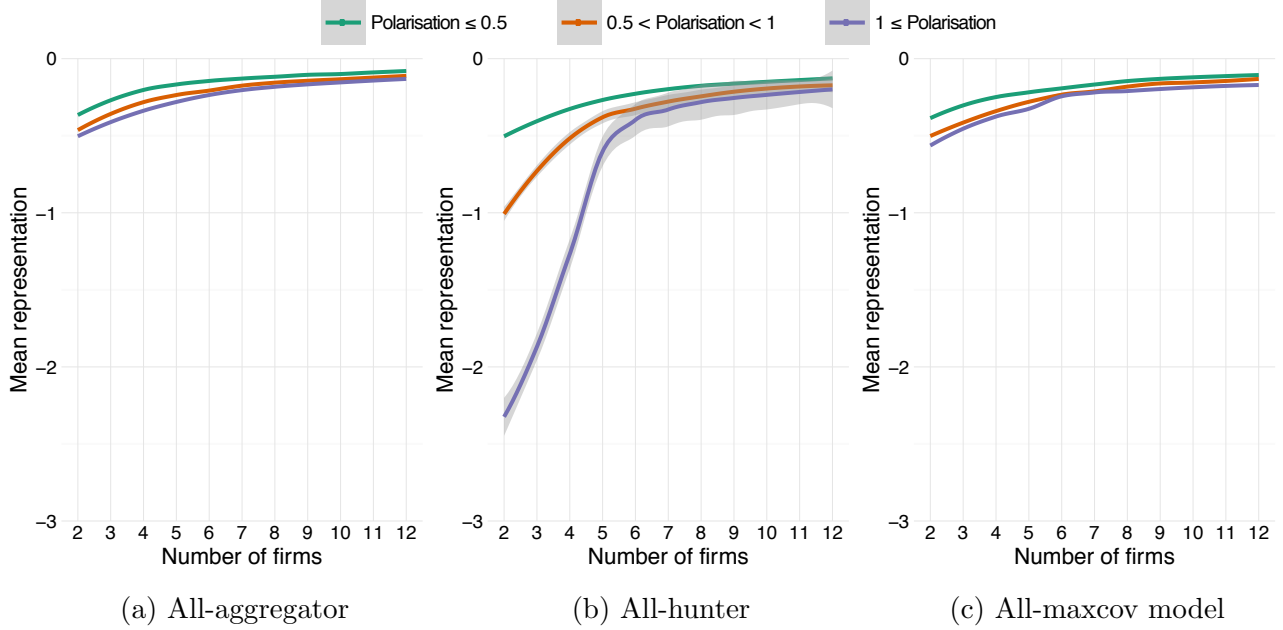
firms are no longer forced to locate at different orbits, which implies that the market is more evenly split among firms. This is why the difference in ENP between *hunter* and *aggregator* model completely disappears as the polarisation of the subpopulations increase.

The market is also relatively even split among the firms in the *maxcov* model. In the market with many firms and a high degree of polarisation the ENP is slightly lower. As earlier noted the firms split into two crowds in the *maxcov* model. However in the *maxcov* model the number of firms in each crowd is not necessarily proportional to the relative size of the subpopulations. Often a few firms manages to capture one-half of the market and this results in the lower ENP. The ENP would be 3.6 if a single firm out of 12 firms captured half the market²³. However with 12 firms in the market we observe an ENP around 10. This tells us that a single firm is unable to capture and maintain half the market by itself in the long-run – it has to be a group of firms. With a low or high degree of polarisation and many firms in the market, the area around peaks tend to overcrowd which forces firms close together and clusters emerge. The firms in the cluster compete for the same consumers which reduces their share of the market. This explains the lower ENP in the model with low and high polarisation. With a medium degree of polarisation the density of consumers is dispersed across a larger area, specifically there is a non-negligible number of consumers with ideal points in between the peaks of the distribution.

Figure 17c presents the mean representation. In the *aggregator* model where the distribu-

²³Calculated using equation 4. With $N=12$ and the market split in two halves: $ENP = \frac{(0.5+0.5)^2}{\left(\frac{0.5}{11}\right)^2 \times 11 + 0.5^2} = 3.6$ when 1 firm has 50% of the market and the 11 other firms share the remaining 50% equally.

Figure 17: Mean representation where $\mu \in [0, 1.5]$ and $n_l/n_r \in [1, 2]$.



tion of consumers is asymmetric, the location of firms still constitute a CVT. And thus this model maximises our social welfare measure – the mean representation. Once again we can use the *aggregator* model as a benchmark. The difference between the mean representation in the *aggregator* model and the *maxcov* model is minuscule.

In the *hunter* model we see that the bimodal distribution of consumers give rise to significant changes when going from 2-3 firms and to 4 or more firms. This is particularly pronounced in the models with a high degree of polarisation. With two or three *hunter* firms the mean representation is around -2, while with four or more firms the mean representation is between -0.5 and -0.2. As earlier noted this is the result of *hunter* firms locating in between the centres of the subpopulations when there are 2-3 firms in the market. Although the firms locate close to the average ideal point of all customers, they also locate in an area with a low density of consumers. Because firms locate far from the densely populated areas it significantly reduces the average utility of customers. On the other hand firms separating into crowds that locate around the centres of the subpopulation, when there are four or more firms in the market. [Laver & Sergenti \(2011, chapter 5, p. 102\)](#) refer to this as the “*sea change*” in mean representation when reaching four or more firms. The effect is a result of a bimodal distribution of consumers.

We are now at the point where we have an understanding of the baseline decision rules; *sticker*, *aggregator*, *hunter* and *maxcov*. The first three rules rely on heuristics, while the last rule deliberately and directly tries to maximises the market share of the firm. We know

how firms, that use these rules, choose to locate – both when the distribution of consumers is symmetric and when it is asymmetric and bimodal. We know which type of competitive environment arises from the location behaviour of the firms. And we know how this behaviour affects the overall welfare of consumers. The following section looks at how firms locate when they take the predicted location of competing firms into considerations.

5.3. *Decision rules with foresight*

When a firm chooses its location, the location of competing firms is unknown. The firm may try to predict the location of competing firms. However if multiple firms use this approach, then the location outcome that each firm is trying to predict will depend on predictions that the firm and the other firms form. As [Arthur \(2014, p. 175\)](#) writes “*predictions are forming a world those predictions are trying to forecast*”. This self-referential loop leads to logical indeterminacy and the maximisation problem of the firm is ill defined and cannot be solved by means of *deductive reasoning*.

The two decision rules discussed in this section rely on *inductive reasoning*. The firm holds several hypotheses and uses these to make predictions on how competing firms will locate. A hypothesis consists of a proposition that might not hold true and so contrary evidence weakens the hypothesis. The firm tests its hypotheses by comparing the predicted location of competing firms to the observed locations. Thereby the firm learns which hypotheses are plausible and the firm is more likely to base its predictions on these. Predictions are made using the hypothesis that worked best in the past. The firm locates – like the *maxcov* firm – it tries to maximise its market share, but uses the predicted location of all competing firms rather than their current location. In the first decision rule the firm is endowed with a fixed set of hypotheses. These hypotheses are exogenously given and do not change over time. Only the accuracy of each hypothesis changes in pace with its predictions being tested. I refer to this decision rule as *maxcov-inductor*, since the firm uses *inductive reasoning*. The second decision rule is an expansion of the *maxcov-inductor* rule, but the firm gradually discards poorly performing hypotheses and forms new hypotheses. If possible new hypotheses should perform at par or better than the existing hypotheses. Therefore new hypotheses are formed through an evolutionary process that mutates and fuses elements from the most fit of the existing hypotheses. Replacing poorly performing hypothesis is another way in which learning takes place – leading firms to make better predictions. Hypotheses are endogenous in this decision rule. I refer to this rule as the *maxcov-inductor-GA*, since a genetic algorithm generates the new hypotheses. I use two decision rules such that I can separate the effects from respectively *inductive reasoning* and the endogenously determined hypotheses. The two decision rules allow the effects to be analysed in turn.

The following method is a modified version of the method developed for the *Santa Fe Institute Artificial Stock Market Model* and described by Arthur (2014, chapter 3) and Arthur, Holland, LeBaron, Palmer & Tayler (1996). In the stock market model multiple agents try to predict the stock price. Each agent faces a single unknown outcome. The agent’s demand for shares depends on the agent’s predicted stock price, i.e. the agent’s action depends on a single prediction. And the actual stock price rely upon the aggregated demand of all agents. In this paper an agent or a firm attempts to predict the future location of all other firms. Each firm faces $N - 1$ unknown outcomes. The firm locates based on its predicted location of competing firms, i.e. the action of the firm depends on multiple predictions. Thus the most significant modification is going from a setting with many-to-one predictions to a setting with many-to-many predictions. Additionally the stock price is one-dimensional – it can go up or down. Whereas the position of the firm is two-dimensional. The firm can relocate in any 360 degree direction. This however only requires a slight modification to the forecasting model used in the method.

5.3.1. *Maxcov-inductor rule*

A firm with the *maxcov-inductor* decision rule maintains several hypotheses on how competing firms locate. The firm uses the hypothesis that fits the current state and that previously proved most accurate to forecast the future location of a competing firm. When the firm chooses its own location it relies on the predicted location of all competing firms.

The firm is endowed with M number of hypotheses. While each hypothesis might only be relevant to a narrow set of situations, together the array of hypotheses cover a wide range of different situations. At every iteration the firm only considers the hypotheses specific to the current state and ignores the remaining hypotheses. This makes the firm capable of recognising different location patterns and applying the appropriate forecast.

Each hypothesis consists of two parts jointly forming a *condition/forecast* rule. The condition part specifies which situations trigger the forecast. And the forecast part contains the specific estimate used to make a prediction about the future location.

To describe the current state of a firm, we use a 13-bit descriptor. The descriptor J_j summarises the location behaviour of firm j , e.g. the fourth bit in J_j relays whether or not *firm j is more than 0.6 standard deviations away from the mean ideal point of all consumers*. The ninth bit in J_j relays whether or not *the position of firm j along the agreement dimension (y -axis) is above the average of the last 16 periods*. Each of these bits take the value 1, if the state occurred, and takes the value 0, if the state is absent. The current state of firm j could for instance be summarised by the following descriptor: 1110010001010. We refer to the first 5 bits as the *fundamental bits*. They relay whether or not the distance from the firm to the mean ideal point of all consumers is greater than respectively 0.1, 0.25, 0.4, 0.6,

or 1.2 standard deviations. These bits measure the degree to which the location of the firm is fundamentally different from the ideal point of all consumers. Bits 6-9 are the *tendency bits*. The bits 6-7 relay whether or not the position of the firm along the disagreement dimension (x-axis) is above the average of the last respectively 4 and 16 periods. And the bits 8-9 relay whether or not the position of the firm along the agreement dimension (y-axis) is above the average of the last respectively 4 and 16 periods. Thus these bits measure trends in the relocation pattern of the firm. Bits 10-11 are the *oscillation bits*. These bits use information about the position of the firm in the last four periods to determine whether or not there is oscillation along respectively the disagreement dimension (x-axis) and the agreement dimension (y-axis). More specifically it calculates the autocorrelation between consecutive periods and determines that oscillation takes place if the estimated first period lag operator is below the lower bound of the 85% confidence interval²⁴. The last two bits are respectively always on and always off. These are experimental controls. By construction they contain no information about the current state, and thus they allow us to analyse to what degree firms act on useless information.

Each *condition/forecast* rule attempts to recognise the current state. Therefore the condition consists of 13 corresponding positions each taking the value 1, 0, or #. The condition is met if all 1 and 0 match the current state descriptor. The # is a wildcard character that matches either case, e.g. the condition **###1####0####** is satisfied, if the state described by the fourth bit occurred, and the state described by the ninth bit did not occur. In other words the condition will match any state, where *firm j is more than 0.6 standard deviations away from the mean ideal point of all consumers, while its position along the agreement dimension (y-axis) is not above the average of the last 16 periods.* The condition **###1####0####** is not fulfilled if the current state descriptor is 1110010001010, but the condition is satisfied, if the current state is 1111010001010. More 1 and 0 values in the *condition/forecast* rule imply that the hypothesis is more specific, while a *condition/forecast* rule with many # will match more states and thus represents a hypothesis that is more general.

All the *condition/forecast* rules that matches the current state of firm j , are said to be active. Among these active *condition/forecast* rules, the rule with the best accuracy is used to forecast the future location of firm j . In the case where several rules tie for the best accuracy, one of the rules is selected randomly and used to forecast. The accuracy of all active *condition/forecast* rules is updated once all the firms relocate and the actual location of each competing firm is revealed. The forecast error is the distance between the actual

²⁴With four periods the lower bound of the 85% confidence interval is $\underline{ci} = \frac{-1-z\sqrt{N-k-1}}{N-k} \simeq -0.8151$, where the sample size is $N = 4$, the lag is $k = 1$, and the test statistic is $z = 1.022$ (Meko, 2015; Chatfield, 2016).

location of the firm and predicted location of the firm²⁵. The accuracy measurement is based on the forecast error variance – a lower forecast error variance imply better accuracy. The accuracy is updated using the inverse of the moving average of squared forecast errors (details in [Appendix D](#)). Over time the firm learns which hypothesis work well in a given situation. Thus the continuous updating of the accuracy of the *condition/forecast* rules facilitates learning. We will refer to this as *learning through experience*²⁶.

Using a *neural network* is an alternative to the *condition/forecast* approach by which firms could form predictions and learn from observations. The precision of forecasts generated by a *neural network* will be just as good, if not better, however the process by which a *neural network* generates predictions is a “black box” ([Arthur, 2014](#), chapter 3), i.e. we would not be able to see what type of information gave rise to specific predictions. Unlike the *condition/forecast* approach, where we can analyse the 13 positions in the condition to see if the *fundamental bits*, *trending bits* or *oscillation bits* are most frequently fulfilled and when the firm acts on useless information.

The firm forecasts the future location of the competing firm j using a linear forecasting model:

$$\begin{pmatrix} x_{t+1,j} \\ y_{t+1,j} \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} + \begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix} \begin{pmatrix} x_{t,j} \\ y_{t,j} \end{pmatrix}$$

The six parameters of this model come from the most accurate active *condition/forecast* rule and take the form $(C_1 \ C_2 \ A_1 \ B_1 \ A_2 \ B_2)$, e.g. the full *condition/forecast* rule might look like `###1#####` / $(0.001 \ 0 \ 1.01 \ 0 \ 0 \ 0.995)$. The rule in this example states that if *firm j is more than 0.6 standard deviations away from the mean ideal point of all consumers, then the predicted location along the x-axis is 1% further right, and along the y-axis 0.5% less north relative to the current position, and then shifted an extra 0.001 standard deviation right along the x-axis.*

The *maxcov-inductor* firm makes predictions on the future location of all competing firms. Each competing firm j has its own unique current state descriptor J_j . But the *maxcov-inductor* firm uses the same set of M hypotheses for all competing firms. Thus we make the assumption that the *condition/forecast* rules are not specific to the location behaviour of a particular competing firm, but generally applicable to any competing firms that exhibit a

²⁵Instead of basing the forecast error on location, the forecast error could be based on the difference between predicted profit (or market share) and the actual profit. While the latter approach is appealing because it aligns with the objective of the firm, we opt for the former approach in this paper. Most likely the precision of hypotheses improves fastest when using the location forecast errors, meaning that less computational resources are needed to generate results.

²⁶*Learning through experience* is related to the concepts of *learning by using* ([Rosenberg, 1982](#)) and *learning by doing*, although not the deductive part ([Arrow, 1971](#)).

particular location behaviour²⁷.

The *maxcov-inductor* firm is endowed a set of one hundred hypotheses (i.e. $M = 100$). All but one hypothesis is randomly generated by the following procedure. Each position in the condition is randomly set to 1 or 0 both with probability 0.1, or set to # with probability 0.8. The forecasting parameters are drawn uniformly random. The forecast parameters C_1 and C_2 are drawn from a uniform distribution with range $[-0.005 \ 0.005]$, i.e. with a mean value of 0. The parameters A_1 and B_2 are drawn from the range $[0.99 \ 1.01]$. And A_2 and B_1 are drawn from $[-0.01 \ 0.01]$. The initial accuracy or forecast error variance of each *condition/forecast* rules is set to zero. The last hypothesis is the default or fallback option in case none of the other hypotheses match the current state. This *condition/forecast* rule only consists of # so it matches any current state and insures that the firm is always able to make a prediction. Each of the forecast parameters for this special rule is set to the average parameter values of the other $M - 1$ *condition/forecast* rules. Because hypotheses are randomly drawn for each *maxcov-inductor* firm this implies heterogeneity in their decision process. *Maxcov-inductor* firms are heterogeneous in terms of their expectation model and knowledge, and not just in terms of the location they choose.

5.3.2. *Maxcov-inductor-GA* rule

A firm with the *maxcov-inductor-GA* decision rule behave as described above. The firm is still endowed with M hypotheses that are randomly generated. But for every τ number of iterations the firm replaces the 20% least fit hypotheses. The fitness of the hypothesis is based on its accuracy and the specificity of the *condition/forecast* rule. The specificity of a *condition/forecast* rule is measured as the number of 0 and 1 in the condition. Accounting for the specificity implies that rules with wider applicability (i.e. many # in the condition) are evaluated as more fit. Fit hypotheses are thus less likely to be discarded and more likely to form the basis of the new hypothesis. This implies that the model has a slight drift towards more general *condition/forecast* rules. The new hypotheses are created using a *genetic algorithm* (GA). This algorithm mimics an evolutionary process, i.e. hypotheses are developed from earlier hypotheses with randomly occurring mutations and by crossbreeding “parent” hypotheses. The genetic algorithm uses either *mutation* or *crossover* to create a new hypothesis. [Appendix D.1](#) details the values, equations and the specific probabilities

²⁷Alternatively one could choose to model the firm such that it only makes predictions on the future location of neighbouring competing firms. The neighbouring competing firms will most likely have the greatest impact on the market share of the firm. Each competing firm j , which the firm currently shares a market boundary with, would have a current state descriptor J_j . The firm would assume that all other competing firms remained at their current location. Local predictions such as this would free up computational resources which instead could be used to allow each firm to hold hypothesis specific to a particular competing firm.

used. Each new hypothesis requires two parent hypotheses. The pair of parent hypotheses are drawn randomly and with equal probability from the set of hypotheses not discarded, i.e. from the 80% most fit hypotheses.

With mutation the new hypothesis only inherits traits from the most fit parent. The method mutates the condition inherited from the parent by randomly flipping the 1, 0 and #, and by randomly replacing or altering the forecast parameters.

With crossover the new *condition/forecast* rule is a mix of both parents. The condition part is mixed by randomly selecting a donor parent for each of the 13 positions, e.g. the value of the first position might come from one parent, and three following positions might come from the other parent, etc. This way the 13 positions are passed on from one of the two parents, and for each position the donor parent is randomly selected. Crossover of the forecast parameter values happens by either 1) component-wise crossover of each value, 2) using the weighted average of the parents or 3) randomly picking a parent that passes on all parameter values. The method used is randomly selected, and the three methods have equal probability of being selected. In the component-wise crossover the 6 parameter values is mixed by randomly selecting a donor parent for each of the values. The weighted average of the parents parameter values uses the accuracy as weights.

To insure that the new hypotheses have a reasonable chance of being used, each new *condition/forecast* rule inherits the average accuracy of its parents. In the case where the parent rules have never matched a state – and thus never been active – the new hypothesis takes the median accuracy of all the non-discarded hypotheses. The firm always maintains the default *condition/forecast* rule that only consists of #. However its parameter values are updated such that these equal the weighted average of all new and non-discarded rules, where the accuracy is used as weight.

The process of discarding the poorly performing hypotheses and forming new hypotheses based on the fittest is another way in which the firm learns. The firm learns to make better predictions by gradually refining its hypotheses. We refer to this as *learning through adaptation*. *Maxcov-inductor-GA* firms are heterogeneous, since their initial hypotheses are randomly drawn and because new hypotheses are formed based on the unique experiences of the firm.

5.4. Models with foresight

In the following models the free parameters are the number of firms in the market, the polarisation of mean ideal points and the relative size of the subpopulations. To parametrise our models we use the Monte Carlo parameterisation method.

A *run* of the model still constitutes a stochastic process. However the process does not fulfil the Markov property, because of the *tendency bits* and *oscillation bits* in the current

state descriptor J_j . These two sets of bits calculate respectively the moving average and the autocorrelation based on the recent location history of the firm. This implies that the future state of the process depends on past states of the process, which violates the Markov property. Additionally the process does not have *stationary transition probabilities* — even if we excluded the *tendency* and *oscillation bits* — and so does not constitute a *time-homogenous Markov chain*. The firm may act differently every time it faces the same situation, because of the experience it has accumulated in the meantime. Over time the firm continuously updates the accuracy of its hypotheses, and so the *transition probability* will depend on the experience of the firm. In any model with a *maxcov-inductor* or *maxcov-inductor-GA* firm the underlying process of a *run* is stochastic, but we do not know much else about the process. We have no prior knowledge about convergence or steady states of the process. The literature does not yet provide answers on how to obtain representative estimates of our output variables when the process does not constitute a *time-homogenous Markov chain*. The tournament model by Fowler & Laver (2008) and the *Santa Fe Artificial Stock Market* by Arthur (2014, chapter 3) face similar issue. In both papers they nonetheless run several *repetitions* of the model for an extended number of *iterations*, discard *burn-in* iterations to remove the effects arising purely from the arbitrarily initial setup of the model, and estimate the output variables using the *ensemble average*, i.e. taking the average over all *repetitions*. This will also be the approach of this paper. We run several test repetitions of the most extreme cases, and visually inspect the trace plots for each output variable to determine when the process appears to have burnt off the effects of the initial position of firms and the initial learning by firms. The estimates of the output variables are calculated using the *ensemble average* once the process has burnt in. Each run produces an estimate. The estimates are smoothed for respectively low, medium and high degrees of polarisation. In the following figures for mean eccentricity, ENP and mean representation each dot represents the estimate from a *run*.

5.4.1. *Maxcov-inductor model*

We start by comparing the results from the *maxcov-inductor* model with the results from the *maxcov* model to get a sense of how firms with *inductive reasoning* and an endogenous set of hypothesis form predictions about competing firms. And how the interaction of such firms influence the overall market and location behaviour of firms. Through out this and the next subsection various examples are used to highlight the inner working of the decision rules.

The *maxcov-inductor* and *maxcov* firms locate at similar distances to the ideal point of all consumers, for all degrees of polarisation between the subpopulation. With a high degree of polarisation, mean eccentricity increases as the number of firms in the market increase, see figure 18. Mean eccentricity tops with around seven firms in the market, and thereafter the

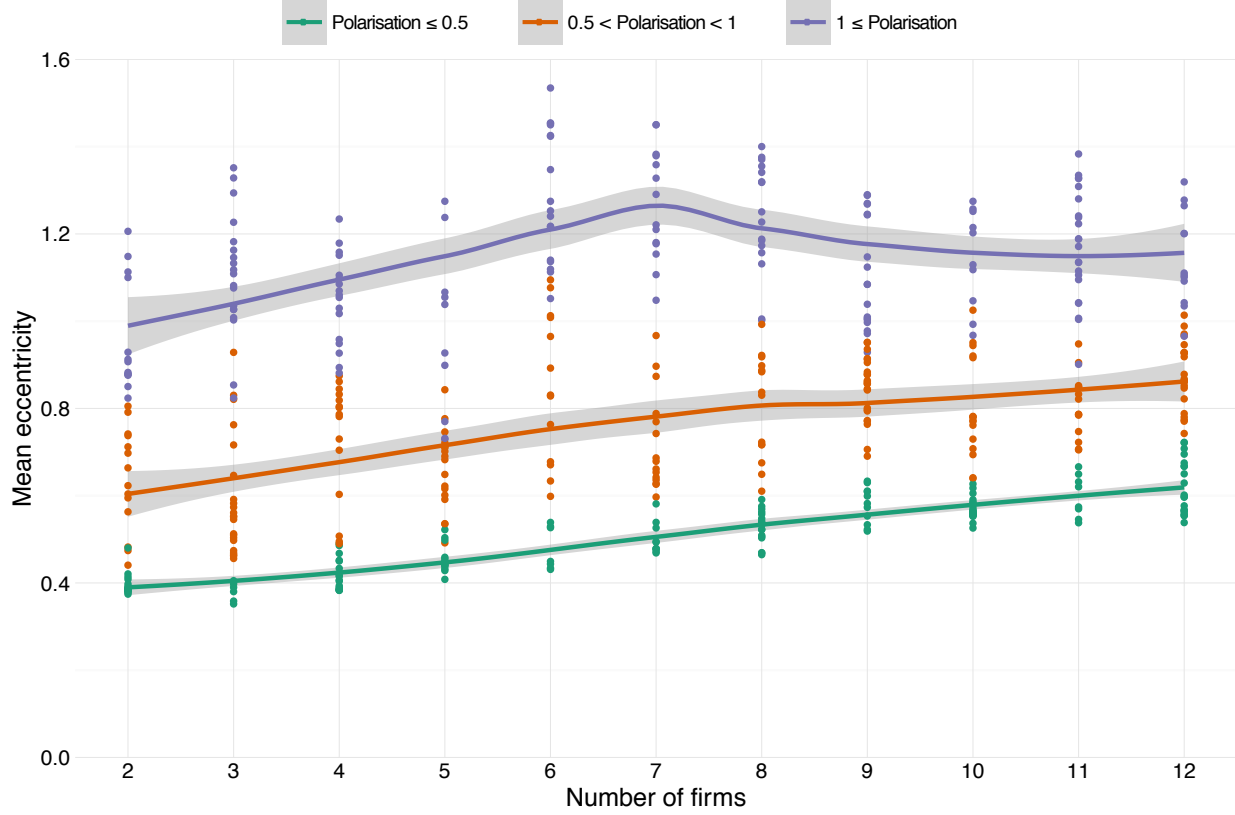


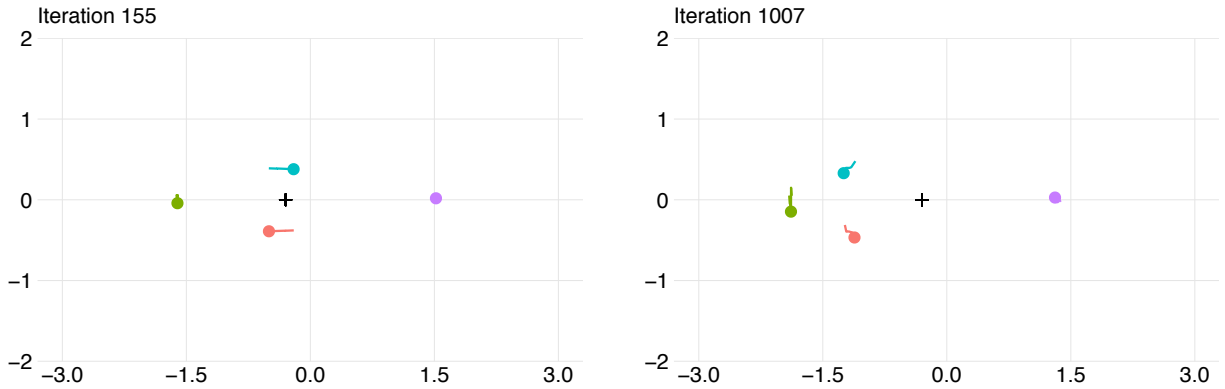
Figure 18: Mean eccentricity for *maxcov-inductor* model.

Note: Uses Monte Carlo parameterisation method, the *ensemble average*, and is executed with 500 runs, 50 repetitions and 1000 burn-in iterations. Each dot represents the estimate from a *run*. Estimates are smoothed using a LOESS method.

mean distance of firms to the average ideal point of consumers remains at approximately 1.2 standard deviations. Although this was also observable in the *maxcov* model (figure 15c), it is slightly clearer in the *maxcov-inductor* model. Recall that *maxcov* firms tend to split into two crowds when the consumer distribution is bimodal. Each crowd locates close to the peak of a subpopulation, and thus far from the average ideal point of all consumers. Consequently the mean distance to the ideal point increases as the number of firms increase. The firms in each crowd compete with one another and compete less with firms in the other crowd. With less than eight firms in the market firms are unlikely to switch its location from one crowd to the other. As the number of firms increase beyond seven firms the probability that a firm switches crowds increases. The decrease in mean eccentricity is due to firms moving between the peaks of the subpopulation, and thus temporarily locating closer to the average ideal points of all consumers. This also seems to indicate a limit as to how many firms the crowds can sustain, before firms from different crowds start to compete more directly with one another. When firms use the *hunter* rule there is no such limit, for $N \leq 12$ (figure 15b).

Although estimated mean eccentricity is quite similar in the *maxcov* and *maxcov-inductor* model, there are slight differences in how the firms choose to locate. This is perhaps best illustrated by looking at an example with four firms in a highly polarised market ($\mu = 1.5$). When subpopulations are equally large ($n_l/n_r = 1$) then two peripheral firms locate close to the peaks of each subpopulation, while the two centre firms oscillate left and right in opposite direction of each other (example of such oscillation in figure 19a). When the left subpopulation is twice as large as the right, the right peripheral firm locates close the peak of its subpopulation, while the three remaining firms locate close the peak of the left subpopulation in a triangular configuration (example in figure 19b). And so, increasing the relative size of the left subpopulation shifts the locations of the centre firms further left. At some critical relative size the location of the centre firms comes sufficiently close to the peak of the peak of the left subpopulation, and the centre firms switch from oscillation left-right and into the triangular configuration. These location behaviours are observed in both the *maxcov* and *maxcov-inductor* model. However peripheral *maxcov-inductor* firms observe and predict the oscillation of the centre firms, and in response move closer to the centre. While centre firms shift further left to counterbalance the invasive peripheral firms, increasing the likelihood of the triangular configuration. This implies that the *maxcov* model requires more uneven subpopulations, before the switch from oscillation to triangular configuration takes place, compared to the *maxcov-inductor* model. In figure 19, for each of the two models, we plot the trajectories of four firms with $\mu = 1.5$ and $n_l/n_r = 1.5$. When the left subpopulation

Figure 19: Trajectory plots for four firms in market with highly polarised subpopulation ($\mu = 1.5$) and where left subpopulation is half as large as right subpopulations ($n_l/n_r = 1.5$). The black cross indicates the mean ideal point of all consumers.



(a) Location of four *maxcov*-firms after 155 iterations, where the lines show the location in the four preceding iterations. The two centre firms oscillate in opposite direction of each other.

(b) Location of four *maxcov-inductor* firms after 1007 iterations, where the lines show the location in the six preceding iterations. The three firms on the left form a triangular configuration.

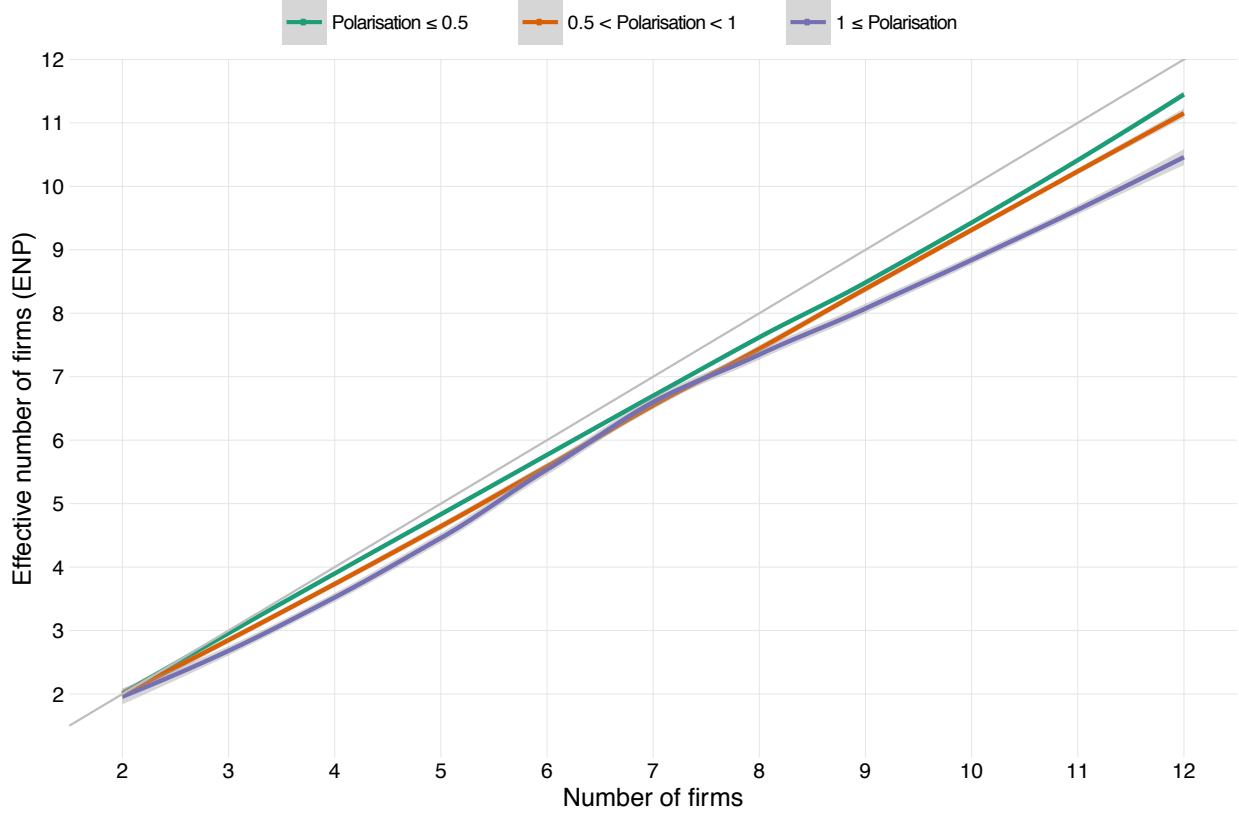


Figure 20: Effective number of firms (ENP) for *maxcov-inductor* model.

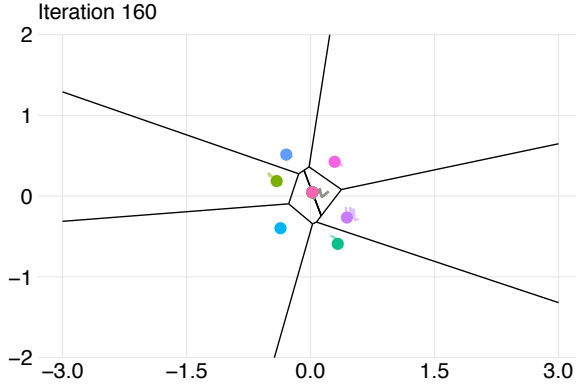
is half as large as the right, the centre firms in the *maxcov* model still exhibits oscillation, while three *maxcov-inductor* firms locate in the triangle configuration, with the right peripheral firm located close to the peak of the right subpopulation. We are observing how predictions lead to strategic considerations that ultimately influence the location of firms. Similar type of behaviour is also observed when the number of firms is not four. However these case are harder to describe, since there are not just two different location configurations, but often many different configurations. Predictions do not lead to new location configurations, but they affect the circumstance that lead to specific configuration.

Figure 20 presents the *effective number of firms* (ENP). Comparing the ENP in the *maxcov-inductor* model to the *maxcov* model (figure 16c) reveals similar ENP with less than eight firms in the market or with a medium degree of polarisation. However with more than seven firms in the market and with a low or high degree of polarisation the ENP is higher in the *maxcov-inductor* model. In these cases the fraction of consumers that each firm captures is more evenly distributed than in the *maxcov* model.

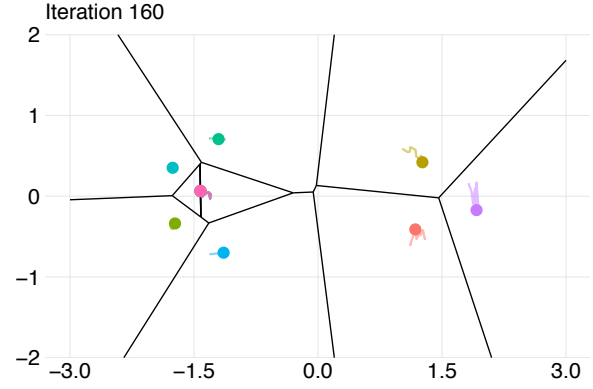
In the *maxcov* model with many firms and a low degree of polarisation, a handful of firms will cluster together at the same location, as seen in figure 21a. The clustered firms move in tandem. Each of the clustered firms assume that the other firms remain at their

current position, and given this relocates to maximise its market share. However all the other firms in the cluster make the same decision, and they all end up at the same location. Once a cluster of firms is formed, it sometimes absorbs additional firms, but seldomly breaks apart. Similarly with a high degree of polarisation and uneven sized subpopulations, see figure 21b. This reduces the ENP in the *maxcov* model with many firms, since the clustered firms compete for the same customers reducing their average market share. The clustering of *maxcov* firms in the models with low or high degrees of polarisation occurs, because there is

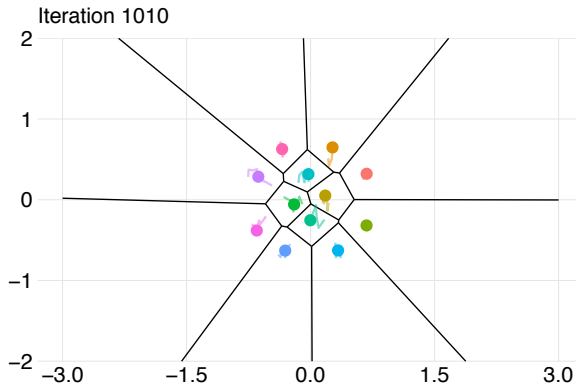
Figure 21: Trajectory plots for twelve firms. The top panels show the *maxcov* model, where dots indicate the location of the firm after 160 iterations. The bottom panels show the *maxcov-inductor* model, where dots indicate the location after 1010 iterations. Panels on the left show market with no polarisation ($\mu = 0, n_l/n_r = 1$). Panels on the right show market with high polarisation and uneven sized subpopulations ($\mu = 1.5, n_l/n_r = 2$). The coloured lines show the location in the ten preceding iterations. The black lines are market boundaries after respectively 160 and 1010 iterations.



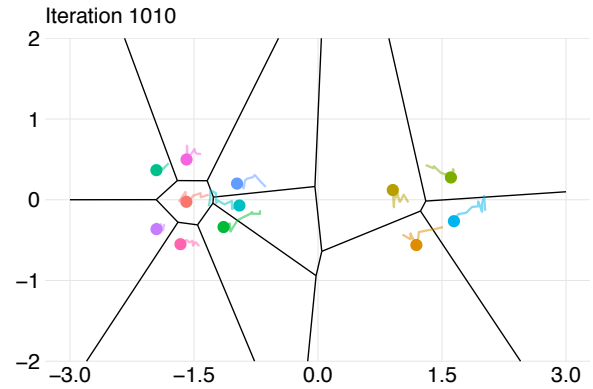
(a) *Maxcov* firms ($\mu = 0, n_l/n_r = 1$). Six of the firms have clustered at the centre of the market.



(b) *Maxcov* firms ($\mu = 1.5, n_l/n_r = 2$). Five of the firms have clustered near the peak of the left subpopulation.



(c) *Maxcov-inductor* ($\mu = 0, n_l/n_r = 1$). No clustering of firms.



(d) *Maxcov-inductor* ($\mu = 1.5, n_l/n_r = 2$). No clustering of firms. The red firm located at the peak of the left subpopulation previously paired with another firm, but only temporarily.

a high concentration of consumers around only one or two points. Competition for these high density areas forces firms closer together and clusters emerge. In the *maxcov-inductor* model two firms may temporarily pair up, but any pairing of firms is short-lived and eventually breaks up, see figure 21c and 21d. In turn, the long-run share of customers is more evenly distributed among firms and we observe a higher ENP. Notice how the *maxcov-inductor* firms in a market with a low degree of polarisation locates in orbits similar those described in the all-aggregator model with many firms, figure 21c. But comparing the mean eccentricity in the two models reveal that *maxcov-inductor* firms locate closer to the population centre than *aggregator* firms, when there are many firms in the market. Thus the *maxcov-inductor* firms on the outer orbit capture almost the same share of the market as firms on the inner orbit. This increases the ENP. The estimated ENP in the all-aggregator model told us that at least one firm has a share of the market that is less than half the share of a competing firm, while the estimate in the *maxcov-inductor* model does not fall below this threshold level. One or both of the paired firms will predict the future location of its competitor and distance itself to increase its share of the market. With a medium degree of polarisation, clustering and pairing of firms is short-lived in both models. There is a non-neglige number of consumers with ideal points in between the peaks of the distribution function. Consumers are not just concentrated around one or two points, but the density of consumers is spread over a larger area. Competition no longer forces firms together which eliminates the clustering of firms.

The difference in mean representation between the *maxcov* and *maxcov-inductor* model is minuscule. The mean representation plot for the *maxcov-inductor* model is therefore not included in the main text (see [Appendix E](#)).

From the comparison above we learn that introducing firms with the ability to forecast the future location of competing firms has a limited effect on the how far away firms locate from the average ideal point of all consumers. This distance is largely determined by the method with which the firm determines its optimal location, in this case the firm locates in the Delaunay triangle that encompasses most consumers. Forecast however play an important role in shaping the location behaviour of firms in cases with oscillation. We see that *maxcov-inductor* firms observe the oscillation of competing firms and uses this information to form its predictions. This precludes the most rudimentary oscillation patterns. Oscillation does not completely disappear with the introduction of *maxcov-inductor* firms. In this paper firms are only able to recognise periodic oscillation with period 2. That is firms only use information about the autocorrelation between two consecutive periods. If firms were also able to recognise oscillation with a higher period, it would likely reduce oscillation even further. Furthermore we learn that several *maxcov-inductor* firms are unlikely to cluster at the same location. Two firms may pair up, but this pairing is short-lived. One or both firms

will anticipate the future movement of the other firm, and realise that it can increase its share of the market by distancing itself.

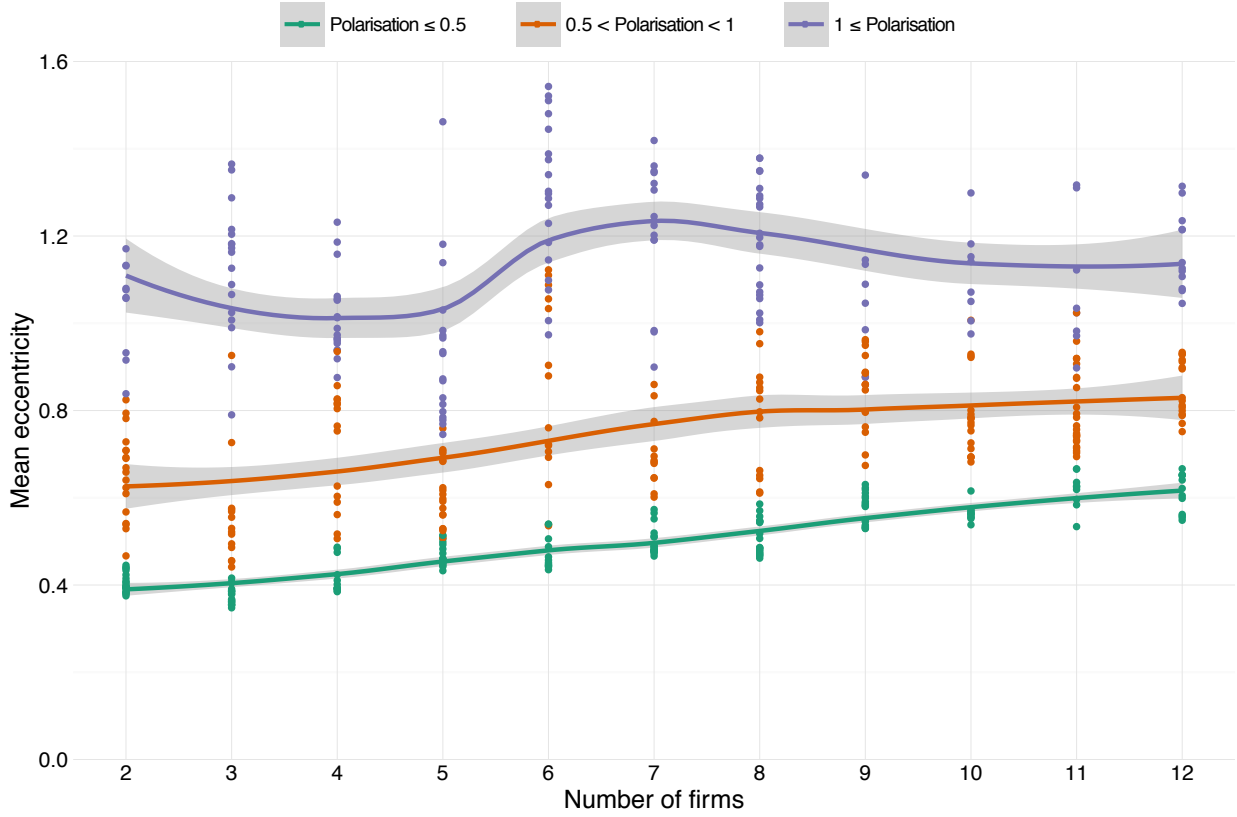


Figure 22: Mean eccentricity for *maxcov-inductor-GA* model.

Note: Uses Monte Carlo parameterisation method, the *ensemble average*, and is executed with 500 runs, 50 repetitions and 2,500 burn-in iterations. Generic algorithm runs every 50th iterations. Estimates smoothed using LOESS method.

5.4.2. *Maxcov-inductor-GA* model

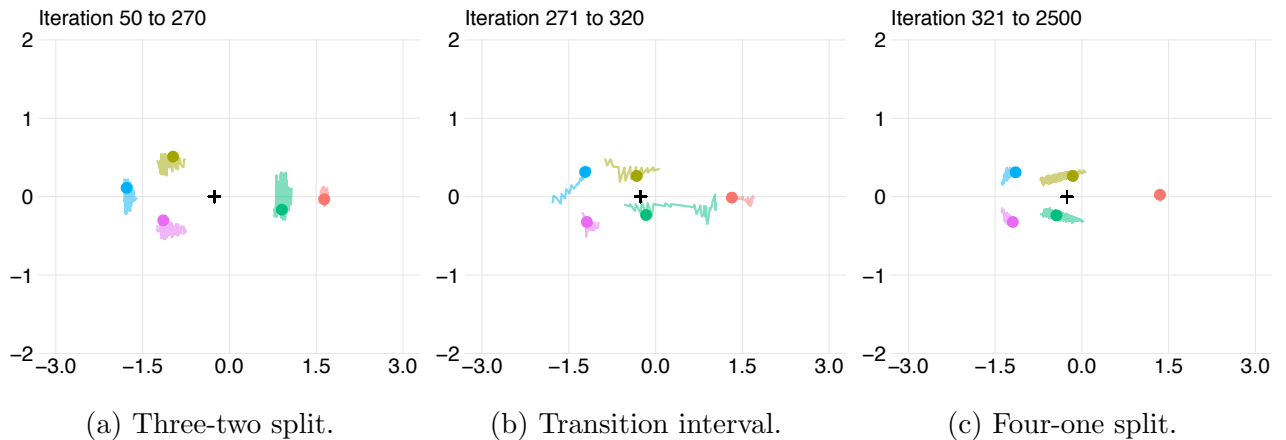
Maxcov-inductor-GA firms locate in the Delaunay triangle that maximises its share. The firms use *inductive reasoning* to forecast the future location of competing firms. And the firms gradually discard poorly performing hypotheses, and form new hypotheses using a generic algorithm. Thus hypotheses are endogenously formed.

Figure 22 presents the mean eccentricity. Once again we see very little change in the mean distance to the ideal point of all consumers with a low and medium degree of polarisation. *Maxcov-inductor-GA* firms locate in the Delaunay triangle that maximises its share. This is the decisive factor in determining the mean distance to the ideal point. With a high degree of polarisation, mean eccentricity is significantly lower in a market with five firms compared with the *maxcov-inductor* model.

As mentioned earlier firms tend to split into two crowds when there is a high degree of polarisation. With five firms and the degree of polarisation close to 1.5, the firms tend to locate in a three-two split. When the degree of polarisation is close to 1 and the distribution is sufficiently asymmetric (n_l/n_r not too close to 1), then the split tends to be four-one. This is the case in both the *maxcov-inductor* and *maxcov-inductor-GA* model. However the *maxcov-inductor-GA* model requires less polarisation among the subpopulations for the four-one split to occur. In figure 23 we plot the location of five *maxcov-inductor-GA* firms for $\mu = 1.3$ and $n_l/n_r = 1.5$ at various iterations. The firms quickly converge to the three-two split, but this location configuration is not sustained and after 270 iterations the firms transition into a four-one split. For comparison we re-execute the model with the same initial positions and the same set of endowed *condition/forecast* rules, but where firms use the *maxcov-inductor* decision rule instead of *maxcov-inductor-GA*. The *maxcov-inductor* firms maintain the three-two split also after 270 iterations and locate as in figure 23a. This explains why we observe a lower mean eccentricity in the *maxcov-inductor-GA* model with five firms compared to the *maxcov-inductor* model. With a four-one split the firms locate closer to the average ideal point of consumers than firms that locate in a three-two split.

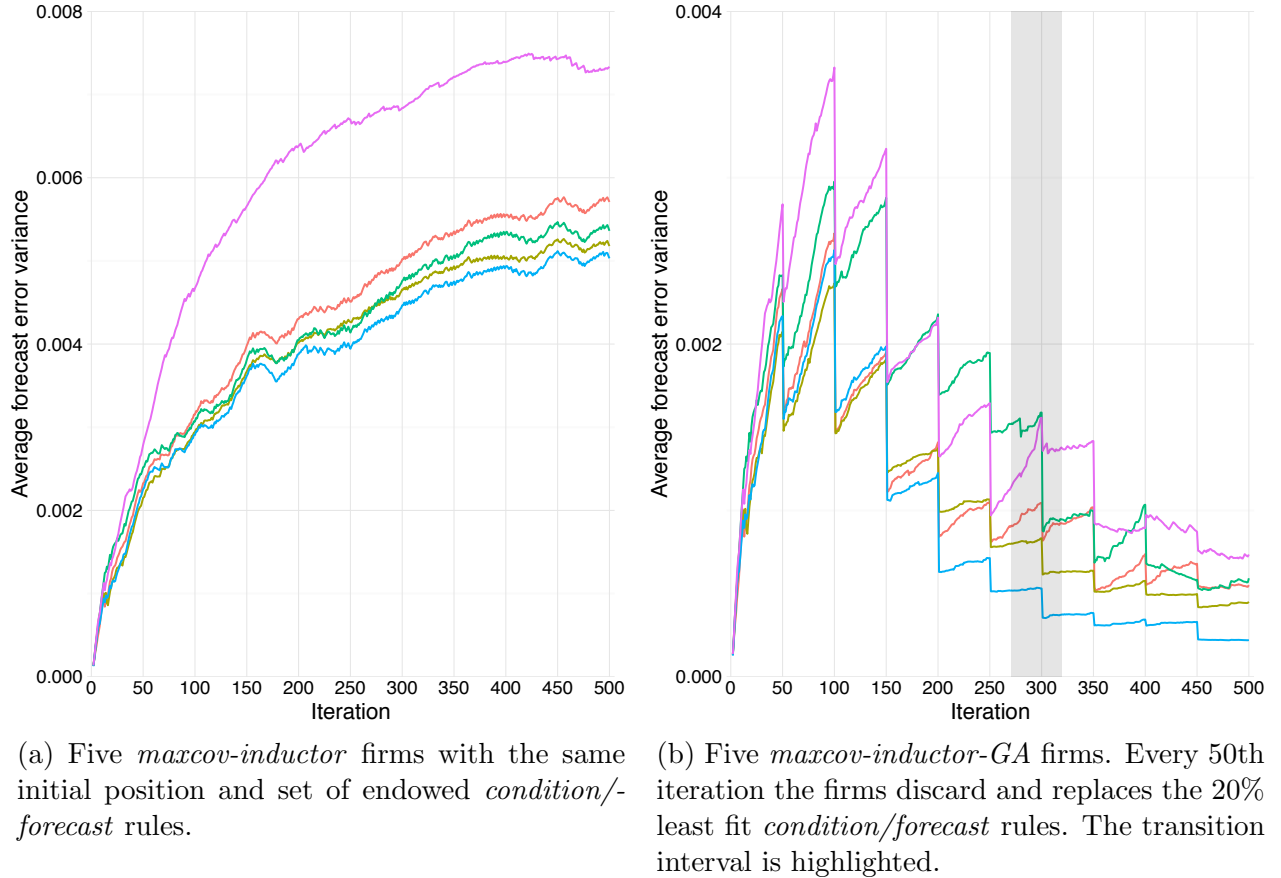
Furthermore we can compare the accuracy of the two cases. The *maxcov-inductor-GA* firms replace the 20% least fit hypotheses every 50th iteration, while the *maxcov-inductor* firm uses the same set of M hypotheses indefinitely. In figure 24 we plot the average forecast error variance of all the *condition/forecast* rules for each firm at each iteration. We disregard *condition/forecast* rules that have a forecast error variance of zero, and thus have not previously been active. A low forecast error variance indicates that the hypothesis provided accurate predictions in the past. We focus on the first 500 iterations. For the first 50

Figure 23: Trajectory plot for five *maxcov-inductor-GA* firms at different iteration intervals ($\mu = 1.3, n_l/n_r = 1.5$). The dots indicate the location after respective 270, 320 and 2,500 iterations and the lines indicate the location in preceding iterations. The black cross indicates the average ideal point of all consumers.



iterations the forecast error variance is the same in both cases, since the firms are making predictions using the exact same set of endowed hypotheses. In the *maxcov* model the average forecast error variance increases steadily, but at a steadily decreasing rate, as firms learn which hypothesis provide the most accurate predictions, see figure 24a. Firms learn through experience. This is also the case in the *maxcov-inductor-GA* model, but in additions firms learn through adaption. Discarding poorly performing hypotheses and replacing them with new hypotheses that are based on the most fit hypotheses significantly increases the accuracy, as seen in figure 24b. After 270 iterations the firms have discovered a sufficiently accurate location pattern of their competing firms, allowing the firms to move from the three-two split to the four-one split. Notice that the three firms that move furthest in the transition interval are not simultaneously experiencing an increase in their average forecast error variance. The predictions they make during this transition are at least as precise as previous predictions. Thus the transition cannot be attributed to synchronised and systematic break down in forecasts, but rather the transition is the result of a deliberate and calculated response to the predicted future locations of competing firms. These are strategic moves. While the two

Figure 24: Average forecast error variance of all the activated *condition/forecast* rules of the firms.



other firms experience an increasing average forecast error variance during the transition, they quickly adapt to the environmental change. And we see that shortly after the transition the firms are quite similar in terms of the accuracy of their hypotheses.

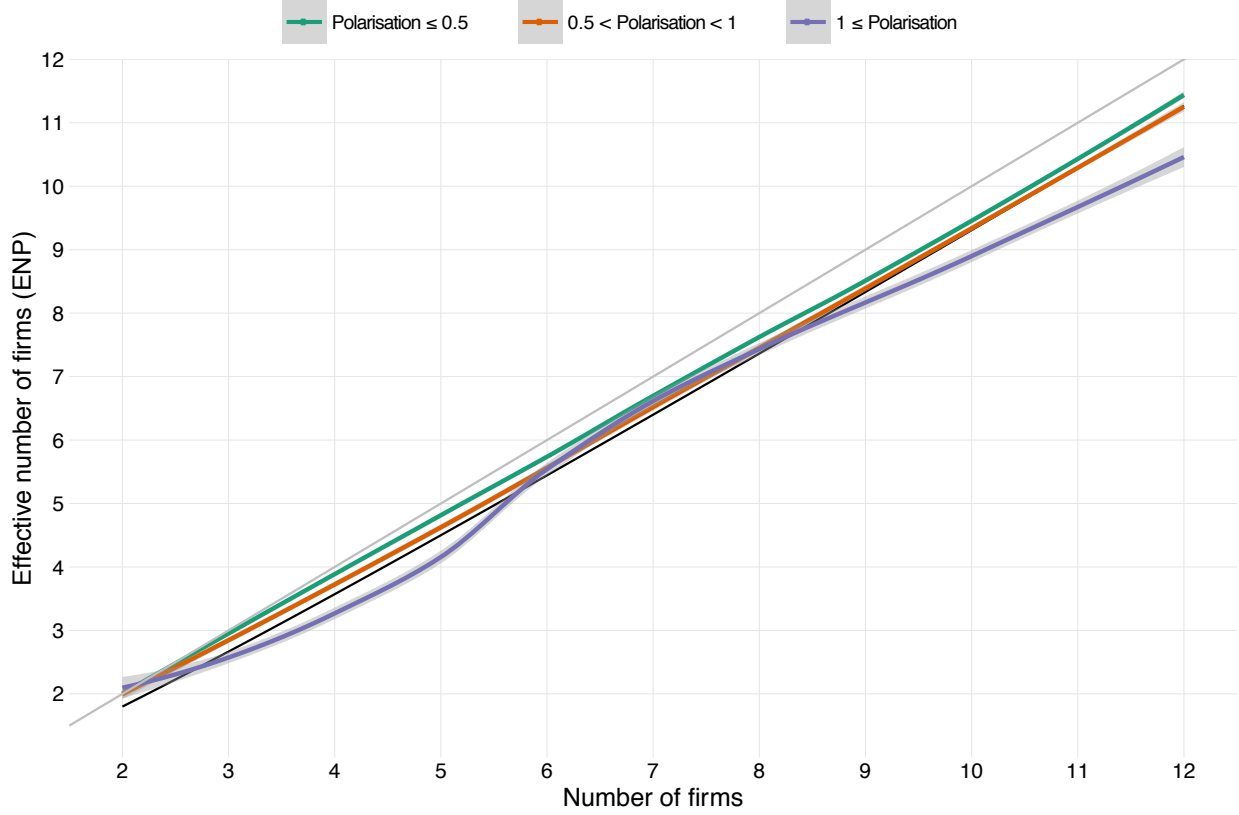


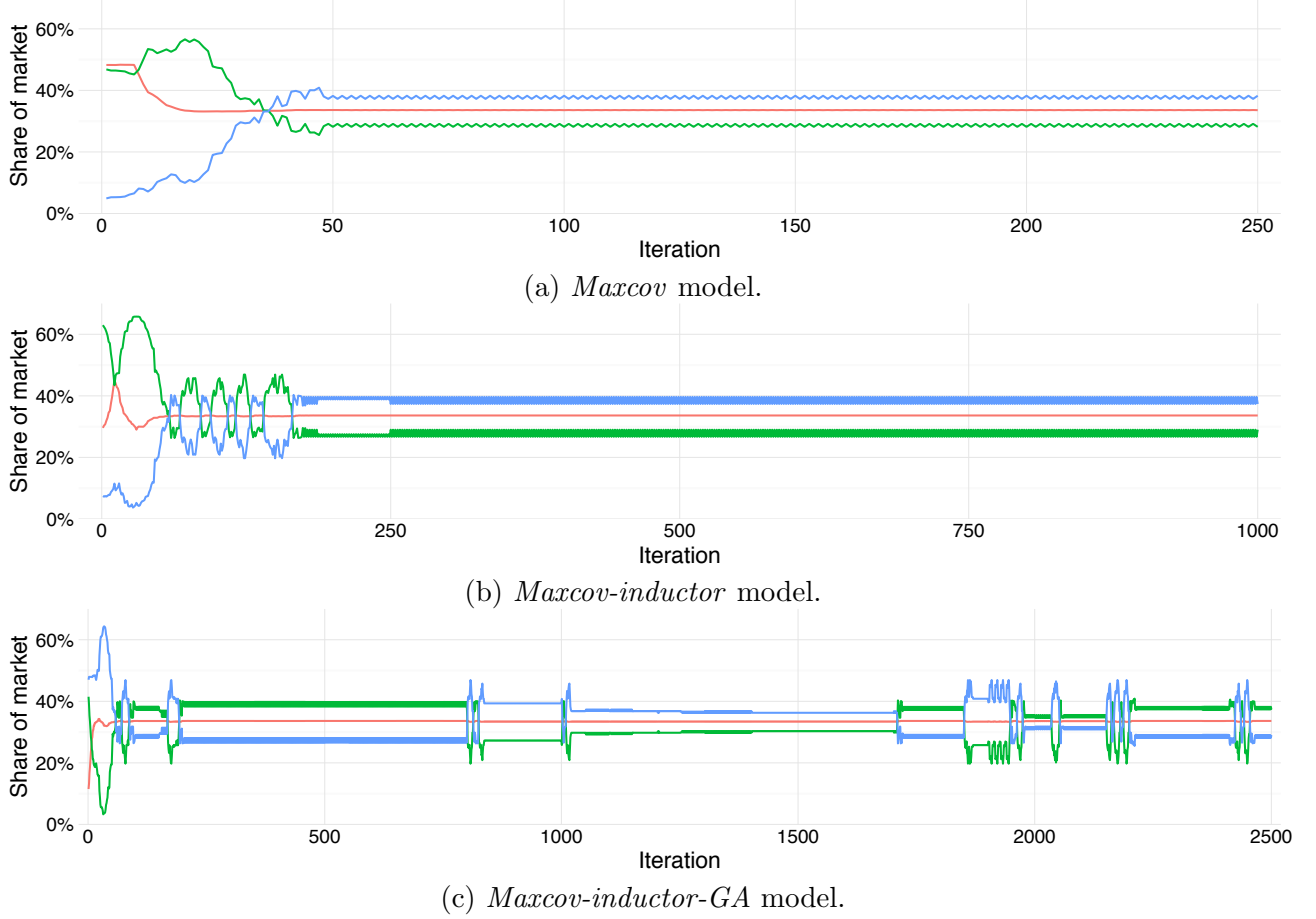
Figure 25: Effective number of firms (ENP) for *maxcov-inductor-GA* model.

Note: Below the black diagonal line at least one firm captures a share that is less than half the share of a competitor.

Figure 25 presents the effective number of firms (ENP). The ENP is largely unaffected by the extension of the decision rule. When the ENP is below the black diagonal line then at least one firm has a share of the market that is less than half the share of one competitor. The fair can pair with the competitor capturing at least half of the competitors share. We see that the firms capture a much more even share of the market when firms use foresight. In all the models without foresight, a firm consistently captures a share that is less than half of the share of its competitors, in the market with many firms. In the *maxcov-inductor-GA* model this is the case with a high degree of polarisation and 3-5 or more than 8 firms in the market. It might also be the case with a low or medium degree of polarisation, but it cannot be determined by the estimated ENP alone. In the model with five firms the ENP is lower than the *maxcov-inductor* model, since the four-one split is more likely and a the single firm captures the entire right side of the market for itself. Examining the market share

of each firm over time reveals differences in terms of the long-run competitive advantage one firm may have over its competitors. This is best illustrated in a market with three firms, by comparing the *maxcov*, the *maxcov-inductor* and the *maxcov-inductor-GA* model. For each model we plot the market share across iterations in a market with a highly polarised and uneven sized subpopulations ($\mu = 1.5$ and $n_l/n_r = 2$), see figure 26.

Figure 26: Share of market with three firms, $\mu = 1.5$ and $n_l/n_r = 2$. Red line is the firm located at the right peak. While the blue and green line are the two firms located close to the left peak.



In the *maxcov* model, after the initial burn-in periods, the location of the firms lock in. One firm locates close to the right peak of the distribution function, while the two other firms locate around the left peak. The location of the right firm is consistent across different initial locations, while the location configuration which the two firms on the left lock into vary depending on the initial position of firms. Additionally the two firms on the left often oscillate. The firm on the right captures $1/3$ of the market. One of the firms on the left captures a slightly larger share, and the other a slightly lower share. While our estimate of the ENP is independent of the initial position of firms, the initial position determines how each firm ranks in term of market share. It is evident from figure 26a, that certain initial

positions gives one firm a clear long-run competitive advantage over its competitors in the *maxcov* model.

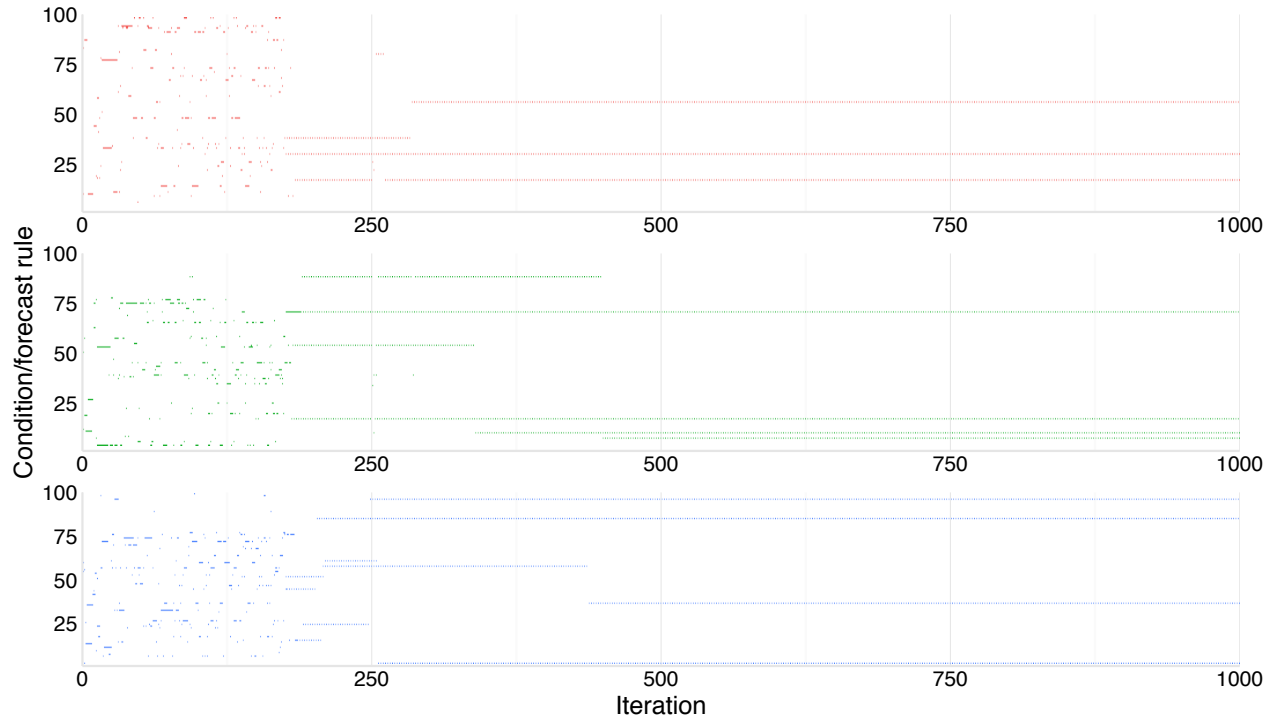


Figure 27: The *condition/forecast* rule used at each iteration by each of the three *maxcov-inductor* firms to forecast the future location of competing firms. After approximately 200 iterations all firms use a limited set of *condition/forecast* rules. After approximately 400 iterations the set does not change. The firm on the right is red, the two firms on the left are green and blue respectively.

In the *maxcov-inductor* model firms predict the future location of competing firms, using the hypotheses that match the current state and proved most accurate in the past. After several dozen burn-in iterations the firm on the right once again locates close the right peak and captures 1/3 of the market. The two firms on the left will not immediately lock in, but switch between different location configurations. The ranking in terms of market share changes when the two firms switch between configurations²⁸. The firms do not immediately lock in, since firms are able to predict and respond to the future location of competing firms, such as oscillation. Over time the firms learn which of their hypotheses give the most reliable predictions. And the firm often ends up using a limited set of hypotheses, see figure 27. The two firms on the left lock into a specific configuration or a specific few configurations, once all firms converge to using a limited set of hypotheses. Once the firms lock in, so does their share of the market, as seen in figure 26b. The initial position of the firm does not by

²⁸These configurations are the same as those that arose, when using different initial positions, in the *maxcov* model.

itself create a long-run competitive advantage in the *maxcov-inductor* model. Instead it is a combination of initial position and the set of hypotheses which the firms is endowed with, that can give the firm a long-run competitive advantage. While the firm is able to update the accuracy of each hypothesis, the conditions and forecast values remain fixed. If a firm is endowed with a set of poorly performing hypotheses, then the firm is unlikely to capture more than $1/3$ of the market in the long-run.

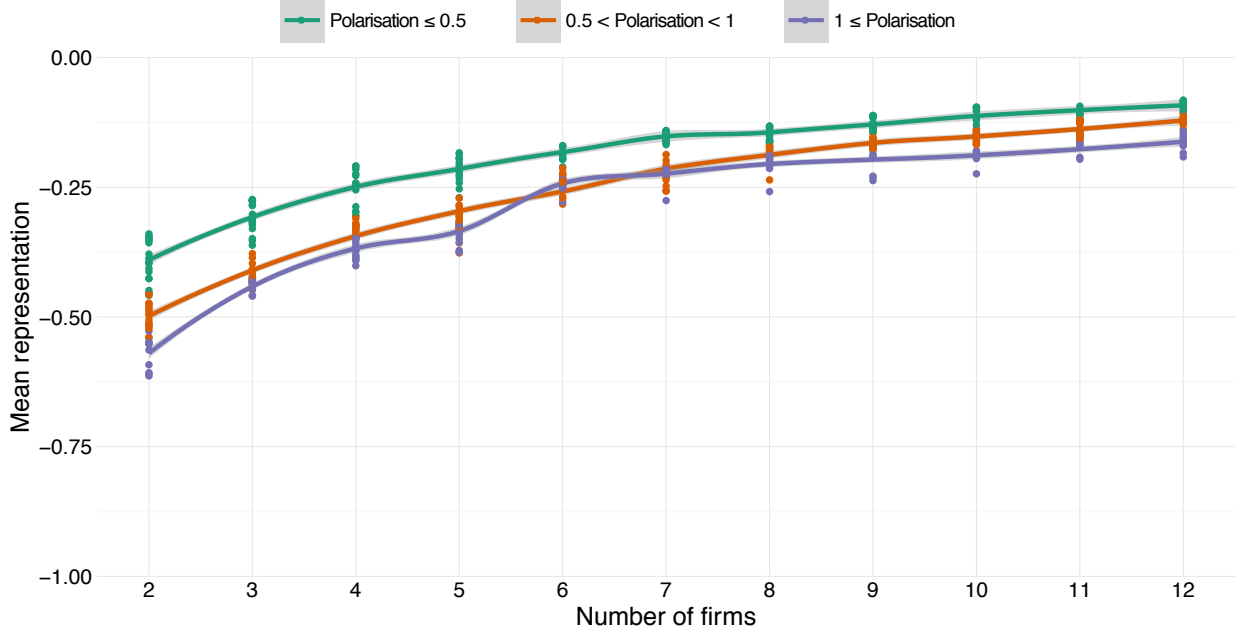


Figure 28: Mean representation for *maxcov-inductor-GA* model.

In the *maxcov-inductor-GA* model the two firms on the left never lock in, but continue to switch between the different location configurations. The firms may temporarily get stuck in a specific configuration, where all firms use a limit set of hypotheses, but eventually the firms form new hypotheses that change how the firms locate. The set of hypotheses which the firm is endowed with does not give the firm a long-run competitive advantage over its competitors. Instead there will be prolonged periods where one of the two firms on the left capture more than $1/3$ of the market, followed by periods where the reverse is true, as seen in figure 26c. This highlights how one firm may use tactics that are superior to its competitors and how competing firms gradually learn and develop counteracting responses. The *maxcov-inductor-GA* firm continuously adapts to changes in its environment. This type of behaviour is not unique to a market with three firms and highly polarised and uneven sized subpopulations. The same is true with a different number of firms in the market and

different parameter values²⁹.

The mean representation in the *maxcov-inductor-GA* model is almost identical to the *maxcov* and *maxcov-inductor* model, as seen in figure 28.

6. Conclusion

This paper set out to reintroduce foresight and thus strategic consideration into the competitive location model, something that has thus far been missing from the agent-based models. The paper argues why it is unreasonable to assume that firms use *deductive reasoning* to deduce their optimal location. Instead we argue that firms derive principles and conclusions from observations and viewing conclusions as probable, rather than certain, i.e. firms use *inductive reasoning*. The firms possess numerous hypotheses, each relevant to specific set of situations. Hypotheses are updated in response to new information or observations that falsifies the currently held hypotheses. The firms predict the future location of competing firms, using the hypothesis viewed as most probable in the given situation. When a firm chooses its own location it takes the future location of competing firms into account. The firm acts strategically, and is no longer oblivious to the simultaneous relocation of competing firms.

The model in this paper covers the “natural” setting of competitive product differentiation or location differentiation. That is a setting with few firms in the market that is not necessarily restricted to a single dimension and where the distribution of consumer preferences might be asymmetric and multimodal. The conclusions from the previous literature on competitive location models mainly concern one dimensional markets and duopolies. Yet in most markets more than two firms compete, and they often choose along several product characteristics not just one. While it is convenient to assume that consumers are uniformly distributed, this assumption is highly unrealistic. The case where the preference of all consumers is uncorrelated must be considered a special case. More generally consumers are likely to lump together in subpopulations. Possibly due to social interactions among consumers, such as desire for social conformity or other network effects. These subpopulations then give rise to an aggregated distribution of consumer preferences that is multimodal and quite possibly asymmetric.

This paper finds that predictions by themselves have a limited effect on the overall location pattern of firms. We do not observe locations that are fundamentally different when using

²⁹The polarised and uneven market with three firms contains a lot of different location configurations while the low number of firms makes it easy to describe. And so we used this for the basis of the example. With two or four firms in the market there are only a few location configurations, all of which are fairly stable, and so the described behaviour is less pronounced, since there is less changes to the environment of the firm.

decision rules with foresight. This contrasts the finding by [Eaton & Lipsey \(1975\)](#). When they consider a strategy with foresight (*minimax strategy*) they find a location outcome that is fundamentally different (from the *zero conjectural variation* location outcome). The main driver in determining the location patterns is the method used by the firms to determine the optimal location given its predictions. In this paper the firms located in the Delaunay triangle with most consumers. While predictions do not affect the overall location pattern, they do influence specific location behaviours such as oscillation, clustering of firms, and it changes the probability of certain location patterns. [Eaton & Lipsey \(1975\)](#) also find that clustering of firms disappears when firms use foresight. In a market with many firms predictions lead the firms to capture a more even share of the market, namely because any pairing of firms is short lived. This is somewhat surprising given the fact that we introduce additional sources of heterogeneity for firms. In the models with foresight, firms are also heterogeneous in terms of their expectation model and knowledge.

In the models with foresight we find that it is less like that a firm captures less than half the share of a competitor, despite the fact that firms only temporarily pair up with a competing firm. Since the firms make predictions using the hypothesis that matches the current state, the firm is better adept to changes in the environment. And even more so when the firm continuously forms new hypotheses. This implies that no firm has a competitive long-run advantage over its competitors, when all firms use the same decision rule.

In the paper we also consider the social welfare of consumers. The interacting effect of firms competing for consumers seems to result in locations that are more or less on par with the social optimal location (except in the cases with less than four *hunter* firms and a bimodal distribution). This is a quite striking result considering the conclusions of earlier papers. In the line market, the *principle of local clustering*, lead firms to locate far from the social optimal locations. And in the two-dimensional Hotelling model with a uniform distribution of consumers, the two firms locate at the centre, thus also far from the social optimum. The main reason that competition leads to more or less socially optimal locations in this paper is due to the assumed distribution function. We assume that each subpopulation follow a *bivariate normal distribution* with standard deviation (0.5, 0.5). This gives a high concentration of consumers around the peaks of the aggregated distribution function. Increasing the standard deviations would decrease the concentration of consumers around the peaks, and we would most likely observe a decrease in the average utility of consumers that would be more in line with previous papers.

The paper is a first attempt at laying down the foundation for a competitive location model with foresight. There is still ample room for improvement and many possible extensions. An obvious extension is entry and exit of firms such that the number of firms is

endogenously determined. Or evaluating the competitive advantage of the *maxcov-inductor-GA* decision rule in a tournament type of setting, where firms using the rule are pitted against firms using different decision rules. Similarly incorporating price competition into the model seems to be a natural next step. The lack of methods for finding the optimal location meant that in all the *maxcov* decision rules used a rough approximation. Developing new techniques for finding the optimal location of new entrants in the market, as well as the optimal location of existing firms, would be of benefit to the models in this paper, and the field in general.

Appendix A. Limitations on the effective number of firms (ENP)

Eaton & Lipsey (1975) provides the upper and lower bound on the market share, when no firm has a total share of the market that is less than half the share of any competing firm. The upper bound is $\bar{s}_i = \frac{2}{N+1}$, where firm i captures a majority of the share, and the share of all other firms is $\bar{s}_{-i} = \frac{1}{N+1}$. Substituting into equation 4 we find that:

$$ENP_{upper} = \frac{(\bar{s}_i + (N-1)\bar{s}_{-i})^2}{\bar{s}_i^2 + (N-1)\bar{s}_{-i}^2} = \frac{\frac{2}{N+1} + (N-1)\frac{1}{N+1}}{\left(\frac{2}{N+1}\right)^2 + (N-1)\left(\frac{1}{N+1}\right)^2} = \frac{(N-1)^2}{3+N} \quad (\text{A.1})$$

The lower bound is $\underline{s}_i = \frac{1}{2N-1}$, where firm i captures a minority of the share, and the share of all other firms is $\underline{s}_{-i} = \frac{2}{2N-1}$. Substituting into equation 4 we find that:

$$ENP_{lower} = \frac{(\underline{s}_i + (N-1)\underline{s}_{-i})^2}{\underline{s}_i^2 + (N-1)\underline{s}_{-i}^2} = \frac{\frac{1}{2N-1} + (N-1)\frac{2}{2N-1}}{\left(\frac{1}{2N-1}\right)^2 + (N-1)\left(\frac{2}{2N-1}\right)^2} = \frac{(2N-1)^2}{4N-3} \quad (\text{A.2})$$

Note that $ENP_{upper} < ENP_{lower}$. When the estimated effective number of firms is below the upper limit ($ENP < ENP_{upper}$) then at least one firm has a share of the market that is less than half of the share of a competing firm. When the estimated effective number of firms is above the lower limit ($ENP > ENP_{lower}$) then no firms has a share that is less than half of the share of any competing firm. This might also be the case when $ENP \in [ENP_{upper}, ENP_{lower}]$, but we cannot determine this from the estimated ENP along.

Appendix B. Model overview

This section contains an overview of the models, and the arguments for the underlying process of a *run* are provided. We abbreviate *time-homogenous Markov chain* as THMC.

All-sticker:

The underlying process is a *non-ergodic* deterministic THMC. The firms never move so the initial state space distribution, π_0 , is stationary and solves $\pi_{t+1} = \pi_t$. The process is fixed in a single state, which depends the initial position of the firm, thus the process is *non-ergodic* and we use the *ensemble average*.

All-aggregator:

The underlying process is a *non-ergodic* deterministic THMC. The process does not contain a random component, since firms always locating at the centroid of their current market. When all firms use the *aggregator*-rule then a *repetition* is an implementation of the Lloyd-algorithm, which always leads to a *centroidal Voronoi tessellation* (CVT). Thus the process

converges to a single state. There is no guarantee that the CVT is unique, a different initial position of the firms might result in a different CVT, which implies that the process is *non-ergodic* and we use the *ensemble average*.

All-hunter:

The underlying process is a ergodic stochastic THMC. Whenever a *hunter* firm experiences a decrease in its market share, the firm turns around and heads in a randomly drawn opposite direction. This random component, makes the underlying process *ergodic*. With a random component every state in the state space has strictly positive probability of being reached in a finite number of iterations. Thus the process, figuratively speaking, avoids getting stuck in a blind alley. Eventually the process will converge to its unique distribution vector, π_∞ . In the model with a symmetric distribution of consumers the *time average* provides a representative estimate of ψ , i.e. for all output variables the R-hat statistic is below 1.05. The model with an asymmetric and bimodal distribution of consumers contains two additional free parameters. In this model we find that several R-hat statistics that are above 1.05 — even when executing the test repetitions with 20,000 iterations. For this model we conclude that the *time average* does not provides a representative estimate of ψ , and we instead use the *ensemble average*.

All-maxcov:

The underlying process is a *non-ergodic* deterministic THMC. The process does not contain a random component, since firms always locating at the centroid of the largest Delaunay triangle. Unlike the other deterministic THMC, this process does not converge to a single state, since firms often oscillate between different locations. Furthermore we observe that different initial locations of firms converge to different locations (or converge to different oscillation patterns, i.e. different sets of locations). So the process is *non-ergodic*, i.e. the stationary state space distribution, π_∞ depends on the initial state space distribution, π_0 . The paper assumes that firms oscillate between a few locations that are not too far apart. It also assumes different initial position only lead to minor differences in the location which firms converge to. With these assumptions in mind we use the *ensemble average* to estimate ψ (see discussion in subsection 4.2.5).

All-maxcovrnd:

The underlying process is a ergodic stochastic THMC. The process contains a random component, since the speed parameter is randomly drawn. Instead of the firm moving 0.1 standard deviations, the move is drawn uniformly random from the range $[0, 0.2]$. This insures that every state in the state space has strictly positive probability of being reached in a finite

number of iterations. The process will eventually converge to the unique distribution vector, π_∞ . Every *stochastic THMC* with a finite state space is *ergodic* (Laver & Sergenti, 2011, chapter 4, p. 64). We estimate ψ using the *ensemble average*, and compare the results with the *all-maxcov* model to see if our assumption in that model regarding oscillation and the steady state distributions are reasonable.

Maxcov-inductor:

The underlying process does not fulfil the Markov property. The *trending bits* and *oscillation bits* calculate respectively the moving average and the autocorrelation using the location history of the firm. And so the probability of the future state will depend on past states, and only the current state, which violates the Markov property. Furthermore since the firm updates the accuracy of it hypotheses the *transition probability* is not stationary. The *transition probability* depends on the experience of the firm, and since the experience evolves so will the *transition probabilities*. Even without the *trending bits* and *oscillation bits* the underlying process would not constitute a THMC. While it has not yet been proven if the *ensemble average* provides a representative estimate of ψ , this is none the less the method used in previous papers, and will therefore also be the method employed in this paper.

Maxcov-inductor-GA:

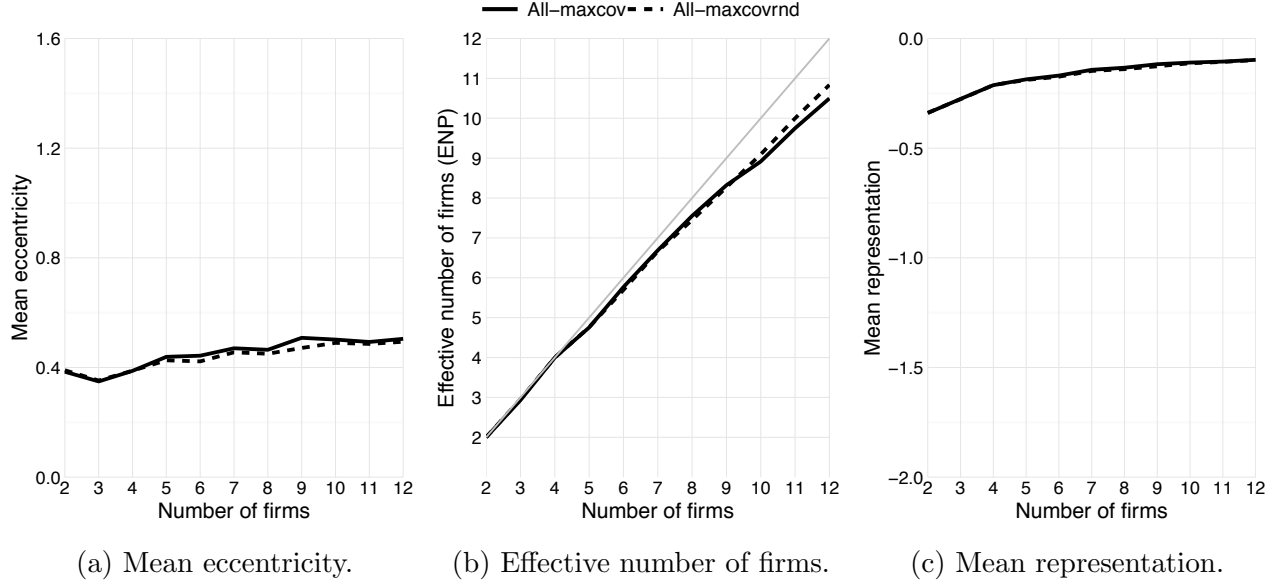
The underlying process does not fulfil the Markov property either. The reason is the same as in the *maxcov-inductor* model. The genetic algorithm which the firm uses to construct new hypotheses with random mutation and crossover, means that there is a random competent in the underlying process.

Appendix C. Maxcovrnd decision rule

To test our assumptions and see if it is reasonable to use the *ensemble average* to estimate ψ in the *maxcov* model we create a new decision rule *maxcovrnd*. Firms using the *maxcovrnd* decision rule will not move 0.1 standard deviations, but uniformly random draw the speed parameter from the range $[0, 0.2]$. On average the two decision rules have the same speed parameter. By randomly drawing the speed parameter we introduce a random component into the process. In the model where all firms use the *maxcovrnd* rule, the random component in the process, insures that the *ensemble average* provides a representative estimate of ψ . We execute both models and estimate using the *ensemble average*. If the two models provide similar estimates, then firms only oscillate between a few similar locations in the *maxcov* model, and different initial position only leads to slight differences in the locations which firms converge to. If this is the case we conclude that our assumptions are reasonable, and

we feel confident using the *ensemble average* in the *maxcov* model. Figures C.29 and C.30 plots the estimates from the models with respective a symmetric unimodal distribution and an asymmetric bimodal distribution of consumers. The difference between the *maxcov* and *maxcovrnd* model is minuscule.

Figure C.29: Estimates from market with symmetric and unimodal distribution of consumers ($\mu = 0$ and $n_l/n_r = 1$). Solid lines is *maxcov* model and the dashed line is the *maxcovrnd* model.



Note: Both models use grid sweep method. Each model executes 11 runs, 1,000 repetitions, and 99 burn-in iterations.

Appendix D. Further details on condition/forecast rules

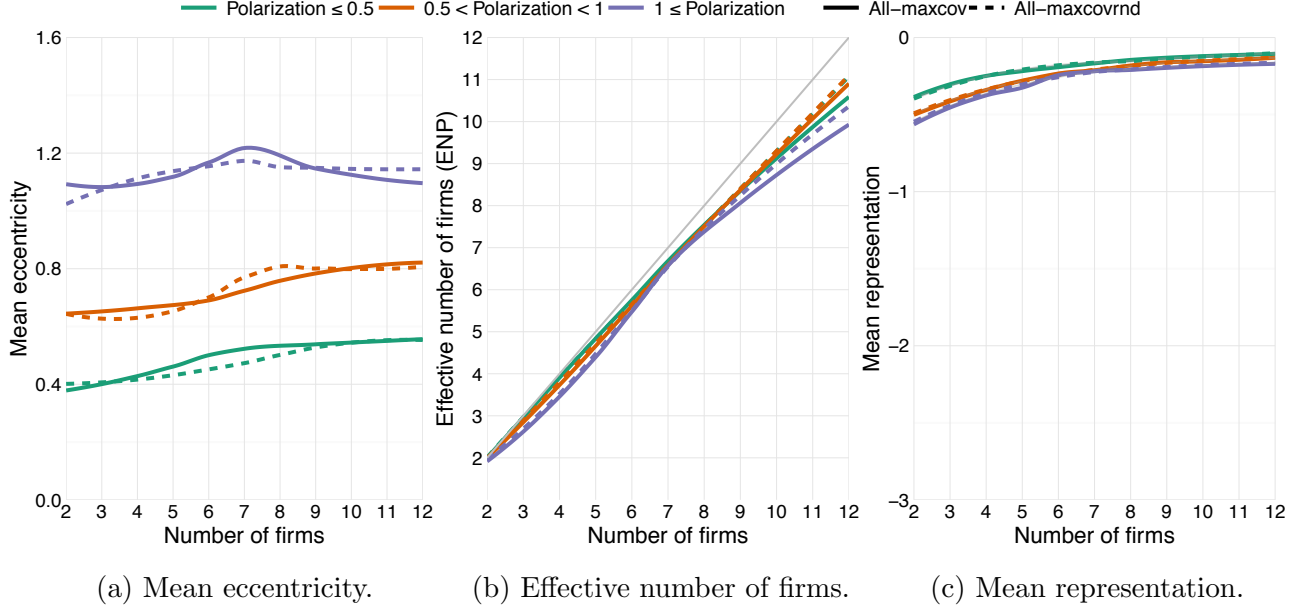
This subsection provides further details — such as equations, parameter values and probability values — used in relation to the *condition/forecast* rules and the genetic algorithm.

Each firm holds a set of M hypotheses or *condition/forecast* rules. While each hypothesis might only be relevant to a narrow set of situations, together the array of hypotheses cover a wide range of different situations. We follow Arthur (2014, chapter 3) and set the number of hypotheses to one hundred, i.e. $M = 100$.

Accuracy:

The firm predicts the future location using the hypothesis it finds most probable given the current state of the competing firm. If the condition part of the rule is satisfied by the current state descriptor, J_j , of competing firm j , then that rule is said to be active. Several *condition/forecast* rules may match the current state of competing firm j and the firm uses the same set of M *condition/forecast* rules for all competing firms, so in any given iteration

Figure C.30: Estimates from market with asymmetric and bimodal distribution of consumers ($\mu \in [0, 1.5]$ and $n_l/n_r \in [1, 2]$). Solid lines is *maxcov* model and the dashed line is the *maxcovrnd* model.



Note: Both models use Monte Carlo parameterisation method, the *ensemble average*, and is executed with 500 runs, 50 repetitions and 150 burn-in iterations. Estimates smoothed using LOESS method.

multiple *condition/forecast* rules are likely to be active. The accuracy of all the active *condition/forecast* rules is updated once the firms relocate and the actual location of each competing firm is revealed. The forecast error of each active rule is the difference between its predicted location and the actual location. To update the accuracy of *condition/forecast* rules we use an inverse moving average of squared forecast errors (Arthur, 2014). We update the forecast error variance of a given active *condition/forecast* rule m at iteration t using:

$$e_{t,i,m}^2 = \alpha_a e_{t-1,i,m}^2 + (1 - \alpha_a) d(X_{t+1,j}, E_{t,i,m}[X_{t+1,j}])^2 \quad (\text{D.1})$$

where firm i uses rule m to make the forecast $E_{t,i,m}[X_{t+1,j}]$ about the future location of competing firm j . The actual future location of competing firm j is $X_{t+1,j}$. The forecast error, $d(\cdot, \cdot)$, is the *Euclidean* distance between the actual and the predicted location. We assume that all firms use the same memory parameter α_a to update the accuracy. If the *condition/forecast* rule m is not active at iteration t , then the forecast error variance is unchanged:

$$e_{t,i,m}^2 = e_{t-1,i,m}^2$$

We choose to smooth out the accuracy and therefore set a high memory parameter; $\alpha_a = 74/75$. Thus new information has a relatively low impact on the accuracy of the

condition/forecast rule, and we avoid that one bad prediction suspends the future use of the rule.

Appendix D.1. Genetic algorithm

Every τ number of iterations the firm discards the 20% least fit *condition/forecast* rules and replaces these with new *condition/forecast* rules created using a genetic algorithm through a process of mutation and crossover of the non-discarded rules. We choose a quick exploration rate and set $\tau = 50$.

Fitness:

The fitness of each *condition/forecast* rules is based on the accuracy and the specificity of the rule. The specificity of a rule is calculated as the number of 1 and 0 in the condition part of the rule (i.e. all # are not counted). We calculate the fitness of rule m at iteration t using:

$$f_{t,i,m} = M - e_{t,i,m}^2 - s_{t,m}c \quad (\text{D.2})$$

where firm i hold rules m . The total number of *condition/forecast* rules held by firm i is M . This is constant and identical across firms, so the term can be left out. The forecast error variance of rule m is $e_{t,i,m}^2$. The specificity of rule m is $s_{t,m}$, and c is the cost levied on the specificity. Inaccurate rules have a higher forecast error variance which reduces the fitness of the rule. A more specific rule also reduces the fitness of the rule. Thus accounting for the specificity of the rule creates a drift towards more generally applicable rules that fit a wider range of situations (i.e. rules with many # in the condition). The condition consists of 13 positions so $s_{t,m} \in [0, 13]$ and the average forecast error is below 0.1, so we set the cost of specificity to $c = 0.0005$ creating a slight drift towards more general rules.

Crossover:

Each new hypothesis is based on two “parent” hypotheses. The pair of parent *condition/forecast* rules are randomly selection from the set of non-discarded rules, i.e. the 80% most fit *condition/forecast* rules. A new *condition/forecast* rule is created by either the mutation procedure or crossover procedure. For each new rule the procedure is randomly selected. The crossover procedure is selected with probability p , and the mutation procedure with probability $1 - p$. This paper follows [Arthur \(2014, chapter 3\)](#) by using $p = 0.3$. With crossover the new *condition/forecast* rule is a mix of both parents. The condition part is mixed by randomly selecting a donor parent for each of the 13 positions. Crossover of the forecast parameter values happens by either 1) component-wise crossover of each value, 2) using the weighted average of the parents or 3) randomly picking a parent that passes on all parameter values. The three methods are randomly selected with equal probability.

Mutation:

With mutation the new hypothesis only inherits traits from the most fit of the two parents. Each position in the condition part is mutated with probability 0.03 and takes the value of the most fit parent with probability 0.97. The position is mutated by flipping its value. The probability that a 0 or a 1 is flipped to a # is 2/3. The probability that 0 is flipped to 1, or visa versa, is 1/3. The probability that a # is flipped to a 1 or 0 is 1/3 respectively, where the remaining 1/3 is the probability that # is not flipped. These flip-probabilities can also be written as a matrix:

$$P = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad (\text{D.3})$$

Taking the limit of P^x for $x \rightarrow \infty$ reveals that the mutation of bits does not lead to a drift in the specificity of the rules:

$$\lim_{x \rightarrow \infty} P^x = \begin{bmatrix} 0.25 & 0.25 & 0.5 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \quad (\text{D.4})$$

Repeatedly flipping a value in the condition will lead to a 50% chance of it taking the value 0 or 1, and 50% chance of it taking either the value #. The steady drift in the specificity of the *condition/forecast* rules occurs because we levy a positive cost, i.e. $c > 0$. The forecast parameter values are also mutated. Each value is either replaced or altered, each with probability 0.2. The forecast parameter takes the same value as the most fit parent with probability 0.6. If the value is replaced, then the new parameter value is drawn randomly from the same ranges as the initial parameter values (range of parameter values presented in subsection 5.3.1). A altered value is equal to the value of the most fit parent, altered with a randomly drawn amount in the range $[-0.5\%, 0.5\%]$ times the length of the respective initial parameter range. For example let A_1 be the third forecast parameter value of the parent (A_1 has been drawn from the range $[0.99, 1.01]$ which has a length of 0.02) — if the randomly drawn amount is -0.5%, then the new altered value becomes: $A_1 - 0.5\% \times 0.02 = A_1 - 0.0001$.

Appendix E. Mean representation in maxcov-inductor model

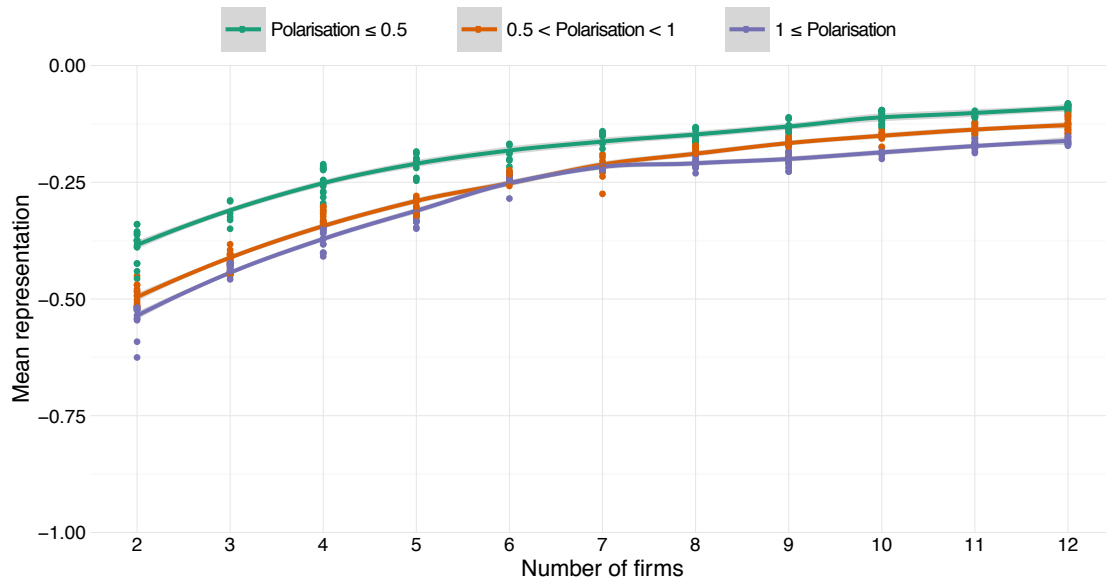


Figure E.31: Mean representation for *maxcov-inductor* model.

References

- Adams, J. & Somer-Topcu, Z. (2009). Policy adjustment by parties in response to rival parties' policy shifts: Spatial theory and the dynamics of party competition in twenty-five post-war democracies. *British Journal of Political Science*, 39(04), 825–846.
<http://dx.doi.org/10.1017/s0007123409000635>.
- Ahn, H.-K., Cheng, S.-W., Cheong, O., Golin, M., & Van Oostrum, R. (2001). Competitive facility location along a highway. In *Computing and Combinatorics* (pp. 237–246). Springer.
http://dx.doi.org/10.1007/3-540-44679-6_26.
- Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1988). *Network flows: Theory, algorithms, and applications*. Prentice Hall Englewood Cliffs.
- Arrow, K. J. (1971). The economic implications of learning by doing. In *Readings in the Theory of Growth* (pp. 131–149). Springer Science Business Media.
http://dx.doi.org/10.1007/978-1-349-15430-2_11.
- Arthur, W. (2014). *Complexity and the Economy*. Oxford University Press.
- Arthur, W. B., Holland, J. H., LeBaron, B. D., Palmer, R. G., & Tayler, P. (1996). Asset pricing under endogenous expectations in an artificial stock market. *SSRN Electronic Journal*.
<http://dx.doi.org/10.2139/ssrn.2252>.
- Averbakh, I., Berman, O., Kalcsics, J., & Krass, D. (2015). Structural properties of voronoi diagrams in facility location problems with continuous demand. *Operations Research*, 62(2), 394–411. <http://dx.doi.org/10.1287/opre.2015.1354>.
- Balakrishna, N. & Lai, C. D. (2009). Bivariate normal distribution. In *Continuous Bivariate Distributions* (pp. 477–561). Springer Science Business Media.
http://dx.doi.org/10.1007/b101765_12.
- Biscaia, R. & Mota, I. (2013). Models of spatial competition: A critical review. *Papers in Regional Science*, 92(4), 851–871. <http://dx.doi.org/10.1111/j.1435-5957.2012.00441.x>.
- Brenner, S. (2001). Determinants of product differentiation: A survey. *Humboldt University*.
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.198.1240>.
- Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4), 434–455.
<http://dx.doi.org/10.2307/1390675>.
- Cabello, S., Díaz-Báñez, J. M., Langerman, S., Seara, C., & Ventura, I. (2010). Facility location problems in the plane based on reverse nearest neighbor queries. *European Journal of Operational Research*, 202(1), 99–106. <http://dx.doi.org/10.1016/j.ejor.2009.04.021>.
- Chatfield, C. (2016). *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.
- Cheong, O., Efrat, A., & Har-Peled, S. (2007). Finding a guard that sees most and a shop that sells most. *Discrete & Computational Geometry*, 37(4), 545–563.
<http://dx.doi.org/10.1007/s00454-007-1328-5>.

- Daskin, M. S. (2008). What you should know about location modeling. *Naval Research Logistics (NRL)*, 55(4), 283–294. <http://dx.doi.org/10.1002/nav.20284>.
- d’Aspremont, C., Gabszewicz, J. J., & Thisse, J.-F. (1979). On hotelling’s” stability in competition”. *Econometrica: Journal of the Econometric Society*, 47(5), 1145–1150. <http://dx.doi.org/10.2307/1911955>.
- Dehne, F., Klein, R., & Seidel, R. (2002). Maximizing a voronoi region: The convex case. In *Algorithms and Computation* (pp. 624–634). Springer. http://dx.doi.org/10.1007/3-540-36136-7_54.
- Eaton, B. C. & Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies*, 42(1), 27–49. <http://dx.doi.org/10.2307/2296817>.
- Ehrentreich, N. (2008). *Agent-based modeling: The Santa Fe Institute artificial stock market model revisited*, volume 602. Springer Science & Business Media. <http://dx.doi.org/10.1007/978-3-540-73879-4>.
- Eiselt, H. A. (2011). Equilibria in competitive location models. In *International Series in Operations Research & Management Science* (pp. 139–162). Springer Science & Business Media. http://dx.doi.org/10.1007/978-1-4419-7572-0_7.
- Eiselt, H. A. & Laporte, G. (1997). Sequential location problems. *European Journal of Operational Research*, 96(2), 217–231. [http://dx.doi.org/10.1016/s0377-2217\(96\)00216-0](http://dx.doi.org/10.1016/s0377-2217(96)00216-0).
- Eiselt, H. A., Laporte, G., & Thisse, J.-F. (1993). Competitive location models: A framework and bibliography. *Transportation Science*, 27(1), 44–54. <http://dx.doi.org/10.1287/trsc.27.1.44>.
- Eiselt, H. A. & Marianov, V. (2011). *Foundations of location analysis*, volume 155. Springer Science & Business Media. <http://dx.doi.org/10.1007/978-1-4419-7572-0>.
- Fekete, S. P. & Meijer, H. (2005). The one-round voronoi game replayed. *Computational Geometry*, 30(2), 81–94. <http://dx.doi.org/10.1016/j.comgeo.2004.05.005>.
- Ferreira, R. D. S. & Thisse, J.-F. (1996). Horizontal and vertical differentiation: The launhardt model. *International Journal of Industrial Organization*, 14(4), 485–506. [http://dx.doi.org/10.1016/0167-7187\(95\)00486-6](http://dx.doi.org/10.1016/0167-7187(95)00486-6).
- Fowler, J. H. & Laver, M. (2008). A tournament of party decision rules. *Journal of Conflict Resolution*, 52(1), 68–92. <http://dx.doi.org/10.1177/0022002707308598>.
- Graitson, D. (1982). Spatial competition a la hotelling: A selective survey. *The Journal of Industrial Economics*, 31(1/2), 11–25. <http://dx.doi.org/10.2307/2098001>.
- Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153), 41–57. <http://dx.doi.org/10.2307/2224214>.
- Kilkenny, M. & Thisse, J.-F. (1999). Economics of location: a selective survey. *Computers & Operations Research*, 26(14), 1369–1394. [http://dx.doi.org/10.1016/s0305-0548\(99\)00041-6](http://dx.doi.org/10.1016/s0305-0548(99)00041-6).

- Kress, D. & Pesch, E. (2012). Sequential competitive location on networks. *European Journal of Operational Research*, 217(3), 483–499. <http://dx.doi.org/10.1016/j.ejor.2011.06.036>.
- Laakso, M. & Taagepera, R. (1979). "effective" number of parties: A measure with application to west europe. *Comparative political studies*, 12(1), 3. <http://dx.doi.org/10.1177/001041407901200101>.
- Laver, M. & Sergenti, E. (2011). *Party Competition: An Agent-Based Model*. Princeton University Press. <http://dx.doi.org/10.1515/9781400840328>.
- Masuyama, S., Ibaraki, T., & Hasegawa, T. (1981). The computational complexity of the m-center problems on the plane. *IEICE Transactions (1976-1990)*, 64(2), 57–64.
- MathWorks (2016). *Floating-Point Numbers*. The MathWorks Inc. http://se.mathworks.com/help/matlab/matlab_prog/floating-point-numbers.html.
- Meko, D. (2015). Autocorrelation. Technical report, University of Arizona. http://www.ltrr.arizona.edu/~dmeko/notes_3.pdf.
- Plastria, F. (2001). Static competitive facility location: an overview of optimisation approaches. *European Journal of Operational Research*, 129(3), 461–470. [http://dx.doi.org/10.1016/s0377-2217\(00\)00169-7](http://dx.doi.org/10.1016/s0377-2217(00)00169-7).
- Prescott, E. C. & Visscher, M. (1977). Sequential location among firms with foresight. *The Bell Journal of Economics*, 8(2), 378–393. <http://dx.doi.org/10.2307/3003293>.
- ReVelle, C. S. & Eiselt, H. A. (2005). Location analysis: A synthesis and survey. *European Journal of Operational Research*, 165(1), 1–19. <http://dx.doi.org/10.1016/j.ejor.2003.11.032>.
- Rosenberg, N. (1982). Learning by using. In *Inside the black box: technology and economics* (pp. 120–140). Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511611940.007>.
- Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, 10(1), 141–156. <http://dx.doi.org/10.2307/3003323>.
- Shaked, A. (1975). Non-existence of equilibrium for the two-dimensional three-firms location problem. *The Review of Economic Studies*, 42(1), 51–56. <http://dx.doi.org/10.2307/2296818>.
- Smith, H. K., Laporte, G., & Harper, P. R. (2009). Locational analysis: highlights of growth to maturity. *Journal of the Operational Research Society*, 60(S1), S140–S148. <http://dx.doi.org/10.1057/jors.2008.172>.
- Strogatz, S. H. (1994). *Nonlinear dynamics and Chaos: with applications to physics, biology, chemistry, and engineering*. Addison-Wesley Pub.
- Weisstein, E. W. (2002a). *Bivariate normal distribution*. Wolfram Research, Inc. <http://mathworld.wolfram.com/BivariateNormalDistribution.html>.
- Weisstein, E. W. (2002b). *Normal distribution*. Wolfram Research, Inc. <http://mathworld.wolfram.com/NormalDistribution.html>.