An agent-based modelling approach to spatial economic theory

by

Dan Olner

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds

School of Geography

March 2014

For Sue

Paper cup of a boat
Heaving chest of the sea
Carry both of us
Carry her, carry me

Elbow

Acknowledgements

First, to my ever-patient supervisors, Andrew Evans and Alison Heppenstall: you managed the supervision equivalent of herding cats. Thank you so much for getting me through to the end in one piece. Thank you to Emanuel Gloor and Kirk Harland, who at different times carried out supervision tasks both before and after officially 'joining the team'. My research support group managed the unenviable job of making me focus in the right direction: Graham Clarke, Stuart Hodkinson and Evan Fraser. Thanks also to Evan for inviting me to join discussion groups in the School of Earth and Environment at Leeds. Members of the 'Multi-agent Simulation Society' in the School of Geography have offered wonderful insights and ideas over the years, and listened to many rambling presentations: thank you. Thanks to Malcolm Sawyer for so willingly offering advice on the economic issues, and to Andrew Neil Ross for putting up with my maths questions. Thank you to Matthew Handy for stoically dealing with my maths problems and for pointing me in the right direction. Andy Goldring's enthusiasm helped me on many occasions to remember why I was doing this, and fed me a plethora of ideas. My office mates made up an unofficial PhD support group: thank you for everything. Thank you to everyone in the School of Geography, and especially to Jacqui Manton who, among other things, made it possible for me to spend time working in Australia. Thank you to Professors Mark Birkin and Seth Bullock for examining the thesis and especially for all your support getting me through the final stages. And to every one of my friends who got me through those last unexpected months: I don't know where I would have been without you. Lastly, thank you to the ESRC for giving me the opportunity to do this PhD: I'm glad I don't have to do it again, but it's been the most amazing experience.

Lastly, this thesis is dedicated to Susannah Bird. For years, you never failed to be there for me. You were my unofficial supervisor and talked me through seemingly impenetrable PhD thickets. The many, many times when I couldn't see a way forward, you told me to trust you. You were heroically patient and kind, and you worked so hard to provide me with the space to get it done. Every day, you told me: 'you *can* do it'. You never stopped believing in me, even when I argued with you and said you shouldn't. I won't forget.

Abstract

This thesis presents an agent-based modelling framework, leading to a ‘transmission belt’ model connecting supply and demand in a spatial setting, as a step towards understanding the “complex, integrated, tightly-coupled global fabric of exchange” (Korowicz 2010 p.2) that makes up the modern spatial economy. Informed by a careful unpacking of economic modelling ideas, the ‘transmission belt’ model succeeds in producing a stable equilibrium of consumption and production across space. The motivation for creating the model framework is as follows. While some argue ‘it is better to assume that moving goods is essentially costless’ (Glaeser and Kohlhase 2004) because space costs are so low, others point out that a future of ‘oil depletion’ (Sorrell *et al*. 2009) and high carbon prices due to climate change mean that the implications of future cost changes are not well understood. The thesis examines the most prominent economic approach to space costs, geographical economics (GE), which finds ways to avoid key modelling problems imposed by space but has a tight ‘mathematical straightjacket’ (Martin 1999) of assumptions. This thesis keeps to simple utility functions for describing actors’ preferences, but uses an agent-based modelling approach to break out of these assumptions. While spatial agent models have dealt with a huge range of actor-environment interactions, very few examine traditional spatial economic problems. As a consequence, simple but powerful spatial economic ideas have been neglected. Much can be learned from issues that faced economists throughout the twentieth century. By closely examining these ideas, the thesis asks: what obstacles have stopped agent-based modelling from tackling the ‘big questions’ of spatial economics? The answers suggested are: a lack of research directly tackling the space of ‘dependencies’ in agent model development and a broader sense that it has broken all ties with history and can learn nothing from past modelling efforts.

Contents

List of Figures

List of Tables

Glossary

Bertrand competition A competitive situation where one firm can take the whole market by lowering their price a tiny amount.. 122, 137, 185

core model Shortened name for Krugman's core-periphery model of regional location.. 2, 3, 6, 8, 14, 17, 23, 25–29, 33, 35–41, 43, 44, 46–53, 55, 56, 58, 63, 64, 67, 71, 81, 83, 93, 104, 107, 108, 130, 131, 134, 137, 142, 145, 183, 186, 194, 196, 199, 201, 205, 207, 209, 215, 219–224, 228–230, 233, 235

Dixit Stiglitz model Monopolistic Competition model first published by Dixit and Stiglitz in their 1977 paper, 'Monopolistic Competition and Optimum Product Diversity'. 20, 33, 35, 37, 46, 48, 56, 86, 201, 215, 219, 221, 222

economies of scale First of the two forms of 'increasing returns' used in the thesis: a larger input into a production function means a more-than-proportional increase in output.. 22, 23, 25, 33, 35, 36, 112, 113, 133, 134, 137, 180, 185–188, 190, 195–197, 199, 201, 202, 205, 207, 209, 212, 215, 219–225, 227, 228

elasticity of substitution Another way of describing the economic dynamic described by the 'Constant Elasticity of Substitution' (CES) function. The other, inversely, is 'love of variety'. The more two goods are substitutable for each other, the higher the elasticity of substitution between them, and the lower the 'love of variety'. In this thesis, elasticity of substitution only applies to people's utility functions, not inputs into firms' production, so 'love of variety' is used.. 19, 20, 42

increasing returns Any situation where, for a given set of inputs, the output is more than proportionally larger. Increasing returns are thus non-linear. This thesis uses two types of increasing returns, 'love of variety' and 'economies of scale' (see other glossary entries).. 18, 19, 22, 23, 26, 33–36, 48, 58, 92, 108, 126, 134

love of variety Second of the two forms of 'increasing returns'. More formally, elasticity of substitution. High love of variety is low elasticity of substitution: for a given total quantity of goods, consumers prefer a mix of different varieties, and thus are less willing to substitute one for the other.. 17, 19–21, 31, 32, 35, 36, 38, 41, 42, 46–48, 50, 82, 102, 113, 142, 149, 173, 179, 188, 197–199, 201, 202, 205, 207, 209, 212, 215, 222–225, 228, 234

monopolistic competition In this thesis, unless otherwise stated, this refers to the approach to imperfect competition developed by Dixit and Stiglitz (1977) in which each firm is guaranteed by their increasing returns to produce one unique variety of good. They have a monopoly on that single good due to increasing returns, but there are enough firms in the market for the price-setting behaviour to not affect others.. 22, 27, 33, 48, 49, 56, 82, 83, 102, 103, 127, 224

representative agent Term used to describe the use of a single agent as a stand-in for all agents in an economic model. The 'core model' uses this.. 44, 46, 98, 99, 109, 221, 231

Acronyms

ABM Agent Based Modelling. 1–5, 7–11, 13, 14, 17, 28, 33, 35–37, 58, 72–78, 80–82, 84, 86, 90–92, 94, 97, 98, 100, 104, 109, 117, 136, 138, 145, 162, 215, 217–220, 224, 225, 228, 230–234, 236

ACE Agent-based Computational Economics. 10, 76–80, 97, 101, 102

AI Artificial Intelligence. 91, 101

CAS Complex Adaptive Systems. 10, 76, 77, 107

CES Constant Elasticity of Substitution. 19, 20, 22, 33, 38, 39, 46, 47, 49, 113, 122, 127, 137, 139, 141, 145, 146, 149, 151, 152, 167, 185, 202, 203, 218, 226, 229, 231, 233, 234

DMR Data Meaning Response. 6, 127, 135, 136, 139, 141, 177, 180, 184, 194, 197, 222, 227–230, 233

GE Geographical Economics. 2, 4, 6, 8–10, 12–14, 17–20, 26–29, 31–35, 37, 39, 40, 57, 58, 63, 66, 68, 71, 77, 78, 80–82, 102, 108, 109, 130, 183, 215, 217–220, 223–225, 230, 232, 233

OOP Object-Oriented Programming. 3, 5, 7, 8, 10, 71, 74–76, 84, 92, 100, 111, 128, 225, 232

Chapter 1

Introduction

This thesis presents an agent-based modelling framework as a step towards understanding the "complex, integrated, tightly-coupled global fabric of exchange" (Korowicz 2010 p.2) that makes up the modern spatial economy. The most pressing spatial economic questions require a way to explicitly model this 'fabric of exchange'. For example, in adapting to climate change and rapidly shifting energy futures, how will changing costs affect economic welfare? Are the spatial components of these changes well understood? Once understood in more depth, can this knowledge be used to aid the development of a sustainable energy and climate future?

This can be framed as a broader theoretical question: how do changing space costs (the costs of moving people and goods across distance) affect the morphology of the spatial economic landscape and how does the resulting morphology feed back into those choices?

This thesis examines the theoretical factors of this problem in depth by comparing agent-based modelling and economic ideas, using the comparison to build its model framework and presenting results from that framework. Agent-based modelling is highly flexible but its theoretical foundations are still developing. The assumptions of more traditional economic theory are brittle but it has a deep well of tested, nuanced theoretical knowledge. By examining both in depth, the thesis hopes to contribute to modelling approaches that capitalise on the best of both worlds.

1.1 Introductory discussion and aims

1.1.1 Agent-based modelling and spatial economics

In Agent Based Modelling (ABM), the 'agents' are distinct code objects, programmed to interact with their environment and each other. The actor-centred nature of ABM makes it perfectly placed for investigating some of the most fundamental questions about spatial economics. It should, in theory, be ideal for tackling two of the most challenging problems that space imposes.

First, it provides a way to model unique individual decisions and how they interact. The costs of moving people and goods across geographical distances change, as do the costs and benefits of proximity. As these changes take place, each agent makes many choices based on the unique set of factors their location imposes on them. Real-world agents have different locations; from where each of them stands, the costs they face differ.

Secondly, these interactions are not bought together into neat auctions that happen at a set time and place. These two points together break market clearing assumptions central to many economic analyses (section 3.3.1). The aggregate result of all these decisions over time is written into the morphology of the landscape as a pattern of settlements and dynamic flows. Isard identified these vital qualities of the spatial economy in the 1950s. As he put it:

> "Each individual commodity market including its labour, capital and land orientation possesses a particular structure which offers a certain resistance to change. Some change frequently, others slowly. Some are active, others highly inactive. By definition those markets of a relatively permanent nature, of persistent inactivity, are grouped together as the essential ones... Their combined structure determines the basic structure of the gestalt whole, of the space economy under question." (Isard 1956 p.39)

Using contemporary language, this 'gestalt whole' is the emergent structure of the spatial economy - something ABM is, in theory, primed for. Yet while there is a large body of ABM work examining a range of spatial and economic ideas, as chapter 3 discusses, ABM as a discipline seems to have stayed clear of tackling these more 'general' spatial economic problems.

To frame this as a question that might be suitable for an ABM research agenda: what happens to the spatial economy as a whole when costs change, given that space itself imposes heterogenous costs and rules out a single auction-style clearing market? This heterogeneity is *the* essential element of space and a tempting target for agent modelling. ABM can enable agents to make spatial decisions outside of those assumptions, independently of others' actions.

This thesis presents a set of spatial economic agent models leading to a 'transmission belt' model connecting supply and demand in a spatial setting. The goal is much narrower than Isard's 'gestalt whole': it succeeds at producing a stable equilibrium of consumption and production, but the process of getting there involves a series of modelling simplifications and compromises. This process is a key focus for the thesis, used to ask: what obstacles have stopped ABM from tackling this classic spatial economic problem? The development of the thesis model is used to address this question in two ways.

Firstly, the goal for the thesis has been to start from first principles: to create a model that explores one possible route to a foundation for the broad question above. That requires starting 'from the ground up': with agents able to make choices based on the costs they face in their particular time and place. The agents are kept as simple as possible to keep the focus on the more important factor (so it is argued) in this type of model: its detailed coding structure and the dependencies between components (see below). The separate components of supply and demand in a spatial setting are examined and methods used that best suit ABM.

Secondly: while building up from these first principles, another central goal has been to recognise and *use* the huge contribution made by older modelling methods and thinking, especially that of economics and the relatively new field of Geographical Economics (GE) that began with Paul Krugman's **core model** in 1991 (section 2.5). GE works in the 'general' tradition of economics: it "aims to explain the formation of various forms of economic agglomeration in geographical space using a general equilibrium framework" (Fujita and Mori 2005, p.377). Though an array of spatial economic ideas have been used (see chapter 2), the thesis pays particular attention to GE. It has two claims to be the most relevant modern approach to spatial economics, each of which makes GE very useful for the thesis. First, a highly self-aware approach to model-building and, due to this, a deep sense of the history of its discipline. Second, a canonical model (the core model) that attempts to boil space down to its absolute essentials, within the strictures of a neoclassical framework. The thesis 'disassembles' the core model to examine its workings and compare to the agent model built here (section 2.5).

Attempts to 'convert' elements of GE into an ABM approach do already exist; Christopher Fowler's work on this has been invaluable (see section 3.3.3). But where Fowler has gone from the core model towards ABM, the 'direction of travel' for this thesis is different. The starting point is actors themselves: the costs they face from their point of view and the decisions they make given those costs.

Using 'first principles plus history' provides a way to carefully examine the components involved in building an agent-based spatial economic model - not just in terms of practical coding choices, though that is a vital part of it, but also in theorising what those components and the model as a whole are doing and what they are for. It is also a way to find a balance between the common ABM rejection of neoclassical economic ideas in favour of complexity (section 3.3; see below also) and what Mirowski describes as a "respectful subordination to the neoclassical profession, discouraging anything that might have been perceived as wandering too far beyond the pale" (Mirowski 2007b).

Chapter 3 examines relevant methodological and modelling philosophy issues using this historical approach - but the use of history has not been an abstract process. These issues have arisen in attempting to solve real modelling problems during the process of model development. Milton Friedman's 1953 paper on the methodology of economics (Friedman 1953a) has been particularly useful. It has been dismissed by many social scientists and (as section 3.4.3 discusses) some agent modellers, but they seem to have caricatured its meaning in the process. As Hoover notes, Friedman was interested in "the quotidian practice of economics, not abstract epistemology" (Hoover 2009 p.303) and this makes it an excellent resource. This thesis is written in the same spirit: it is about the 'nuts and bolts' process of model-building.

The rest of the introduction is organised as follows. Each of these sections provides cross-references to the rest of the thesis to give an overview of the upcoming material:

- The following section summarises the above discussion into a set of aims and objectives, to be used as reference for the conclusion;

- Section 1.2 explains the models presented in the thesis, discussing how they map onto the standard economic distinction between 'partial' and 'general' models. It also explains how the concept of 'dependencies', borrowed from Object-Oriented Programming (OOP), is used to frame the thesis argument;
- Section 1.3 sets the context for the comparison of ABM and other economic ideas by examining how different theories work under different 'streetlights'; it also argues that ABM has a more acute problem than just the standard streetlight effect;
- Section 1.4 highlights some of the empirical reasons that more attention paid to agent approaches may benefit spatial economics, setting up a comparative discussion in the conclusion;
- Section 1.5 gives a brief chapter summary and an overview of terminology.

1.1.2 Aims and objectives

To summarise the above in the form of aims and objectives, this thesis' aims are as follows. Specific reference to these aims and objectives will be made in the conclusion (chapter 7). The aims are:

- **Develop an agent-based spatial economic model framework in which agents with heterogenous locations can optimise welfare given changing costs**
- **Use the development process of the model framework to answer the following: what obstacles are in the way of ABM exploiting its potential as a tool for spatial economic analysis?**

The following objectives are connected to the first aim:

- Develop a spatial agent model framework and explain its workings in detail in the thesis;
- Compare agent-based and economic approaches to model-building and use this analysis to inform the thesis' model development.

The following objectives are connected to the second aim:

- Identify obstacles revealed in modelling development at the coding level and a route through them;
- Identify obstacles revealed by the comparison of different approaches to model-building in ABM and economics, with specific emphasis on GE.

The following chapters describe the route the model found through these obstacles at the coding level. The details of the obstacles themselves are raised throughout, and these are then bought together for a synthesised analysis in the conclusion. The second 'obstacles' objective is more theoretical: the important differences in approaches to model-building, it will be argued, are about how ABM has developed a dysfunctional relationship with assumptions.

To pre-empt the conclusion, two key obstacles have been identified: a lack of research directly tackling the space of 'dependencies' in ABM model development (see section 1.2 below); a broader sense that ABM has broken all ties with history and can learn nothing from past modelling efforts (see section 1.3). The conclusion analyses these obstacles in depth, building on what has been presented throughout the thesis, and presents an analysis of the model framework's dependencies and assumptions in order to highlight some possible ways forward.

1.2 The models

A series of agent-based spatial economic models are presented, built with a common framework. Chapter 4 explains this framework in detail. As well as a series of testing models (in the first sections of chapter 5) the thesis contains three main model set-ups. Note that capitalised 'People' and 'Firms' are used to describe model object representations, as per the OOP norm of capitalising objects, to distinguish them from 'people' and 'firms' generally. These three modelling scenarios are:

1. 'People' agents are mobile, able to optimise their locations as part of their larger economic decision set, with Firms' locations static (section 5.4).

2. 'Firm' agents are mobile, able to seek revenue-maximising positions in relation to an immobile market of People (section 5.5).

3. A 'transmission belt' model linking supply to demand (chapter 6): demand signals from People feed through to Firms. Using price signals, Firms are able to reach a market-clearing level of stock production that takes advantage of production-level economies of scale and allows for agents with heterogenous locations - thus keeping the 'transmission belt' between supply and demand turning (see below for more on this).

The first two model set-ups examining mobile People and mobile Firms are 'partial' and the transmission belt model an attempt at a 'general' analysis; the next section explains this distinction.

What is the rationale for these two 'partial' scenarios? They have been chosen because they provide the most insight for understanding assumptions and dependencies. During the process of model development, many options were tested. Some failed to work at all and some partially worked; a number of these are discussed in section 6.2. The two presented in the thesis prior to the transmission belt model are the most informative examples.

The ‘mobile People’ models achieve two things. Firstly, they are used for testing a ‘density cost’ proxy. This is a very simple way to consider the effect of a range of proximity costs from congestion to land markets, very suited to an ABM approach as the cost is centred at each agent’s own location. Section 5.4 explores its ability to produce a range spatial equilibrium outcomes, while section 2.7 provides some theoretical rationale. These spatial outcomes were a surprise, but a very useful one, providing an effective example for thinking about assumptions in ABM.

Second, the ‘mobile People’ models show that (within the model framework) commuting and the movement of goods cause the same agent reaction when changed. Equally, the ‘base’ part of good costs (that is, without any spatial component) and ‘base’ wage cause the same reaction as each other. ‘Commuting’ is thus reframed as the spatial part of the wage, in the same way that goods have a spatial cost and a base cost. This idea is useful in the transmission belt model (see below).

In the ‘mobile Firms’ model, two Firms attempt to maximise their revenue by optimising their location on a line of People. Creating Firms able to optimise good prices and wages was one of the problem areas for the framework; these models use the more transparent case of location optimisation to explore why this optimisation was so difficult. It uses an idea adapted from Jane Jacobs (2001), framing Firms’ decision process in terms of ‘Data Meaning Response (DMR)’ (see section 4.7.7). The main lesson is that independently acting agents in a spatial setting can cause signalling confusion that make emergent spatial outcomes less likely. Using a canonical GE utility function and a noise function, an analysis of success and failure is given.

The transmission belt model presents one solution that allows agents to coordinate economic activity to a successful dynamic supply/demand equilibrium, capable of moving between production ‘regimes’ where more-efficient producers can emerge. Section 2.2.5 gives the production function used; section 4.7.5 explains how problems of strategy were avoided in the coding of production. Reframing person-moving and good-moving costs into separate spatial and non-spatial elements helped by providing a rationale for People to buy goods directly with time (or labour) from a range of Firms. Thus a single distance cost, rather than two (distance moved by goods and commuting) can be varied to investigate its impact. The DMR investigation leads to a novel way of allowing price signals to work that avoided a range of the thornier timing issues (discussed in chapter 4). Leijonhufvud’s concept of economic ‘laws of motion’ (section 3.3.1) has been adapted, with the addition of oscillation damping.

1.2.1 ‘Partial’ versus ‘general’ models

General equilibrium models (of which the core model is a geographical example) must link supply and demand to produce an equilibrium outcome that can be claimed to represent a whole economy. ‘Partial’ models concentrate on sub-components of this, looking only at supply or demand and keeping much of the model fixed by assumption.

Overman explains the importance of the difference between partial and general:

> "Time and again economists have found that this is vitally important because partial equilibrium reasoning (just looking at a small part of any problem) often fails to provide the right reasoning in a general equilibrium context (when taking the economy as a whole)" (Overman 2004 p.506).

Birkin and Wilson define the difference between 'partial' and 'general' *location* models similarly:

> "Partial theories are concerned with the location of a small number of production centres in relation to a given distribution of other centres and of markets. General location theory, on the other hand, aims to determine simultaneously an optimum distribution of all production centres and markets." (Birkin and Wilson 1986 p.178)

Combining 'general' theories with the heterogeneity imposed by space is challenging. This problem has been usefully avoided in neoclassical economics through *a priori* equilibrium assumptions. Analytic approaches can use models that assume an equilibrium endpoint has been reached. Agent models, however, are stuck trying to find a path to stable outcomes through interaction, unable to rely on those assumptions. ABM as a field has taken this coordination problem to be one of its main research aims: showing how 'emergent' economic phenomena can arise through interaction. Fowler captures the tangle of relations involved in these economic interactions:

> "the amount a firm can offer in wages is dependent on the amount it can produce and the price it receives for its goods. These quantities are affected, in turn, by consumer's wages, which depend on which firm employs them (and whether or not they have found employment at all.)" (Fowler 2007 p.277)

The transmission belt model finds an agent-based way to link economic production and welfare in a spatial setting. Agents coordinate through decentralised actions. Production takes place and this feeds through into collective welfare. For the thesis, 'welfare' refers to the aggregate level of utility agents are able to get. The term is useful for distinguishing discussion of model-wide optima in comparison to the utility of individual agents. The transmission belt model equilibriates supply and demand and can move between production equilibria as distance costs and utility are varied. In doing this, a range of varied prices can emerge based on differences in both location and production scale: no 'market clearing price' is necessary for the equilibrium to sustain. As with the other models, however, there were compromises to achieve this end; these are discussed in-depth in the chapter presenting the model (chapter 6).

1.2.2 Dependencies

The thesis argues that model 'dependencies' are unavoidable. In the process of attempting to move from a 'partial' to 'general' supply-demand link, the issue of dependencies arose naturally. Choices made in one area of a model determine those required in others. These choices are often mutually exclusive: one set can rule out another. This language of 'dependencies' is taken from OOP. In OOP terms, dependencies are "the villain of the piece"[1]: the goal is to achieve 'loose coupling' (see e.g. Jana 2005) where objects are, as far as feasible, not reliant on any particular code choice elsewhere in a program. This philosophy, while useful for developing code itself, presents problems when applied to the process of modelling social systems. As section 3.4.4 discusses, it helps tilt ABM towards a 'virtual world' philosophy (see below also). There is a sense that if good OOP practice has been followed, loosely coupled agents can interact as independent beings in their provided environment and a pseudo-empirical world *in silico* will naturally result, leaving the modeller an observer-god of their own creation.

While an OOP concept, the language of dependencies is a very useful way to see the connection between older and more modern modelling methods. The spatial issues described above are, in reality, dependencies: spatial economics imposes a series of strictures on what modelling assumptions can be used (these are explained in section 2.3). ABM's error has been in thinking it can transcend these dependencies.

In economic approaches like GE, these kinds of compromises, simplifications and choices are accepted and openly discussed. Section 2.5 explains the core model in detail; as with all economic models, it is a tightly knitted set of arguments built on mathematical links in a chain - all fully 'dependent' on each other. It could be argued this is anathema to an ABM way of working. However, this thesis has provided an opportunity to explicitly examine how dependencies arise in the logical journey of model framework development. It has involved threading a path through these dependencies and understanding the modelling compromises required to do so. Rather than being an incidental obstacle on the path to creating an ideal ABM 'virtual world', this thesis argues that these dependencies are an important part of the actual meaning of any model.

Section 3.4.1 makes a distinction between 'descriptive' and 'functional' mapping in models. Because of its deep roots in OOP, ABM has developed a implicit philosophy of mapping the real world descriptively onto code structures. As chapter 3 argues, the pursuit of good OOP practice in framework development is one thing; applying the same OOP principles to the modelling philosophy itself quite another and may in fact damage ABM as a social modelling approach.

By carefully breaking down the elements of an agent-based spatial economic model and comparing to the philosophy and methodology of GE, this thesis hopes to show that the mapping of agent models is tangled into the minutiae of the code itself. OOP is still hugely powerful and, if carefully applied, can indeed be mapped onto the structures the modeller is interested in. But much of the mapping happens into the spaces in between the code where implicit meaning can

[1] Scitovsky (1954 p.144), discussing the role of externalities in neoclassical economics; see section 2.7.

exist.

These dependencies and compromises determined the form of each set of models. A clear examination of optimal Person choice versus optimal Firm choice requires fixing the locations of the other agent type in their respective models. By enabling and disabling certain aspects of the model framework, it is able to produce a series of spatial economic results that resemble classic spatial economic model outcomes (see section 2.4) while also displaying novel behaviour.

This section ends with a note on the way the model framework has been used. The vast majority of the model outputs presented in the thesis run ‘live’, with selectable parameters the user can interact with (see the appendix for more information on this). This is not incidental to the message of the thesis. Once a framework is tested enough to trust that it is robust, model behaviour can be quickly tested with this interactive approach and novel behaviour discovered. There is a cycle of experimentation and deduction. (This is discussed below also, in relation to the role of computation.)

There is a vital caveat to this. As section 3.4.7 discusses, the strong ABM tendency to treat emergence and explanation interchangeably means that, often, agent analyses stop at the point a model appears to do something interesting. This should be just the start, however. The discovery of interesting model properties raises the question, *why* is it happening? For the results presented here, this led to a deeper analyis of the underlying mechanism, often relying on a more analytic approach or creating simpler examples with complex assumptions stripped away.

To the extent that this found useful answers, these did not come solely from either the interactive agent framework, analytic calculations or more basic ‘toy’ models, but an iterative combination of all of these, with each playing a key role. The role of visual interaction in *discovering* potentially interesting dynamics was essential, however.

1.3 The streetlight effect

ABM has positioned itself as the natural successor to what it sees as the ‘moribund Newtonian world-view’ of neoclassical economics (see section 3.3.1). At present, however, ABM is stuck in a limbo of its own creation. Its methodology has played too strong a role in determining its research agenda. This is related to a common methodological problem called the ‘streetlight effect’. As Kirman explains, this “corresponds to the behaviour of the person who, having dropped his keys in a dark place, chose to look for them under a street light since it was easier to see there” (Kirman 1992 p.134). The streetlight effect shapes what answers can be found as well as what questions can be asked. Maslow calls this ‘method-centring’ in contrast to ‘problem-centring’ (Maslow 1966 p.15); he summed it up in the now-famous aphorism: “it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail” (ibid).

But ABM seems to have a particularly severe case of this problem - to a greater extent than the economic ideas it wants to supersede. Its methodology has not merely restricted what it can 'see' but has heavily influenced its research questions. In combination with a prevalent rejection of everything neoclassical, this has left spatial economics surprisingly empty of agent-based analyses.

In contrast, GE is well-aware of its own limitations. Fujita *et al.* explain that GE works with -

> "... assumptions that reflect not so much a realistic view of how the world works as a judgement about what will make the analysis of geographic issues manageable without doing too much damage to the relevance of that analysis." (p.6)

More recently, Krugman - discussing his current approach to macro-economic modelling - notes that it:

> "rests on some patently untrue assumptions about reality... models are an enormously important tool for clarifying your thought. You don't have to literally believe your model - in fact, you're a fool if you do - to believe that putting together a simplified but complete account of how things work... helps you gain a much more sophisticated understanding of the real situation. People who don't use models end up relying on slogans that are much more simplistic than the models." (Krugman 2010a)

Overman also acknowledges "the tendency to privilege particular economic forces purely because they are more amenable to the theoretical and empirical tools used by mainstream economists" (Overman 2004 p.504). Summer notes the result: "it is all too easy to confuse what is tractable with what is right" (Summers 1991 p.145). Krugman, as is often the case, puts it best: "the methodology of economics creates blind spots. We just don't see what we can't formalise" (Krugman 2005). The result is that clear empirical realities often find themselves ignored if they fail to fit into existing theory. GE's openness about this is one of its main strengths: it imbues the approach with a sense of caution.

ABM theorists, on the other hand, often argue they are immune from the streetlight effect by default. The promise of a much deeper realism aids this sense that it has broken free of the 'mathematical straightjacket' (Martin 1999) of neoclassical theory. However, this thesis argues that ABM has developed a unique straitjacket of its own making. The methodology of OOP itself, as well as the way Complex Adaptive Systems (CAS) theory perfectly marries to OOP, are at least partly the cause. This is not a new thought: to mix metaphors, some argue the Santa Fe Institute has been overly dominant, "retailing the snake oil of 'chaoplexity' " (Mirowski 2007b, p.361 paraphrasing Horgan 1998). That is going too far: the ideas within the CAS canon have clearly opened up huge new research avenues and continue to be fruitful. The complexity agenda has been driven by a desire to understand social systems through treating them as composed of interacting parts. However, in Agent-based Computational Economics

(ACE) (an economic sub-discipline of ABM; see section 3.3.1) this has dampened the pursuit of many of the traditional economic research problems, leading to it carving out a quite separate theoretical territory.

Two recent examples highlight this notion that ABM has the power to declare a theoretical 'year zero'. Reflecting on the nature of agent models, one concludes that ABM is coming into its own in a time when "the classical canons of scientific inquiry melt away into the vestiges of history" (Batty 2012 p.4). Another argues:

> "half a century ago, the idea of a model was in its infancy. Scientific theory essentially was based on formal and systematic theories, often represented mathematically, whose testing was confined either to controlled experiments in the laboratory or to various categories of thought experiment. Computation changed all that. The idea that a scientific theory could then be translated into an intermediate form - called a model - represented a way of enabling controlled experiments to be carried out not on the actual system of interest but on a computable abstraction of that system." (Batty *et al*. 2012 p.21)

The idea of this kind of model goes back to at least the turn of the twentieth century, and possibly to Bacon[2]. But this notion that a *computational* model is something historically unique ties to the belief that computers have allowed modelling to transcend that history. It is irrefutably true that computational modelling has radically transformed research. As Berlinski puts it:

> "The computer has changed the very nature of mathematical experience, suggesting for the first time that mathematics, like physics, may yet become an empirical discipline, a place where things are discovered because they are seen." (Berlinski 1997, quoted in Gold and Simons 2008 p.36).

Beyond the hyperbole of this quote there is a useful seed of truth. This is precisely the point made above about the cycle of experimentation and deduction: computation has made that possible. For ABM, though, this power can be a trap, precisely because it empowers the modeller to create very plausible-seeming 'virtual worlds' (section 3.4.4) that are both compelling for the researcher and appealing to those outside of the discipline. In its place, compelling or appealing ABM work are not bad things - but they must not be confused with the scientific job that models can do.

As section 3.4.4 discusses, Friedman saw exactly the same tendency in his day: a clash between attempts at 'photographic reproduction' versus 'engines of analysis' (Friedman 1953a p.35). This thesis argues that the fundamental nature of models has not changed since Friedman

[2]The Oxford English Dictionary has references from 1901 for 'model' as "a simplified or idealized description or conception of a particular system, situation, or process, often in mathematical terms, that is put forward as a basis for theoretical or empirical understanding, or for calculations, predictions, etc.; a conceptual or mental representation of something" (OED 2002)

wrote this. The classical canons of scientific inquiry are most emphatically not melting away. This is a positive thing for ABM: as this thesis explores, it means that the rich well of thinking about modelling in economics is still highly relevant.

The models in this thesis are only asked to pass 'face validation' (Oeffner 2008 pp.35): does it work as a self-consistent set of relations, able to explore the question of how changing costs affect actors in a heterogenous, abstract geography? Is it a useful tool for "probing the internal consistency of a theoretical position?" (Di Paolo *et al.* 2000 p.1). The impact of these differences in modelling philosophy affect every facet of model building, however. For this thesis, they impact most keenly on the way dependencies are considered and how they relate to assumptions about reality. The issue is examined by comparing how agent modellers and earlier economic theorists have thought about what they are doing. What do the two approaches think their models are 'explaining'? How do they think about simplicity and realism? The way model assumptions are theorised is key. There are very few theorists without a personal pet-hate assumption that they see as - in Hayek's phrase - self-evidently disregarding "everything that is important and significant in the real world" (Hayek 1945 p.530). No assumption can be dismissed so easily since what is 'important and significant' is in the eye of the beholder. In assessing assumptions, "everything depends on the problem" (Friedman 1953a p.7).

1.4 The spatial economic 'problem'

As discussed above, many theorists of have argued for the pursuit of a 'general theory' of the spatial economy capable of understanding the 'gestalt whole' of its interactions. While the focus of this thesis is methodological, not empirical, this section outlines some of the empirical motivation for pursuing new spatial economic ideas.

Spatial economics has developed during a particular historical period of constantly dropping costs through both an increase in production efficiency and a drop in transport costs. Modelling approaches have been shaped by this, sharing much the same general view of space costs as Glaeser and Kolhase, who have claimed:

> "reduced costs, and the declining importance of the good-producing sector of the economy, means that in our view it is better to assume that moving goods is essentially costless than to assume that moving goods is an important component of the production process" (Glaeser and Kohlhase 2004, p.199).

They also argue that "there is little reason to doubt that this decline will continue" (Glaeser and Kohlhase 2004, p.197). This assumption of space costs inevitably dropping is a widespread part of the literature. Lang points out that GE models aim to:

> "examine how the spatial distribution of economic activity changes as transportation costs slowly decline. The slow decline is meant to mirror the actual decline

> in transportation costs observed over the modern era of civilisation." (Lang 2010 p.191)

As Neary notes (Neary 2001 p.536), this is a bold claim of telling the 'history of the world part one', spanning "first caravels, then steamships and railroads, then air freight..." (Fujita *et al*. 2001 p.253) This is a somewhat 'Whig history' approach that assumes the direction of costs is inevitably one of decline - but is that a problem?

Twenty-first century challenges are breathing new life into spatial economic questions. Theorists are rediscovering that spatial costs may go up as well as down. During the next few decades an energy revolution must take place to avoid the worst effects of climate change. What price must carbon be to keep within a given global temperature? How long will any switch to new infrastructure take? (Kramer and Haigh 2009; Jefferson 2008) Will the cost of fossil fuels impact on the spatial economy as much as climate change? (Wilkinson 2008; Bridge 2010; see Sorrell *et al*. 2009 for a thorough analysis of the issue of oil depletion). More broadly, what role might a deeper understanding of spatial economics play in reshaping a post-carbon economy?

The more qualitative approach of 'economic geography' examines many of these questions. The idea of 'transition' - the need to consciously move to a low-carbon economy through planning and civic engagement (Hopkins 2008) - has gained a great deal of research territory in recent years. But it brings many assumptions with it. Economic analysis is eschewed; in fact it is common to find a polar view of an " 'Industrialised World' of standardised-generic production... associated with commercialism, efficiency and branding" versus an " 'Interpersonal World' of specialised-dedicated production... associated with trust, local renown, and spatial embeddedness" (Morgan *et al*. 2006 p.22). Many geographers have already taken sides in a battle where localities fight to 'respatialise' social, economic and environmental goods lost to globalisation (Glasmeier 2007).

While many see 'relocalisation' as the self-evidently correct normative response to a perceived future of rising space costs, the most common view from the economics profession is summed up nicely by Harford: 'buying local is simply a lifestyle choice - not the difference between environmental salvation or damnation' (Harford 2007). The problem with accepting Harford's argument at face value is - as discussed in detail in this thesis, and argued by Paul Krugman (see section 2.3) - economists have been spectacularly good at ignoring geography, because the 'dependencies' involved in modelling space makes many central economic assumptions unworkable. As section 2.3 points out, critiques of this "wonderland of no dimension" (Isard 1956 p.26) go back at least to Isard.

'Transition' analyses argue that rising oil prices will have severe impacts, both spatial and economic. Economic takes on oil price shocks, though in-keeping with the economic avoidance of space, raise questions about this assumption. Oil price *volality* may be more important than price change itself (Bachmeier *et al*. 2008 p.528; see section 2.8.3 also), though, matching an argument often also used by transition thinkers, recent oil price effects on economic output may by "caused by strong demand confronting stagnating world production" (Hamilton 2009 p.215).

The oil shock economic literature also offers alternative theories to 'transition' accepted truths. For example, in the transition literature, it is commonly suggested that oil price rises correlate to recessions, thus indicating its central role as the lifeblood of the economy. However, Segal suggests that the oil-price/recession connection has been driven mostly by monetary policy response: raising interest rates when high oil prices feed through to inflation, thus helping to cause recession without necessarily addressing the root cause of the problem (Segal 2011 p.169-170).

A quantitative spatial modelling approach may help shed some light on these different views - and indeed, there are relevant quantitative examinations of aspects of the 'transition' problem. For example, Kerschner and Hubacek, in a very rare occurrence, actually take on the 'oil depletion' question directly (Kerschner and Hubacek 2009). The MARKAL model approach (see e.g. Seebregts *et al*. 2001) attempts to integrate national energy economics into one framework. As with the oil shock literature, however, these avoid any of the fundamental distance effects that changing space costs cause. This throws into sharp relief the fundamentals of spatial economics being missed: it is a web of interconnected agents making decisions separated by both time and space, in which changing costs can alter morphology, feeding back on those costs to create Isard's 'gestalt whole'. GE, aware of its own limitations, makes no claim in this regard. No ABM analysis directly addresses it, leaving these questions to qualitative geographers. This thesis hopes to open up some space for exploring quantitative approaches to the issue using agent modelling. This will be returned to in the final section examining future work (section 7.5) where it will be asked: is ABM well-positioned to deal with understanding how the 'gestalt whole' spatial economy changes?

1.5 Organisation of the thesis

1.5.1 Terminology

A glossary and list of acronyms is provided after the table of contents. The first time the glossary terms appear in the thesis, they will be in bold. The first time acronyms are used, the full phrase will precede them.

A brief note on the terminology used to describe fields of research. 'Spatial economics' is used as a top-level heading to describe all theories that attempt to say something about how economic activity and geography interact. GE has already been mentioned: it is a specific approach beginning with Krugman. As Brakman *et al.* say, it is "an attempt to put more geography into economics" (Brakman and Heijdra 2011 p.xxii). GE's other common name, the 'new economic geography', is no longer really apt since it is twenty years old now (ibid). Confusingly, 'geographical economics' is completely different to 'economic geography': the latter a much broader church, often qualitative, case-study-based and empirical, and sometimes openly hostile to GE. Almost all current academic work on 'transition' issues comes under the 'economic geography' banner. This point is discussed more in section 2.3.

1.5.2 Chapter outline

Chapter 2 explains the key economic and spatial ideas used in the thesis. It presents the utility and production functions chosen to build the model. The theoretical context for the economic and spatial ideas used is given. The core model is given its own section. The way in which the thesis theorises the costs of moving people and goods[3] across space is presented. The chapter ends with discusion of some strategic simplifications.

Chapter 3 has two aims. Firstly, it introduces the idea of ABM in more detail. Secondly, it looks critically at how ABM has developed and puts it in the context of some key ideas from more traditional economic modelling theory. The chapter ends by discussing the use of utility and production in an agent modelling context.

Chapter 4 explains how the model framework has been put together. This covers: how the model setup works; how actors have been set up; an explanation of the model scenarios used; finishing with a more detailed discussion of some of the key framework areas that need more background detail.

Chapter 5 firstly presents a series of testing models to demonstrate the framework achieves its fundamental economic outcomes. It then presents the two 'partial' model set-ups, examining 'mobile People' and 'mobile Firms'.

Chapter 6 presents the transmission belt model. It begins by discussing how the model resulted from a series of strategic choices; this section also outlines some of the approaches that are excluded by the the particular set of dependencies the model has. The transmission belt model is then put through a series of runs, each introducing successive elements: a 'damping' method for successfully finding prices; non-spatial models showing the equilibrium outcome; an examination of the interaction of 'love of variety' and 'economies of scale'; and a series of models introducing a spatial element.

Chapter 7 concludes by discussing the aims and objectives in the light of the thesis model results. It then examines the set of specific dependencies and assumptions revealed through the model development process and uses them as the focus for highlighting both the limitations of ABM and new opportunities these limitations reveal.

The appendix provides a detailed overview of the model code structure and a full description of the downloadable model location and contents.

Next, chapter 2 outlines the economic and spatial ideas used in building the model.

[3]Glaeser talks of cities eliminating "transport costs for goods, people and ideas" (Glaeser 2008 p6.). This thesis sticks to goods and people.

Chapter 2

Economic and spatial concepts

2.1 Introduction

This chapter outlines the economic and spatial concepts central to the thesis model framework. Section 2.2 begins laying out the foundation for the thesis model approach by explaining the utility and production functions that are used. The economic concepts in the thesis model framework have a strong connection to those used in GE. The **love of variety** utility function is central to the thesis, providing agents with a generic way of demanding goods across space while being able to control the mix of that demand. Instead of using the GE approach to production, however, a different function is used that allows the transmission belt model to successfully link supply and demand and move between production equilibria.

Later sections introduce spatial economic ideas. These begin in section 2.3 by combining an overview of the 'big questions' of spatial economics with an examination of the particular way that GE has approached these problems and the role that dependencies and assumptions play. This section also discusses a number of other classical location theory ideas relevant to the models presented in sections 5.4 and 5.5.

Section 2.5 explains the core model of GE in detail, with the aim of carefully disassembling its 'engine' and laying its parts out in a logical order. The purpose of disassembling the core model is to aid the comparison of model-building methodology - and also philosophy - between GE and ABM. Chapter 3 concentrates on the more philosophical side of this problem and discusses the tendency of much ABM to be rather too literal in its interpretation of engine-building.

After this, sections 2.6 and 2.7 present context for the choices made in the thesis model framework. These are based on breaking down the catch-all term 'space costs' into two elements: 'distance costs' (the costs of moving people and goods across space; section 2.6) and the 'costs and benefits of proximity' (section 2.7): that things being nearer to each other may be both beneficial or detrimental to them in ways pure distance costs cannot capture.

The chapter ends with an outline of the concepts that have been deliberately placed outside the scope of the model framework. This sets up the discussion for 'strategic simplifications' in the following chapter.

2.2 Fundamental economic concepts

2.2.1 Factors

At the most fundamental level, economic action is all about exchanging factors for other factors. Factors are often defined as broad abstractions. The most traditional of these is to identify capital (K) and labour (L) as the two inputs into aggregate production, where capital is a catch-all term variously meaning everything from factory plant to finance (Robinson 1971 p.597). Arguments abound about the validity of this; some are discussed in chapter 3. The abstraction most common to GE is to separate manufacturing (M) from agriculture (F), although, as will be explained, the distinction comes down to which sector has **increasing returns** and which is the unchanging base for comparison as values change.

The utility and production functions outlined below allow people and firms to get the most out of the factors at their disposal[1].

The most important point to start with is that actors' time is given prominence. In GE, this is equivalent to the labour they put into production. Section 2.3.5 looks at one particular critique of the way factors are treated in economics. The next section's outline of the utility functions used in the thesis will explain in more detail how factors are used.

2.2.2 Utility

At its simplest, utility is a very effective way of describing when an economic actor is better or worse off. It can abstract away from any details about the nature of 'better off' in any psychological sense. This is a strength of utility theory, though it is often portrayed as a weakness. Chapter 3 examines these arguments in more depth, since they impact on the way that agent modellers have chosen to approach economics.

As with GE, the forms of utility used here are all reasonably simple and tractable. The simplest of these is the Cobb Douglas. A consumer gets utility from two goods, labelled F and M. (For consistency, the variable names here are the same as those used later when discussing geographical economics.)

$$U = F^a M^b \tag{2.1}$$

The sum of a and b control the effect of scale. Summing the exponents to one gives constant returns to scale, whereas if $a+b > 1$, the function has increasing returns: increasing the amount of F and/or M causes a greater than proportional increase in utility (see section 2.2.5 for a full

[1] It is an axiomatic economic idea that actors want to do this, though not a problematic one. This is discussed in section 3.5.

explanation). Most commonly it is assumed that, for utility, $a+b=1$. For this, an equivalent function can be used:

$$U = F^{\delta} M^{1-\delta}, (0 < \delta < 1) \tag{2.2}$$

This Cobb Douglas is particularly popular because of the convenient relationship that exists between a consumer's budget and the optimal mix of F and M. Let Y be the consumer's budget. Assuming the cost of both F and M is unity (they both have a price of one), total consumption cannot be more than the budget:

$$F + M = Y \tag{2.3}$$

What, then, is the optimal mix of the two goods for the consumer? Maximising 2.2 with the budget constraint 2.3 gives -

$$F = \delta Y \tag{2.4}$$

$$M = 1 - \delta Y \tag{2.5}$$

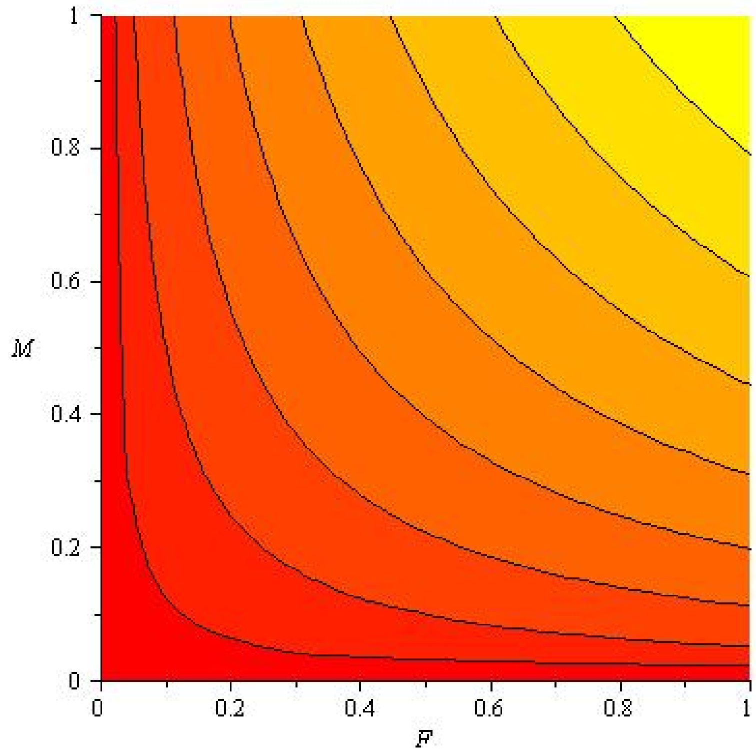


Figure 2.1: Cobb Douglas function, δ is 0.5. Contour lines show isoquants. Lighter colours are higher utility values.

The Cobb Douglas exhibits, in the simplest way possible, a basic feature used by GE: love of variety - a mix of the two goods is preferable to either good alone; see below for a fuller description of this concept. The optimum mix can be controlled with δ, which can be referred to as the 'share parameter'. Figure 2.1 shows contour lines when $\delta = \frac{1}{2}$, indicating isoquants: lines of equal utility, where a consumer would be indifferent between the mix of F and M, if their budget didn't matter. However, the budget does: according to (2.4) and (2.5), their optimal choice if $\delta = \frac{1}{2}$ is equal. That is, $F = M$ - a positive, straight line. So, for example, if the budget $Y = 10$, their utility-maximising mix would be five for each good. These preferences would also be homothetic - if the budget changes, the proportionate mix of optimal goods remains the same: F always equals M.

The strength of this simple function, then, is that δ can so easily be used to find the optimal mix of two goods given any budget. The next section looks at a slightly altered Cobb-Douglas where income level does matter.

2.2.3 The CES function and 'love of variety'

The Constant Elasticity of Substitution (CES) function, and the economic ideas it represents, are used heavily throughout the thesis. Love of variety and **elasticity of substitution** are two sides of the same coin: a few moments on the intuition of what they represent will help. If goods are completely substitutable (and thus have the highest possible elasticity of substitution) consumers will not mind which they have: each gives them the same utility. As elasticity of substitution drops, however, love of variety comes into play: they increasingly prefer a mix of goods to a smaller range (while consuming the same overall quantity). At a very different scale, consumers prefer - in aggregate - a choice between (for example) four models of car than two, though in total they buy the same number of cars. (This example is returned to in section 2.5.) To the extent that consumers prefer a mix, they have high love of variety and low elasticity of substitution. The more indifferent they are between goods, the higher the elasticity of substitution and the lower the love of variety.

The CES function is able to represent this dynamic with one parameter. First defined by Arrow et al. (Arrow *et al.* 1961), it is central to GE. It is an economic swiss army knife; "perhaps the most frequently employed function in modern economic analysis" (Durlauf and Blume 2008 p.737) and is used to represent both production output and utility. The first appearance of the CES function, in its two-input form, saw it being applied to cross-country data. Dixit and Stiglitz give a clear illustration of the basic principle, using a function able to reflect that:

> "... a consumer who is indifferent between the quantites $(1,0)$ and $(0,1)$ of two commodities prefers the mix $(\frac{1}{2},\frac{1}{2})$ to either extreme." (Dixit and Stiglitz 1977 p.297)

It is easiest to see how this works in a simple two-good function that looks like Dixit and Stiglitz's example, before moving on to the n-good case. Consider a consumer with the utility

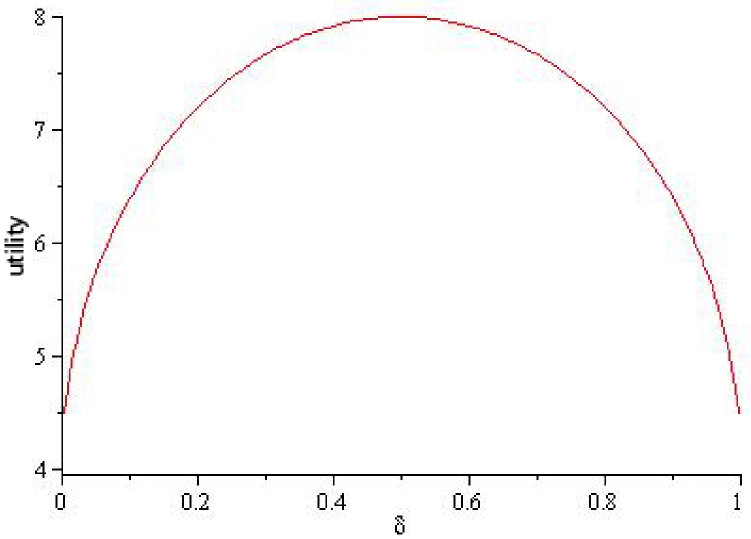


Figure 2.2: Two-good CES function, $\rho = \frac{1}{2}$, constrained by δ, $0 < \delta < 1$

function:

$$U = [F^{\rho} + M^{\rho}]^{1/\rho} \tag{2.6}$$

The first thing to note is that, if we suppose $M = 0$, and only F has a positive value, then $U = (F^{\rho})^{1/\rho}$. This is always just $U = F$, regardless of the value of ρ. If both goods have positive values, and $\rho = 1$, this reduces to $U = F + M$ and utility is also entirely linear. But for $0 < \rho < 1$, the consumer is increasingly better off with a mix of the two goods as $\rho \to 0$. So, for example, let $F = 4$ and $M = 4$. If $\rho = 1$,this means utility is just 8. If $\rho = \frac{1}{2}$, however, F and M both become 2. So utility is then $(2+2)^2$or 16 - double the linear result. In effect, ρ is the love of variety. For values of $0 < \rho < 1$, the two goods are imperfect substitutes. This can be seen if (2.6) is altered slightly to:

$$U = [\delta F^{\rho} + (1-\delta)M^{\rho}]^{1/\rho} \tag{2.7}$$

Again, δ is a share parameter, a built-in budget, and the consumer must choose a limited mix of the two goods. Figure 2.2 gives an example of the function with an equal amount of each good at the two extremes (four), spreading the share between the two. ρ is 0.5. This underscores Dixit and Stiglitz's basic point: four of either good alone gives less utility than an even mix, two of each.

There is also only one unique maximum for δ, at 0.5. A consumer is best off with a perfect mix of the two, rather than either extreme. This is just a restatement of Dixit and Stiglitz's original point; however the simple principle illustrated here holds for all other versions of the CES.

The many-good CES is just an extension of the above:

$$U = \left(\sum_{i=1}^{N} c_i^{\rho} \right)^{1/\rho} \tag{2.8}$$

In this version, c denotes the quantity of good i and can be set individually, so that utility is dependent on all goods. (This is a feature of the CES put to good use in the thesis model framework.) The continuous-varieties version of the CES function is discussed in section 2.5 in the context of the core model. The next section outlines how these utility functions can be combined with budget constraints.

2.2.4 Constrained optimisations

A Lagrange multiplier is used to constrain a utility function by a budget equation. The essential point is that utility and budget equations are superseded by a third that describes demand for goods, given an actual budget amount and a set of goods prices.

Actors with a Cobb Douglas utility function (2.2) want to buy a mix of two goods F and M, where p_F and p_M are the price of F and M respectively. The constrained optima including price are as follows:

$$F = \frac{\delta Y}{p_F} \tag{2.9}$$

$$M = \frac{(1-\delta)Y}{p_M} \tag{2.10}$$

For the CES function, a constrained optimum can be defined for each single good, given a budget and the cost of each good. The optimal amount for good g_j is given by:

$$g_j = \frac{P_j^{\frac{1}{\rho-1}} Y}{\sum_{i=1}^{n} P_i^{\frac{\rho}{\rho-1}}} \tag{2.11}$$

- where p_j is the price of good c_j, Y is the budget, and the denominator sums the price of each good raised to $\rho/\rho-1$. (Capital P is used later for distinguishing the full good price including space costs from its base price, lower-case p.)[2]

2.2.5 'Increasing returns' and production

The concept of increasing returns describes all economic situations where *per-unit* output is higher for each extra amount of input (so where the first derivative of output with respect to input is always more than zero). In this thesis, increasing returns in *production* are called **economies of scale** to distinguish them from increasing returns in utility as described by the

[2]This single-good optimum was derived from Brakman *et al.*'s constraint of a CES function for a single good by substituting the price index back in and rearranging. See Brakman *et al.* 2009 pp.94.

CES function. The relevance of this distinction is explained in section 2.3.5. This section explains how increasing returns enter into production functions.

Single-input production functions can use simple equation forms for turning an input into some larger output quantity. The core model uses a straight line equation, where output (x) is defined in terms of the amount of labour (l) required to produce it, and where α are fixed costs and β are marginal costs (see the next section for an outline of the use of 'marginal'):

$$l = \alpha + \beta x \tag{2.12}$$

Usually, 'average labour' is the focus (which is labour per unit of output; this declines as total labour increases in increasing returns functions.) For the purposes of this discussion, it is easier to see the role that α - the fixed labour requirement - takes if instead the axes are switched and output is the focus:

$$x = \frac{l - \alpha}{\beta} \tag{2.13}$$

The production function used in the thesis model framework also takes in a single input - people's time (or 'labour'; these two terms refer to the same thing here). Many different people can input into a single firm, increasing its production efficiency. Section 4.7.5 sets out how this is done in the framework. However, due to vital differences between agent modelling and the analytic approach represented by the core model (also outlined in section 4.7.5), the function used is of a different form.

The production function used in the thesis is 'smooth': it does not have a fixed and marginal component, but instead a nonlinear and linear component, each separately scalable. To achieve the small increase in productivity per unit of input time required, a nonlinear increment to linear production is added as labour input increases, such that the second derivative of the function is positive (it is zero in the core model's production function). The basic version is this:

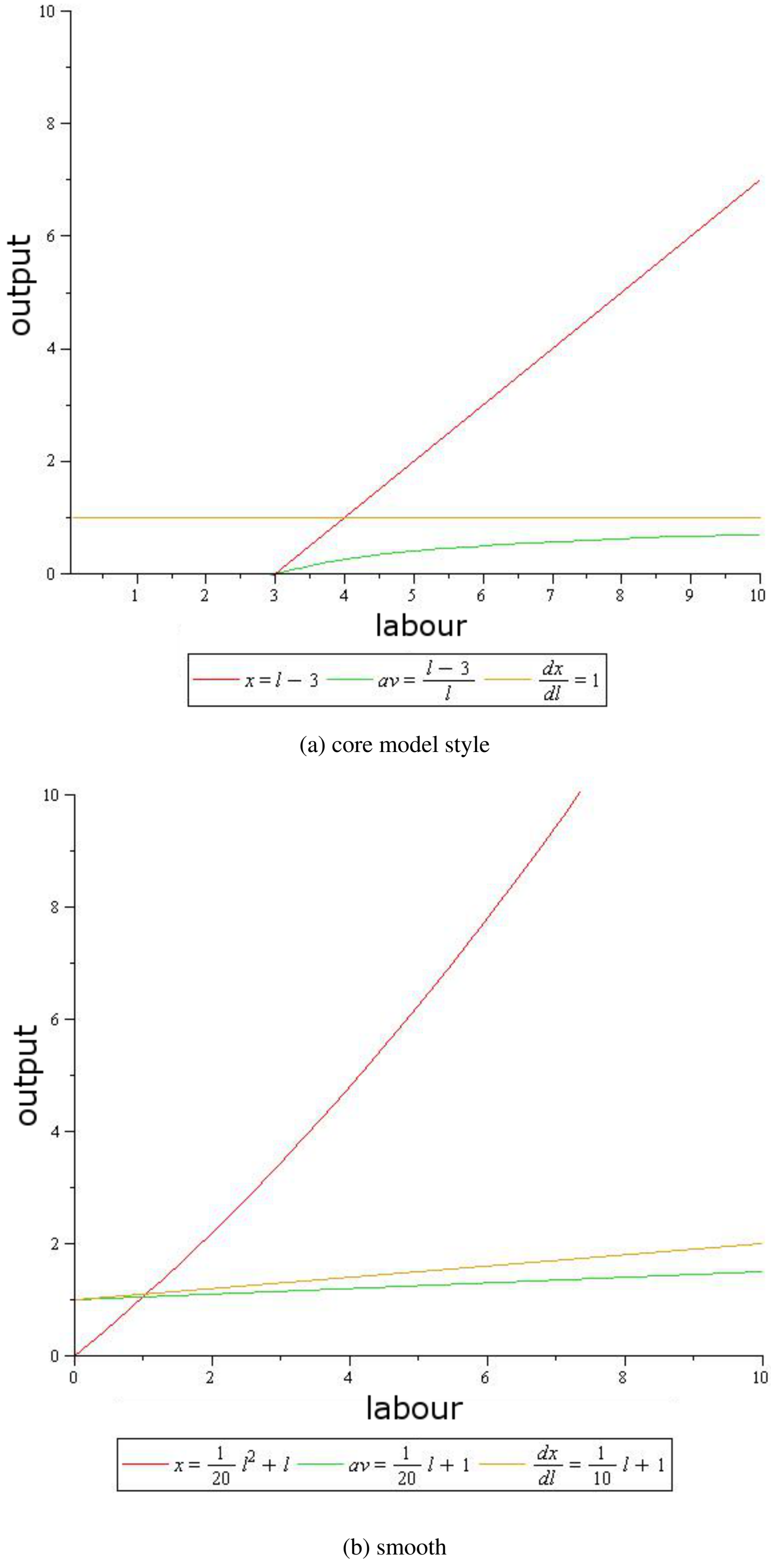
$$x = l + l^2 \tag{2.14}$$

In the thesis models, each of the two terms is given its own parameter as a way to control scale in particular model runs:

$$x = ml + vl^2 \tag{2.15}$$

- where v is the 'curve' parameter controlling the scale of returns and m is the magnitude of the linear component. This gives a simple smooth curve increasing output per unit of labour input as labour goes up.

Figure 2.3 graphs both the core model production function (for $\alpha = 3$ and $\beta = 1$) and the 'smooth' economies of scale function (for $v = \frac{1}{20}$ and $m = 1$). Their different structure shows



(a) core model style

(b) smooth

Figure 2.3: Two ways to achieve increasing returns for a single input. (a) core model style production function; (b) 'smooth increasing returns' function. Both show total, average and marginal output per unit of labour.

up: both marginal and average output increases for the smooth function (2.3b). The version used in the core model (2.3a) is built on fixed and marginal quantities (see the following section) so it already has a constant marginal output built in. The only other difference is that average output for a unit of labour in the core model increases, but at a decreasing rate (its second derivative is $-6/l^3$), whereas the smooth function's average increases constantly.

So does it matter which is used? In the analytic approach of the core model, a fixed and marginal labour requirement makes sense, since they can both be treated algebraically (this is discussed further in section 7.1). In an agent model, however if a 'fixed labour' amount is required, it needs to actually be supplied. This is not impossible, but it impacts directly on whether or not such a model functions correctly, given its set of components. The smooth economies of scale function allows for any amount of input to produce output; in contrast, the core model function produces negative output for fixed labour values below α. These issues are discussed in section 4.7.6, and models using these functions are discussed in section 6.3.

2.2.6 The meaning of 'marginal'

A quick note on terminology. Classical Economics has been described as the 'marginal revolution'. The term 'marginal' just means the first derivative of some quantity. Here is one example from utility and one from production. To start with the Cobb Douglas function discussed in section 2.2.2: if one quantity is held constant, the other's diminishing marginal utility can be seen from graphing the partial derivative, $\frac{\partial U}{\partial f} = \frac{\sqrt{M}}{2\sqrt{F}}$ - see figure 2.4, where M is set to one. The change in utility per unit diminishes (so $U'' < 0$); as more is consumed, each extra unit becomes less 'satisfying'. Note, this makes sense given what has been said about the Cobb Douglas exhibiting basic 'love of variety': if M is held constant as F increases, 'variety' is actually dropping away from a perfect mix of the two.

As discussed, the same kinds of functions are used for firms and so the same principles hold. The canonical result for firms is that output should be set where marginal cost equals marginal revenue ($MC = MR$). This is a straightforward intuition: a firm increases output by one unit. If the extra cost is lower than the extra revenue, they can still make money. If the cost is higher than the extra revenue, they would lose money. So this directs them to the point where any change in cost exactly matches the change in revenue: $MC = MR$. For this to work, the firm's change in output cannot affect prices: see section 2.3.2. This completes the explanation of the utility and production functions used in the thesis model framework, though the nature of the assumptions involved and how the functions are used is discussed throughout the thesis.

2.3 Bringing space into economics

This section examines a set of key issues for understanding how spatial economics has been theorised and the form it has taken in GE. The discussion is shaped by thinking about the assumptions and dependencies involved. Section 2.3.1 gives an overview of the spatial economic

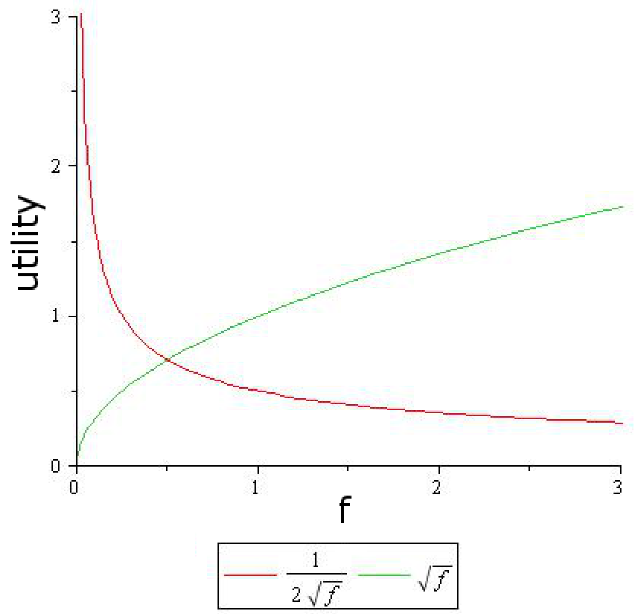


Figure 2.4: Total ($\sqrt{f}$) and marginal ($1/2\sqrt{f}$) utility for a $\delta = \frac{1}{2}$ Cobb Douglas utility function, holding M constant at 1.

'problem' and offers an initial thought regarding why ABM struggles with spatial economics, setting the scene for chapter 3. The following three sections explain the central market structure dependencies that determine what is considered a 'valid' spatial economic explanation in GE. Section 2.3.5 looks at the nature of the 'increasing returns' assumption as context for understanding how love of variety and economies of scale (two different forms of increasing returns) are treated separately in the thesis model framework.

2.3.1 Overview

Spatial economics, both classical and modern, asks the broadest of questions about economic activity: "who produces what, where and why?" (Ohlin 1933 quoted in Brakman *et al.*, p.81) Spatial structure is one of the most obvious facts of economics, and one of the most difficult to explain. Duranton and Puga (2003) put it well: "Only 1.9% of the land area of the United States was built up or paved by 1992. Yet, despite the wide availability of open space, almost all recent development is less than one kilometre away from earlier development." From the point of view of much economic theory, this shouldn't happen at all -

> "If we postulate only the usual list of economic forces, cities should fly apart. The theory of production contains nothing to hold a city together. A city is simply a collection of factors of production, capital, people and land - and land is always far cheaper outside cities than inside. Why don't capital and people move outside, combining themselves with cheaper land and thereby increasing profits?" (Lucas

1988 , p.38)

GE is the most prominent modern spatial economic theory to offer an answer. Paul Krugman, the original creator of the core model at the heart of the method, asks, “why and when does manufacturing become concentrated in a few regions, leaving others relatively undeveloped?” (Krugman 1991b, p.484) The success of GE stems from a framework able to address each of Ohlin’s ‘what, where and why’. What is produced? A variety of generic products. The more variety, the better-off the consumers. Where is it produced? There is a ‘cumulative causation’ between the location choice of firms and people. Combined with transport costs, the core model (section 2.5) claims to answer the ‘why’ - a positive feedback flowing from the decisions of rational actors. The model shows how it is possible for one region, with no natural advantage, to end up a core producing area, while another is periphery.

The core model is such a vital point in the history of spatial economics because it connected ‘general’ and ‘spatial’ for the first time. This and the following sections explain why this was so important. Calls for a ‘general location theory’ to parallel general equilibrium models in (non-spatial) neoclassical economics go back at least as far as Isard (Isard 1956). For Isard, general equilibrium was a “wonderland of no dimension” (ibid p.26). In the period between Isard’s and Krugman’s work, however, general equilibrium continued to avoid explicit space. As Isserman notes (Isserman 1996), Krugman repeated Isard’s claim almost verbatim, pointing out that “in the late nineteen-eighties mainstream economists were almost literally oblivious to the fact that economies aren’t dimensionless points in space” (Krugman 2010b, p.1).

Krugman’s stated aim in developing his model was to build a bridge between neoclassical economics and geography, precisely in order to encourage ‘dimensionless’ neoclassical economists to take geography seriously. What was the bait used to lure neoclassical economists out of their dimensionless ‘wonderland’ onto new geographical pastures? A model able to preserve all the key neoclassical features: a self-contained general equilibrium outcome reached through maximising utility, by building on his insights in trade theory that had used **monopolistic competition** to solve the ‘similar-similar problem’ (Krugman 2009 p.561) of international trade in similar sectors (see section 2.5.1). As section 2.3.2 outlines, the vital element is provided by the new approach to market structure of the Dixit-Stiglitz method: the ability to construct a non-perfect-competition model, required by the neoclassical approach for allowing spatial economic activity to happen at all (see section 2.3.3 for more on this).

The success of Krugman’s project, however, throws into sharp relief the absence of other attempts to examine these fundamental components of spatial economics. This may be, as Overman notes (quoting Martin 1999) that many see -

> “ - individual optimising behaviour, equilibrium outcomes and mathematical modelling as a regressive step for economic geographers who have ‘long since abandoned location-theoretic and regional science models’.” (Overman 2004, p.505)

The analytic approach of GE is - as many GE theorists openly acknowledge - a ‘mathematical straitjacket’ (Martin 1999). The difference between GE and ‘economic geography proper’ (as Martin puts it) is in their evaluation of whether the benefits of the straitjacket outweigh the costs.

Martin argues that GE is unable to properly explain the real internal dynamics that make localisation more productive. Many theorists agree with this while remaining “squarely on the side of those who consider that general frameworks are necessary to scientific progress in geographic economics” (Storper 2010 p.315). They want to be able to “pinpoint the basic sources of urban dynamism, which lie in the selective geographical matching of productive resources, skills and institutions of coordination” (Storper and Scott 2009 p.164). Duranton and Puga have done just that, concentrating on the underlying microfoundational mechanisms of ‘sharing, matching and learning’ taking place (Duranton and Puga 2003), as well as broadening out to investigate the “conditions for a configuration in which diversified and specialized cities co-exist to be a steady state” (Duranton and Puga 2001 p.1). See also more recently Delgado *et al.* (2010) for a subtle empirical look at the interaction of clustering, productivity and startups: “while at a (narrow) industry level firms may compete for the given pool of resources, the cluster environment that surrounds an industry will increase the pool of competitive resources and reduce the barriers of entry for new firms” (ibid p.514).

Fowler argues that ABM is perfectly positioned to help “bridge the gap between geographical economics and economic geography” (Fowler 2007 p.266; section 3.3.3). However, agent modelling seems to have a default affinity with “economic geography proper”’s outright rejection of everything neoclassical (section 3.3.1), leaving GE and ABM in quite incompatible positions. ABM also tends to chime with Martin’s argument about GE: that valid explanations require digging into a system’s internal dynamics (section 3.4.8).

The rationale for ‘disassembling’ the core model in section 2.5 is to gain the opportunity to carefully compare the assumptions of GE and ABM in an attempt to question the validity of this common ABM rejection of everything neoclassical. The two cannot easily be directly translated into each other (as Fowler’s work has revealed; section 3.3.3) and each has very a different model assumption space. Laying these out clearly enables a more informed set of choices for building spatial ideas into the thesis model framework.

2.3.2 Market structure

This and the following section examine the ‘dependencies’ at the heart of market structure arguments and explain how they have determined the role of space in economics. There are three broad categories of favoured market structure in analytic economics: perfect competition, monopoly and monopolistic competition. Why ‘favoured’? Each of these are analytically tractable, (relatively) simple and allow some strong economic conclusions to be drawn from static equilibrium analyses. The last of these - monopolistic competition - is vital to the ‘engine’ of the core model (section 2.5).

Perfect competition and monopoly have straightforward mathematical interpretations: in the former, the size of the market is so large that no one economic actor can affect price: everyone is assumed to be a 'price-taker'. In practice this means prices can be treated as constants. (This is generally the case, though general equilibrium theory has developed an argument for perfect competition that does not rely on assuming all actors are price-takers; see e.g. Starr 1997 pp.157.) Not only does this mean consumers can have no effect on price, but firms cannot either - and so strategic interaction is ruled out. In a monopoly, on the other hand, the monopoly firm can set a price that gives them the best revenue, based on the supply-demand curve they face for their product. There are no other firms, so they face no strategic considerations, and the revenue-maximising calculation is straightforward.

Perfectly competitive models must work with linear output - no increasing returns are allowed. To see why, consider a (rather artificial) situation where many firms have linear output, but one firm, producing the same good, is able to utilise economies of scale. They can thus produce each unit more cheaply than their competitors - and so this is no longer a perfect competition situation. A single price can no longer be assumed, nor can one say that no actor can influence price. In this case, that firm can immediately affect demand for other firms' goods; the more goods are substitutable (in the CES function, as $\rho \to 1$), the more monopoly power that firm can gain.

Increasing returns also create 'indivisibilities': if a particular product can be produced more efficiently at a larger scale, there is no incentive to 'divide' this productive capacity into smaller strands, each producing less efficiently. Those efficiency gains provides firms with market power. This idea is important for GE and the core model; it is used to justify individual firms producing a single variety and is the basis for its market structure argument (see section 2.5.4).

Between these two poles, analyses of 'imperfect competition' have produced a sprawling tangle of approaches over the twentieth century, with some famous early attempts at synthesis (e.g. Robinson 1933). The arrival of the **Dixit Stiglitz model** is such a pivotal moment in the history of market structure models because it offered a (relatively) simple way of keeping a stylised form of imperfect competition while saying useful things about the economy as a whole. The monopolistic competition model first published by Dixit and Stiglitz in 1977 is one of the most prominent approaches to imperfect competition. It is the 'engine' of GE (a description discussed in more depth in section 2.5). As Brakman *et al.* note, in providing a tractable method able to do away with many of the complexities facing theorists of imperfect competition, it neatly sidesteps "getting bogged down in a taxonomy of oligopoly models" (Brakman *et al*. 2009 p.93). In particular, it can account for economies of scale, and it finds a way around problems of strategic interaction. The workings of this 'combustion engine' as it is applied to the core model are explained in section 2.5.

2.3.3 Preconditions for spatial economic activity

The development of the core model was motivated by the fact that previous economic theory, built on perfect competition models, was unable to deal with space. The problem begins with the relationship between perfect competition and increasing returns explained in the previous section. Perfect competition cannot work with increasing returns, because any firm able to gain them will have a cost advantage: that is immediately no longer perfect competition. It would also nix the possibility of assuming agents are price-takers and that there is one price. Economics in space, however, cannot work *without* increasing returns, because if output is linear, gains from trading across space can never be enough to compensate for losses incurred transporting goods or people. This issue led to space being largely ignored by mainstream economics. As Fujita put it:

> "since the traditional general equilibrium analysis abstains from the consideration of indivisibilities or increasing returns, it fails to capture the essential impact of transport and spatial costs on the distribution of economic activities." (Fujita 1999 p.376)

The way in which perfect competition rules out spatial economic models has been dubbed the 'spatial impossibility theorem' by Fujita. Starrett (1978) is credited with first making this point formally. As Fujita and Thisse describe it -

> "Starrett has shown that if space is homogenous and transport costly, then... the economy degenerates into separate single-location groups of agents with all trades taking place within, rather than between, groups. Consequently, the perfectly competitive price mechanism alone is unable to deal simultaneously with cities and trade." (Fujita and Thisse 2002b p.13-14)

To put it another way: if an actor is able to produce 'in their own backyard', why would they choose to trade or collaborate in a world where all producers' output is linear? Any gain they would receive can never outweigh transport costs. As a consequence, "people in each of these locations, like Robinson Crusoe, will produce all goods at a small scale for self-consumption." (Duranton and Puga 2003 p.1) However, with increasing returns introduced to the picture, the minimum conditions for spatial economic activity now exist. So increasing returns serve a dual purpose: they allow actors to break out of this 'spatial impossibility' trap, as well as guaranteeing uniqueness of varieties due to creating indivisibilities (a vital assumption for the core model; section 2.5.4).

The overlap of market structure arguments and the preconditions for spatial economic activity are responsible for the key 'dependencies' that have driven the development of economics and its separation, until recently, from spatial economics. GE models are built to make sure space is not impossible: locations do not lapse into autarky, and economic activity does not collapse into a single point 'black hole' (Fujita 1999 p.58).

Through the use of an 'iceberg' conception of trade costs (see section 2.5.3) the core model also gets to use the Dixit Stiglitz model while tactically avoiding the second key problem that space imposes: market interactions are "not all brought together in a single location and at the same time" (Leijonhufvud 2006 p.1633). This simple fact about spatial trade breaks the assumption that exchanges happen in a centralised 'auction' way: there is no single site where prices can equilibriate to 'clear' supply and demand efficiently. For this reason alone, Isard argues, "the generally accepted principle of pure competition is not applicable to the analysis of spatial economic processes" (Isard 1956 p.489-490). This point is taken up in more detail in section 3.3.1 as something ABM is in theory primed for.

2.3.4 Endogenisation

The drive towards 'general' economic models is underpinned by arguments for 'endogenising' economic dynamics into those models. Section 2.5 explores one example of this, where the use of the Dixit Stiglitz model provides an explanation of international trade that does not require any exogenous imposition of differences between countries. From this point of view, many classical location theories fall short, requiring many exogenous impositions to function; section 2.4 looks at a number of these relevant to the thesis models.

GE brings the same philosophy to spatial economics, seeking to model spatial morphology "as an endogenous outcome of the economic process" (Storper 2010, p.315), rather than being caused by 'first nature' effects, such as "political power (the State), trade (rivers and ports), and finance" (ibid p.316), or the presence of natural resources. These may determine starting conditions, but the goal is to identify and understand 'second nature' effects - those that flow from the interaction of people and firms in space, not some external spatial feature. As Glaeser argues (Glaeser and Kohlhase 2004, Glaeser 2008 p.7), as transport costs decline, any exogenously imposed distance factors resulting from 'first nature' landscape features decline relatively in importance. On this basis, Fujita argues that -

> "To test whether a model can explain the endogenous formation of agglomerations, it is best to consider the simplest case of a perfectly homogenous geographical space (and free choice of location by all agents), for if any concentration of economic activities occurs in such a homogenous space, it must be due to endogenous forces." (Fujita 1999 p.376-7)

The need for increasing returns to allow spatial trade to happen at all also stems from the same motivation. Many older classical location theories (discussed in section 2.4) breach the 'spatial impossibility' condition by not explicitly modelling increasing returns, thus providing no endogenous motivation for trade to occur at all.

In the thesis models, the difference between an exogenous and endogenous model shows up in sections 5.4.6 and 6.4.2, which present two results where an exogenous difference in the first is found endogenously in the second, more 'general' model. Partial models with exogenously fixed features still have their uses, however; chapter 5 makes use of partial models to make specific points relevant to the thesis. The use of a 'density cost' proxy in the models in section 5.4 is based on a particularly agent-based take on endogenisation, where all costs of being 'here' are endogenised; this is discussed in section 2.7.4.

There is another potentially disruptive point mentioned in passing by Fujita at the end of his speech honouring Walter Isard in 1999:

> "For the development of human society in the long-run, economic aspects represent merely a part of the phenomenon, perhaps not even the most important ones. In other words, economic development may represent merely a possible result of the more fundamental process of social and cultural development." (Fujita 1999 p.380)

If, in reality, the spatial economy is a complex mix of economic and other factors, this may suggest this 'endogenise everything' philosophy will always produce incomplete models if it relies on purely economic forces. This point is returned to in the conclusion.

The previous sections suggest the following: an ABM 'general' theory of spatial economics would require solving the following two problems. First, an effective market structure would be required, capable of allowing genuine competition dynamics, or at least developing a specifically ABM-friendly argument about competition. Second, there would need to be no imposed condition about when or where economic exchanges are initiated. Or, less strongly, economic exchanges would be required to avoid the need to happen at a single time and place, 'auction' style (this last point is given a more detailed treatment in section 3.3.1).

The transmission belt model presented in chapter 6 is the result of finding a series of modelling compromises in pursuit of this goal. It is most successful in producing stable supply/demand equilibria for agents at differing locations, through decentralised price signals. There is no single auction site or process: equilibrium is the result of individual Firms' price signals responding to market signals from all People, with both Firms and People having heterogenous locations. However, it does this with no explicit competition dynamic between firms, even though they do use price adjustment to sell goods and effectively clear the market. The next chapter, in examining ABM, discusses how the kind of coordination issues involved in market structure problems have been a prominent research topic for ABM. Many problems still appear unresolved: this key point is returned to both in the chapter presenting the transmission belt model results and in the conclusion.

2.3.5 The 'ecological' criticism

The discussion of market structure in sections 2.3.2 and 2.3.3 explained that increasing returns are a vital component of the GE rationale for spatial activity. As section 2.5 explains in detail, increasing returns have a rather murky presence in the core model: economies of scale are, ostensibly, present through a specific production function but, through a series of mathematical and verbal arguments, production scale itself becomes fixed and increasing returns manifest themselves instead through the (implicit) entry and exit of firms into the market. Each firm, producing a single good, adds to the models' capacity for love of variety - and thus increasing returns become a function of utility, not production.

GE theorists are conscious of this but consider the benefits of using the 'engine' of the Dixit Stiglitz model offset this murkiness. The kind of production and utility functions used in GE get plenty of criticism, aimed in one way or another at their intrinsic lack of realism. The question of realism of assumptions is a central theme for chapter 3, but the issue is important on a more prosaic level when building a spatial economic agent model. Increasing returns from production may imply very different quantities of material, with differing weights and properties. In comparison, increasing returns in utility (e.g. a gain in love of variety from adding another firm to the market) appear 'weightless'. This section looks at a particular critique that helps frame this problem.

From an ecological economics point of view, some see in the Cobb Douglas function a 'Garden of Eden' world (Georgescu-Roegen 1975 in Daly 1997) where natural resources are optional. Daly sums this up nicely, if somewhat loquaciously. He begins by quoting Solow saying, "if it is very easy to substitute other factors for natural resources, then there is in principle no 'problem'. The world can, in effect, get along without natural resources" (Solow quoted in ibid p.261). Daly responds that economists seem to believe in -

> "making a cake with only the cook and his kitchen. We do not need flour, eggs, sugar, etc., nor electricity or natural gas, nor even firewood. If we want a bigger cake, the cook simply stirs faster in a bigger bowl and cooks the empty bowl in a bigger oven that somehow heats itself." (ibid.)

More formally, he quotes Georgescu-Rosen's critique of the Solow-Stiglitz production function - a Cobb Douglas with the addition of natural resources. If Q is quantity produced, K and L are again capital and labour, and R are natural resources and exponents $a1$ to $a3$ control share:

$$Q = K^{a1} R^{a2} L^{a3} \tag{2.16}$$

He points out this means that $R^{a2} = Q/K^{a1}L^{a3}$. So, holding labour constant, one could get any quantity of output with a vanishingly small amount of R, as long as $R > 0$ and K is large enough to compensate for that tiny amount. Solow's reply to this highlights a recurring argument used to defend the abstract nature of many economic models: critics are taking them too literally, and not considering how the models are used:

> "We were trying to think about an interesting and important question: how much of a drag on future growth, or even on the sustainability of current production, might be exercised by the limited availability of natural resources and the inputs they provide? ... The role of theory is to explore what logic and simple assumptions can tell us about what data to look for and how to interpret them in connection with the question asked" (Solow 1997 p.267/8).

Solow goes on to point out that the argument should be about how substitutable renewable and non-renewable resources are, given that the former are likely to be highly capital-intensive. (Ibid.) He appears to be saying that his critics have mistaken economists' models for their actual understanding of the world, rather than tools that aid that understanding.

Two issues flow from this. The first is about how one judges whether a model approach is 'valid' and what can push it over into invalidity. Confusion between a model and the modeller's understanding seems to be particularly blurry in ABM; this theme is picked up in chapter 3. The second is directly relevant to the thesis model framework: how, actually, to build in functions for increasing returns? Using an agent approach offers an opportunity to very explicitly separate economies of scale from love of variety, and this opportunity is taken up. But as will be argued, whilst there are gains from doing this, there are also new compromises. The second results chapter (6) presents these findings.

2.4 Model parallels to specific classical spatial economic ideas

Prior to explaining the core model in depth, this section explains how the two 'partial model types in the first results chapter ('mobile People' and 'mobile Firm' models; chapter 5) have some parallels from classical location theory. The aim of this section is to explicitly state what those connections are; these are picked up again in the sections where they are relevant in the results. This section also discusses some classical location theories placed outside the framework.

Agents in the thesis model framework retain the same fundamental decision structure across models; the difference between the presented models comes down to a few very simple choices between settings. Chapter 4 explains these different settings in detail, but they all function with the same basic logic: agents making decisions based on the costs they face from their location.

The two 'partial' models both have clear parallels to classical location theory and urban economics. Before discussing those, the model framework's most basic starting point is covered. Agents with locations each face unique costs. Using an agent modelling approach makes this simple to implement - at least in terms of defining cost functions. The complications arise when interaction is introduced. The rationale for these simple functions is discussed in section 2.6 and the detail of how they are implemented in the framework presented in sections 4.3 and 4.4.

This approach of providing individual agents with methods for assessing costs directly is actually rather rare in both classical and more modern variants of spatial economics. This is for a good reason: without the ability to produce disaggregated models of the sort that agent modelling excels at, the best route to analysis is through analytic deduction involving assumptions about *groups* of agents. For example (see below) the assumption of spatial equilibrium can be used to replace many-agent interaction by deducing what the end-point of that process may look like if agents act 'rationally' (see section 3.4.3). Where agent interaction is important (e.g. with Hotelling's competing firms; see below), simple arguments can be used to make the necessary deductions.

Weber's work on optimal location is the clearest parallel to treating agents as separate decision-making units (Weber 1909): for instance, the Weber-Fermat problem (see e.g. Brimberg 1995) of optimising a plant location involves balancing inputs from two firms with the location of a market being sold to. The optimisation takes place on an otherwise entirely featureless and costless surface, with the results depending only on distance factors. This Weberian tradition of dealing explicitly with distances and weights for specific firms and goods is uncommon. In the core model, distance costs are combined with other costs of trade into a single metric between discrete regions (see section 2.5). Krugman only ever mentions 'Weber-type stories of transport cost minimisation' to dismiss them as good only for optimal plant location type problems (Krugman 1991c p.49)[3]. In contrast, using an agent approach starts from 'Weber-type stories' by default: distance costs between discrete actors presents itself as the obvious approach.

The 'mobile People' models have strong parallels in the history of 'monocentric' models that assume a fixed central point and investigate optimal land and utility outcomes surrounding that point. The 'Eve' of monocentric models is von Thunen's 'Isolated State' (Thunen 1826). Von Thunen's own exposition was purely verbal; it was Launhardt (Launhardt 1885) who first produced a mathematical version (Blaug 1997 pp.600). In this, a settlement is represented by a single point and farmers must optimise their location by trading off distance against the rent they pay for the spot of land they use for production.

For a given crop, farmers have a set yield, from which is deducted both their production costs and the distance to the central settlement, leaving their surplus production. This surplus can be represented with a simple negatively sloping line based on the distance cost they incur: at a certain distance from the centre, the yield becomes zero and no surplus is possible. Assuming linear distance costs, each farmer's theoretical surplus is a negatively sloped line. The surplus becomes zero at different distances for different yields and production costs. This surplus line represents the maximum amount each would be able to pay in rent for land at a given distance. If a series of crops of differing yields are considered, each has points where this surplus line is either highest for that location or the only remaining crop, closer to the settlement edge. Higher production-cost goods will slope faster to zero, but if they are higher-yielding, will also have higher surpluses near to the centre. Competition for land is then assumed to end with the farmer able to spare the highest surplus to rent.

Haggett *et al.* note that von Thunen's model is a rather unique combination of highly abstract assumptions and a deep empiricism stemming from von Thunen's own insights as a land owner and manager in the early nineteenth century (Haggett *et al*. 1977). He has a strong claim on producing the first equilibrium-based economic model of any kind and, as Fujita *et al.* say, this was "an ingenious and quite deep analysis [and] a striking example of the power

[3] In Fujita, Krugman and Venables' later book (2001 p.41) Weber is relegated to a footnote.

of economic modelling to generate unexpected insights" (Fujita *et al*. 2001 p.17). They make three other pertinent points about the von Thunen model; as they say, it is easy to forget how radical these points were at the time. First, that the model showed "the spontaneous emergence of a concentric ring pattern", that this result comes about without any single actor needing to know about the larger pattern, but just pursuing their own ends and - what they consider most 'startling' - that this " unplanned outcome is efficient, is indeed the same as the optimal plan" (ibid).

While this outcome began life in von Thunen's original as a verbal deduction and, via Launhardt, a geometric deduction, this vital concept of spatial equilibrium made its way into modern urban economics via Alonso transposing it into a 'central business district' setup (Alonso 1964). As Glaeser notes (2008 p.4), spatial equilibrium is "the bedrock on which everything else in the field stands... essentially, there must be no arbitrage across space". As with other economic equilibria (see section 2.5), the urban economic approach is to assume that the system has become static at $\delta u/\delta d = 0$ (where u is utility and d is distance) as one moves from $d = 0$ at the centrepoint to the edge. That is, utility is always equal for all agents at the end of their deliberations.

For GE theorists, urban economics is of limited value. Fujita *et al.* argue the approach is limited because "if your question is not simply how land use is determined when the location of the town or towns - indeed their number and size - is itself endogenous - the von Thunen model offers no help" (Fujita *et al*. 2001 p.18). For the thesis models, however, not only is fixing firms' location a useful analytic tool, it is able to produce an interesting take on the spatial equilibrium concept as well as suggest a novel way to analyse cost change. Section 5.4.3 compares the effect of changing a series of space costs and, as discussed in detail there, the modelling process revealed parallels between changing good- and people-moving costs. In short, the non-spatial part of cost changes (such as the base price of goods before delivery costs, or the wage itself, without costs related to commuting) have an opposite morphological effect to the spatial parts.

The second important aspect of the 'monocentric' section is in relation to the 'density cost' idea itself as a subject for investigating the nature of 'valid' model assumptions. Is it a justified simplifying innovation, suitable to an agent modelling approach to spatial equilibrium, or does it "disregard everything that is important and significant in the real world" (Hayek 1945 p.530). Chapter 3 addresses the issue of assumptions in the context of ABM; the conclusion assesses the density cost assumption in detail.

The models in section 5.4 provide 'People' agents with a simple proxy for a range of proximity costs. This has a parallel in how GE and the trade cost literature (see section 2.8.2) use a single metric to summarise the full range of costs involved in moving goods across space. The idea of proximity and its relation to externalities is discussed in section 2.7.

In the model runs in section 5.4, the density cost is the medium of interaction between agents. Without it, each would make decisions entirely independently of the others, with no actual difference between running a model with one actor a hundred times, or a hundred actors once (i.e. it would be ergodic). Using density cost as a proxy makes the connection between land and distance costs vital to monocentric models. Density cost supplies the same thing that Alonso required consumption of land quantity for[4]: an economic reason for distance to exist at all in a model trying to explain spatial morphology “as an endogenous outcome of the economic process” (Storper 2010, p.315). The density cost avoids a ‘black hole’ outcome where actors do away with space altogether if they can, collapsing into a single point (Fujita 1999 p.58; see below).

A last note on the ‘monocentric’ models. The spatial equilibrium outcomes presented in section 5.4 are referred to as Nash equilibria. This not robust; as section 4.6.1 explains, the ‘Bundle’ method used by People agents takes random spatial samplings. Over time these converge on stable equilibria, but they are still at root probabilistic. The outcomes cannot be considered game-theoretically robust (cf. Vives 2001 pp.13). Nevertheless, they do manifest the essential Nash equilibrium property: no agent can take an action that will unilaterally improve their utility.

Section 5.5 presents a model set-up with two Firms optimising their location on a line of evenly spaced People. Hotelling’s model investigating competition (Hotelling 1929) is the closest parallel. Previous analyses (e.g. Palander 1935; Hoover 1937) had examined the issue of optimal market area, producing top-down graphical representations of their idealised boundaries. Hotelling’s contribution was to introduce a proto-agent question: was the optimum from the point of view of consumers actually optimal for firms, given that they compete? He showed that optimal market areas were often unstable: any firm had a competitive motivation to cannabilise other firms’ market areas. The stable result of this process for two firms, for completely substitutable goods being sold to evenly spread consumers on a line, was for both firms to edge closer until they reached the centre. One firm moving, taking more than half the market, forces the other firm to respond and a feedback begins.

Firm agents in Hotelling’s model are, at least in theory, optimising from moment to moment, not deducing some top-down idealised optima. The Firm location line model in section 5.5 sees firms competing in the same way, with the result dependent on both love of variety and distance cost. As well as linking to the Hotelling outcome, the central point for the thesis is what the Firms are able to do given the information they have access to.

The model shows interesting outcomes when love of variety is varied, allowing a break from considering just the perfectly substitutable goods of the Hotelling-style approach. But equally relevant for the thesis is the question of Firms’ ability to act. What data do they use to make decisions? How do they use it to respond? What does their apparent failure to sometimes read its meaning correctly imply? These questions are considered by framing the Firms’ decision structure as ‘data, meaning, response’, based on an idea from Jane Jacobs (Jacobs 2001) and

[4]“If the only criteria for residential location are accessibility to the centre and the minimising of the costs of friction, and considerations of the size of the site are excluded, all residence would be clustered around the centre of the city at a very high density.” Alonso 1964 p.9

presented in more detail in section 4.7.7. This ties this to some of the most vital issues of agent coordination covered in chapter 3; the issue of coordination and agent interaction is also discussed in section 7.3.

There is another set of classical theories worth mentioning, just to avoid any confusion. As section 5.4 explains in depth, the way People react to density costs causes them to form a clear hexagonal patterning. This is due to them seeking out the cost-minimising interstitial points: the result is the emergence of a set of highly structured density radii that evoke the classical hexagonal geometry of the 'central place theory' of Christaller (Christaller 1933) and Losch (Losch 1940). The two have no relation to each other, however. The emergent pattern in the thesis models is solely the result of People agents finding cost-minimising points, where central place theory assumes a plane of evenly spread people and a geometric arrangement of demand curves serving them. While the result is a striking feature of the self-organising behaviour of the thesis model, for the purposes of this thesis, it is not relevant: as section 5.4 discusses, the 'density cost' approach is instead useful for providing a way to mimic, in an agent-friendly way, a range of proximity costs. Section 2.7 below explores the theoretical rationale for this approach in more depth.

Lastly, another set of theories left out of the thesis involve spatial diffusion processes, of the kind most famously analysed by Hagerstrand (1953) in relation to innovation diffusion, but equally important for market signals and trade flows in a spatial setting. These kinds of processes are intrinsically about the complex relationship between space, time and distance, overlapping with two of the 'important spatial ideas placed outside the framework' in section 2.8. As the model framework chapter explains, agents in the thesis models make decisions either 'now' based solely on information immediately present, or over two timesteps, taking an action and waiting for a 'sonar' signal back from the market. This is a huge simplification, allowing the model to avoid concerns of information diffusion, the speed of trade flows over time and all associated problems of risk and uncertainty (section 2.8.3). The conclusion returns to this point (section 7.5) as it is an obvious area for further work and a key assumption to consider.

2.5 The 'core model' of geographical economics

This section presents a detailed explanation of the core model. The aim is to get as deeply as possible into the web of assumptions involved. It is compared both to an 'engine' and a 'scaffold'. This mixing of metaphors is surprisingly useful. Leading GE author Masahisa Fujita sees the core model (and especially its use of the Dixit Stiglitz model; section 2.5.4) as providing "location theory with a powerful combustion engine" (Fujita 2010 p.8). Indeed, he argues that its absence meant that von Thunen (see section 2.4) could not have produced anything other than the partial theory he did, any more than one could expect "de Vinci, upon one's first view of his amazing 'flying machines' drawn in the fifteenth century, to invent a real airplane" without access to the internal combustion engine.

The ‘scaffold’ metaphor is used to describe the set of mathematical deductions and simplifications in the core model, combined with verbal arguments, that are stripped away once the central model equations have been deduced. Christopher Fowler’s work on translating the core model into an ABM (see section 3.3.3) demonstrates why this metaphor matters. An engine implies a series of specific, interconnected working parts in synchronised motion. This is a natural way for an agent modeller to think as they code: model components must also be connected together explicitly and the whole must ‘run’. The first of Fowler’s two papers on producing an agent-based core model has a highly valuable rationale: do not deviate from the engine components specified in the original core model (Fowler 2007 p.266). It did not produce the core model dynamic - precisely because many of the elements responsible for the core model’s functions have already dropped away once one arrives at its ‘operational’ equations.

In the following sections, the ‘story’ of the core model is told first, followed by a broad overview of its structure, before diving into the detail of how it is constructed. This begins with the end process of the ‘scaffold’ first in order to make clear what the scaffold actually helps create. The final section puts this together into a working example of the core model’s dynamics, showing how it is used to determine a spatial ‘general’ equilibrium between two regions.

2.5.1 The core model’s story

The central story of the core model builds on two older stories about trade. Each of these tells its story by thinking about two regions and asking, what leads them to trade with each other? What do they gain by doing so? Before the monopolistic competition era, this trade was theorised to happen because each region has its own unique set of factors. Each would thus benefit by some trade, even if one actually had access to both internally and an advantage in production technology in both. This is the central idea of comparative advantage, and had remained fundamentally the same since Ricardo (1817) first formulated it. This factor-difference explanation is also compatible with the linear-production assumptions of perfect competition discussed in section 2.3.2.

Empirical reality, however, appeared to suggest another possible dynamic driving trade. Evidence showed trade taking place intra-industry; for example, two regions both selling each other very similar cars, or apparently similar intermediate factors (Brakman and Heijdra 2001 p.27).

The CES function offers a very simple way to think about this phenomenon, as section 2.2.3 has already introduced. If two regions both contain consumers with a love of variety, having access to a variety of cars to buy will increase utility overall. Krugman (1979) has a simple example. Two regions in autarky (not trading) each build three types of car, and their consumers thus gain from that variety - able to access those three. What happens if trade is opened up?

Here, some additional assumptions can enter into the simple picture to explain the outcome.

Firstly, the total amount of labour across both regions does not change: the number of workers is always the same. Second, each of the car industries - six in total across both regions - has increasing returns. The more workers each has, the more efficient their output. So what happens if each isolated region moves to open trade with their neighbour? One possibility is as follows: the highest utility for both regions comes about by each region producing two varieties. Four in total are now made, each at a higher output efficiency.

That story contains a number of crucial steps. First, opening trade restructures the economy of both regions. In this case, one car-maker in each region disappears. Secondly, no exogenous difference between the regions is required to show how trade can be beneficial. Thirdly, it highlights the important job that increasing returns is playing: any firm making a single variety of car will have no incentive to make anything other than its own. If it did, it would lose the efficiency of larger-scale production. Each is thus 'locked in' to its own variety. The assumption of a '**representative agent**' is also very useful here: while individuals may only buy a single car, at the level of international trade, the aggregate effect of love of variety is more clearly applicable.

Krugman's trade theory showed how trade patterns seen in the real world can come about with no fundamental initial differences and gave a way to think about intra-industry trade. With the core model, he introduced this one extra simple factor: transport costs between the two regions. As Brakman and Heijdra say:

> "In standard trade models location does not matter, because when transport costs are absent there is nothing to gain from being in specific location" (Brakman and Heijdra 2001 p.27).

The impact of adding transport costs is dramatic. It demonstrates a mechanism: as transport costs drop, one of these two regions enters a positive feedback and becomes the productive 'core' containing all manufacturing workers, relegating the other to the 'periphery'. Just slight differences in initial conditions determine which is which.

The core model also introduced a number of ideas into a general equilibrium model more familiar to complexity theorists. There could be multiple equilibria, some of them less optimal than others, and each of them path-dependent. Some are stable, others will shift to a new regime with the slightest perturbation (Brakman *et al*. 2009 p.127-8; see section 2.5.5 also). These regime changes can occur 'catastrophically': a marginal change in transport costs can trigger the start of a feedback process that leads to a core emerging (Fujita *et al*. 2001 p.94). (This sensitivity to initial conditions is also a factor in the transmission belt model; chapter 6.) The core model also illustrates a mechanism for the 'home market' effect, where a feedback between domestic economies of scale and local demand make those products the cheapest - and thus the best exports.

As Krugman himself points out in the original core model paper (Krugman 1991b p.486), a number of these ideas are not new. Mrydal's theory of 'cumulative causation' (1957) introduced the idea that a feedback could lead a 'core' to separate itself from a 'periphery'. Jane Jacobs' work (see e.g. Jacobs 1986 pp.147) can be seen as a detailed empirical investigation of the home market effect: she identified the tendency for places to export those goods most strongly demanded locally. In fact, she argued this was the central mechanism for economic development more generally. However, the purpose of the core model was to demonstrate it is *mathematically* possible for these dynamics to emerge endogenously.

That is some way from saying that it is either economically possible, or that the dynamic actually exists in reality. But one goal of the core model was clarity: to produce a model where everything possible is stripped away to reveal the essential dynamics. This is seen in GE as a strength of the model, not a weakness, though plenty of critiques exist of particular assumptions. This ties to the issue of realism of assumptions taken up in chapter 3.

2.5.2 Outline of core model explanation

This section gets into the workings of the core model in detail. Two main sources are used. First, Brakman *et al*. 2009, which contains a detailed explanation. Brakman is also used by Fowler to build his agent version. (See section 3.3.3 for more information regarding Fowler's take on the core model.) The second source is Fujita *et al*. 2001. As will be explained, this takes a slightly different approach (continuous rather than discrete goods) but provides much useful extra information. There are a number of other outlines of the core model in the literature including Krugman's own original paper (Krugman, 1991b). These will be used to clarify certain points. To keep terminology and choice of parameter names consistent, those from Brakman *et al*. are used, so where quotes are given from Fujita *et al*. 2001, any mathematical labels have been changed to match. While various additions and extensions to the core model have been developed over time (and indeed Brakman *et al.* say a number of these have "gained the reputation of being a 'core' model of GE", p.134), Krugman's original model structure is the basis for the core model this thesis discusses.

The usual approach to describing the core model in the GE literature is to follow the logical path of the mathematical argument, beginning at the micro level and building up towards the model's conclusions. While this is perfectly sensible mathematically, it is not a particularly good way to open the bonnet and get a clear view of its workings. Section 2.2 already dealt with a number of the key economic concepts that the core model uses. Where relevant, those will be added to here.

This section attacks the problem in reverse, starting with the three simultaneous equations toward the endpoint of the analytic argument, which will be called the '**region equations**'. These equations are the 'operational' elements - that is, they can be used to produce simulation code to run the numerical model calculation. Once these are reached, a great deal of the 'scaffold' has been analysed away. So, the top-level structure is discussed first, followed by a more detailed look at the scaffold that supports it. After this, the 'operational' elements are presented to show how they are used to identify the 'core / periphery' dynamic[5].

As Neary notes, the core model can end up "degenerating at times into a near-impenetrable soup of CES algebra" (Neary 2001 p.537). The goal here is to get to the crux points of the model as clearly as possible first, before dealing with that soup. Though it cannot be completely avoided, much of it can be compartmentalised (as Brakman *et al* do with a series of technical derivations for model subcomponents). It can be particularly confusing trying to locate exactly where the economic mechanisms are; the GE approach to model-building (as with much of neoclassical economics) is built on a web of verbal argument, equilibrium assumptions and algebraic rearrangements. These end in very clean results but make the internal workings hard to track down.

The following section lays out the top-level structure of the core model; section 2.5.4 examines the 'scaffold'.

2.5.3 Top-level structure

The core model defines **two discrete regions**, each of which is a dimensionless point, and with transport costs set between them. There are also:

- **Two types of workers**: manufacturing and agricultural. The number of each is set exogenously. (As Brakman *et al* note (p.88), the distinction between *manufacturing / agricultural* is in a sense arbitrary: what matters is the difference between *mobile / immobile* sectors.)
  - **Agricultural workers** are immobile - they cannot move between regions.
  - **Manufacturing workers** *can* move between regions, seeking the best wage.
- **Two types of goods matching the workers**:
  - **Agricultural goods** are a static base for the core model where prices cannot change. They are assumed to be produced under perfect competition. Production is linear - there are no economies of scale. Trade happens freely across regions, so wages and prices are assumed to equilibriate and everyone is a price taker.
  - **Manufactured goods**, along with manufacturing workers, are the dynamic part of the model. A single firm makes one variety (see section 2.5.4). The range of varieties produced is the key determinant in workers' utility, through their love of variety. The sector has economies of scale (though the somewhat murky nature of these in the core model is something that will be discussed in detail).
- **Manufacturing firms**; these are assumed to be the mirror of workers, and vice versa: they don't make separate decisions or appear explicitly as individual agents in the model.

[5]The code used to produce the accompanying core model diagrams is downloadable via the link given in the code appendix (A).

They have a vital implicit presence, however, responsible for the key increasing returns dynamic that drives the model. This is explained in more depth below.

Each region has its own set of the three 'region equations' - these are given in table 2.1 with a number subscript indicating regions 1 and 2. This section deals with the simplest two region version of the model, and so will define the region equations separately for each of the two regions (as Brakman *et al.* do, p.119).

region 1	region 2
$Y_1 = \lambda_1 \delta W_1 + (\frac{1}{2})(1-\delta)$	$Y_2 = \lambda_2 \delta W_2 + (\frac{1}{2})(1-\delta)$
$I_1 = \left[\lambda_1 W_1^{1-\varepsilon} + \lambda_2 T^{1-\varepsilon} W_2^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$	$I_2 = \left[\lambda_1 T^{1-\varepsilon} W_1^{1-\varepsilon} + \lambda_2 W_2^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$
$W_1 = \left[Y_1 I_1^{\varepsilon-1} + Y_2 T^{1-\varepsilon} I_2^{\varepsilon-1}\right]^{1/\varepsilon}$	$W_2 = \left[Y_1 T^{1-\varepsilon} I_1^{\varepsilon-1} + Y_2 I_2^{\varepsilon-1}\right]^{1/\varepsilon}$

Table 2.1: 'Regional equations' for two-region core model

Before coming on to an explanation of the parameters, what do the 'region equations' achieve? Krugman's question for the core model was mentioned in the introduction: "why and when does manufacturing become concentrated in a few regions, leaving others relatively undeveloped?" (Krugman 1991b, p.484). In the model, a **difference in real wages between regions** is the mechanism for answering this: workers are attracted to where real wages are higher. Note that the concept of 'real wage' is different here from the more common understanding of the term, which would be 'what quantity of goods I can buy with my wage?' Instead, the 'real wage' is 'what quantity of **utility** can I buy with my wage?'

There are *two equilibrium stages* to the model. The first ('short-run') finds the **nominal wage in each region**: this is the job of the 'region equations'. This nominal wage can then be used to work out each region's real wage (see below). This sets up a tension to be resolved in the second ('long-run') equilibrium stage: workers move to regions with a higher real wage until regions end up with the same real wage, or all workers move to one region, leaving the other with zero real wage (and thus unable to attract anyone).

The following explains how all this is defined mathematically. To start with, these are the main variables in the 'region equations' from table 2.1:

- Y: total income for each region, from manufacturing and agricultural workers.
- I: the price index - the 'real' value of manufactured goods, or "what utility can I buy, given the price of manufacturing goods and my love of variety?" Its full meaning needs a detailed explanation; this is done below.
- W: The nominal wage rate (not the real wage; see below).

The four parameters used to find those values are as follows. These are all set exogenously, though as will be explained, they can be parameter-swept to understand the model dynamics:

- δ (delta): the share of income spent on manufacturing, as well as the fraction of workers in the manufacturing sector. It is not obviously the case that these two values should be same; this is discussed in the section on normalisations below. Note in the equations for total income (Y_1 and Y_2) the term $(\frac{1}{2})(1-\delta)$: this gives the number of agricultural workers in each region. The total number of workers is assumed to be 1 and so the number of agricultural workers is $1-\delta$. They are then split evenly between the two regions.

- λ (lambda): the fraction of manufacturing workers in each region. Note that $\lambda_1+\lambda_2=1$. This means that the number of manufacturing workers in the each region is $\lambda_1\delta$ and $\lambda_2\delta$.

- ε (epsilon): elasticity of substitution; as explained below, this relates to ρ and love of variety, but plays a broader role in the model.

- T (tau): transport costs between regions; the cost of goods is multiplied by this. If a good is being sold where it is produced, transport costs would be 1. This represents there being no internal costs for moving goods. (In the regional equations above, T is thus left out for the home region, since 1 raised to any power is 1). For other regions at a distance, $T>1$. T is defined by Brakman *et al.* as "the number of goods that need to be shipped to ensure that one unit arrives." (p.107) This use of 'iceberg' transport cost is, again, explained in more depth below.

The analytic goal of the core model is to find out where mobile manufacturing labour will migrate to. Will it agglomerate in one region - for example, $\lambda_1=1$ and $\lambda_2=0$? If all workers are in one region, the wage there will be equal for everyone. Or will it spread across the regions, perhaps evenly, where λ_1 and λ_2 both $=1/2$ and the real wage is equal in both regions? As mentioned, solutions to the regional equations do not supply this answer. Instead, they find the short-run equilibrium: how real wages in each region differ, given exogenous values for λ, δ, ε and T. The regional equations only supply the nominal wage - the real wage is as follows (where r indicates region; the derivation of this is explained below):

$$w_r=\frac{W_r}{I_r^{\delta}} \tag{2.17}$$

As can be seen from table 2.1, the equations are inter-dependent. Intuitively enough, income (Y) is dependent on wages (there are no other sources of income in the core model apart from wages). The price index I for manufactured goods is also dependent on wages since - as will be explained below - I roughly translates as "what utility can I buy for my wage?" Finally, the wage itself (W) is dependent on both the price index and income. Ultimately it is the wage that is of interest, as this will determine migration between the regions - but how to solve for it? For the original core model, this can be done numerically due to the complexity of the equations[6]. Fujita *et al* set out a simple explanation of their numerical method in a footnote[7]; Brakman *et al* expand on this (p.137). A start value is guessed, fed into the system of equations and iterated until a stable value emerges (for some given definition of 'stable' based on how much the nominal wage changes between iterations). Thus, if a guess is taken for W, both Y and I can then be calculated - and these can then be used to re-calculate W until a desirable level of stability is reached[8].

Once this stable result is found, the real wage can then be worked out. This is the short-run equilibrium, in which real wages between regions can differ. The core model can then use those values in its next stage to examine wage migration. The migration stage is actually very straightforward in comparison: the number of manufacturing workers in each region (as determined by the λ fractions) simply flows to where the real wage is higher.

The mathematical form of this migration is in a way arbitrary as long as the value of λ is changed to reflect the idea of workers moving to where real wages are higher. Section 2.5.5 presents how the dynamics work in the model to produce the 'long-run' equilibria. The move from short-run to long-run equilibrium does not have the same 'micromotive' foundation as the short-run part does (see section 3.2 for more on 'micromotives'). It is specified in Fujita *et al.* as a simple 'ad hoc' dynamic equation (p.62) that returns disequilibrium values of the difference between the two wages back to zero. They note it "has no deep justification" (p.77); it is just a way to describe what workers are doing when they move to seek higher wages.

As will be explained below when the dynamic outcomes are given in section 2.5.5, the final results themselves in their essential form require only knowing whether, on each side of an equilibrium point, these real wage forces will return back to that point (meaning that point is stable) or move away from it towards a new equilibrium. (This idea is also used in section 6.3.3 to dig into the transmission belt model's dynamic.)

By starting with a top-down view of the structure, it is hoped that the central dynamic of the core model is clear, as well as the way in which equilibrium solutions are reached. The next section examines the 'scaffold' used to build up to the regional equations.

2.5.4 Scaffold

This section explains how the 'regional equations' in table 2.1 are put together. There are the following elements to cover:

[6]Extensions to the core model were mentioned above; one of these makes some alterations that allow for a solveable version of the model, but require getting into quite different basic core model structures. See Brakman *et al*. 2009 pp.151

[7]It is very brief: "although it is not necessarily the most efficient procedure, simple iteration generally works: that is, start with guesses at W_r, calculate in sequence the implied Y and I vectors; calculate new values of W; and repeat until convergence" (p.77).

[8]This is skipping over a great many questions about showing that a stable solution is even possible for this system, but this discussion is beyond the scope of the thesis. It will be assumed that known solutions exist and the algorithm can reach them.

1. How utility for consumers is worked out and used to find the 'real wage'
2. How production is conceptualised
3. How transport between regions is defined

Some time will also be spent discussing how this scaffold is removed to leave the 'region equations'. This will leave two further elements to be discussed once the link between 'scaffold' and region equations has been made:

1. How manufacturing workers migrate between regions
2. What the results of all these are when put together - the dynamics of the core model.

Utility, real wages and the 'price index'

Real wages are at the heart of the core model. Consumers in the model want to know, 'what is the maximum utility I can buy for my wage?' To answer that, an explanation is needed of how utility is worked out, and how it is then used to derive the real wage. Utility for the model as a whole derives from a Cobb Douglas function, formulated as described in section 2.2.2 and equation (2.2):

$$U = F^{1-\delta} M^{\delta}, (0 < \delta < 1) \tag{2.18}$$

The core model uses a representative agent: this one utility function stands in for **all** actors' utility. A vital feature of the Cobb Douglas for the purposes of the core model is that demand for F and M are completely separate from each other. As the budget optimisations of the Cobb Douglas show (again in section 2.2.2), the equations describing the optimal amount of either F or M only contain their own price. This means there is no interdependence between them: changing the price of F can have no effect on demand for M and vice versa. The demand split between F and M is completely controlled by δ, and this shows up in the optimised budget equations. If the price of one of the goods drops, demand rises in exact proportion, keeping the spend on that sector exactly the same. Goods in each of the two sectors are 'unit elastic'. This is a feature of the Cobb Douglas that the core model relies on, as it allows M to be optimised separately, keeping it self-contained, unable to effect demand for F. (Section 5.2.2 looks at this point in more depth, as it has implications for agent decision-making.)

Whereas F is straightforwardly the quantity of food, M is actually *the utility from manufactured goods*, not the quantity consumed. Intuitively, it is obvious that a single value M could not represent a range of different manufactured goods directly, but this is a potential source of confusion, since - for example in Brakman *et al.* - it is sometimes described as straightforward "consumption of manufactured goods" (p.90). Fujita *et al.* (p.47) label M a 'composite index of the consumption of manufactured goods' (ibid p.46) but also note it can "can be thought of as a utility function". To keep things as simple as possible, that is exactly how it will be treated here.

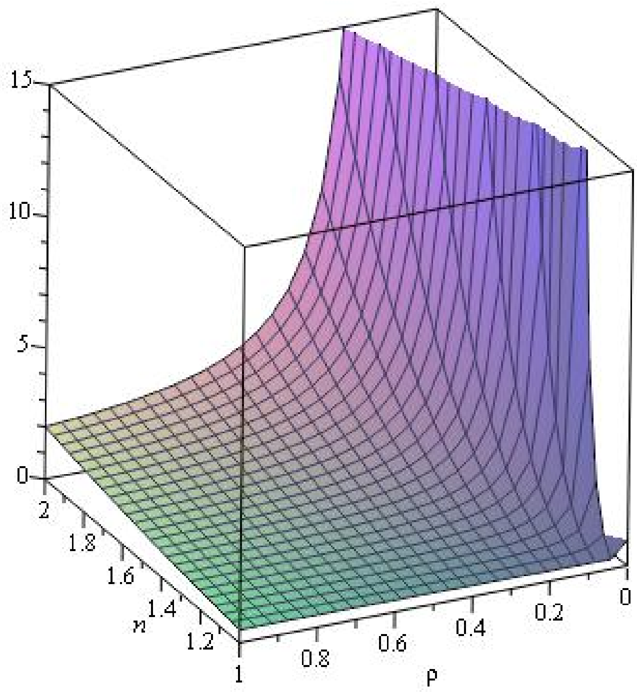


Figure 2.5: Continuous CES function, $0 < \rho < 1$, $1 < n < 2$; n is the number of varieties. Vertical axis shows utility.

The actual mathematical definition of M is a CES function containing each variety of consumed manufactured good, of the form described in equation (2.8) and repeated here. N is the total number of varieties consumed; $c(i)$ is the quantity of a single variety:

$$M = \left(\sum_{i=1}^{N} c_i^{\rho} \right)^{\frac{1}{\rho}}, (0 < \rho < 1) \tag{2.19}$$

So, for a given quantity and number of goods, increasing love of variety (as $\rho \to 0$) increases M.

A vital assumption of the Dixit Stiglitz model is that goods enter symmetrically into demand. As Fujita *et al.* note, this "lets us have our cake in discrete lumps while doing calculus on it, too" (Fujita *et al*. 2001 p.7). What this means in practice is that, in the many-good CES function (2.8), all c_i have the same value. Sometimes, this is guaranteed mathematically by defining it as an integral (see e.g. Fujita *et al*. 2001 p.46), where n is the (now continuous) number of varieties, and $m(i)$ stands in for the consumption of each variety. (Note that M is still manufacturing; in the Dixit Stiglitz model, M and F go into a Cobb Douglas, and so demand for each sector is independent of the other: changes in the size of M cannot affect demand for F).

$$M = \left[\int_0^n m(i)^{\rho} di \right]^{1/\rho} \quad 0 < \rho < 1 \tag{2.20}$$

It helps to get a quick visualisation of what this means by reframing (2.20), assuming one unit of each variety:

$$M = \left[\frac{n^{\rho+1}}{\rho+1} \right]^{1/\rho} \tag{2.21}$$

The starting point for the Dixit Stiglitz model approach has already been outlined in discussing the CES function. That section explained how the CES can be used to describe an 'optimum product diversity' from a utility point of view. It was explained that ρ controls love of variety: the closer to zero, the more variety is preferred. If the number of varieties is a continuous quantity, ranging from one to two, ρ looks like figure 2.5. M increases either as ρ approaches zero or as the number of varieties n goes up. Where $\rho = 1$, goods are perfect substitutes. In this case, it can be seen that M increases linearly with n. As will be explained, however, the number of varieties is endogenous in the core model.

The budget constraint for the representative agent consuming food and manufactures is as follows, where again $c(i)$ is a single variety and $p(i)$ its cost:

$$F + \sum_{i=1}^{N} p_i c_i = Y \tag{2.22}$$

The separation imposed by the Cobb Douglas applies to any pair of terms in the budget constraint, so as well as knowing that $F = 1 - \delta Y$, it is also the case that $\sum_{i=1}^{N} p_i c_i = \delta Y$. This second term can be used as the budget constraint for M - see below. The price of F does not change endogenously in the model; it is set to one, leaving just F in the budget constraint. The job of the agricultural sector is to provide demand in both regions so that it does not collapse to a point if and when manufacturing workers all move to one region. As mentioned above, in the region equation for Y, each region has $(\frac{1}{2})(1-\delta)$ agricultural income. It is $1-\delta$ across both regions, and this cannot change. Having positive demand in both regions serves a more prosaic role in the algorithm used to find the long-run equilbrium: it cannot 'complete' without it (the iteration algorithm described in section 2.5.3 is unable to reach a stable value).

One of the three 'region equations' above is the 'price index' I. This serves a very useful simplifying role in the core model. M has already been defined above as 'utility from manufactures' using a CES utility function. When M is optimised by its budget constraint ($\sum_{i=1}^{N} p_i c_i = \delta Y$) it becomes possible to rearrange the optimisation[9] to find that $M = \frac{\delta Y}{I}$, where:

$$I = \left(\sum_{i=1}^{N} p_i^{\frac{\rho}{\rho-1}} \right)^{\frac{\rho-1}{\rho}} \tag{2.23}$$

I is a function of the price of each variety, but also contains ρ, the love of variety parameter. As Brakman *et al* say, the price index is a "mirror image" of M (p.99) where more varieties - or higher love of variety - will *lower I*. Noting again that $M = \frac{\delta Y}{I}$, it can be seen that lowering I increases utility from manufacturing, as would be expected if either variety number or love of variety were increased. I thus roughly translates as "what utility can I buy, given the price of manufacturing goods and my love of variety?" This definition of I thus simplifies the optimal quantity of M to the budget for manufacturing varieties (δY) and the price index I, and nothing else.

The equation for real wages (2.17, $w_r = \frac{W_r}{I_r^{\delta}}$) can now be explained in more detail. If W is the nominal wage, the real wage is 'what utility I can buy with W'. For the core model's whole economy, Fujita *et al* use a 'cost of living index' (p.48; see Brakman *et al* p.99 also) to describe what it costs to buy a unit of utility: $I^{\delta} p_f^{1-\delta}$, where p_f is the cost of food. δ sets the spend on manufactures and so I^{δ} separately describes the cost of a unit of utility from manufacturing. As the price of F is set in the model to one and, since $1^x = 1$, the 'cost of living index' for the whole model reduces to I^{δ}. Real wages are then simply the nominal wage divided by this 'cost of living index', giving total utility. Differences in utility between regions can then be used to determine where mobile workers will be attracted to.

Production

Section 2.3.2 explained the difference between perfect and monopolistic competition. In the core model, agriculture is put in the 'perfect competition' category, and made spaceless. Or

[9] Brakman *et al* (p.94-5) and Fujita *et al* (p.47) piece their logic together slightly differently but (apart from the latter using integrable continuous varieties, not discrete) they arrive at the same place.

rather, is it argued that no transport costs exist between regions, which amounts to the same thing economically, though each region has its own fraction of agricultural workers.

Manufacturing is responsible for the model's dynamic outcomes. Two key stages occur to get to the region equations. The first, already discussed in section 2.3.2, is the monopolistic competition outcome that changes in production scale means **changes in the number of firms**, not in an actual shift in any firms' own production function or price. (This is where the externalities take effect in the

> "the size of the market affects neither the markup of price over marginal cost nor the scale at which individual goods are produced. As a result, all scale effects work through changes in the variety of goods available."

In terms of market structure, the Dixit Stiglitz model can guarantee that each firm only produces a single variety. Increasing returns take care of that: any firm producing one good will be locked in to producing it more cheaply than any other can. As a firm, it is always better to devote all of one's resources to that single product than attempt to split over more than one, since this allows full exploitation of increasing returns, and thus one's market power.

Because firms are locked into one product, the relation to love of variety means that economies have an equilibrium number of firms, and thus varieties. Net firm profits are set to zero as an equilibrium condition, underpinned by arguing that profits attract firms, thus lowering the markup for other firms until the zero profit condition is met, and losses drive them away.

So: one firm always produces a single variety, since no firm using economies of scale in a market where goods are differentiated can gain from producing more than one variety. Reducing changes in the market size to a growth or shrinkage in the number of firms (and thus varieties) is - as Fujita *et al* note - "a rather strange result" (ibid). They do note that other competition effects obviously exist, but simplicity wins out: reducing market changes to firm/variety number is "a dramatic simplification, allowing us to model cleanly issues that might otherwise seem intractable." As will be explained below, a key simplification is that price elasticity of demand ends up constant. Brakman *et al* point out (p.96), this is "the main advantage of the Dixit Stiglitz model aproach". It transforms an otherwise extremely complicated mathematical problem (the feedback between shifts in production function, output and prices) into a manageable model that, so it is argued, still captures the essential feedback mechanism that is key to understanding the core-periphery dynamic.

A series of simplifications then take place. The first of these is on the demand-side. In section 2.2.4, equation (2.11) was given as the optimal amount of a particular good resulting from optimising a CES function with a budget constraint. As discussed above, it is also possible to define a sub-equation, I, to make that optimisation more useful in the core model. In the core model optimisation, using I to simplify, the optimal amount for a single good is (Brakman *et al*. 2009 p.94):

$$c_j = p_j^{-\varepsilon}[I^{\varepsilon-1}\delta Y] \tag{2.24}$$

Monopolistic competition assumes that the market is ‘large’ - meaning that no single firm is able to effect it by changing price. Given this assumption, I takes on a vital role: it contains the price choices of all other firms ($p_{i...n}$) and, since a single firm is assumed unable to effect it, it can be treated as a constant (Brakman *et al*. 2009 p.96; Fujita *et al*. 2001 p.51-2), along with the split of income going to manufacturing, δY. Brakman *et al.* include the square brackets in equation (2.24) to indicate that the whole term can be treated thus. So, they define $constant_1 = [I^{\varepsilon-1}\delta Y]$. Substituting this into (2.24) means the optimal output of a variety becomes the following. This definition is key to how firm profit and number are worked out next.

$$c_j = p_j^{-\varepsilon} * constant_1 \tag{2.25}$$

On the demand side, then, the optimal amount of any single good is defined. This can then be plugged into the *supply* side. This is one point in the core model (though not the only one; see the construction of W below) at which demand and supply are linked together into a production equilibrium. The stages for this are as follows:

- Define the production function in terms of l (the straight-line equation $l = \alpha + \beta x$): “the amount of labour necessary to produce x” (Brakman *et al*. 2009 p.102). Brakman *et al* use x_i to label ‘quantity of variety i’. This is different from c_j only in that the latter is the *optimised* quantity of good. α is the fixed amount of labour required, and β is the marginal amount. (This is equation 2.12 for ‘fixed’ economies of scale, discussed in section 2.2.5.)

- Define a firm’s profits π using the production function. Profits are equal to total takings (price times quantity) minus total outgoings (wage times amount of labour):

$$\pi = px - W(\alpha + \beta x) \tag{2.26}$$

- Firms want to maximise their profit. This is done by maximising the profit equation using what has been deduced about demand - thus linking optimal supply and demand. So, substitute in equation (2.25) (the optimal amount of good, given what economic facts firms take as constant) to replace x. If this altered profit equation is differentiated and set to zero, this is describes the first order point of maximum profit. Through rearrangement, it ends up as a simple optimal price equation, known as ‘markup pricing’ (Brakman *et al*. 2009 p.103-4). This is:

$$p = \beta W/\rho \tag{2.27}$$

- Given that the love of variety parameter $0 < \rho < 1$, if $\rho = 1$, the markup price becomes just $p = \beta W$. So, where there is no love of variety and all goods are substitutes, price is equal to marginal cost. As $\rho \to 0$ and goods become increasingly differentiated, so the price each firm charges increases above marginal cost.

- This potential for profits is where the dynamic for changing firm numbers comes from, as section 2.3.2 explained. Profits attract firms, losses drive them away. So if profits in equilibrium are zero, the profit equation (2.26) can be set to zero. Also, the optimal 'markup price' has been deduced, so this can be substituted into the profit equation, replacing p. Optimised demand and supply are now linked, and a number of key mathematical deductions then follow that each lead on to the next.

- First: an optimal amount of good in equilibrium is found:

$$x = \alpha(\varepsilon - 1)/\beta \tag{2.28}$$

  Second: this optimal quantity for x, when plugged into the production function (2.12) gives the "amount of labour required to produce this much output" (Brakman *et al*. 2009 p.105): $l_i = \alpha\varepsilon$. Lastly, the equilibrium number of firms N, which is just the number of manufacturing workers divided by the number required to produce one variety (since each firm makes only one variety). Taking into account that λ_1 and λ_2 describe the split of workers across two regions, it becomes $\lambda_n\delta/l_i$ or, using the full definition for l_i:

$$N = \lambda_n\delta/\alpha\varepsilon \tag{2.29}$$

As mentioned above, causally, firms follow workers in the core model: their equilibrium number is a direct result of the quantity of labour and demand in a region. Fujita *et al* are explicit about this as well: "we are assuming that the entry and exit of firms occurs very fast - so profits are always zero - but relocation of workers among sectors or locations occurs more slowly, with a dynamic we model explicitly" (Fujita *et al*. 2001 p.53). The latter dynamic is the long-run equilibrium of workers moving between regions.

Piecing the 'region equations' together

The equations deduced above can now be used to build the 'region equations', with one extra addition: transport costs. Before coming onto these, however, the equation for regional income Y is the most straightforward of the three and easy to clear out of the way first. There is only wage income in the core model: the zero-profit conditions for both sectors rule out any income for firms or owners (Brakman *et al*. 2009 p.113). The agricultural sector is entirely static, its income set to $1 - \delta/2$ for each region. Income from the manufacturing sector is just its wage W

in each region, times the number of manufacturing workers there: $\lambda_n \delta W_n$. Y is the sum of these two:

$$Y_n = \lambda_n \delta W_n + (\frac{1}{2})(1-\delta) \tag{2.30}$$

As already mentioned, T defines the transport costs between regions[10], having a value of one for 'here' and $T > 1$ for the distant region. T is used to simply multiply the cost of goods, so they remain the same for 'here' and have a transport markup for elsewhere. The effect on demand is described by Brakman *et al.* thus (p.115): "If there are positive transport costs, that is $T > 1$, demand in region 2 for products from region 1 is lower than without transport costs, because transport costs make them more expensive." These transport costs can be plugged directly into equation (2.27), the cost a firm charges for a variety:

$$Tp = T\beta W/\rho \tag{2.31}$$

Equation (2.31) is the final piece of the puzzle that allows one of the region equations, the price index I (equation 2.23), to be constructed (though note that some normalisations are required to reach the versions in table 2.1; these are discussed below, but the result will be to leave them stripped of any values for α, β and ρ). I uses a sum of the price each firm charges. Equation (2.31) can provide the price each firm charges in all regions it sells to. The equation for firm number (2.29) was given above: if $l_i = \alpha\varepsilon$, the number of firms is $\lambda_n \delta/\alpha\varepsilon$. As Fujita *et al*. 2001 note (p.50), since all varieties in a single region have the same non-transport price, it is possible to simply multiply the number of firms by this price, for each region being sold in - meaning that equations (2.31) and (2.29) can be substituted into (2.23) to produce a value for I in each region[11]. This is the n region version, where I_r is the price index in region r ('here') and s is an index for all regions:

$$I_r = \left(\frac{\beta}{\rho}\right)\left(\frac{\delta}{\alpha\varepsilon}\right)^{\frac{1}{1-\varepsilon}}\left[\sum_{s=1}^{N} \lambda_s W_s^{1-\varepsilon} T_{rs}^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{2.32}$$

This leaves only the regional wage W. This is solved by assuming supply equals demand, and working out both sides of that equation (Brakman *et al*. 2009 p.115). Once more, this process starts by using the equation for demand for a variety (2.24) and substituting. The 'markup price plus transport costs' equation (2.31) stands in for price. The result describes "demand in region r for a product from region s" (ibid. p.116):

$$x_s = \delta\left(\frac{\beta}{\rho}\right)^{-\varepsilon} Y_r W_s^{-\varepsilon} T_{rs}^{-\varepsilon} I_r^{\varepsilon-1} \tag{2.33}$$

[10] Since only the two-region core model is being discussed here, T does not need a subscript. Transport costs for the home region are 1, T otherwise.

[11] Again, Brakman *et al.* provide a note showing these two equations plugging directly into the price index (p.112). Equation (2.32) is simply an algebraic rearrangement of this, keeping W and T tidily in the region summation.

Equation (2.34) is included here because the next step captures where the 'iceberg' part of transport costs come in. The first thing that needs to happen is to find demand for x_s in all regions. (Note that W_s can be moved outside the summation as r is being iterated.)

$$\delta \left(\frac{\beta}{\rho} \right)^{-\varepsilon} W_s^{-\varepsilon} \sum_{s=1}^{N} Y_r T_{rs}^{1-\varepsilon} I_r^{\varepsilon-1} \tag{2.34}$$

As well as rearranging, the whole has been "multiplied by T_{rs} to compensate for the part that melts during transport" (ibid). Thus, equation (2.34) is what demand firms face, which must include the full good - hence the key difference in demand from the point of view of the consumer and the firm, seen in the difference in T: $T_{rs}^{-\varepsilon}$ multiplied by the transport cost to include what 'melts away' when consumerd is $T_{rs}^{1-\varepsilon}$. Consumers get x, for a price Tp that includes the transport cost. Producers have to *make* $x * T$.

The supply quantity, the 'optimal amount of good in equilibrium', was already given above in equation (2.28): $\alpha(\varepsilon - 1)/\beta$. In equilibrium, this must match the demand given in equation (2.34). The two when equated can be simplified and rearranged to give the equation for W (ibid. p.117):

$$W_s = \rho \beta^{-\rho} \left(\frac{\delta}{(\varepsilon - 1)\alpha} \right)^{1/\varepsilon} \left[\sum_{r=1}^{R} Y_r T_{rs}^{1-\varepsilon} I_r^{\varepsilon-1} \right]^{1/\varepsilon} \tag{2.35}$$

Normalisations

As mentioned, the core model equations can be simplified by 'choice of units'. As Fujita *et al.* 2001 say (p.54), "we are free to choose units of measurement - be it units, tens of units, kilos, or tons." Given the fact that only λ actually has any dynamics, this leaves a lot of static variables that can be set as the modeller wants. The marginal labour requirement β can thus be set to equal ρ. This has no economic meaning, but it does allow a number of simplifications. Immediately in I_r, the term β/ρ becomes 1 and disappears. Next, the fixed labour requirement can be set to $\alpha = \delta/\varepsilon$. Again in I_r, this then allows the term $\delta/\alpha\varepsilon$ to become 1 also. This leaves I_r as it is in table 2.1.

The same two normalisations can also be used to reduce W to its simpler form. This is easiest done before rearranging in terms of W with the two supply and demand equations. First, the supply-side, equation (2.28) $\alpha(\varepsilon - 1)/\beta$ reduces to δ by first substituting ρ for β and $\varepsilon - 1/\varepsilon$ for ρ, leaving $\alpha\varepsilon$, then using $\alpha = \delta/\varepsilon$, ending up with just δ. On the demand side, in equation (2.34), the term $\left(\frac{\beta}{\rho}\right)^{-\varepsilon}$ also becomes 1 and disappears. Labelling the summation as a constant and equating supply to demand, that just leaves $\delta = \delta W^{-\varepsilon} C$. Isolating W, δ can cancel, leaving $W = C^{1/\varepsilon}$. Substituting the summation back in leaves W in its 'region equation' form.

One of the normalisations later added by Brakman *et al.* is actually assumed to begin with by Fujita *et al.*, and has already been mentioned: δ is both income for the manufacturing sector and the number of manufacturing workers. This simplification has been used throughout in order to keep the explanation as clutter-free as possible[12], along with the assumption that the total number of workers is 1. Note also, however, that using the normalisations above, the equilibrium supply quantity also became δ.

The above section has pieced together the construction of the 'region equations'. These equations can now be used to examine the dynamics the core model is interested in.

2.5.5 Dynamics

As section 2.5.3 explained, the core model's goal is to identify a dynamic able to produce a manufacturing 'core' and an isolated 'periphery' using a set of simultaneous equations describing the feedback between income Y, a price index I and the wage level W. That section also explained that 'a simple iteration process' is used to find convergent values, resulting in a value for the real wage of each region.

In this section, figures 2.6a and 2.6b are produced from this process, plus a second process for working out where the tension in real wages set up in the short-run leads workers to move. The figures illustrate the data in a way seen in many presentations of the core model: a short-run equilibrium showing points where real wages are out of equilibrium and where they are equal and a full set of long-run equilibria for a range of transport values[13].

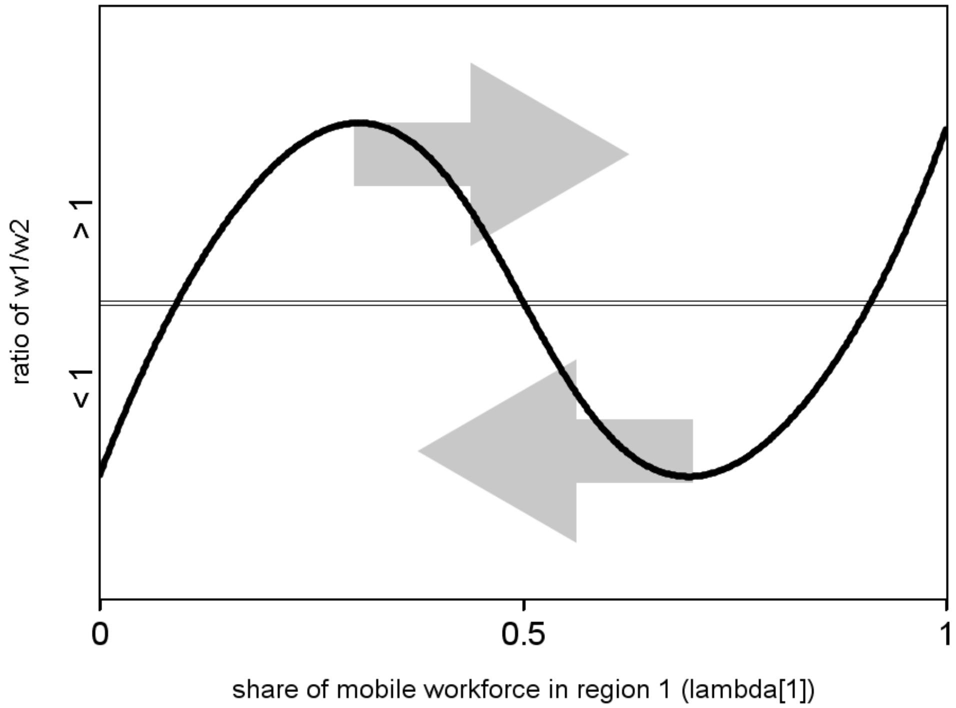
Figure 2.6a shows the first, 'short run' outcome of this process for all values of λ_1 between zero and one. If $\lambda_1 = 0$ on the very left, all workers are in region two. Conversely, at $\lambda_1 = 1$, all are in region one. For each of these points, the thick black line shows the ratio of the real wage across the two regions (w_1/w_2). When this equals one, where the line crosses the horizontal midpoint (marked by the double line) real wages are the same in both regions.

The vertical axis indicates that this ratio is more than one in the top half and less than one in the bottom half. When $w_1/w_2 > 1$, real wages in region one are higher: a tension is presumed to exist that will draw manufacturing workers towards region one. The direction of this tension is indicated by the grey arrow pointing to the right: any points in the upper half (where $w_1/w_2 > 1$) imply workers moving in that direction (towards $\lambda = 1$). The reverse applies for the bottom half, where real wages are higher in region two.

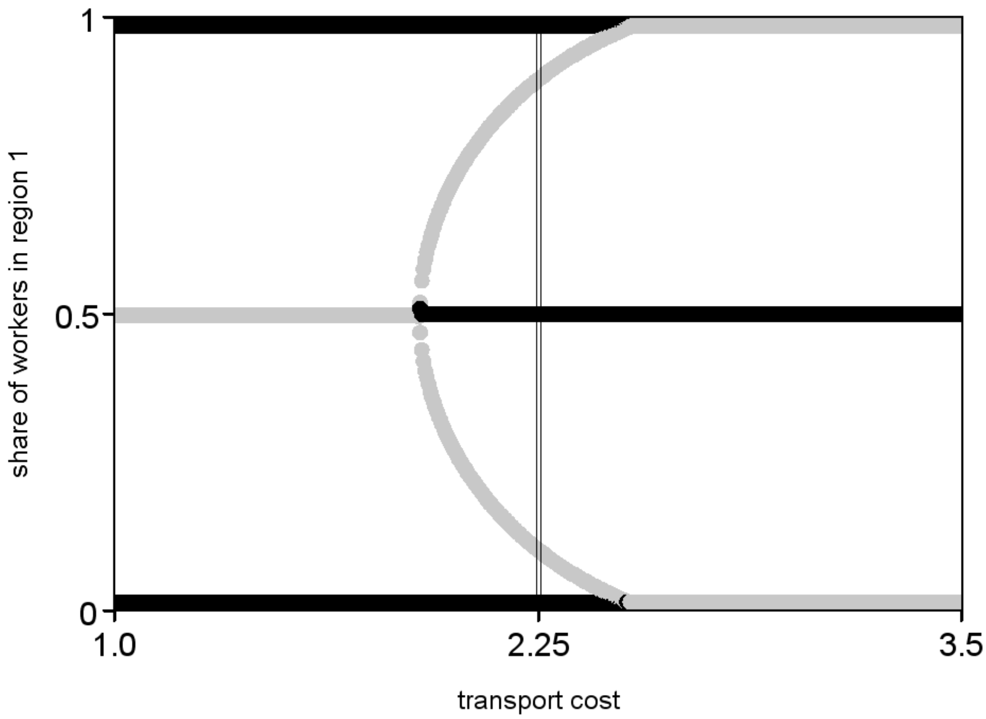
All other values are fixed exogenously. Manufacturing and agricultural workers are divided evenly ($\delta = 0.5$). Transport costs are set to $T = 2.25$. As the 'long run' dynamic explanation below will show more clearly, this is in a particular part of the variable space that produces five possible equilibrium points. With the understanding that workers move to higher real-wage regions, it is possible to identify which points are stable and unstable. Where $\lambda_1 = 0.5$, a stable equilibrium results, evenly splitting workers between the two regions. Pushing real wages to either side of this point sets up a tension that would bring it straight back (i.e. would be 'pushing

[12] Charles van Marrewijk provides an analysis discussing income and worker number being set to δ on the website to accompany the Brakman *et al.* book. This can be found at: http://www2.econ.uu.nl/users/marrewijk/newgeo/answers/newgeo%20answer%203.pdf

[13] Again, the commented code for producing them is available via the download link given in the code appendix (A).



(a) Short-run equilibrium



(b) Long-run equilibria 'tomahawk'

Figure 2.6: Core model short-run equilibrium for $T = 2.25$ showing real wage ratio between the two regions before migration; and long-run equilibria for $1 < T < 3.5$ showing the range of stable (black) and unstable (grey) equilibria. Double line in both crosses the same five equilibrium points for $T = 2.25$.

against' the migration direction indicated by the arrows). Equally, the two 'core' endpoints are stable - though note this is not because the ratio of real wages has equilibriated but rather because all workers have moved to that region and have no incentive to move away.

The final two points where the real wage ratio crosses the line at one are unstable. Moving an infinitesimal amount to the left of either of these points (so below the central line) implies workers moving to the left on the graph - either towards full agglomeration in region two or towards the point of even split. The same applies in reverse if the real wage shifts slightly to the right of those points.

Figure 2.6b identifies a range of long-run equilibria showing where workers ultimately end up as they move toward higher real wages. This is done for a spread of transport values between one and 3.5 (starting at one because, as mentioned, transport costs in the core model describe 'the number of goods that need to be shipped to ensure that one unit of a variety arrives'). The share of workers in region one (λ_1) is shifted onto the vertical axis. The value of $T = 2.25$ used in figure 2.6a is at the middle point of the horizontal axis; the double line in both figures crosses the same five equilibrium points but is on the vertical axis in figure 2.6b.

Five hundred data points were taken across the range of transport costs and, for each of those, five hundred between $0 < \lambda_1 < 1$. Each datapoint is checked to see if it crosses the line where real wages are equal. If two points do so, they are on each side of an equilibrium. Each is then checked for stability.

To be stable, the real wage on each side must move workers back to the same point. When does this happen? The arrows in figure 2.6a help make this clear: where the ratio is more than one for the lower value of lambda (or where values drop going 'left to right' in figure 2.6a). For example, if $w_1/w_2 > 1$ (or real wages are higher in region one, pushing workers 'to the right' towards that region) workers will be pushed towards higher values of λ.

Black lines show where the stable equilibria are found. Grey lines show unstable equilibria (or show where all workers being present in one of the regions is not stable and will move towards a split between the two). The five equilibrium points from figure 2.6a show up at $T = 2.25$ as three stable and two unstable points.

2.5.6 Core model: concluding thoughts

The core model is made up of a series of optimisation stages. These are, in the scaffold stages, all about finding points of no change: Lagrangians for the budget optimisations, a profit-maximisation (including some demand simplifications - the added constant term) and a zero-profit rule (built on the monopolistic competition argument from the Dixit Stiglitz model) that is responsible for the equilibrium firm number. In the final 'long-run' equilibria, an 'ad hoc' approach moves workers to the region with the highest real wage.

The core model is the canonical example of how GE 'solves' the dependency issues of spatial economics described in section 2.3. This section has examined in detail the way GE's assumptions have been implemented, providing a solid base for comparing to those of ABM and preparing the ground for explaining what components and assumptions the thesis model framework uses.

In particular, the transmission belt model in chapter 6 builds up from separate agents making their own economic decisions. It uses methods that allow asymmetric consumption of goods in order to allow a flexible way to examine different spatial setups. Also, rather than assuming increasing returns manifest through changes in the equilibrium number of firms, a model is presented that utilises increasing returns at the production level (section 6.3).

That chapter also picks up on many points covered in this section regarding particular features of the core model's assumptions and construction. This theme is taken up in depth in the conclusion, which includes a direct comparison between the core model and the thesis model framework (section 7.1).

2.6 How space is theorised in the thesis model framework

This and section 2.7 concentrate specifically on the background for how 'space costs' are implemented in the thesis model framework. 'Space costs' is used as a top level term to describe all costs that impact on agent decisions in a spatial setting. In the thesis, there are two distinct types of space cost. 'Distance costs' are the actual costs of moving goods or people across physical distances. 'Proximity costs' are used to describe any costs or benefits that arise from agents being closer to or further away from each other; these are often discussed in terms of 'externalities' (see section 2.7). Each of these has a matching section within chapter 4 where the specific mathematical and coding set-up for each is explained. Proximity costs, in the thesis framework, become a 'density cost' that is incurred by each agent. These are covered below under three broad headings:

- The cost of moving goods (connects to section 4.3.1)
- The cost of moving people (connects to section 4.3.2)
- The cost and benefits of proximity (connects to the density cost outline in section 4.4)

Section 4.3 explains how these three are translated into the model framework itself. The final section in this chapter (section 2.8) examines a number of space cost issues that, while important for a full understanding of spatial economics, have been placed outside the scope of the thesis model framework.

2.6.1 Goods

The thesis model framework takes a very literal approach to distance, in contrast to the 'iceberg' transport costs of the core model (section 2.5), where the goods themselves 'melt away' and a single parameter controls "the number of goods that need to be shipped to ensure that one unit of a variety of manufactures arrives per unit of distance" (Brakman *et al*. 2009 p.109). In GE, distance is used as a broader term that can include general trade-cost impediments to the exchange of goods across space. For an ABM approach, it makes sense to treat distance in the more literal way done here, allowing actors to make decisions about specific goods with given prices.

One useful way of thinking about how this works is to consider the idea of 'value density'. This can be defined as the ratio of product value to physical size or weight (Lovell *et al*. 2005 p.144). What this idea shows is that, the more value is added, the more feasible large space costs become. That is, if a good's price is increased, the good does become more expensive - but space becomes relatively less important. Conversely, as Lovell *et al.* note, "as the value density of a product decreases the percentage of additional distribution costs increases" (ibid.). In the models presented in section 5.4 this effect of space becoming relatively less important manifests through Peoples' spatial choices: as good costs go up, they disperse spatially. Proximity to Firms matters to them less.

More value is added, by definition, with larger increasing returns (see section 2.2.5). A unit of input labour can produce a higher output. That output, however, could be measured either in added value or an increase in physical output. This is an old point of Weber's: he distinguished between processing that added weight or removed it. Lovell *et al.* note Coopers' research on the production of microchips as a classic example of combining massive scale with very high value density (Cooper *et al*. 1990 in Lovell *et al*. 2005 p.144). Glaeser and Kohlhase give a good list of shipped goods with different value densities: in the U.S., wood products' average shipment length was 287 miles at their time of writing and transport costs a fifth of the value. For base metal, a tenth of the value went on transport. In contrast, machinery and electrical goods were always less than 1.2% transport costs (Glaeser and Kohlhase 2004 p.208).

As section 2.8.1 explains, choice of transport mode is excluded from the model, but it is worth noting the relationship between value density and mode choice. As Lovell points out, it is only high value-density goods that get air-freighted (Lovell *et al*. 2005 p.153). This highlights how important, in reality, the time element of good transport is.

2.6.2 People

As section 4.5 explains, the models presented in this thesis run for many timesteps, each timestep considered a 'day'. People are assumed to have 'one day per day' to spend and optimise using that time **only** at their point of decision. By reducing 'time' decisions down to single-day optimisations, no agent is required to make any calculations involving other time periods. This is a simplification compared to real time-based decisions spread over shorter or longer timespans. In reality, real-world freight shipment happens over many days, weeks and even months and can be modelled accordingly (see, for example, Hummels 2001). It also underpins problems of risk and uncertainty in trade across space (see section 2.8.3).

There are two rationales for giving agents this simple conception of time. The first is the primacy of time for the decisions people take about space - for example, the relationship between home location and commuting times. The second is that time is unique, in that an uninterrupted flow is given freely to actors, at exactly the same constant rate. It is exogenous to production and space. Equally, once used - either ‘spent’ on space or production - it ‘melts away’. Most of the models presented here give actors a constant ‘fixed income’ in this way; this can equally be considered as a fixed daily income or a ‘free gift’ flow of time to optimise. If there is an intermediary stage of acquiring a wage, it still needs to be bought with time.

The cost of moving people - and especially the time cost - has become increasingly the most important space cost. This is an argument made most forcibly by Glaeser (Glaeser and Kohlhase 2004, Glaeser 2008; see also Beckmann 2010, p.3). The reason is intuitive enough: as transport costs for goods drop, people’s time becomes relatively more important. A knock-on consequence has been to emphasise those elements of human productivity that rely on time proximity. Leamer and Storper explain this by pointing out that -

> “humans remain the containers for shipping complex uncodifiable information. The time costs of shipping these containers is on the rise because of congestion on the roads and in the airports while the financial costs of so doing are also rising due to increases in real wages of knowledge workers who are the human containers” (Leamer and Storper 2001 p.648).

The spatial consequences of this are most strongly felt for those activities requiring the most time - and employment tops this list. Close seconds are access to housing and goods. The spatial solution to this problem does not necessarily require people to stay within a day’s radius: Ojo, for example, studied agricultural ‘commuters’ in 1970s Yorubaland who developed a “fortnightly pattern of periodic commuting to reduce the friction of distance”, moving between more distant settlements (Ojo 1973 p.86; Chapman 1979 p.122). Closer to home, commute patterns that span larger distances than one day’s range are not uncommon, but the vast majority keep to the geographical limit bounded by feasible daily commute times.

The cost of moving people is of paramount importance to firms; the idea of external costs (see section 2.7) comes in particularly useful here. Firms face these costs from people travelling to work or moving to get goods and services. Firms primarily want to locate near to labour: as well as being intuitively obvious, this is a key finding from the literature. For instance, “using the same workers is approximately eight times more likely to increase the degree of co-location than are trade relationships” (Glaeser and Kohlhase 2004 p.223). Similarly, service providers are usually reliant on proximity to customers. In both cases, the same *type* of space cost exists: that incurred by people moving location. A simple example would be a hairdresser: the service provider has externalised one of the major costs - getting the person’s head under the scissors from some distance away. In the case of commuting, the cost is less cleanly external or internal to the firm’s production structure, since wages in part must cover it. (Section 5.5.2 has an example of such an external cost and how two firms cope with them.)

For people, the direct financial cost of transport is very high, relative to how much it costs firms to move goods. Button finds that "after housing and food, transport represents, at about 10 to 15 percent for a single-car-owning household, the largest element of expenditure out of income in industrialised countries" (Button 2010 p.90). In the U.S., Glaeser and Kohlhase note both the money and non-money movement costs people face: they cite a 2001 consumer expenditure survey showing that, in the U.S., "18% of total expenditures for the average household is spent on vehicle purchases, gasoline and other vehicular expenses" - but point out that time costs are not accounted for in that, and that "these time costs are not withering away with technological progress" (Glaeser and Kohlhase 2004 p.208).

2.6.3 The interaction of people and goods

The most obvious difference between moving people and goods is the distance they cover. The prominence of time-cost for people dominates. (As mentioned, while time is a key cost for moving goods in reality, here the time element of good movement is ignored.) Good movement also does not impose the same hard limit as commuting: distance means smaller demand, but not necessarily zero demand. The geographical result of these differences is an economic landscape adapted to minimise both people-costs and good-costs into a series of overlapping plates of economic activity: Isard's 'gestalt whole' (see introduction).

This connects to the problem of analysing the complete impact of costs, not just those that can be measured - a point made by Burstein *et al.*, who note that concentrating just on incurred costs is -

> " - likely to underestimate true transportation costs. This is because we do not observe the cases in which transportation costs were so high that the transaction did not take place." (Burstein *et al*. 2003 p.1198)

While this problem is easy to state, it is harder to work out how those costs could be quantified. After all, it is essentially a counterfactual: what trades would have taken place if costs were lower or demand higher? The fact remains, however, that the observable situation is one where cost minimisation has already taken place. There is an overlap here with the idea of 'demand thresholds', where high costs cut off transactions (Chapman 1979 p.86). This has tended to be applied to consumer demand problems, rather than the economy as a whole. It is certainly easiest to understand Burstein's point by applying it to people: as discussed above, time thresholds mark a clear limit on how far people can travel to work (without needing some strategy to get around the day's time limit). The same applies to firm inputs other than labour, however. Value density and demand combine to limit the feasible range any good can be traded.

The size of the ratio of space costs to other costs is an important determinant for actor outcomes. Consider the service industry: Glaeser and Kohlhase correctly point out that since "transportation costs for people are still high, but transportation costs for goods are not" (Glaeser and Kohlhase 2004 p.220), proximity to people is the most important spatial factor for services. They also note (p.201) that "services tend to involve little freight shipment". But what about transport costs as a proportion of the service sectors' overall costs? For the U.K. at least, Diamond and Spence found they were 9.9%, compared to 4.7% in manufacturing (Diamond and Spence 1989 in Button 2010 p.58). Unsurprisingly, much of the service sector has lower overheads than other sectors. The paradoxical result is that services are more space cost sensitive.

The introduction began with Glaeser and Kolhase claiming that "it is better to assume that moving goods is essentially costless than to assume that moving goods is an important component of the production process" (Glaeser and Kohlhase 2004, p.199). They are certainly right about the relative change in importance for time versus other space costs, but assuming away those other space costs is a large step to take.

This point overlaps with the 'ecological criticism' discussed in section 2.3.5: it seems, mathematically, that the importance of transport costs can be reduced simply by increasing the cost of other inputs into the production process. By making other inputs more valuable, the importance of space can be lessened, with no limit on how much they can be reduced.

A spatial version of the ecological criticism might go like this: it is obviously nonsense to think that distance costs can be reduced to nearly zero simply by increasing the value of other inputs. First, there are hard cost limits to moving goods and people. Second, economic value cannot be magically conjured out of thin air; as Daly put it, one cannot simply 'stir faster in a bigger bowl'. Space is a crucial ingredient. Exactly how little of that ingredient is required? This problem is central to recent attempts to find ways of 'decoupling' transport and economic productivity. Decoupling can be defined as "achieving economic development without a proportional increase in transport activity (and emissions) (Gray *et al*. 2006 p.3). Historically these two have been umbilically linked (Echenique 2002) and, while there is evidence that transport intensity (amount of transport per unit of GDP) has been dropping in the richest countries, an attendant increase in demand for oil in developing countries suggests that outsourcing of high-intensity activities is at least partly responsible (on Peak Oil and Security 2010 p.29).

As chapter 3 discusses, there are no easy routes to dismissing Glaeser's simplifying assumptions out of hand as "unacceptably distorted" (Granovetter and Swedberg 2001 p.9). But ruling out non-time space costs in this way appears to also rule out asking, 'what happens when they change?' It also makes it difficult, if not impossible, to investigate the relationship between space costs and productivity.

2.7 The costs and benefits of proximity

This section explains the rationale for the 'density cost' approach used in the thesis model framework. The implementation of the density cost itself is described in section 4.4. As the density cost results explain (section 5.4), the density cost allows agents' decisions to accrete through external interactions or 'externalities'. The meaning of 'externalities' and their place in spatial economics is discussed first. The density cost is used as a proxy for a 'basket' of proximity costs including congestion and land costs; this idea is discussed in section 2.7.3. Finally, section 2.7.4 explains the idiosyncratically agent-based way in which 'second nature', spatially relative density costs are theorised.

2.7.1 Externalities

Externalities have a central role to play in analytic approaches to spatial economics. Section 2.7.2 looks in more depth at the way externalities are used to define the costs and benefits of being in a specific place. The rest of this section looks at the idea itself in more detail. Button defines externalities as where -

> "... the activities of one group affect the welfare of another group without any payment or compensation being made." (Button 2010, p.161)

The concept of externalities is so central because of the role they play in market structure. The pivotal chunk of Button's definition is 'without any payment or compensation being made'; externalities are outside of the price system. Scitovsky (1954) first clearly formulated their meaning as 'direct interdependence' between economic actors. Scitovsky defined this interdependence in two ways, as technological (or 'pure') and pecuniary externalities. The first, technological, he defined in relation to a firm's production function -

$$x_1 = f(l_1, c_1, ...; x_2, l_2, c_2, ...) \tag{2.36}$$

- where x_1 is its output (Scitovsky 1954 p.145). l_1 and c_1 are its own inputs. By including a second firm's output (x_2) and its inputs (l_2 and c_2), Scitovsky explicitly states the interdependence: any alteration in the second firm's outputs or inputs will impact on the output of the first, changing the technical relations in its own production function. It is more common in spatial settings for externalities to be considered as an aggregrate effect within a region; for example, in Brakman *et al.* (2009 p.38), where "an increase in *industry-wide* output alters the technological relationship between inputs and output for each individual firm. It therefore has an impact on the firm's production function" (emphasis added). This is how external economies work, then: a firm's production function is dependent not only on the inputs it buys, but the actions of other firms. Early on, Scitovsky pointed out the rarity of this kind of 'pure' externality:

> "It is not easy to find examples from industry. Going through the many examples of external economies quoted in the literature, I found only two that fit the above definition: the case in which a firm benefits from the labour market created by the establishment of other firms and that in which several firms use a resource which is free but limited in supply." (p.145)

Examples of the latter he gives are all common pool resources: oil-well production changes due to others drilling nearby; fishermen using the same waters; congestion problems on a public road. As he notes, even labour market pooling can be seen as a common-pool resource.

Moving on to the second type, in similar fashion, Scitovsky defines pecuniary externalities in relation to a firm's profits (Scitovsky 1954 p.146):

$$P_1 = g(x_1, l_1, c_1, ...; x_2, l_2, c_2, ...) \tag{2.37}$$

These profits are, of course, a function of its output and inputs, but also of other firms'. These affect the *profits*, however; they do not alter the production function. Scitovsky's example of a pecuniary externality is the effect a firm's own investment decisions can have on others. One firm investing to increase its efficiency, for example, could give it better economies of scale; it can sell its output for less. Inputs to the firm increase as a consequence - this is where a pecuniary externality can be found. That increase in demand will push the price of that input up. For other firms using it, that is a negative externality. For any firms selling it, it is positive (Scitovsky p.149).

Scitovsky points out that externalities are - for general equilibrium theory and many arguments about market structure - "the villain of the piece and the cause for conflict between private profit and social benefit" (Scitovsky 1954 p.144). One of the key elements of GE's success stems from having managed to combine externalities and equilibrium in a way that avoids this apparent villainy.

For each of the model types in this thesis, externalities play a role. The outcome of the density cost is a form of externality caused by one Person's decisions impacting the choice set of the next (section 5.4.4). The two-Firm DMR model (section 5.5) examines the problem Firms have distinguishing the effect of their own action from that of another Firm - another externality. The transmission belt model's equilibrium outcome is based on price signal adjustments causing both internal and external demand effects (chapter 6).

2.7.2 Externalities and proximity

The obvious benefit of proximity is a simple cost saving: being closer means less distance to cover. This is a point often absent from discussions of space costs. Black and Henderson (1999, p.327) are one of the few to explicitly mention it: "capital good plants agglomerate in locations with high manufacturing employment even though this generates no externalities: it just conserves on transport costs of exchange." More usually, the costs and benefits of proximity are described using the language of externalities. This is how the key dynamic of the core model is often framed: the effect of workers choosing to move to one region is to trigger an externality as demand for goods shifts the production structure.

This kind of regional 'localisation externality'[14] is an effect that derives from being 'here',

[14] The terminology used varies widely and can be confusing. These are agglomeration effects (sometimes

and can be either a cost or benefit. In much of the literature (and in the core model) 'here' is a discrete place, which actors are either inside or out. Section 2.7.1 explained that externalities can be defined by introducing external values directly into an actors' utility or production function, or introducing them after these - for example, into a firm's income level. In a discrete space, external economies are easily implemented. For instance, in a two-region model with regions r and s, firms in r could include in their production function the total output of other firms in that region, nx. Using Scitovsky's way of defining an externality, that would be:

$$x_r = F(l_1, c_1, ...; nx) \tag{2.38}$$

In Scitovsky's terms, this is a pure externality since it enters directly into the production function. Economically, what does it mean? The classic example goes back to Marshall: the number of firms of the same industry in one locality increases the stock of knowledge for all firms there. This can be explained in terms of the number of workers in a region also; their presence acts as a knowledge multiplier as they move between firms. (See e.g. Krugman 1991c pp.36, and Marshall himself, 1895 pp.349.)

These 'Marshall externalities' are often contrasted to 'Jacobs externalities'. These attempt to capture the central thrust of Jane Jacobs' argument: that diversity of production lies at the heart of regional productivity (Jacobs 1970, 1986). A simple way of achieving this would be to insert a CES-type function into (2.38) rather than the raw number of firms: a region's productivity could thus, in effect, have a 'love of variety'. Duranton and Puga (2003 pp.4) use a more sophisticated version of this to model a region's gain from sharing a variety of intermediate inputs. The effect of competition between firms can also be reduced to an externality in the same way (again, as it is in the core model; section 2.5.4).

The common element to all localisation externalities is mathematically describing a relation to help explain the effect of many actors being in close proximity. Doing this for discrete regions has clear advantages. The effect of proximity is definable in a tidy way. One of the clearest benefits of this is empirical comparability. Being able to define a direct mathematical relationship between regional output or diversity and overall firm levels gives a simple route to regression analysis - a fact used by Glaeser *et al.* (1992) to identify Jacobs externalities as the most powerful (a result since challenged by van der Panne 2004).

As the next section explains, this thesis uses a 'density cost' that, as the name suggests, is not based on a discrete absence or presence of other economic actors. The rest of this section looks at the ideas used to argue for a density cost approach.

described in the literature as 'localisation externalities' or 'localisation economies'; e.g. Malmberg and Maskell 2002, Wheeler 2007, Figueiredo *et al*. 2009). Glaeser talks of 'agglomeration economies' that "exist whenever people become more productive through proximity to others" (Glaeser 2008 p.5). Again, by the definition used here, these are agglomeration effects. There are many different terms for the opposite of agglomeration: classically, Weber used 'deglomeration' (Weber 1909); Krugman uses centrifugal forces, opposed to centripetal forces holding places together; Brakman *et al.* (2009) use 'spreading'. This thesis will use dispersion, following Fujita and Thisse (2002a).

2.7.3 Land cost and congestion

The thesis model uses density cost as a proxy for both land costs and congestion. It is not intuitively obvious that these two can be lumped together; this section explains the argument for doing so.

Space has two fundamental kinds of cost. The first comes from distance; this has been the subject of most of this chapter so far. Distance itself is not the cost, but rather it forces actors to spend time and resources to traverse it. It is this space that Weber worked with (Weber 1909).

The second is productive space: land that may be bought, sold, rented and invested in, and can enter into a production or utility function as a quantity[15]. From an economic point of view, it is possible to think about land costs in the same way for farm production, factory or service output, as well as housing. Land and distance costs are of course related: this is the whole point of the von Thunen-style models described in section 2.4, and can be seen in action when spatial equilibrium is discussed in section 5.4. Where no exogenous difference in land quality is imposed, its cost must be considered as part of the other costs and benefits of being on that spot. From this point of view, a farm has the same considerations as a factory: how much land is required to produce a given output, and how does its location affect proximity to other important economic actors? This is the approach taken in the thesis.

The concept of 'congestion' is broader than simple traffic-related problems. It can be used to describe a generic loss of amenity caused by density (see e.g. Glaeser 2008 pp.133). Traffic congestion is an element of this, and not an insignificant one. Glaeser and Kohlhase describe the problem (for the U.S.):

> "Over the past three decades, congestion and delay have been increasing in all size classes of cities, not just the very large urban areas. For all 75 urbanised areas taken together, on average the annual delay increased by over 280%, with the most dramatic increase occurring for large cities - those of between one and three million - about 450 per cent." (Glaeser and Kohlhase 2004 p.209)

The overall result from both land and congestion costs, however, is the same: higher density is expensive. This idea underpins the use of a density cost as a proxy for a range of proximity effects. As mentioned above, section 5.4 looks specifically at the outcome of this, showing that spatial equilibrium dynamics result and suggesting that spatial costs (good movement and commuting) have an opposite morphological effect to non-spatial costs.

[15] Another common way to distinguish land components comes from eco-economics, where 'Ricardian land' has both a location and 'physical substrate', as distinct from its resources and productivity. This is not used here, not least because Ricardo's work on land rent was based precisely on that aspect this definition leaves out - the productive differences between types of land - rather than its location. See Daly and Farley 2003 pp.88.

2.7.4 First versus second nature in the density cost

Section 2.3.4 introduced the idea of 'first nature' and 'second nature' effects, where 'first nature' are those elements already present in the spatial economy. It was mentioned that GE models aim to endogenise as much as possible, avoiding 'first nature'.

Analytically, first and second nature effects can be distinguished thus: first nature geography is determined by absolute location (such as a resource site or port), second nature by relative location - that is, the effect comes about through interaction between locations. But as O'Sullivan and Unwin point out, "it is usually impossible to distinguish these effects in practice simply by observing the intensity of a process as it varies across space." (O'Sullivan and Unwin 2002, p.79)

A stronger argument than this can be made: second nature effects necessarily lead to first nature. In GE, this is reflected in the Myrdal-style cumulative causation the models manifest (see below). Tiny differences in initial conditions determine morphology. For agent-based models, the issue is the same: as soon as two actors have formed a proximity preference, they produce a first-nature effect: their absolute location determines subsequent timesteps, and will likely determine the best choices for other agents.

It is impossible to distinguish first from second nature effects, almost as soon as a model has started. What that means in practice is that supplying some exogenous first nature feature at time t is not all that different to allowing a second nature effect to take place between two agents - leading, as it does, to a first nature feature at time $t+1$. In GE, this is not so relevant: the importance of endogenous outcomes is in the initial conditions, not subsequent timesteps. For agent models, which are almost always multi-timestep, the issue is a little less clear.

In the 'mobile People' models, this point is relevant in two ways. Firstly, section 5.3.2 shows that, with no other spatial costs, People and Firms will find their way to a 'black hole' outcome, locating on a single spot, reducing space costs to zero. This is used to justify fixing Firms' location. While this is a 'first nature' assumption, it will be shown that Firms will locate centrally for these models[16] and the assumption makes the subsequent analysis more straightforward.

Secondly, land costs can be considered second nature effects in the abstract. If land itself has no exogenous differences in quality, its value results from the interaction of actors. This is perhaps obvious, but it contrasts with the intuitive idea of land being 'fixed' at an absolute location, and that location determining land cost. This is taken to its logical conclusion by centering a 'second nature' density cost on each Person. This provides a way to explore how agents trade proximity costs off against other costs.

[16]This is not the case for all types of model using the thesis model framework; it is possible for Firms to seek different locations when elasticity of substitution is high in order to gain their own market share. These outcomes are not the subject of the thesis, however; for the models presented, the assumption that Firms have already found an optimal spot is justifiable. As will be discussed a number of times, for any particular assumption, 'everything depends on the problem'.

2.8 Important spatial ideas placed outside framework

The next chapter, as well as discussing agent modelling in depth, tackles one of the central issues of the thesis: the role of simplifications and assumptions in model-building. 'Strategic simplification' is a vital part of the model-building process and best done as consciously as possible. The comparison with the core model in particular also addresses this problem. This chapter ends with three other areas that, while essential to understanding the reality of spatial economics, have been specifically left out of the thesis model framework. These choices, and other lessons from comparing model simplifications, are considered again both in the following chapter and in the conclusion.

Distance cost (for moving both people and goods) can be theorised as having three components, separate from whatever is being moved across space. First, the raw cost involved in moving people and goods. As section 2.6.2 outlined, time is a key raw cost, but others include fuel costs. Second, technology: this changes the level of raw cost required to 'buy' a given unit of distance (for instance, in the choice of transport mode). Thirdly, infrastructure development, such as road and rail networks or ports, that affect both technology and raw cost.

Each of these is associated with a particular timescale in which change affects actor decisions, and changes within one feeds into the others. Raw costs are the most short-term consideration. For instance, if fuel prices change, actors will immediately alter their behaviour - though fuel is a particularly inelastic commodity, so the changes will be slight (see e.g. Goodwin *et al*. 2004). Technology change is more medium term. Using the same example, technological responses to fuel cost change have been clearly identified: actors alter technology choices in the medium-run by changing car, fleet owners by more rapidly scrapping older vehicles while companies respond to those fuel costs by building more efficient vehicles (Brons *et al*. 2008; Li *et al*. 2008). Short-run and long-run elasticities for fuel costs, then, are reflected immediately in how much movement actors will 'buy', and then in their choice of mode and vehicle. Technological change can also open up entirely new trade doorways: for example, refrigeration for meat opened up new export markets for Argentina and New Zealand (Chapman 1979 p.182). In the longer-run, larger-scale technology changes blend into infrastructure development as the two co-develop.

These kinds of medium to long-term structural and technological change are put beyond the scope of this thesis. As section 7.5 discusses, these connect to the issue of how time and spatial economics are closely intertwined - an important area for further work. The following three sections outline the background to these 'strategic simplifications', starting with the longest-term source of space cost change, infrastructure.

2.8.1 Avoiding infrastructure change and networks

Any fully comprehensive theory of space cost change requires a theory of endogenous infrastructure and route development, that would have to include different tranport modes and their inter-relation. These long run outcomes dominate over cost change or technological change. Andersson and Stromquist make a particularly strong claim: “all the major transitions in the European economic systems were accompanied, or initiated by major changes in transport and communications infrastructure” (Batten and Thord 1988 quoted in Button 2010 p.421).

In the seventies, Haggett *et al.*, after a detailed look at the issue, concluded that “there remains to be specified a comprehensive model of route development” (Haggett *et al*. 1977 p.95). No such model exists today, though perhaps for good reason. Infrastructure development changes proximity between points unevenly. Transport development tends to happen between economically important sites, reducing costs between them while relatively increasing costs for others. As Fowler notes, “uneven power relations” are both a cause and consequence (Fowler 2006 p.1433). Routes themselves are also part of feedbacks: it is, as Chapman says, “equally plausible to regard transport investment as a result of a need for movement or as a generator of movement” (Chapman 1979 p.230).

The obvious modelling approach to this complexity is to avoid it altogether. This is the first (and one of the most important) simplification made in this thesis: using featureless, continuous space to avoid the complications of transport and infrastructure change. This is only in any way plausible, though, for considering short to medium-term change: time-scales where the impact of route development will be only minimally felt. It is also quite a large assumption that fixed networks can be mapped onto a Euclidean plane unproblematically: see, for example, Wilhite’s work on fixed networks (Wilhite 2006) and section 3.3.1. Nevetheless, this approach has one clear advantage: it is very simple. Calculations are trivial and, as an added bonus, visualisation is straightforward and intuitively easy to grasp. This method is used in this thesis; the rest of this section, however, explores the issue, to give this approach some context. It is also returned to in the conclusion, since it represents a perfect example for thinking about model simplification versus complex reality.

Infrastructure change makes no sense in anything other than a two-dimensional space (or higher). Between or within discrete regions, it must be conceptualised as either an externality (see above) or a generic transport cost change between points. Modelling space as a one-dimensional line, the same applies: there is only one route between points. Choosing discrete space or a line, however, does not make the problem go away: it just masks the fact that this uneven change dynamic exists. If, as argued here, it is perhaps the most important long-run dynamic, this means avoiding it leads to wrong results. However, as long as routes in a two-dimensional space are presumed to be static (or near-static), it is possible to assume a mapping of any route system to a Euclidean approximation, making the analysis potentially much easier. Specifically, it becomes possible to use Euclidean distance as a reasonably accurate proxy for actual distance and space cost between points (see, for instance, Cooper 1983.)

A simple way to describe the discrepancy between route distance and straight line distance is with a ‘route factor’ (see e.g. Chapman 1979 p.215; Black 2003 p.68). This is simply the ratio of the distance between points along a network route over the Euclidean distance between those points. This value approaches one as the two measurements converge, implying network density is increasing - and Euclidean approximations can do a better job. There is a well-established correlation between network density and development levels (Chapman 1979 p.220), and this is reflected in the route factor: it approaches one in more developed countries - though even that assumption breaks down at small enough scales.

Early research in this area captured the most extreme manifestation of how competing interests determine route development, describing 'colonial' transport networks in less developed countries. Argentina and Uruguay had railways for shipping meat and wool to Britain; Chile, Bolivia and Peru for copper, lead and tin to the coast (Gilbert 1974 in Chapman 1979 p.231). Melchior found that "Latin America appears as a complex of national space more closely tied to exogenous decision-making centres than to itself" (Melchior 1972 p.88 in ibid p.16). World Bank projects still emphasise export infrastructure.

Haggett *et al.*'s work used network analysis, and this is still a common approach. Network analysis itself is, of course, now a very large part of the economics literature (for an overview, see e.g. Goyal 2009) and has a central role in the development of complexity approaches to economics (see section 3.3.1). There is also network-related discussion in the GE literature, but again, its focus is on *economic*, not spatial or transport, networks; for example, see Johansson and Quigley (2003), which discusses the relation of networks to localisation externalities.

More abstract network analysis relevant to transport optimisation has found an enthusiastic research community in physicists and statisticians. A particularly good recent example of this is Gastner and Newman's network growth model (Gastner and Newman 2006), in which they explicitly look for the balance between shortest routes and total route distance (that is, the sum of all route distance in the model: the more total route, the more expensive the build). They conclude that while "these two criteria are often at odds with each other... real networks nonetheless manage to find solutions to the distribution problem that come remarkably close to being optimal in both senses" and that 'growing' them in their model does as well, or better, as attempts to explicitly optimise them.

This type of analysis, however, tends to be economics-free: cost is reduced to route length alone, and route development is not driven by particular interests. For instance, Gastner and Newman's distinction between shortest route and least total route optimisation matches the difference between optimal networks for users versus builders (Haggett *et al.* 1977 p.218). Users want dense networks to lower costs of travelling between points; builders want as few links as possible to reduce their costs.

Once routes are developed, and considered as static features of the economic landscape, two types of dynamic cost are still present: arbitrage and congestion costs. (Congestion costs are discussed in section 2.7.3.) The 'principle of charging what the commodity will bear' applies to transport routes as any other commodity: actors will look to maximise profits (Behrens *et al.* 2007 p.626). This means that transport costs will, in part, reflect the fact that -

> "high-value goods, for which the cost of transport will add proportionately less to

> total costs, usually pay more to move a given distance whereas goods of low unit value may be charged rates which do not even cover the costs involved" (Chapman 1979 p.117).

Given the spatial nature of transport routes, this can often mean monopoly, monopolistic competition and cartels, as with (until recently) shipping 'Conferences' (Carrre *et al*. 2009 p.17), American trucking prior to the Motor Carrier Act (Glaeser and Kohlhase 2004 p.204), or indeed any of the monopolised large shipping canal routes.

On static routes, the shortest route problem is quite a research topic in its own right. Discussed at length in Haggett *et al.*, more recent interesting examples exist: one looks at whether supermarket delivery route optimisation can reduce the total quantity of traffic, as compared to customers driving to and from stores (Cairns 2005).

So, the fact that route and infrastructure development are the result of a nexus of competing interests makes any generic modelling very difficult. Euclidean space can act as a proxy in a limited way, but the above suggests that it will be less applicable in less developed countries, at smaller scales, and as longer timescales are considered.

2.8.2 Trade costs

Some of the most important space costs are the hardest to measure. The issue of 'trade costs' is a case in point. Duranton and Storper define these as "the sum of all costs incurred to deliver a good to its user" (Duranton and Storper 2005 p.1)[17]. These are much broader than just transport: according to the trade cost literature, "estimated distance effects are about an order of magnitude too large to be explained by shipping costs" (Disdier and Head 2008 p.2). A list of potential mechanisms responsible is given by Anderson and van Wincoop: "transportation costs (both freight costs and time costs), policy barriers (tariff and nontariff barriers), information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs (wholesale and retail)" (Anderson and Wincoop 2004 p.691).

The existence of trade costs are deduced by examining how trade drops off with distance between countries. Gravity models are the key tool used; an argument is then made regarding what this reveals about actors' choices. "Theory looms large" (Anderson and Wincoop 2004 p.692) - and must be made to analyse a range of costs, along a spectrum from data-rich to completely unmeasurable. The decay of trade over distance is thus - in the context of an economically grounded gravity model - taken as 'revealed preference' (see section 3.5.1).

Using this approach has revealed an apparent puzzle: trade costs have been rising since the mid-20th century, despite transport costs dropping. Or rather, distance of trade has been dropping. Carrere *et al.* conclude that if separate groups of countries are considered, distance

[17] Duranton and Storper wrote this working paper in 2005, and an article version was later published in the Canadian Journal of Economics (Duranton and Storper 2008). The original working paper has a rather more detailed introduction looking at the issue of trade costs in some depth.

of trade has only been dropping for poorer countries. There are two possible explanations, they argue: the lowering of trade costs between those low-income countries, or their marginalisation as trade costs with more distant countries increase (thus making nearer trade relatively less costly) (Carrre *et al*. 2009 p.30).

There are two closely related points relevant to this thesis. The first is the problem of mechanism: if the goal here is to model, at the actor level, decisions made given space costs, to what extent does the range of trade costs make this problematic? Any attempt to answer this question will immediately hit the second issue: what counts as an explanation of the costs identified in the real world? Section 3.4.8 discusses this in more depth, as it relates to the problem of micro versus macro analysis.

2.8.3 Risk is expensive

Risk is a major cost for both firms and people, though a complex one to theorise. It is important enough for real-world outcomes that it is worth discussing briefly. Note, however, that while uncertainty does play a part, explicit risk modelling is not part of this thesis (though it is a highly interesting topic for agent modelling).

The importance of minimising risk in space costs is not new. Wallace, for example, found in the U.K. that "where the choice between two modes involves both a considerable price and a considerable quality-of-service differential there is a clear willingness to pay the necessary price of reliability" (Wallace 1974 p.41). Certain goods have always been more time-sensitive than others. Chapman also notes that "consignments which fail to turn up on time impose indirect costs by interrupting production schedules. Indeed, in certain industries, the time factor is critical" (Chapman 1979 p.117).

The importance of stability is clear from the way markets react to changes in the cost of oil. Price volatility is very damaging: "not a simple dependence of the economy on the level of oil prices, as would be suggested by production-function based accounts" (Hamilton 2003 p.33). This interpretation is supported by the fact that investment decisions are delayed both when oil prices increase and decrease: firm do not want to replace expensive capital at times of uncertainty.

A more modern approach to reducing risk aversion is the trans-national corporation. As Leamer and Storper put it:

> "much global trade consists merely of shipping products or components between divisions of the same firm located in different countries - transactions that do not raise the trust and enforceability issues present in arm's length transactions." (Leamer and Storper 2001 p.649)

Because of this, modern optimisation techniques for global production applied in-firm have been very successful in managing spatial risk. Carrere *et al.* (discussing Abernathy *et al*. 1999) note that for apparel "the key to success is no longer solely price competition but the ability to introduce sophisticated information links, forecasting capabilities and management systems" (Carrre *et al*. 2009 p.34). This makes increasingly 'lean retailing' possible, and has led to other geographical effects. Particularly, these lean firms have been bringing production networks closer: U.S. firms networked with South America and European firms with Eastern Europe - which may in part explain the trade cost puzzle mentioned above.

As with route and infrastructure development, this is a hugely important empirical area, but a particularly thorny modelling problem - and again, one this thesis avoids.

2.9 Chapter summary

This chapter has explained the fundamental economic and spatial economic concepts to be used in the thesis. It has placed these in the context of GE, covering the way that market structure has been theorised. The story of how market structure assumptions have changed shows that 'dependencies' in model creation are not just limited to OOP; the idea can be applied more broadly. Classical location theory ideas applicable to the thesis model framework were presented, as well as some specific classical ideas to exclude.

Time was spent laying out the exact components of the core model as a way to get to the heart of the web of assumptions used by GE and give some substance to the discussion of market structure in section 2.3. An overview was then given of the way space costs can be theorised, looking at the costs of moving both people and goods, and the costs (and benefits) of proximity when agents are nearer together. This sets up the explanation of how these costs enter into the model framework, presented in section 4.3 and 4.4.

The chapter ended with an examination of three key issues that are consciously placed outside the analysis of the thesis, though they are discussed again in the conclusion. Throughout, a key goal has been to examine the web of assumptions in economics and spatial economics. The next chapter goes into more depth regarding the theory of assumptions in modelling and puts this in the context of ABM. The conclusion builds on this to make an argument that ABM has not yet developed an effective practical understanding of how assumptions should be understood when modelling.

Chapter 3

Agent-based modelling and economic theory

> "When you explain a 'why', you have to be in some framework that you allow something to be true. Otherwise you are perpetually asking why."
>
> *Richard Feynman*[1]

> "How would you physicists like it if you had to survey a bunch of molecules to find out what they planned to do, only to have most of them change their minds anyway, and the government restructure the laws of physics because of some opinion poll?"
>
> *'Gaz'*[2]

3.1 Introduction to chapter

This chapter introduces the concepts behind ABM. It also puts them in the context of ideas about modelling, concentrating particularly on those that have been developed in economics. The reason for this approach is to understand Richard Feynman's quote above: explanation requires knowing what framework the modeller is currently allowing to be true. This is an

[1] In interview on 'Fun to Imagine,' BBC, 1983.

[2] Comment at http://tamino.wordpress.com/2010/03/11/not-a-random-walk/comment-40731

argument that goes back at least to Hume, who imagined a peasant and an artisan wondering why a clock had stopped. The peasant manages no better than ‘it does not go right’; the artisan knows more of the inner workings, and suspects the springs or the pendulum (Hume 1739 p.132 in Hoover 2001 p.11). But while it is intuitively obvious, trying to work through its implications for model choices is less so. It is a tricky problem and as Friedman notes:

> “There is no magic formula for wringing knowledge about complicated problems from stubborn facts. No method is proof against incompetent application.” (Friedman 1953b p.613)

When discussing economic modelling, examples from the natural sciences present themselves as obvious thought experiments. This chapter uses plenty of them, and looks at arguments made by economists that do the same. A lot is written on the validity of such parallels; critiques of the utility-maximising foundations of neoclassical economics often go no further than pointing out its roots in Newtonian physics as self-evident proof of its absurdity (see section 3.3). But the quote above gets to the heart of the issue. Regularities in human behaviour do exist: how people react to cost changes is central to this thesis. But parallels with physics can only go so far. To pre-empt the conclusions, the argument for modelling people as predictable, reactive objects (perhaps with added noise) must be lightly made. It is unlikely that there will ever be the same robust link between levels of explanation in human behaviour as there are between atomic kinetics and gas theory (see section 3.4.8). Vriend quotes Lucas: economists are “programming robot imitations of people, and there are real limits on what you can get out of that” (Klamer 1984 p.49 in Vriend 1994 p.31).

The structure of the chapter is as follows:

- An explanation of the basics of ABM, introducing OOP, discussing how it has come to inform model building, as well as the roots of the idea of ‘dependency’;
- A brief overview of the ABM literature, setting the scene for the more specific argument that follows
- ‘Mapping the model’, a section looking at how theorists from both economics and ABM points of view have understood how models should be built. The goal is to think through how to distinguish a good from a bad ‘mapping’.
- ‘Production and utility in an agent context’ grounds the previous discussion by asking how agents with utility and production functions can be built.

3.2 What is agent-based modelling?

In ABM, the ‘agents’ are distinct code objects, programmed to interact with their environment and each other. ABM developed in tandem with OOP. Though OOP has a history going back to the sixties, it was not until the nineties that computing power and programming languages like Java developed enough to mark a ‘paradigm shift’ in programming (Robinson and Sharp 2009 p.211). This shift provided the soil for ABM to flourish.

Writing in 2000, O’Sullivan and Haklay noted that ABM was “a rapidly developing field that is already well beyond the scope of any limited survey” (O’Sullivan and Haklay 2000 p.1410), a sentiment shared by one of the few authors brave enough to attempt an ABM textbook (Wooldridge 2009 p.xix). ABM’s use ranges from the most abstract artificial life to ‘autonomous’ agents earning their keep controlling real-world infrastructure. As a field, ABM still seems as fluid and as expansive as it did ten years ago: little in the way of consolidation has taken place, despite pleas for standardisation and various groups arguing their particular framework offers a one-stop-shop for all things agent-based.

Wooldridge defines objects as “computational entities that encapsulate some state, are able to perform actions, or *methods*, on this state, and communicate by message passing” (Wooldridge 2009 p.28). Objects are created from classes; a real-world metaphor would be that classes are the blueprint and objects the physical form. Thus, a model may have a single ‘Firm’ class but many ‘Firms’ created from that blueprint, with their own internal state.

This focus on objects’ ‘encapsulation’ - that ‘objects have control over their own internal state’ - requires the programmer to think of them as separate entities, and to define relations between those entities very clearly. (The procedural element of programming has not gone away, however, and is still of vital importance, particularly with regard to timings; see section 4.7.3.) A popular online java tutorial (1995) has a simple illustration of encapsulation: if a rider of a bicycle attempts to change gear, the bicycle should have a ‘method’ that ensures it cannot exceed its gear number. The rider is denied the ability to *force* a gear-change above those available. In coding practice, this mean providing methods to raise and lower gears: the rider may ‘request’ a gear change, but the bicycle ‘decides’ if it can be done.

The ideal of encapsulation leads to the idea of ‘loose coupling’ (e.g. Jana 2005): if objects can be built that rely is little as possible on needing to know how other objects function, code frameworks can be made highly flexible and reuseable. An example from the thesis model framework would be the data output code, where an earlier iteration hard-coded variable names to be stored. By providing the data store methods with a way to ‘read’ variables from anywhere in the framework, target variables can be identified with a string name only rather than hard-coded - thus replacing a set of objects, one for each variable, with a single re-useable object.

However, as the introduction mentioned, the ‘dependencies’ that loose coupling attempts to avoid presents some problems when used as a philosophy for mapping agent models to real-world features; this theme is taken up in section 3.4.

Many of the coding innovations now associated with OOP that found their way into ABM have a prehistory in procedural programming. Some of these have become emblematic of the agent modelling ethos. For example, Thomas Schelling’s segregation model, first presented in 1969 (Schelling 1969) and later described in ‘Micromotives and Macrobehaviour’ (Schelling,

1978) is often seen as the ‘Eve’ of agent models (see, for example, Schelling’s closing essay (2006) in volume two of the ACE handbook), and is used by Krugman as a compact illustration of how a good economic explanation links micro to macro results (Krugman 1996, p.15). Craig Reynold’s ‘Boids’, while originating from a desire to animate flocking, talked of agents as objects (Reynolds 1987). Schelling initially used only pencil and paper (Schelling 2006); Reynolds used an OOP extension to LISP.

These models pre-empted the later vital importance of encapsulation. In conjunction with a set of supporting OOP coding concepts, this has created a very powerful set of tools for defining how objects interact. This is not just of historical interest: it is hard to overstate the importance of the OOP paradigm’s approach to the subsequent shape of agent modelling theory and practice. Section 3.4 examines this issue in depth. Before that, the next section gives an overview of ABM.

3.3 ABM overview

3.3.1 Agent-based computational economics (ACE)

This thesis uses agents to look for spatial economic outcomes from actors with heterogenous locations. There is very little in the way of ‘complex’ behaviour, and where it does exist (for instance in the interaction of price-setting and production), it is not the focus of analysis. This is not to deny the uses of CAS theory. Evolutionary ideas are particularly important for developing an understanding of how economic growth actually functions (Martin and Sunley 2007, and see section 3.5.2), as well as the evolution of diverse market structures, rather than just postulating a ‘Market’ (see e.g. Mirowski 2007a discussing ‘marketomata’). But a perceived need to keep ABM and complexity umbilically linked cuts off some of the more mundane - yet crucial - uses of heterogeneity. (Section 3.4.8 looks at an example of an ‘argument from complexity’ that illustrates this point.)

Otter *et al.*’s agent paper presenting an economic geography model, for instance, argues it is possible to use “complexity as underlying explanatory variable” (Otter *et al*. 2001 p.1). If ‘complexity’ can be used this way, so can ‘equations’ or ‘statistics’. It is suggested that “regional economics and geography” are apparently “not sufficient to explain the complex spatial patterns, such as clusters and sprawl, that we encounter” (Otter *et al*. 2001 p.1). Their model is in some sense believed to be ‘validated’ if it manifests some element considered complex. In another example, Boschma and Martin ask,“in what sense can complexity theory notions (or metaphors), such as the emergence, self-organisation, criticality and so on, be used to conceptualize the economic landscape?” (Boschma and Martin 2007 p.541) This is perhaps a natural consequence of some of the theories of explanation that have grown along with CAS theory; see section 3.4.7 on Epstein.

As O’Sullivan and Haklay say, ACE is driven by “a view of the ‘economy as an evolving complex system’ promoted by the Santa Fe Institute”, accompanied by a “widespread disillusion with neoclassical equilibrium economics" (O'Sullivan and Haklay 2000 p.1412). Disillusion is a gentle term: the reality for some researchers is a rather more 'year zero' feel, seeing CAS as "a pioneering break from a moribund Newtonian worldview" (Manson 2001 p.412). Mirowski, for example, argues (paraphrased by Blaug) that "the whole of neo-classical economics ever since the has been an attempt to create an economics that emulates all the essential features of of nineteenth-century physics." (Blaug 1997 p.284)

As Blaug notes, this is rather strong; there was no conscious attempt to emulate physics. Grauwin also points out that the economic interpretation is very different from the physical: "scientific fields assume distinct points of view for defining the 'normal' or 'equilibrium' aggregated state": physics uses entropy where economics has agents finding Nash equilibria and, as they note, "the two approaches lead to radically different outcomes" (Grauwin *et al*. 2009 p.20622). Nevertheless, from a CAS perspective, the faults with 'traditional' economics are seen to be self-evident. What Conlisk calls the "strange sacrifices required for the 'ritual purity' of optimisation-only models" (Conlisk 1996 p.686) are considered relics. While understandable at a time when no computers were available, the feeling is - as Mirowski says - economists should "kick the habit of their physics envy and join the 21st century by rethinking the importance of computation and evolution in the way that they approach markets" (Mirowski 2007b p.359). Edmonds goes even further, suggesting that numerical representation itself is questionable. For example, he criticises numerical representation of variety on the basis that not all the dynamics associated with variety (such as evolutionary dynamics) can be described numerically (Edmonds 2004 p.5).

As section 2.3 discussed, GE is descended from this 'Newtonian' lineage. Unsurprisingly, then, there is very little overlap between the research agendas of GE and ACE. ACE as a distinct sub-discipline of ABM takes it name from the broader school of computational economics. Despite arguments that ACE is perfect for re-examining the Marshallian roots of economics (Leijonhufvud 2006), it has tended to follow the 'Santa Fe' research agenda. One editorial contains a good checklist of the concepts of this agenda: "complexity, evolution, auto-organisation... emergence... bounded rationality, inductive reasoning" (Consiglio 2007 p.vi).

The economic coordination problem in particular lends itself to ABM's default focus on interacting agents. Indeed, Tesfatsion defines ACE as "the computational study of economies modelled as dynamic systems of interacting agents". The natural question for ACE theorists thinking about markets is then: "who does the job of the so-called market adjustment?" (Posada *et al*. 2007 p.102). This requires the modeller to "analyze explicitly how agents interact with each other" (Kirman and Vriend 2001 p.460). Howitt, for one, sees this requirement as a key strength of the approach: "one of the virtues of the ACE approach to economics... is that it forces one to make explicit the mechanisms through which individual actions are coordinated, for better or worse" (Howitt 2006 pp.1068). Agents, it is hoped, force theorists to crack open the problem and look at the dynamics inside.

Hayek's short essay, 'the use of knowledge in society' (Hayek 1945 p.529), is something of a talisman for this way of thinking. In particular, his strong aversion to macro-level assumptions fits ACE perfectly. He insisted that "we must show how a solution is produced by the interactions of people each of whom possesses only partial knowledge" (Hayek 1945 p.530) articulating both an assumption that the 'solution' arises from micro-level interactions, and that it must come from 'boundedly rational' actors. Epstein's 'generativist's question' (discussed in-depth in section 3.4.7) - "how could the decentralised local interactions of heterogeneous autonomous agents generate the given regularity?" (Epstein 2006 p.5) - echoes Hayek's take. Some theorists (Vriend 2002 p.2; Miller and Page 2007) can imagine that, had Hayek only been able to access present technology, he would surely have embraced it, and complexity theory along with it.

This Austrian tilt to ACE is further reason for the lack of overlap with the kinds of questions GE asks. A good example of how this manifests itself is the criticism levelled at the concept of equilibrium. As Blaug notes, "the desideratum of any economic theory is the delineation of an equilibrium end-state" (Blaug 2009 p.224). These end-states are analysed in the same way that basic physical models of stationary systems assume all forces balance to zero and are then able to deduce force values required for that stationary state. In-keeping with the attacks on its Newtonian underpinning, Colander and Rothschild argue:

> "the self-correcting 'stability' vision cultivated by economic pedagogy is problematic in several respects. First and foremost, it is simply wrong: stability is not the norm in complex systems" (Colander and Rothschild 2010 p.286).

Critiques of the equilibrium assumption predate ABM, of course. Holub made the case that any new framework would need to declare a 'year zero':

> "The long tradition of equilibrium thinking in economics has lead to an unrivalled, consistent structure of thought... an anti-equilibrium attempt cannot build further on this structure, not even on the ruins of equilibrium theory (should it succeed in toppling this construction), it must on the contrary seek a new site for its own, a new thought structure, in other words, a new central idea" (Holub 1977 p.395).

Many seem to believe ABM is exactly this new central idea. The concept of the 'Walrasian auctioneer', often a target of anti-equilibrium critics, is a good example of this in practice. It is described as an actual mechanism, though in Walras' own work it is used only as a thought experiment to explain the equilibrium outcome (Blaug 1997 p.555-6). Many ABM theorists take issue with a centralised adjudicator capable of mediating a process of 'groping' towards a set of market-clearing prices, and knowing the moment when they are all correct[3]. The auctioneer

[3] See e.g. Tesfatsion 2006 p.834, where it is argued that "equilibrium values... are determined by market clearing conditions imposed through the Walrasian Auctioneer pricing mechanism; they are not determined by actions of consumers, firms, or any other agency supposed to actually reside within the economy"; see also Vriend (1991); Posada *et al*. 2007.

concept is a good summary of everything ACE opposes as implausible. Once such convenient (but, it is argued, impossible) assumptions are shown to be unjustifiable, a Pandora's box is opened:

> "The modeller must now come to grips with challenging issues such as asymmetric information, strategic interaction, expectation formation on the basis of limited information, mutual learning, social norms, transaction costs, externalities, market power, predation, collusion, and the possibility of coordination failure." (Tesfatsion 2006 p.836)

Each of these 'challenging issues' does in fact garner plenty of attention from analytic economics: Stiglitz got his Nobel prize for his work on asymmetric information going back to the seventies (Rothschild and Stiglitz 1976), game theory deals with strategy (and 'predation and collusion' are forms of strategy), and Krugman himself discusses expectation and the role of history as a key element of his geographical work (Krugman 1991a). As regards agglomeration economies, the importance of externalities has been discussed in chapter 2.

ACE models often share one vital feature with the equilibrium methods they reject, however: a market clearing point. This often means that auction-style models are used where markets reach a defined end-point via a bidding process through many interactions, perhaps, but still terminating at the end of trade. An example would be Kirman and Vriend's fish market model, with an interdependent morning and afternoon's trading (Kirman and Vriend 2001 pp.467). The question of what happens if markets do not 'clear' was asked by Hicks, and Isard understood its implications for trade across space. If all actors are making market decisions on the head of a pin, an auction-style clearing market presents itself as the natural way to think about it. But as soon as actors have heterogenous location, that becomes impossible: 'clearing market' conditions do not hold. From an ACE point of view, Dibble (2006 p.1516) notes the effects of space on standard market-clearing assumptions. Ladley and Bullock also point out (2008 p.296) that market actors may be segregated by space, able only to interact with a given subset of others, and that spatial segregation forms natural networks as actors' ranges of interaction overlap. Hamill and Gilbert apply this idea 'in reverse': they use actors with randomised location in an R^2 space, and a fixed radii around each actor determines their social network connections (Hamill and Gilbert 2009, 2010).

Leijonhufvud describes the nature of the kind of market that space imposes (though he is not discussing spatial economics). He quotes Hicks -

> "In a normal, ongoing market, transactors are not all brought together in a single location and at the same time. Without centralisation and synchronisation, the supply-equals-demand condition 'cannot be used to determine price, in Walras' or Marshall's manner'." (Leijonhufvud 2006 p.1633; see also Ladley and Bullock 2007 p.83-4)

While easier to ignore this problem if geography is not an element of the model, one cannot consider spatially varied trade as 'one market' (Isard 1956 p.43). Leijonhufvud makes another suggestion based on this, which he dubs 'Marshall's laws of motion': that agents follow very simple rules in response to market changes, not be presumed to rely on a point on a supply/demand curve. For example, a consumer may say, 'if demand price exceeds the market price, buy more' in the opposite case, buy less', or a producer may say, 'if market price is above supply price, increase production; in the opposite case, cut back' (ibid p.1630). This basic idea is used for a key component of the transmission belt model in chapter 6, though with an important alteration. Consumers use only a constrained utility function; they do not use a 'law of motion' at all. Producers do use a 'law of motion' - but not to change production levels. Instead, they say, 'if stock is more than my target, lower prices to sell more; in the opposite case, increase them to sell less'. This simple idea is the basis for producing a flexible spatial 'general equilibrium.

It is the 'non-clearing' part of this problem that the thesis model attempts to deal with, rather than the network element. Section 2.8.1 already made the argument that transport networks can be replaced with a Euclidean proxy when discussing distance, though only by acknowledging the severe limitations for discussing change this imposes. Ladley and Bullock note Wilhite making a related point: that if network change is slow compared to activity on the network, its structure becomes more important, not less (Ladley and Bullock 2008 p.299 referring to Wilhite 2006). This thesis cannot answer this issue fully as it works within the Euclidean assumptions from section 2.8.1, where it was argued that the success of Euclidean stand-ins is dependent on both the spatial and temporal scale under examination, as well as whether network dynamics are part of the research question. So the model cannot ask, as Epstein does, how "the endogenous connectivity the topology of a social network affects its performance as a distributed computational device, one that... computes price equilibria, or converges to (computes) social norms, or converges to spatial settlement patterns such as cities?" (Epstein 2006 p.17-18).

3.3.2 Spatial versus economic agents

A Venn diagram of ACE and spatial ABM would perhaps reveal just as little overlap as between ACE and GE. Torrens' recent review of 'ABM in the Spatial Sciences' (2010) illustrates the point. One economic example picked up on is not actually about spatial modelling: Torrens cites Farmer and Foley's Nature piece who argue that the economy is complex, ABM can 'do' complexity where equilibrium models cannot, representing something closer to reality; therefore, ABM is better:

> "Agent-based models potentially present a way to model the financial economy as a complex system, as Keynes attempted to do, while taking human adaptation and learning into account, as Lucas advocated. Such models allow for the creation of a

> kind of virtual universe, in which many players can act in complex - and realistic - ways" (Farmer and Foley 2009 p.685-6).

They suggest an effort akin to general climate modelling for the economy is in order, coupling the various social science disciplines in the same way climate models might couple ocean and atmosphere feedback. Their emphasis on building a 'virtual universe' is also present in many of the spatial models Torrens surveys (this idea is the subject of section 3.4.4). The cellular automata models surveyed by Torrens in particular have this quality. Crowd and 'swarm' models loom large, leveraging the interactive properties of agents (a point noted by O'Sullivan and Haklay, 2000 p.1411).

In terms of spatial *economic* models, Torrens cites his own work (among others) - in one example, a technically rich implementation of a combined cellular-automata agent hybrid model (Torrens and Benenson 2005). There appears to be no driving research agenda, however, beyond proving the technical feasibility of doing so, though the ability of disaggregated models to produce 'emergence and self-organisation' is mentioned (p.396).

Agent models using explicitly spatial economic ideas that would seem familiar to a GE theorist are few. Sasaki and Box use ABM to attempt to 'verify' classic spatial analytics like the von Thunen model (Sasaki and Box 2003) by showing how the spatial equilibrium result can come about through individual-level action but, aside from referring to von Thunen himself, it sticks to the ABM literature. The work of Christopher Fowler stands out as a theoretically grounded agent model with the explicit aim of examining the core model; this is discussed in-depth in the next section.

3.3.3 Fowler

Christopher Fowler has attempted to recreate the core model in agent form. This section discusses the two papers where he does this. In some key aspects, his goal is the same as this thesis. Specifically, he keeps the focus on the simple interactions:

> "the effects of economies of scale, the role of preferences for variety in driving trade between regions, and the push/pull relationship between these two factors in shaping patterns of agglomeration and dispersal in economic activity based on transportation costs" (Fowler 2011 p.2).

The title of his first paper 'Taking GE out of Equilibrium' indicates what Fowler initially thought an ABM approach should be good at: "the inability of the deductive models to describe the movement of a system between equilibria represents a major drawback of these models" (Fowler 2007 p.267). An agent approach, he argues, is a good method for achieving this. In his first attempt, he set himself a specific goal: to "re-create as exactly as possible the relationships in the economic [core] model" (ibid p.266) in agent form, meaning that "to the extent possible, the equations used to express the analytic model have been maintained." (p.272) His second paper made a few extra choices and aimed to "explore the capacity of an economic system to identify a stable Nash Equilibrium without it being enforced by assumption." (Fowler 2011 p.14)

Fowler's direction of travel, then, is very much from the analytic core model towards agent modelling. The main conclusion from his first attempt was that -

> "although the equations of the analytic model can be mimicked in a way that is sufficient for a simulation to run, such a simulation cannot be made logically defensible without significantly altering the relationships among workers, firms and cities posited in the analytic model." (p.282)

In initially binding himself to the assumptions of the analytic core model, Fowler was stuck with one of the more thorny abstractions: there are no firms explicitly defined as separate actors. The equilibrium number of firms - a key feature of of the Dixit-Stiglitz monopolistic competition approach, as discussed in section 2.3.2 - is purely a consequence of consumer demand combined with love of variety. As he puts it, "firms appear and disappear in cities based on the full employment of the workforce even to the point of following the lead of workers who move from place to place in order to benefit from increased real wages." (Fowler 2011 p.2)

Fowler's first pass at an agent version of the core model, then, lacking any method for linking firm and consumer behaviour even implicitly, could not produce equilibrium firm levels. Fowler notes that "as a result, the model fails to move towards any of the equilibrium conditions predicted by its analytic counterpart." (p.266) As he notes, the full 'general' conditions present a fairly daunting proposition for the agent modeller:

> "the amount a firm can offer in wages is dependent on the amount it can produce and the price it receives for its goods. These quantities are affected, in turn, by consumer's wages, which depend on which firm employs them (and whether or not they have found employment at all.)" (Fowler 2007 p.277)

In this situation, how can actors, unable to coordinate through any central mechanism, produce stable spatial economic outcomes, especially given the extra new problems of an 'absence of rational expectations' (p.275) stemming from actors' intrinsic inability to predict the actions of others? (This 'true uncertainty' is discussed in section 4.7.5.) He also points out the uncertainty that firms face: they need "some mechanism with which they could predict an appropriate production level for themselves and estimate the levels chosen by their competition in an environment where sufficient labour supply is not guaranteed. The delicate balance among the equations of the analytic model does not allow for the error and uncertainty that are necessary parts of this sort of prediction, and so a new set of relationships needs to be specified" (p.282).

In a statement he later appeared to regret[4], Fowler goes on to say, "as complex as this specification sounds, it is actually relatively simple in an agent-based framework where bounded rationality, learning by doing and other types of decision-making all have substantial supporting bodies of literature" (ibid p.283).

There are two particular points to pick up on. First, to return to the need for explicit 'firm' actors: it could be argued that the original core model *did* have firms as part of an argument built on the monopolistic competition model. The fact that it did not have distinct firm *agents* is what Fowler is criticising: he is arguing for what the next section calls a closer 'descriptive mapping' and, perhaps, that it is self-evident that any model without distinct 'firm' objects is invalid. Certainly, it appears to be what he sees as the most unjustifiable simplification: "geographical economics does itself a disservice by ignoring the labour market dynamics that are arguably the most important set of relationships driving the movement of labour and capital in the real world" (Fowler 2011 p.7).

The second point is the flipside of this: of course, Fowler must make his own simplifying assumptions. For example, he uses what could be described as a 'localisation externality' simplification in the later paper's model. Three actions are taken: workers calculate their individual utility, an average utility level for each region is then worked out, and finally one randomly selected worker compares their individual level with the average between regions before deciding where will give them the better utility. The externality here is the regional average: each worker is allowed to 'know' this value for all cities. This a necessary simplification for his purposes, but nevertheless not a dynamic that emerges from agent interaction.

Again, then, how does one tell the good simplifications from the bad? Which merely (in Hayek's phrase) "assume the problem away and... disregard everything that is important and significant in the real world" (Hayek 1945 p.530; see also Miller and Page 2007 p.87). The next section deals with this question.

3.4 Mapping the model

3.4.1 'Descriptive' versus 'functional' mapping

Miller and Page (2007 p.36) argue that one can construct a 'homomorphism' between model and reality, an exact equivalence between identified real-world and model structures, but that too broad a homomorphism would be "at the cost of lowering the model's resolution and value" (p.40). As with arguments that models should aim for isomorphism[5], these terms have particular mathematical meanings that encourage the idea that object structures should very explicitly

[4]"Economists are largely willing to leave the exploration of the labour market to other models. Given the complexity of the model necessary to replace this assumption, this researcher, at least, has grown increasingly sympathetic to the allure of such a potent simplifying assumption" (Fowler 2011 p.7).

[5]e.g. "Isomorphism is a relation between mathematical structures. If there is a function that maps each element of one structure onto each element of another the structures are isomorphic" (Downes 1992 p.147). The term is used by, among others, Epstein to describe model-to-reality mappings (Epstein 2006 p.24-5).

map onto reality, and be transformable in the same way. They are too strict: the looser idea of 'mapping' that Miller and Page start with can be a more useful way to think about building models.

As Scott says, models and maps share a common purpose, both being "designed to summarise precisely those aspects of a complex world that are of immediate interest to the map maker and to ignore the rest" (Scott 1998 p.87). Baumol and Blinder conclude that modelling means choosing the best map and that will depend on the purpose (Baumol and Blinder 2005 p.12). Miller and Page point out it is intuitively obvious why a more realistic map may not be a better one (Miller and Page 2007 pp.36). Krugman takes this idea further: in the earliest Colonial explorations, map data consisted of verbal reports sometimes apocryphal, spatially inaccurate, but still very useful ('six days south of the end of the desert you encounter a vast river flowing from east to west'.) However, as more formal mapping took place, "the improvement in the art of mapmaking raised the standard for what was considered valid data". For a time, much useful information was lost (Krugman 1993). In the end, more accurate maps resulted but in the transition to formalism and rigor, descriptive information gained may temporarily be lost. This is a map version of the streetlight effect: "the methodology of economics creates blind spots. We just don't see what we can't formalise" (**?**).

One of the supposed strengths of ABM is precisely that it promises to combine both descriptive accuracy and formalism, thus avoiding the streetlight effect. 'Descriptive mapping' can appear a natural extension of the structure of OOP, which makes defining a close link between object structures and real-world counterparts an obvious thing to attempt. This is true both for objects themselves and the structural relations that OOP uses. For example, Tesfatsion uses OOP's distinctions between public, private and protected methods to define private behaviours and to allow agents to "communicate with each other through their public and protected methods" (Tesfatsion 2006 p.837).

This section critiques this kind of 'descriptive mapping' by looking at the reasoning of ABM theorists and more traditional economic thinkers, in particular Milton Friedman. The argument developed is that while ABM's flexibility makes it appear feasible and even desirable to attempt 'more realistic' mappings, the focus should be more on 'functional mapping', where the function is closely tied to the purpose of the model.

Wooldridge notes from an early commenter a "tendency to think of objects as 'actors' and endow them with human-like intentions and abilities." (Inc. 1993 p.7 in Wooldridge 2009 p.28; see also Franklin and Graesser 1996). As regards structure, Tesfatsion argues that -

> "encapsulation into agents is done in an attempt to achieve a more transparent and realistic representation of real-world systems involving multiple distributed entities with limited information and computational capabilities." (ibid p.838)

It is an entirely understandable and intuitive belief: the closer the match between code and real-world - the more 'realistic' - the better. But in what way should real-world systems be represented? It is obviously true that any effective model must in some way represent the dynamic it wants to examine; what Craik (1967 p.51) calls a 'relation structure' (see Stafford 2009 p.3) and Suarez *et al.* (2003 p.225) a 'mapping of source to target'. But to what extent must this representation actually map directly? How important is 'realism?' The goal is, as Hoover puts it, that -

> " the idealised model capture the essence of the causal structure or underlying mechanism at work... Models are not, of their nature, cleanly idealised; they must involve particular properties, whose only function is to make them operable or realisable in a manipulable form." (Hoover 2010 p.346)

Two examples of actual *physical* models illustrate the point graphically. Craik argues that models "need not resemble the real object pictorially; Kelvin's tide-predictor, which consists of a number of pulleys on levers, does not resemble a tide in appearance, but it works in the same way in certain essential respects" (Craik 1967 p.51). MONIAC, the Newlyn-Phillips hydraulic model of the economy, makes the same point. It was built mainly as an educational tool, using water to represent money-flows and various levers and wheels to control flows. Some argue it actually influenced some Keynesians (see Wood 1994 p.249). There were counter-arguments about the 'misleading reduction of economics to hydraulics' (Shackle 1983 p.189 in Wood 1994 p.249). Newyln himself was well aware of this limitation: "once the model has served its purpose... the student will need to return to the literature for the complications and refinements - hydraulics is no substitute for economics" (Newlyn 1950 p.119).

This is a perfect illustration for the upcoming section: one needs to know the purpose of the model before attacking it as misleading. The next section starts to unpick this issue by examining the place of both simplicity and realism in models.

3.4.2 What role for simplicity?

The underlying motivation for descriptive mapping is, perhaps, that a perceived closer match to reality is its own 'validation'. Conversely, this offers an easy line of attack for anyone unhappy with particular simplifying assumptions: they are unrealistic. But how can perceived lack of realism be used to judge a model, given that all models are simpler than reality? How to distinguish good from bad simplifications?

'Occam's Razor' is the idea that, given two theories that may explain some phenomenon, the simpler one is likely to be the better explanation. The centrality of complexity for ABM has tarnished the idea of simplicity, though the two should not really be opposed to each other. A particularly severe example of this sees ABM theorist Bruce Edmonds attacking Occam's Razor because 'simplicity is not truth-indicative' (Edmonds 2007) He is very blunt about what simplicity should mean to modelling:

> "if I am right, model selection 'for the sake of simplicity' is either: simply laziness; is really due to pragmatic reasons such as cost or the limitations of the modeller;

> or is really a relabelling of more sound reasons due to special circumstances or limited data. Thus appeals to it should be recognised as either spurious, dishonest or unclear and hence be abandoned." (Edmonds 2007 p.78)

There are grounds for simplicity that go beyond laziness or dishonesty on the part of the modeller, however. Occam's Razor is not an 'appeal' to simplicity, and Edmonds is right: simplicity by itself, is *not* truth-indicative at all. Occam's Razor is a shortcut for finding successful theoretical needles in among the haystack of competing ideas. Friedman nicely outlines the role it has played in the physical sciences: "the theorist starts with some set of observed and related facts, as full and comprehensive as possible" (Friedman 1953b p.282-3). There are, he argues, then an infinite number of theories consistent with the facts. Some 'arbitrary' method is needed to choose between them - such as Occam's Razor. One could just as arbitrarily choose the more complex theory, but it so happens that simple explanations have done better in the physical sciences.

Plenty of the time, however, simple assumptions do not get tested in any meaningful way. These can sometimes be seen in the wild, appearing as 'heroic assumptions'[6] in the literature. An heroic assumption is characterised by two things: being highly unrealistic, but opening up new avenues for analysis. As such, they have always been an easy target for critics, and ABM theorists have certainly taken aim at them.

Is there any way of telling what sort of assumption is valid? What distinguishes 'heroic' from 'useful' from 'silly'? The next section looks at this question.

3.4.3 What role for realism?

Moss and Edmonds' paper on 'good social science' argues that:

> "The essential feature of software agents devised for purposes of social simulation is that they should be validated as good descriptions of the behaviour and social interaction of real individuals or collections of individuals" (Moss and Edmonds 2005 p.10).

They want a social science that "coheres with directly observable evidence in as many ways as possible" (ibid p.5). In seeking this coherence, they say that "evidence and observation have priority over theory" and "when evidence and theory disagree the theory is changed" (ibid p.4). Their goal here is to firmly fix the causal arrow from reality to theory and reject any approach that points in the other direction. As they put it:

> "There are many such cases in the natural sciences where observation and experimentation lead to conceptualisation. We know of no such cases in the core of

[6] One reviewer of the English translation of Weber's 'Theory of the Location of Industries' (Weber 1909) in 1930 warns, "in approaching this book the reader must be prepared to meet a very abstract treatment and some very heroic assumptions" (Fetter 1930 p.233). There is a more recent example from Brakman *et al.*, who see the same type of heroism underpinning the success of the Dixit Stiglitz model (Brakman *et al*. 2009 p.93).

> mainstream economics or sociology, where the conceptualisation has tended to come first" (Moss and Edmonds 2005 p.10).

In this view, valid theory must flow from reality by a process of accreted, validated induction. Theirs is only a subtle tilt in emphasis away from theory-building, but its impact on what counts as 'valid' agent modelling is huge. It is suggested that Einstein's theories, built up from Maxwell's and a number of other theorists, embody the approach they espouse, being "driven by experiment and observation of natural phenomena" (ibid p.3). However, Einstein's view of the matter seems to have been rather different:

> "Physics constitute a logical system of thought which is in a state of evolution, whose basis cannot be distilled, as it were, from experience by an inductive method, but can only be arrived at by free invention. The justification (truth content) of the system rests in the verification of the derived propositions by sense experiences. The skeptic will say: 'it may well be true that this system of equations is reasonable from a logical standpoint. But it does not prove that it corresponds to nature'. You are right, dear skeptic. Experience alone can decide on truth." (Quoted in Kaldor 1972 p.1239.)

The point here is not that 'free invention' has free reign - 'experience alone' still determines the truth-content of theory but that reality does not automatically supply the descriptive elements of theory.

The search for an understanding of gas theory (discussed below in depth) shows much the same role for 'free invention' being thrown against the wall of reality. The many scientists involved went down different experimental routes, depending on prior assumptions about the nature of matter going back to Aristotle, and especially about atomism and whether a vacuum was possible. As Webster notes, for instance, Boyle got ideas about the elasticity of air from Descartes, who suggested "air was analogous to a pile of wool fleeces" (Webster 1965 p.445). Ultimately, reality was the arbiter of which theories were corroborated and notice that reality provided the metaphor in this case - but induction never had the unique priority Moss and Edmonds want to give it.

Three issues flow from this. The first is whether social modelling *needs* to meet the same criteria as physical modelling. The second, related, issue is what it should take to reject any particular theory as beyond the social-scientific pale. Any conclusive answer is beyond this thesis, but the above should at least give rather more slack to the use of Einsteinian 'free invention' in model experimentation. Third is that, as regards realism, a straightforward 'coherence' test of the sort quoted above would appear to be problematic.

In economic models, probably the most frequent subject of scorn is the idea that people are *homo economicus*: infinitely rational utility-maximising machines. This serves as a good subject for thinking through the role of realism. Sociological critiques of *homo economicus* abound: Hollis (1975; 1994 pp.52) takes on 'rational economic man', and for Granovetter and Swedberg, the starting point of economic sociology is precisely that "while interests are central to any explanation of economic activities, a purely interest-driven model is unacceptably distorted" (Granovetter and Swedberg 2001 p.9).

As mentioned above, a model is by definition simpler than the system it aims to represent. So what counts - in Granovetter and Swedberg's terms - as an acceptable distortion, and what is unacceptable? Moss and Edmonds' take is again a good representation of the starting point for many agent modellers. Talking about financial prediction (or lack of), they attack utility-based models:

> "The standard, naïve response... follows Friedman's classic claim that the descriptive accuracy of assumptions is irrelevant and all that counts is predictive accuracy... its *ceteris paribus* conditions fly in the face of common observation, common sense and experimental evidence." (Moss and Edmonds 2005 p.9)

Milton Friedman's argument - called the 'F-twist' by Samuelson (see Blaug 1992 pp.91 for an overview) is indeed a classic critique of the 'basic confusion between descriptive accuracy and analytical relevance'. Friedman says that "a theory cannot be tested by the 'realism' of its 'assumptions' " (Friedman 1953a p.23) Note, however, that he is not quite arguing assumptions are irrelevant. This idea has become something of a caricature to the point where Moss and Edmonds can claim the argument is now 'naıve'. Friedman makes an interesting case, however, and it is a very useful way into picking apart how models should map to real world.

Friedman spends some time criticising "necessarily unsuccessful attempts to construct theories on the basis of categories intended to be fully descriptive" (ibid p.34). A model may succeed in 'descriptive accuracy', but what about analytical relevance? It is true that Friedman focuses on prediction as the ultimate arbiter:

> "Complete 'realism' is clearly unattainable, and the question whether a theory is realistic 'enough' can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories." (Friedman 1953a p.41)

A vital question is then: what is meant exactly by prediction? This is dealt with in section 3.4.5. Before that, however, some context for Friedman's thinking. His argument is built on keeping clear blue water between a self-contained theoretical structure (like Einstein's take on physics as a 'logical system of thought') and the problems it is used to analyse. For Friedman, theory in and of itself is nothing more than a tautologous filing system that, while internally consistent, by itself has no 'substantive content' (Friedman 1953a p.7):

> "The objective is to construct a language that will be most fruitful in both clarifying thought and facilitating the discovery of substantive propositions." (Friedman 1962 p.8)

Internal to any particular system, it does not make sense to single out any one element as an unrealistic assumption. They cannot, by themselves, be used to accept or reject a theory, because 'everything depends on the problem' (Friedman 1953a p.7). That includes whether or not something is an assumption: Friedman concludes, much as Blaug does, that "the logical distinction between 'assumptions' and 'implications' disappears in a perfectly axiomatized theory" (Blaug 1992 p.143). The two are distinguished from each other only by the particular question under examination. This does not mean, however, that Friedman thinks 'assumptions are irrelevant'. As he says

> "if this were all there is to it, it would be hard to explain the extensive use of the concept and the strong tendency that we all have to speak of the assumptions of a theory and to compare the assumptions of alternative theories. There is too much smoke for there to be no fire." (Friedman 1953a p.41)

What counts as a 'crucial' assumption will depend on the problem at hand, and is something that Friedman thinks is beyond the scope of any simple methodology to determine (ibid p.25). There are, in fact, situations where assumptions "can be used to get some indirect evidence on the acceptability of the hypothesis in so far as the assumptions can themselves be regarded as implications of the hypothesis" (ibid p.28). This will, of course, depend on the hypothesis being proposed.

In a simple thought experiment, he suggests that one can state 'leaves seek to maximise the sunlight they receive'. This is, of course, an egregious simplification of the processes a tree goes through to achieve that maximisation, but it is nevertheless "more compact and at the same time no less comprehensive" than a list of particular rules would be (Friedman 1953a p.24). Friedman goes on to a human thought experiment: how one would go about modelling billiard players? A successful model that is, one able to make good predictions of game outcomes, on average might assume that they "knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc... could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas" (ibid. p.21). Conlisk asks (discussing life cycle decisions):

> "But what of a beginner taking the first shot, in poor light, on a badly warped and randomly moving table, with assorted friends and relatives guiding the cue stick?" (Conlisk 1996 p.684)

Then the assumptions would be poor ones for this situation: the model would not work. A relevant thought experiment of Friedman's in this case goes as follows. The equation $s = 1/2gt^2$ is a good representation of the way bodies fall under gravity (where s is distance, t time and g a gravity constant). An assumption is that this happens in a vacuum: air resistance is left out. This leads to several examples where the formula fails: for feathers, or for objects dropped from thirty thousand feet. The point is

> "... under a wide range of circumstances, bodies that fall in the atmosphere behave *as if* they were falling in a vacuum. In the language so common in economics this would be rapidly translated into: the formula assumes a vacuum. Yet it clearly does no such thing." (Friedman 1953a p.18)

There exist only certain conditions where this simple model works, but that does not nullify its use in those conditions, and requires understanding the sort of error one might expect.

In sum, Friedman's argument is run throughout with a strong vein of 'it depends on the problem' and 'nothing is set in stone'. Focusing purely on isolated assumptions makes it easy to dismiss (for example) *homo economicus* as obviously wrong, but this is an error, stemming from a misunderstanding of the type of model-building Friedman argues to be useful. Later, section 3.5.1 examines the issue of utility in more depth, and looks at how the difference between assumptions and implications depends on what is being asked. Next, section 3.4.4 looks at two model styles 'virtual worlds' and 'engines of analysis' and argues they closely resemble the distinction between descriptive and functional mapping.

3.4.4 Virtual worlds versus engines of analysis

Di Paolo *et al.* contrast two positions on what simulations should be: "maximally faithful replicas" versus "thought experiments: unrealistic fantasies which nevertheless shed light on our theories of reality" (Di Paolo *et al*. 2000 p.4). This argument in ABM has a strong parallel in economic comparisons between two nineteenth century theorists, Walras and Marshall. Friedman was a defender of Marshall's approach, saying consistently that he "took the world as it is; he sought to construct an 'engine' to analyse it, not a photographic reproduction of it" (Friedman 1953a p.35). Walras, in comparison, built what Blaug calls a "a peculiar vision of a sort of 'realistic utopia' " (Blaug 1997 p.569). He was the first theorist to construct a full 'general equilibrium' model, in an attempt to work out how all elements of the economy managed to interact to produce an apparently stable outcome. (His use of the concept of tatônnement or 'groping' towards clearing prices has already been mentioned in section 3.3.1.) This approach became the foundation for arguments that such equilibria were also optimal. As Blaug notes, over time Walras displayed "an increasing tendency... to fit the world to the model rather than the model to the world" (Blaug 1997 p.569) something carried on by further refinements to general equilibrium over time. Cast into modern language, Walras could be said to have created a 'virtual world'.

A model's status as 'virtual world' or 'engine' is not an intrinsic property of the model itself, though via an accretion of theorists' choices, one or the other can become set in. As Friedman put it

> "... by slow and gradual steps, the role assigned to economic theory has altered in the course of time until today we assign a substantially different role to theory than

> Marshall did. We curtsy to Marshall, but we walk with Walras." (Friedman 1953a p.89)

Building virtual worlds is sometimes an entirely reasonable goal. Weather models are an obvious example: the goal of meteorology is precisely to create as close a mapping as possible between virtual and real, where the virtual model can be played out faster than reality to make predictions about the future. But it is a mistake to think this is always a laudable, or even feasible, goal for human systems.

ABM is perhaps more structurally suited to virtual world building, insofar as it can be used to make simulacra that can appear to be worlds *in silico*, but are these really different from the sort of virtual world Walras made? In Artificial Intelligence (AI) research, the 'strong versus weak a-life' argument sees some saying virtual worlds can indeed be considered as more than just 'simulations of living systems': actually 'realisations of living systems'. Miller wonders whether, if the strong case were true, "we would have to add a sixth kingdom of life to the current five... and databases of biological phylogenies would have to be updated every time a new Ph.D thesis in a-life was written" (Miller 1995 p.21). It is hard to imagine this argument even arising in the same form for purely equation-based models but ABM does not represent quite the break from purely mathematical methods that some claim.

While computers are iterative by default (the system clock imposes a discrete structure), this does not necessarily mean that agent models must be discrete, or indeed that they must be algorithmic: approximations to continuous and algebraic[7] functions are possible. ABM, however, has tended to integrate discrete time-steps into its approach, and methods are employed to design agent models with appropriate timing.

It is true that ABM is algorithmic rather than equation-based[8]. While equation-based elements may be present (as they are in the thesis models) they are used by agents as part of their discrete decision-making. In a comparison of equation-based versus ABM methods for supply-chain management, Van Dyke Parunak *et al.* highlight the fundamental differences. Equation-based methods using differential equations start with observables; in contrast, the ABM modeller -

> " - begins by representing the behaviours of each individual, then turns them loose to interact. Direct relationships among the observables are an output of the process, not its input." (Van Dyke Parunak *et al*. 1998 p.10)

Again, this sounds like a virtual world: the 'direct relationships among observables' can only be examined through a pseudo-empirical observation of the model as it plays out. In

[7] Symbolic computing is capable of producing good analytic results: for example, where iterative approaches to integration like Runge Kutta functions can only approximate, symbolic computing can generally do as well as analytics in producing exact results where they exist.

[8] Epstein argues this distinction is logically false; see Epstein 2006 p.27. Epstein argues all agent models must logically have an equation-based counterpart. In practice, however, this logical equivalence is less important than the way algorithmic and OOP approaches affect the choices modellers can make.

reality, there is less difference between equation-based and algorithmic approaches. Krugman's own work illustrates one aspect of this: it took him many years to work out what the full details and consequences of his modelling approach were:

> "Why exactly I spent a decade between showing how the interaction of transport costs and increasing returns at the level of the plant could lead to the 'home market effect' and realizing that the techniques developed there led naturally to simple models of regional divergence remains a mystery to me." (Krugman 1999)

Part of the answer is that equation-based modelling can be just as opaque as ABM, and that working out the implications of the interactions is equally challenging. As Di Paolo *et al.* note: "in general, and qualified claims of the superiority of one style of modelling over another are not compelling." (Di Paolo *et al*. 2000 p.3) Ultimately, however, the difference between virtual worlds and engines of analysis does not come down to their structure, but to the way they are interpreted by modellers. The one defining characteristic of 'virtual worlds' is that they claim to explain something about this world purely through their own existence. It is taken to be self-evident that virtual and real world dynamics are identical in some key aspects. Arguments for model 'realism' lend weight to this interpretation: virtual and real world dynamics are more alike.

A useful way to put some meat on the bones of this discussion is to examine an especially popular take on agent modelling, the idea of 'generative' social science. Before that, the next section outlines three types of prediction: these will be used in the following sections to help shed light on the difference between 'engines of analysis' and 'virtual worlds'.

3.4.5 Three types of prediction

Three types of prediction can be defined. The first is forecasting: making a claim about something that will occur at a future point. The second is closely related, and uses precisely the same methods: in modern terms, it is 'back-casting'. As Friedman puts it, prediction -

> " - need not be forecasts of future events; they may be about phenomena that have occurred but observations on which have not yet been made or are not known to the person making the prediction." (Friedman 1953a p.9)

The third is what Betz calls 'ontological prediction' (Betz 2006). This is a claim about a phenomenon that has always existed (that has occurred, is occuring now, and will in the future) but was not known about or looked for. Einstein's work on the relationship of gravity and space, for example, describes something that always happened. The theory led Eddington to carry out his photographic test confirming that light 'bent' near the sun as it followed a straight line in gravity-warped space. (See Almassi 2009 for a recent discussion of this example.) Betz also notes the prediction of Neptune's existence from Newton's laws, which showed other planetary bodies' orbits to be incorrect without another planet to explain it.

The discovery of tectonic plates is a more down-to-Earth example, and one used by Epstein (2008). This example highlights the clear relation between the different types of prediction: understanding tectonic plates places a boundary around possible future outcomes that were unknown before. Earthquakes will mostly happen on or near lines of tectonic plates, and understanding of tsunamis can also be built on that knowledge (Thompson and Derr 2009, who critique Epstein's approach).

'Ontological predictions', then, can lead to discoveries, if the predictions are good, as well as provide methods for placing bounds around forecasts. Putting it in 'streetlight effect' terms, they shine a light into previously dark areas. Krugman's core model, it could be argued, is an economic example of this, opening up novel ways to think about international trade that do not rely on factor differences, as well as enabling a much more clear conception of the possibility of emergent 'core' and 'periphery' structures. If a theory can facilitate the 'discovery of substantive propositions', these may take the form of ontological predictions.

3.4.6 Applied versus theory

Webber, writing in the eighties, identified some confusion from mixing up the purpose of models. He distinguishes between applied and theoretic (or 'scientific') models (Webber 1984). Theoretic models are "designed to increase our understanding... and to integrate theoretical and empirical research." Applied models, in contrast, are tools used in the day-to-day running of organisations and institutions, and in longer-term forecasting and planning. They -

> " - regard the form of society as given and are produced to make that society operate more profitably; since scientific research examines the conditions under which the society operates and the manner in which it is changing, such research ought not to regard the form of society as constant." (p.149)

Webber argues this comes to down to 'a difference between the need to understand and the need to prescribe' (p.151). Traditionally in economics, this is the distinction between positive and normative analyses (Friedman 1953a).

Questions from an applied or theoretic perspective are completely different. Transport economics is a good example of an applied approach; as Button says, it takes a "given land-use pattern and looks at methods of providing efficient transport services within this constraint" (Button 2010 p.51). He also notes the same 'applied' ethos, taking existing structures as given in the short-term, characterises supply chain analysis and operations research (ibid p.327). Compare this to Haggett's search for a 'comprehensive model of route development' discussed in section 2.8.1: a transport economist may need to consider, for example, how the risk from fuel cost changes could be mitigated through infrastructure design, but that is quite different to asking, how would the spatial economy as a whole be affected by these cost changes in the long term?

ABM, as with any form of social modelling, has feet in both camps, but the boundaries between them are more blurred than were the urban models Webber examined. Some do identify a similar distinction; for instance, Brown and Xie (2006 p.941) talk of two ‘modes’: instrumental and representational. Instrumental agents are carrying out some real-world task, whether collecting information on the internet or optimising container port flows, whereas representational agents are doing just that: representing some object. But these do not seem to count as either ‘applied’ or ‘theoretic’: either can be used in applied settings, and either could have a role in operations-related work. Optimisation of container ports is a typical case. The gap between models of good flow and reality is shortened; the model not only represents the target system, it is uploaded directly the port’s computer systems, in effect making the model reality (Steenken *et al*. 2004). If models are maps, these maps are reflexive: it not only describes, but is used to act upon the world (Scott 1998 p.87).

The need to model systems that require regulation is not new; Conant and Ashby (1970), for instance made an early argument for this. One might reply that agent-based operations models are qualitatively different, because cognitively more sophisticated. Regardless of whether this is true (and some cyberneticians would probably disagree; see e.g. Beer 1994), there are other reasons why the distinction between applied and theoretic models has become blurred in ABM that tie back to the type of mapping a model is meant to achieve.

This mirroring of real and virtual systems is underpinned by arguments that, as Epstein says, “certain social systems, such as trade networks, are essentially computational architectures.” (Epstein 2006 p.16) Epstein can thus develop a theory that computations carried out *in silico* can be considered as a version of the physical computation under study. Similar lines of thought about theoretic agent modelling have created quite a tangle as regards the model-reality relationship. For instance, an argument has developed that such models can actually be “synthetic sources of empirical data” (Di Paolo *et al*. 2000, who criticise this approach). In combination with the idea of ‘emergence’, this has been followed through to its logical conclusion: what Epstein calls ‘generative social science’, in which a model that generates a particular phenomenon can be said to explain it.

3.4.7 Generation versus explanation

Joshua Epstein’s work on ‘growing artificial societies’ (Epstein 1996) first presented his notion that *in silico* emergence could be a form of explanation in and of itself. The motto became, ‘if you didn’t grow it, you didn’t explain it’ (Epstein 2006 p.xii). Epstein calls this ‘generative explanation’ and styles himself has a ‘generativist.’ He sees computational models as ‘a new scientific instrument’ (ibid. p.xv) able to generate explanations about the social world almost without reference to physical reality. This stems from the idea that both ‘worlds’ are in some sense producing the same dynamic. As Epstein puts it, “the agent-based approach invites the interpretation of society as a distributed computational device, and in turn the interpretation of social dynamics as a type of computation” (ibid. p.11). Epstein (perhaps accidentally) gives a perfect illustration of the blurring between virtual and real worlds this argument leads to:

> "No-one would fault a 'theoremless' laboratory biologist for claiming to understand population dynamics in beetles when he reports a regularity observed over a large number of experiments. But when agent-based modellers show such results indeed, far more robust ones there's a demand for equations and proofs." (Epstein 2006 p.28)

They might fault them, however, if they claimed their experiments had in fact explained the behaviour of beetles in the wild. More than that, there are plausible reasons for thinking live beetles might be a better model than an artificial life version, even in a laboratory setting. No-one would fault them for using either digital or live models to explore the problem but this is a much less grandiose goal than explanation. Yet for Epstein, any model 'sufficient to generate' a target macroscopic phenomena has succeeded in explaining it. This may be in part hyperbole, since it appears to be contradicted by statements elsewhere (see below), but it is still striking.

Epstein picks on the central dogma of neoclassical economics, discussed in section 3.3.1, as a basic illustration of his argument: demonstrating the existence of equilibria says nothing about the process of reaching it. Epstein says of equilibrium -

> "To the generativist, this is unsatisfactory; to explain a pattern, it does *not* suffice to demonstrate that under this ensemble of strictures *if* society is placed in the pattern, no (rational) individual would unilaterally depart. Rather, one must show how the population of boundedly rational (i.e. cognitively plausible) and heterogeneous agents... could actually arrived at the pattern on time scales of interest" (ibid. pxiii).

The generative approach, then, sounds very Hayekian: "it is irrelevant that equilibrium can be computed by an economist external to the system... The entire issue is whether it can be attained generated through decentralised local interactions of heterogenous boundedly rational actors" (ibid. p.27).

Schelling's segregation model, discussed by Epstein as a classic model of emergence, illustrates the problem of privileging this kind of 'generation'. As a virtual world, one makes the conclusion Epstein does: it 'explains' segregation, and the model's use stops there, 'brushed under the carpet of emergence' (Di Paolo *et al*. 2000 p.8). As an engine of analysis, it presents a hypothesis: a possible dynamic factor, among others, that may explain something about real-world segregation. Putting aside the hidden assumption - that all social macro phenomena must be the result of a process of interaction between micro-elements it would need to be put into the context of a fuller theory, that might include (for example) the racial impact of transport infrastructure (e.g. the policies of Robert Moses, see Winner 1980), regeneration policy (Gaffikin and Morrissey 2011) or the effect of estate and letting agents' steering (Phillips and Karn 1992).

A last example gets to the nub of the issue with the generative approach. Epstein considers the 'confusion between explanation and description'. He looks at an example from his 'Sugarscape' models (Epstein 1996) where agents generated a sine-like oscillation of population change over time, and asks, "could you not get that same curve from some low-dimensional differential equation, and if so, why do you need the agent model?" He suggests an equation to describe the model, just a simple, stable sine-based oscillation. He then asks (italics in original) -

> "now, what is the explanatory significance of that descriptively accurate result? It depends on one's criteria for explanation. If we are generativists, the question is: how could the spatially decentralised interactions of heterogenous autonomous agents generate that macroscopic regularity? If that is one's question, then the mere formula $P(t) = A + B \cdot sin(Ct)$ is *devoid of explanatory power despite its descriptive accuracy.* The choice of agents versus equations always hinges on the objectives of the analysis" (Epstein 2006 p.28-9).

Epstein is absolutely right: it does depend on one's criteria for explanation. His example takes the story back to Feynman's quote at the start of the chapter: whether something counts as an explanation depends on what level of explanation is required. The most vital point is that nothing is implied by accepting one particular level of explanation; certainly, it does not mean the theorist has failed to understand that deeper levels of explanation may exist. (The next section examines this issue in detail.)

Epstein appears to be completely in agreement that a model may represent many possible theories about the world, and need testing appropriately, and that the 'generativist' is merely subscribing to one particular 'criteria for explanation'. Yet his central argument is still that 'growing it explains it'. Why this far-reaching claim, rather than perhaps more modestly proposing his models as a way to develop hypotheses? The only way it makes sense is if the model is, in some sense, an empirical world of its own.

The story in this section has been about ABM's tendency toward building of virtual worlds, and treating them as sources of empirical data. As Di Paolo *et al.* say, however, "simply treating non obvious patterns or entities as 'emergent' is not an explanation at all, but rather the statement of a problem" (Di Paolo *et al*. 2000 p.8). Epstein's mistake, then, is to water down the concept of 'explanation' so much that it cannot be distinguished from 'hypothesis'. Epstein seems to tacitly acknowledge the distinction when he argues that, if more than one candidate 'microspecification' is found that could potentially be applied to the problem, "as in any other science, one must do more work, figuring out which of the microspecifications is most tenable empirically," which may involve the need for new experiment or data collection (p.9). It is perhaps an issue of semantics then, but the impact on the outcome of models is not trivial.

3.4.8 What level of explanation?

The theory of the ideal gas law often arises in the literature as a useful thought experiment for examining the difference between 'levels' of explanation, and is particularly well-suited to thinking through ABM issues. Flake puts it well:

> "Collections of gas molecules behave in very predictable ways. Knowing only the temperature and pressure of a gas tells you enough about the whole ensemble of molecules that you can effectively ignore what the individual molecules are doing." (Flake 1998 p.134)

So, atoms are 'not like' their macroscopic behaviour might suggest, but this does nothing to change the validity of describing collections of atoms in terms of temperature and pressure. There are theories for understanding the behaviour of gases at both the aggregate and atomic level, and each is useful in the correct context. Heating a gas while keeping the volume constant will lead to an increase in pressure.

The thought experiment reveals a number of different issues in separate areas of economic theory. The first relates to ACE's understanding of complexity. Take Tesfatsion's definition of complexity for example. Two requirements are given: the system is "composed of interacting units" and "exhibits *emergent* properties, that is, properties arising from the interactions of the units that are not properties of the individual units themselves" (Tesfatsion 2006 p.836). By this definition, temperature and pressure exhibit complexity, and yet clearly they can be described using simple known equations.

Secondly, as Hoover suggests, many economists are unhappy with having several explanatory levels, each qualitatively different from the other. He argues some believe that 'aggregates are nothing else but summary statistics reflecting individual behaviour'. In comparison -

> "... those who believe that the ideal gas laws reduce to statistical mechanics do not claim that the ideal gas laws should be abandoned for practical purposes." (Hoover 2010 p.331)

Hoover charts the change in economics from a Marshallian emphasis on 'individual *and social* action' (emphasis added) to the micro-economic focus on 'human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins 1935 p.16 quoted in Hoover 2001 p.108). Many economists, like Hayek, think aggregate properties have no real existence a 'fallacy of misplaced concreteness'. From this point of view, aggregate or emergent properties cannot be understood in the straightforward way that temperature and pressure can in gases:

> "Hayek thus argues that aggregates exist, but derivatively rather than fundamentally, and that... they do not exist objectively (i.e. unconstituted by the representations of theory." (Hoover 2001 p.108)

The argument is not unique to economics: physicists also tussle over whether certain emergent properties 'are not reducible without loss to the behaviour of the particles that constitute their substance' (ibid p.112). Yet economists often have a more troubled relationship with the connection between explanatory levels. In his 'Self-Organising Economy' (Krugman 1996, p.15), Krugman asks "what constitutes an 'explanation' from the point of view of economists?" The title of Thomas Schelling's 'Micromotives and Macrobehaviour' (1978) is, for Krugman, a compact answer to that question: a good economic explanation shows how micromotives link to macro results. A common argument from ABM-sympathetic economists is that, while the goal of linking micromotive and macrobehaviour is laudable, mainstream economists have built only fallacious arguments.

What arguments are these? Returning to Flake, he makes a second point: "notice that the properties of temperature and pressure cannot be attributed to a single gas molecule but only to collections of molecules" (Flake 1998 p.134). Making this sort of attribution is known as the 'fallacy of division'. The reverse assigning properties to macro-level entities because its components possess them - is a fallacy of composition. In all these cases, the fallacy is not that such claims are *a priori* unjustifiable, but rather in presuming them to be true without evidence to support that presumption. Howitt believes that -

> "these twin fallacies play an even bigger role in a macroeconomist's education than they did a generation ago; the difference is that instead of being taught as pitfalls to be avoided they are now presented as paradigms to be emulated." (Howitt 2006 pp.1069).

A particular target for critics is the use of the 'representative agent' approach: a single agent, with one utility function, stands in for all agents. Kirman (1992) has made a strong critique of the representative agent, claiming it is nothing more than 'pseudo-microfoundations' (p.125). It is, he argues, a necessary lynchpin for claiming that equilibria are unique, and thus comparable. One of his most effective points is that many economic problems can actually be understood much more simply if the economy is treated as many separate agents. As he notes, "erratic individual demand behavior may give very smooth aggregate demand behavior, if individuals are different enough" (p.129). An intuitive example is stability of crop prices over a season's production, as farmers individually look to maximise profit: by selling when they judge they will get the best price, and avoiding over-supply points where prices will be lower, the market is smoothed. More formally, Kirman points out that a 'representative' consumer with non-convex preferences[9] would see demand "jump from one bundle to another at certain prices" (ibid). In contrast, many heterogenous actors with non-convex preferences may well produce a smooth overall market response in aggregate[10].

[9]The useful property of convex preferences is that a budget line drawn through them will only ever provide a single unique point of consumption. In contrast, other sorts of 'not well-behaved' preferences may mean demand 'jumping' as income or prices change. See e.g. Varian 2006 p.77.

[10]The model in section 6.3 has oscillatory behaviour: it may be the case that heterogenous agents with non-convex preferences would help to smooth out those market shifts.

This fallacy, Conlisk notes, might well be an "ironic misspecification problem" (Conlisk 1996 p.677) that suggests (in Heiner's words) -

> " a reversal of the explanation assumed in standard economics: the factors that standard theory places in the error term are in fact what is producing behavioural regularities, while optimizing will tend to produce sophisticated deviations from these patterns. Hence, the observed regularities that economics has tried to explain on the basis of optimization would disappear if agents could actually maximize" (Heiner 1983 p.586).

In conclusion, there are arguments over just about every element of 'levels' of explanation. What seems to separate them all is a view of the nature of the connection between levels. Some, like Hayek, dismiss macroscopic explanations completely, while one of his critics suggests he fails to analyse the connection between levels adequately: he "merely invokes the magic words *the price system* without examining its entrails. It is as if correctly sensing the importance of sunlight for life on earth, we were to merely worship the sun rather than study astronomy or photosynthesis (Desai 1994, p.47).

On the other hand, many economists are, allegedly, "scandalised to discover how cavalier physicists are in making conjectures that lack any fundamental justification" (Ball 2007 p.647 paraphrasing Miller and Page 2007), preferring theories with what they perceive as solid micro-foundations. This is very evident in the way gravity models are treated by economists. Whereas gravity model use in trade cost flows has proved to be very effective (e.g Carrere *et al.* can use a gravity model to conclude that "distance impedes trade by 37% more since 1990 that it did from 1870 to 1969"; Carrre *et al*. 2009 p.6), Fujita *et al.* (2001) can lament the 'limitations of regional science', noting that "the general sense of loose ends left hanging prevented it from becoming a well-integrated part of mainstream economics." (p.33) However, they also point out that it *did* become a 'toolbox for practical analysis' used to guide policy, despite its purported lack of a 'rigorous framework'. Brakman *et al.* are equally happy to reject 'macro' theories like market potential and gravity models on the same basis. They argue these are theories that -

> "... try to come to grips with a spatial regularity but that lack a convincing economic-theoretical foundation. In contrast to (neoclassical) economic theory, there is a tendency to merely give a representation using, for example, simple equations, of the regularity without a connection to a model of the underlying individual behaviour by economic agents." (Brakman *et al*. 2009 p.48)

Yet they acknowledge that a gravity equation can accurately capture the drop-off of trade flows between regions; one study of Germany they mention manages an r-squared value of 0.915. This sort of result explains why such models end up in planners' toolboxes.

It also suggests a question: why can't gravity models be just as appropriate a description *at their level* as temperature and pressure are for gases? Wilson, for one, consciously acknowledges the parallels between them (Wilson 2000 p.151). In the development of gas theory, physicists attacked the problem at all levels and, as argued, no approach *a priori* nullified the other. Newton's original theory of gravity illustrates that one level of explanation can be entirely successful indeed, in this case, is the exemplar of a scientifically robust, generalised theory while lacking 'microfoundations'. As Newton said: "I have not been able to discover the causes of those properties... and I frame no hypothesis" (quoted in Silver 2000 p.44). The search for gravity's microfoundations carries on to this day at CERN.

Given all that confusion, the next section makes some attempt to see how the connections between levels of explanation might be important for ABM.

3.4.9 Connecting levels of explanation

The previous sections have discussed that different levels of explanation exist. But how does one 'get between levels'? This section looks at some examples. Below, the nature of firms' decision-making is discussed. Before that, the focus remains on thinking through how the microscopic level of actor interaction connects to macro outcomes.

O'Sullivan and Haklay are absolutely right to point out the default state of ABM as one of 'methodological individualism' (O'Sullivan and Haklay 2000 p.1413): it is concerned almost exclusively with how atomistic actors produce macro-level regularities. This often "unacknowledged assumption" (ibid) is a natural consequence of OOP guiding the modeller to individual objects.

There is, clearly, a connection between individual behaviour and aggregate outcome, but how can it be theorised? When building agents, what cognitive resources are required to connect them? The problem can be thought about as two poles: *metis* versus zero-intelligence. The former is Hayek's position. For example, he argues any kind of socialist economy is *a priori* impossible precisely because complex, embedded human knowledge makes up the microscopic components of economic activity: his argument is "not so much that a socialist economy could not transmit the necessary data, but rather that it could not generate it to begin with" (Chamberlain 1998) because it originates in the complex web of embodied experience in each economic actor. The knowledge involved, according to this argument, cannot be summarised or modelled.

Scott calls this kind of knowledge *metis*: situated knowledge, of the kind a tug captain has who knows how to pilot through one specific harbour. Metis is a form of knowledge which "represents a wide array of practical skills and acquired intelligence in responding to a constantly changing natural and human environment" (Scott 1998 p.313). The metis argument implies that exchange of information between humans cannot be reduced to 'information' in a computer. As Hodgson says

> "the information held and transmitted in the form of a symbol is thus embedded in a network of interconnected meanings, related to and produced by social structures. Genetic or computer information does not have this quality; it is at most

> indexical. In contrast, human information is structured and cultural; it is entwined with institutions." (Hodgson 1996 p.253)

Putting aside that genetic information is obviously entwined with its entire past evolutionary history of interacting in given environments, the point is the same: if the metis argument is right, it implies not that models of human behaviour must fail, but that generating valid macroscopic results from modelling at the actor level would require cognitively sophisticated objects able to develop their own 'metis' for example, their own particular, contextual understanding, to be summed up in a choice of price.

Relatedly, but not quite the same argument, is that (as Tesfatsion describes Penrose) "there is something fundamentally non-computational about human thought, something that intrinsically prevents the algorithmic representation of human cognitive and social behaviors" (Tesfatsion 2006 p.844). Tesfatsion brings up Franklins' 'first AI debate'; as she frames it, the problem is this:

> "in any purely mathematical model, including any ACE model in which agents do not have access to 'true' random numbers, the actions of an agent are ultimately determined by the conditions of the agent's world at the time of the agent's conception. A fundamental issue... is whether or not the same holds true for humans" (Tesfatsion 2006 p.844).

The question for the first AI debate is, 'can we, or can we not, expect computers to think in the sense that humans do?' (Franklin 1997 p.99). The Penrose view, as Franklin says, is that "not only can computers not experience the things we experience consciously, they can't *do* the things we do consciously" (ibid). Tesfatsion opts for a fudge: "lacking a definitive answer to this question, ACE researchers argue more pragmatically that agent-based tools facilitate the modelling of cognitive agents with more realistic social and learning capabilities (hence more autonomy) than one finds in traditional Homo economicus" (Tesfatsion 2006 p.844).

The point that seems to be missing from Tesfatsion's argument is that Penrose could be right and agent modelling still valid. The existence of the simplest supply and demand dynamic shows this: regardless of whether some irreducible nugget of consciousness places human buying decisions beyond the reach of any model's initial conditions, those decisions can still produce aggregate order. If they do, there are definitely valid reasons for disaggregating into 'unrealistic' agents where aggregate regularities exist.

The whole enterprise of microsimulation (e.g. Ballas and Clarke 2001) is based on using those regularities to cycle between macro and synthetically generated micro models and back as a way of leveraging the information in different datasets. The key point is this: there is no 'fallacy' of composition or division taking place, despite the fact that the model actors are being treated as microscopic statistics-processing machines. The unrealism involved - the lack of coherence to their actual internal sophistication or embeddedness - does not *a priori* invalidate the approach.

In some senses, this makes Penrose's argument *less* problematic than metis, to the extent that metis means all actor decisions have massive interdependence. Actors with metis may sit in the middle of the 'complexity curve' between equation-based and statistical predictability (Flake 1998 p.135), and thus summary statistics may be of no use. If predictive power were the goal, attempting what Friedman calls 'photographic representations' in this situation would not be sensible.

Tesfatsion's insistence that, regardless, agent models can manage more cognitively plausible agents suggests the persistent implicit preference for descriptive mapping: human cognitive sophistication does not necessarily require model agents to attempt to mimic it, and doing so does not make them better models. Ross summarises why: "you... have enough in common with economic agents, especially in modern institutional settings, that non-trivial predictions about your individual behaviour can be had by modelling as if, within temporal and institutional constraints, you were such agents." Ross 2008 p.130)

At the other end of the scale are 'zero intelligence' agents: these can "bid randomly subject only to budget constraints [and] may achieve near perfect market efficiency" (Conlisk 1996 p.675). Their success suggests that, for important classes of problem, metis is of no relevance. Rather, it is the structure that actors must work within that constrains outcomes so much so that random actions, properly structured, can reach optimal macroscopic outcomes.

As section 3.5.2 discusses shortly, the modelling of firm decision-making has all the same issues as for human beings. Yet, again, collosal simplifications are common. Taking the example of the 'engine' of GE, the monopolistic competition model, the equilibrium number of firms is completely constrained by the level of demand, as well as its love of variety nature. If a disaggregated version were modelled, would there be any difference between allowing firms entry and exit randomly, and giving them random decisions, or building a more sophisticated 'cognitive structure'? Firms' internal and external processes are obviously much more accessible than the human mind: shouldn't this make such simplicifications a clear violation of reality?

The function of the monopolistic competition model is to allow a simple analysis of what happens in a market of that sort which, as an added bonus, allows a spatial model to be created. It is not meant to be realistic. So would developing a disaggregated model supersede it? What if that model failed to recreate the monopolistic competition model? Would that falsify it? The Hayekian/Epstein-style argument would be, if the dynamic in the monopolistic competition model could not be produced through the interaction of agents, it is nothing more than an unacceptable distortion of reality, an assumption too far that claims to know the end process of monopolistic competition with no justification.

Friedman's 'as if' argument is often used in relation to this question. Firm entry and exit is standing in for a complex process of adaptation to economic reality. As Friedman says, "the process of 'natural selection' thus helps to validate the hypothesis or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgement that it summarises appropriately the condition for survival" (Friedman 1953a p.22). Firms are acting 'as if' evolutionary pressure is selecting them: "they will evolve to the extent that selection processes quickly eliminate poorly administered behaviour" (Heiner 1983 p.586) In a book written specifically as a guide to action in business, Baumol makes the same point in more concrete way:

> "It is at least possible that shear business acumen and experience permit management and other economic units to arrive at decisions which come close to being optimal. Moreover, in business, competition may soon eliminate firms whose decision-making is consistently poor. To the extent that these assertions are valid, optimality analysis should serve as a relatively good predictor of economic behaviour; that is, it should provide a reasonably good explanation of actual economic decisions and activities." (Baumol 1976 p.5)

The fundamental point from the monopolistic competition model is that, for a given demand level, only a set amount of firms can survive. The *process* leading to that number is obviously completely glossed over but for its purposes, that is taken to be a strength, not a weakness.

This is a useful example because it highlights one of the major divides between the physical and social sciences. Baumol wants to say that economists' relationship to operations research is "somewhat analogous to the physicist's relation to the engineer" (Baumol 1976 p.5) But as the quote at the start of the chapter suggested - no physicist need attempt to educate atoms on how best to follow physical laws. McLuhan's aphorism applies: "we shape our tools and thereafter our tools shape us" (Mcluhan 1964). In mundane modelling terms, the problem is intractable: a model is unlikely to be able to accurately map a reality that includes itself.

To put it in rather more down-to-earth terms, Baumol is conflating applied and theoretic models here. It may be entirely feasible to treat firms in the way the core model does for gaining understanding, but arguing, as Baumol does, that such knowledge could actually be used by firms to 'engineer' outcomes seems a much more questionable proposition.

The parallel to ideal gas theory is also instructive in what it says about the *difference* between physical and social modelling. The connection between atomic motion and pressure needs to be robust. Such a solid link between layers of social analysis might seem an appealing goal, but the failure to achieve it would not automatically throw any particular level of explanation into doubt. Indeed, Summers suggests that "attempts to make empirical work take on too many of the trappings of science render it uninformative" (Summers 1991 p.130).

3.5 Production and utility in an agent context

3.5.1 Using utility in an agent model

The idea of utility is a simple way of thinking through how people react to cost changes, and should be considered as just that. As discussed in section 3.4.3, it is often seen to be umbilically linked to what Howitt calls an 'irrational passion for dispassionate rationality', which makes it "easy to dismiss as *ad hoc* or poorly grounded any theory that starts with behavioural rules not explicitly derived from rational foundations" (Howitt 2006 pp.1610) But pinning utility onto rationality is not necessary any more than the theory of gravity requires planets to love each other.

Section 3.4.3 discussed the question of realism of assumptions. Friedman cites an early critic, Thorstein Veblen, pouring derision on the idea of people as "a lightning calculator of pleasures and pains, who oscillates like a homogenous globule of desire of happiness" (Veblen 1898 p.389 quoted in Friedman 1953a p.30). It was argued that pointing out people are simply not like that is not, by itself, a strong enough reason to reject the whole approach.

The ABM literature contains opposing views, though the overall bias towards complexity means utility is ignored more often than attacked. Vriend is a rare theorist, both an agent modeller and a careful critic of utility. As he points out, "economic behaviour simply means that an individual agent chooses (one of) the most advantageous options, given their preferences, in their perceived opportunity set" (Vriend 1994 p.33). Arguments for rational underpinnings are, for Vriend, just "another name for economic behaviour; a question of rhetorics" (Vriend 1996 p.265-6). Feynman's 'levels' argument helps here: this is just working in a framework where something is allowed to be true. As Vriend puts it:

> "Abstracting from an explanation of the individual agent's preferences, and from the mental processes by which he arrives at choices, economics is just a very specific abstraction from reality. Whether these fundamental abstractions are good approximations of reality depends upon the usefulness of the explanatory discourses one can build on it." (Vriend 1996 p.265-6).

This is just reiterating Friedman's point: the theory is essentially tautologous. As Blaug puts it, it does not need a 'hedonistic premise' (Blaug 1997 p.338): there is no need to gain access to people's internal states for utility to be a useful tool for understanding the effect of cost changes. But treating utility theory as a self-consistent tautology is hardly unproblematic. Becker's take on rationality illustrates the issue well:

> "When an apparently profitable opportunity is not exploited, the economic approach does not take refuge in assertions about irrationality... Rather it postulates the existence of costs, monetary or psychic, of taking advantage of these opportunities that eliminate their profitability costs that may not be easily 'seen' by outside observers" (Becker 1976, p.7).

In this way, any action taken by economic actors becomes their 'revealed preference': if one accepts *a priori* that the action is rational, the action itself must logically be the outcome of a rational choice. This way of thinking helped Becker produce ideas like rational addiction theory, where the focus becomes precisely about 'rationality as explanation'. This leads, unsurprisingly, to incredulous critiques that argue it "raises the question of how they can be taken seriously"

(Rogeberg 2004 p.264). (For a survey, see Melberg and Rogeberg 2010.) Becker's argument also seems to make this approach to utility completely immune to empirical testing. As Conlisk puts it: "whatever the truth about the particular case, economic research often seems to work backwards from empirical findings to whatever utility maximisation will work. Where the empirical arrow falls, there we paint the utility bullseye" (Conlisk 1996 p.685).

This is certainly how Becker's approach sounds: a classic case of allowing *ad hoc* modifications (see e.g. Chalmers 1999 pp.75) to keep the theory of rationality through revealed preference coherent. A simple thought experiment should illustrate. If a model of allotment production versus growing food in one's backyard were built, how would utility be used? The factors can be reduced to time and distance. Does an actor stand to gain more utility from travelling to the allotment and putting time in there, or by staying at home and growing, thus eliminating travel costs? (This is a similar approach to optimising time used in the models in this thesis.) For a set of actors who can choose between allotment and home-garden plots of similar size, rationally, it seems they should choose to eliminate travel costs and produce in their own backyard. But what if reality does not conform to this - there is a set of people who choose allotments over their own, perfectly suitable garden. Now one might propose an alteration to the theory, making sure to keep the assumption of rationality, but assuming a previously hidden element: utility is gained from leisure use of the home garden that would be lost by turning it into a vegetable patch. That would need to be outweighed by the disutility of distance before an actor would dig up their begonias.

From a philosophy of science point of view, this makes utility theory especially of the type grounded in rational choice beyond the falsificationist pale. It is immune to almost any challenge, since all questions posed by reality can simply be tidied up as a hidden cost or benefit. However, should utility theory be treated as a scientific theory in this way?

Thinking about how people react to *space* cost changes actually helps give the discussion some concreteness. Transport economics' use of revealed preference is an excellent illustration of the usefulness of the idea in practice. Approaches to utility in the field have used the idea of revealed preference successfully to say useful things about people's choices. Button points out that "the general conclusion about the idea that some overall budget mechanism governs individual travel decisions, however, must be that, to date, the evidence available still leaves many questions unanswered and the theory is still largely unproved" (Button 2010 p.92). Yet, despite this, many directly practicable ideas have emerged through studying people's reaction to space costs. One that stands out is finding revealed preferences for travel time: "if a person chooses to pay \$x to save y minutes then he/she is revealing an implicit value of time equal to at least \$(x/y) per minute." This has led to an understanding that

> "savings in walking and waiting times are valued at between two and three times savings in on-vehicle time - parameters that have proved to be remarkably robust over the years." (ibid. p.104)

This is striking: how people value travel time is not simply a function of the time taken. Those willing to commute by car for an hour each way would be much less likely to walk for that length of time. As Button points out, these findings go back to at least 1967 (Quarmby 1967 p.297). One might counter: utility describes but does not explain this. But again, it is clear that the difference between the theory being a description or an explanation comes down to one's particular purpose, and the level of analysis. Certainly, the underlying explanation would require a deeper unpicking of the factors affecting mode choice, but it would allow one to develop a predictive theory.

One last issue is worth mentioning. A particularly troublesome problem is the issue of comparability. Much of the early arguments related to utility revolved around this issue. This included early attempts in experimental economics: even searching for 'utils' and marginal utility by taking milk and bread away from people (Fisher 1927 in Blaug 1997 p.314). As mentioned in section 4.7.3, the development of Pareto optimality avoided the whole problem by ruling out inter-actor utility comparisons. In the thesis models, the situation is somewhat peculiar, in that from a 'modellers-eye' point of view, actors all use the same utility functions and share the same tastes, yet no direct comparisons are ever required of the model actors. The only ability actors are *required* to have is the basic microeconomic ones: complete, reflexive and transitive preferences. These are as follows. Complete: any two bundles of goods can be compared; reflexive: any bundle is at least as good as itself. Transitive: of three bundles of goods: x,y and z. if $x > y$ and $y > z$, then $x > z$ (Varian 2006 p.35). Section 4.6 explains how model actors go about assembling and comparing bundles to achieve this.

3.5.2 Production

Agent models of production face all the same issues as individual action and utility, not least because often precisely the same functions are used to describe both, and both are required to maximise those functions given a limited quantity of inputs. The problem of abstracting from the reality of firms' decisions is as thorny as for people. As Blaug notes, utility "no more 'explains' an individual's choices than a production-transformation curve 'explains' the state of technology" (Blaug 1997 p.337). Analysis of production, however, is clearly a more accessible problem than building testable models of people's internal mental processes. From Adam Smith's analysis of pin factories (Smith 1776) to modern work on the collective cognitive processes of productive activities (e.g. Hutchins 1996, a detailed study of a navy navigation crew), the mechanics of production have been a key focus for economics. Innovation is an ideal subject for CAS theory, and is often analysed as an evolutionary process. (For a recent overview see Safarzynska and Bergh 2010 pp.347; or earlier, Dawid 2006.)

So the simple approach to production taken in the thesis model needs even more careful caveats than for utility. As Storper notes of the core model, attempting to explain agglomeration and growth through "an indeterminate, simultaneous dance of firms, consumer-workers and product varieties and scales... is not very convincing" (Storper 2010 p.317). In particular, linking the quantity of labour input directly to an instantaneous increase in output efficiency clearly does not capture the process of developing those efficiencies. It is important, therefore, to keep claims about modelling that development separate.

Ellerman's essay on Jacobs captures the difference between growth and development nicely: development is not just 'growth' but "differentiation, diversification, and transformation in the products and in the underlying processes of production - all of which might be hidden in the black box of "total factor productivity' ". (Ellerman 2002 p.4). Referring to the common use of capital and labour inputs into production functions (K and L), Ellerman sees Jacobs' take on development as "more like the process of epigenetic transformation, not blowing up a small balloon - with more K and L - to make a big balloon" (ibid).

From this point of view, reducing development to a difference between Jacobs and Marshall externalities (see section 2.7.2) is dubious: Jacobs' argument goes beyond measuring the diversity of sectors in a region, as Glaeser does. Again, however, it comes down to the level of analysis: it is perfectly possible that Glaeser's measuring of diversity is a sensible proxy for just the kind of development Jacobs talks about, without denying the fact that the underlying mechanics are much deeper.

Firm dynamics go beyond internal structure. Section 2.8.3 described a facet of this dynamic: the trans-national corporation as a method of managing risk over space and time. As Conlisk says, production structures are "critically shaped by a need to economise on various transaction costs" (Conlisk 1996 p.675): a force for consolidation. Jacobs' take on cities as "symbiotic nests of suppliers" (Jacobs 1986 p.76) pulls the other way, making productivity a function of the 'tangled bank' of firm inter-relations (Ellerman 2002 p.6). This points more towards a focus on the combination of evolutionary dynamics and the traditional 'forward and backward linkages' between diverse firms (see e.g. Brakman and Heijdra 2011 p.5).

The approach to increasing returns in both the core model and the production model presented here is thus a huge simplification. But the goal is to examine the connection between production and welfare in as simple a form as possible; any conclusions must be made in the light of the complications described here.

3.5.3 The link between production and utility

The problem of modelling production starts with demand. If constrained optimisation is used, a utility function is constrained by a budget. As outlined in section 2.2.4, this produces a set of equations giving the 'objective' optimal quantity of goods, given that budget. The approach in GE (and many other economic models) is to assume the objective demand implies the correct level of production to meet it, and thus the correct number and scale of firms. In terms of agents, as Leijonhufvud puts it -

> "in this theory utility or profit maximisation is a statement about actual performance, not just motivation... The theory does not leave room for failures to realise

> the relevant optima." (Leijonhufvud 2006 p.1628)

It is this kind of assumption in particular that exercised Hayek. Production is assumed as the mirror image of consumption: "consumers in evaluating ('demanding') consumers' goods *ipso facto* also evaluate the means of production which enter into the production of these goods.' " (Schumpeter 1942 p.175 in Hayek, ibid.) Hayek's reasoning for rejecting the 'implicit production' assumption differs markedly from how Friedman says assumptions should be judged. Hayek refuses to allow these assumptions, labelling them a figment of the modeller's imagination imposed after the fact.

Jane Jacobs took a very similar line towards any descriptive production theory not grounded in the underlying mechanics. She attacked arguments that comparative advantage originates from the division of labour as teleological; "one might as well say rain is beneficial to plants and that is why it rains." (Jacobs 1986 p.70) For Jacobs, as well as Hayek, the logic of causation is wrong. As she says, of the definition of efficiency intrinsic to increasing returns, claiming a country is more efficient because of specialisation "is to stand reality on its head" (ibid p.71). Jacobs wants to build a causal argument about the role of 'nests of symbiotic suppliers'; any black box approach makes this impossible.

A similar argument occurred over the earliest use of the Cobb Douglas function to describe output for the whole economy, reducing capital and labour to single terms. As Fisher put it: "the suggestion is clear, however, that labour's share is not roughly constant because the diverse technical relationships of modern economies are truly representable by an aggregate Cobb-Douglas but rather that such relationships appear to be representable by an aggregate Cobb-Douglas because labour's share happens to be roughly constant" (Fisher 1971 in Robinson 1971 in Wong 1973 p.324). Wong replies: "of course, the way is open for the rebuttal that the Cobb-Douglas is not really to explain but to describe the empirical relationships found in economies" (ibid p.324).

There is a common thread that has found its way into ABM modelling: if the macro result has not been produced by interaction alone, it must be a *post hoc* arbitrary assumption hiding the true causal structure of the system.

It seems a little too easy to simply point to Feynman's quote and say 'it depends on what level a model wants to analyse', but following Friedman there is very little else to do if one accepts a model's assumptions alone cannot condemn it. At any rate, the approach to production used in the thesis model is the same as Fowler's: to stick with objective demand. As he puts it, "workers select their optimal bundle of goods without reference to the actual supply of the good" (Fowler 2011 p.10). This avoids the need to solve rationing of limited goods. It also sends "clear signals to firms about the actual level of demand for their goods at the current price" (ibid). As section 6.3 explains, it is entirely possible to do this without violating the zero stock limit. Section 5.2.3 outlines that, while it is possible to create a utility structure that *can* deal with subjective demand (where stock may not be available), it is not especially useful to do so for the current model goals.

3.6 Summary

How model mapping is theorised has a profound effect on the kind of approach to building models deemed ‘valid’. This chapter has examined Friedman’s argument in some depth and compared it to the current way in which agent modelling is done.

The argument about ‘levels’ suggests that, if many economists were left to investigate gas theory, they would conclude that all atoms must be temperature and pressure maximisers. On the the other hand, many agent modellers would insist temperature and pressure were of no use as concepts, and the only theoretically important feature is how atoms interact. It has been argued that there are valid reasons for treating components of a system as small ‘divisions’ of the larger system (as in microsimulation). GE demonstrates that under some limited circumstances, treating a collection of agents as a single representative agent is a useful modelling trick, but persuasive arguments have been presented that suggest doing so may make potentially simple behavioural models more complex than they need to be.

The result is that giving simple utility and production to disaggregated agents is a potentially useful thing to try. The next chapter presents the model framework developed for the thesis for doing this, and is followed by the results of that framework.

Chapter 4

Model framework

4.1 Introduction

This chapter has two aims:

- Explain the logical structure of the model framework used in thesis and how it is presented in the results chapters;
- Cover in more depth some of the specific issues arising from the logical journey of model development (section 4.7).

The model framework uses the economic and spatial ideas from chapter 2 and provides a structure for agents using them to interact. It is built using object-oriented programming, as discussed in section 3.2. Table 4.1 gives a key for the diagrams used to describe the logic of the model; these diagrams use the code relations between objects to do this. (Appendix A describes the code more traditionally, using UML, and explains the code in more detail.) The rationale for this approach is to present the object structure and the logic of the algorithms in one diagram as clearly as possible, but they should be used in conjunction with the text description. As mentioned in the introduction, an OOP coding norm used here is to capitalise objects, when they are being discussed as objects. For instance, 'Good' is a particular object, distinct from goods generally.

4.1.1 Chapter overview

The chapter covers the following topics:

- An overview of the model framework;
- How the theorisation of space costs in section 4.2 is implemented in the framework (section 4.3);
- The (non-Agent) structural elements of the framework are explained (section 4.2);

- The functioning of the agents themselves is then described (section 4.6);
- The 'framework discussion' section picks up some essential elements of the model framework that make more sense to explain after the details of the framework itself have been laid out. These include key ideas on the way information is theorised and how economies of scale is structured (section 4.7).

4.2 Framework overview

This section gives an overview of the framework and what variables it uses. To start with, the following types of space are used in different model set-ups:

- Spaceless (a dimensionless point);
- A two-region space consisting of two discrete points separated by a defined distance;
- A continuous, one-dimensional line (referred to as R^1);
- A continuous, two-dimensional plane (referred to as R^2)

From the point of view of the thesis models, the differences between them boils down to how agents' location choices are restricted. All other model features remain the same. (This also means that discrete spaces can be 'simulated' by simply restricting agents to binary location choices.) Agents make their distance calculations based on Euclidean distance from other agents' locations.

The size of both R^1 and R^2 is kept to a value of 1 in all models except those using the density cost in section 5.4, where it is set to 2 to allow People a more variable space to disperse into. In R^2, width and height are always kept the same. Some very basic differences exist between R^1 and R^2 when costs are changed. Most simply, whereas in R^1, a doubling of distance leads to a doubling of space covered, in R^2, the area covered is quadrupled ($\pi(2r)^2$ is $\pi 4r^2$). If, for example, potential economic agents are evenly spread across space, there is thus a four-fold difference in the number of new interactions the doubling enables.

This space is used by two types of agent: People and Firms. The key list of features used by each of these is as follows:

- **People:**
  - Location within one of the space types;
  - Time to spend: one 'day' per model iteration. For the models in chapter 5, People optimise this one day on distance ('commuting') and giving labour to employers in exchange for a wage. Or to frame it another way (see section 4.3.3 below), People 'pay for' their wage and their commute cost with this one day. For the transmission

belt model in chapter 6, People's one unit of time is exchanged directly for goods (which can include the distance cost for goods);

  - A 'love of variety' / CES utility function as described in section 2.2.3 with ρ controlling their preference for a mix of goods;
  - A density cost; the background for this was given in section 2.7 and section 4.4 below outlines its application in the model framework. This is used in the models in section 5.4 and the final model set-up in section 6.4.2. Otherwise, it is set to zero, thus effectively 'turned off'.

- **Firms:**
  - Location within one of the spacetypes;
  - A wage they offer to People in exchange for their time; is fixed to 1 unless otherwise stated. If a Person incurs a 'commute cost', the time it takes them to commute is accounted for in the wage they get (see below for how this is formulated). As mentioned, the transmission belt model functions differently: People buy goods directly with time. The goods are thus their direct wage (see section 6.3 for the full explanation of this mechanism);
  - A price for the good they sell, not accounting for distance; so if a Person was at 0 distance from a Firm, this is what they would pay. These are set to 1 unless otherwise stated; the distance cost for moving this good is identical for all Firms in any given model run, unless otherwise stated;
  - A production function, used in the transmission belt models, that turns People's input into a single-good output quantity using the smooth economies of scale function given in section 2.2.5;
  - A level of stock, again used in the transmission belt models.

The 'partial' models in chapter 5 keep most variables fixed, selecting only specific ones to vary, depending on the analysis. The details of these are given in the relevant results sections. For the more 'general' model in chapter 6 agents have fixed locations (except for two cases, for reasons explained in that chapter). The main variable for the transmission belt models is good price: Firms are able to change this in order to target a stock level. In the transmission belt models, People buy goods directly with time, so there is only one distance cost to be incurred. Again, unless otherwise stated, all other parameters are fixed to 1.

The rest of this chapter deals with the specifics of how these agent features are implemented in the thesis model framework. The next section explains the implementation of space costs, how they enter into the wage and good costs and how these can be reframed in order to create the idea of one spatial and one non-spatial cost. This idea is then used in the final transmission belt model in chapter 6.

4.3 Implementation of space costs

4.3.1 Cost of moving goods

A simple mathemathical way to describe the cost of moving goods across space is as follows. Let P_g be the full cost of one unit of good g, p_g the base per unit cost, before any space-costs; c the cost to move one unit of this good over one unit of space, and d the distance moved. Then -

$$P_g = p_g + (c \cdot d) \tag{4.1}$$

On a given model iteration, this will be the cost faced by agents wanting to buy the goods. In line with the simple approach to time, no distance will be too large for goods to travel during one iteration, but the total cost, including cd, may be too high. c represents a catch-all variable for the generic costs of moving goods, whether due to weight, bulk or some other factor. In equation (4.1), the addition or subtraction of weight would be reflected in c, and changes in value in p. This is also where 'value density' (see section 2.6.1) can be defined. The ratio of product value to physical size/weight bcomes p/c, with c standing in as a proxy for generic per-unit shipping cost.

4.3.2 Cost of moving people

There is a hard limit to people's time: a flow of 'one day per day' must be spent. Because of this, one way to model distance for People in the model is to build up from the idea of one 'day' and the distance People can get in that day. This approach is as follows:

- For a given distance d, the time is takes to cover it is t_d. So, generically, $t_d = f(d, tech)$, where $tech$ is some technology that has an impact on the time taken. For many of the models, time to cover a given distance will be a function of the distance alone. However, the generic factor $tech$ here indicates that the amount of time to cover a distance may change. This is used as a way to consider the impact of varying the cost of commuting.

- $tech$'s role in this function is separate from its cost. For example, if $t_d = d/tech$, the time taken to travel d goes down as the level of $tech$ increases. It says nothing about $tech$'s cost. From the point of view of agents in the thesis framework, this cost is ignored: they pay only for the distance cost itself, not any further transport overheads. Relatedly, in the thesis model, agents do *not* need to 'buy' migration. That is, if an agent moves their permanent location, this is free. It costs them to move goods or move themselves to a workplace.

4.3.3 The cost of moving people as a parallel to moving goods

Section 4.6.1 outlines in detail one approach to how People's time can determine their wage. Simply, the amount of time remaining to contribute to an employer will be less if a Person is further away. If an employer pays a given wage for a unit of time contributed, at $d = 0$, the Person gets a full unit of wage. At the one day limit, however, they will get no wage at all. (This is, of course, another unrealistic approach: see below for further discussion on this.) The model can be set up so that People select employers on this basis. However, as section 5.4.3 will later explain, it is useful to understand the parallel between good movement and People movement when it comes to describing their effect on spatial outcomes.

In the situation just described, if a Firm was paying a wage of $w = 1$ and $tech = 1$, the Person's wage will just be $w = t - d$, with the imposed condition that if d is over the hard time limit, the Person cannot consider that Firm as an employer. This is easy enough to code but not a very satisfactory analytic approach. It also masks the difference between the non-spatial and spatial elements by subsuming them in one function.

It can be reframed in a way that parallels the description of good costs used in section 2.6.1 above. To do this, the cost element needs to be separated from the budget element, as follows. A Firm is not offering a wage: rather, they are selling a factor in this case, money - in precisely the same way they sell goods at a given price. (This will also come in useful later when agents are exchanging factors other than money.) The amount supplied is just $g = f/p_g$, where g is the supplied factor, f is the payment and p_g the price of g.

In these terms, People are just buying money with time. The price a Firm will charge is as just stated: p_g for one unit of money. What about the space cost element? The uniqueness of moving People, it has been claimed, is the hard limit to time, beyond which no wage can be bought. The good-cost function above would not do this job: regardless of how high the price goes, some small quantity is always demanded. The only way such a hard limit could work is if the cost of money becomes infinite at the hard limit. This can be done as follows, using the variable definitions already given above. At the point where $d/tech = 1$, p_g is divided by zero and the cost becomes infinite:

$$P_g = \frac{p_g}{1(d/tech)} \tag{4.2}$$

A Person with a time-budget of $t = 1$ must optimise given these costs. Assuming they are only 'buying money' from one Firm (an assumption changed later) , they will want to spend their whole budget on the single cheapest supply. The amount they can buy in this situation is just their budget divided by the full price described in (4.2). Assume t and $tech$ equal one, then the amount they can buy is just their time divided by the price:

$$g = \frac{t}{p_g/(1-d)} = \frac{(1-d)t}{p_g} \tag{4.3}$$

This brings the description full circle: if the 'price of money' is one, the amount the Person gets is proportional to what is left after distance has been accounted for.

What was the point of this long roundabout trip? Three things: firstly, to highlight that the more algorithmic description of People getting their wage can be reframed into a simple optimisation problem with a price and a budget. Secondly, to illustrate that different cost functions can be made to achieve particular spatial outcomes. Thirdly, it helps in reframing all movement costs as one combined cost in chapter 6, where People are paid directly in goods using the less hard-cutoff cost equation (4.1) above.

4.4 Density cost

As section 2.7 explained, the model framework uses a density cost as a proxy for both land costs and generic congestion, covering all 'costs of proximity'. The density cost is imposed on agents in proportion to the number and proximity of others within a set radius from that agent's current location. This section outlines how the density cost is calculated and applied as well as providing a rationale for this approach. This use of density cost is very simple, but would be prohibitively difficult in anything but an ABM setting. It requires each agent to make their calculation individually while taking account of others' location.

The density cost uses three parameters. First, a radius around each Person within which density costs are incurred. Second, a variable capturing the density of People within that radius, returning a normalised value between zero and one. The resulting value would be one if all agents are on exactly the same spot as 'me' (this never happens in practice; see below), zero if none are within the density cost radius, and between zero and one otherwise. It is worked out for each Person as follows:

- Using the density cost radius, normalise each other Person's distance from 'me' to between one and zero, where one is the same location as 'me', and zero is on or beyond the radius.
- Sum these normalised distances, then divide by the total number of People in the model run.
- Thirdly, this normalised density cost variable is then multiplied by a factor that determines its overall effect.

This final density cost value is imposed directly by subtracting it from People's wage. Note that the density cost is independent of the wage, so it becomes proportionately smaller if wages are higher.

The radius is kept to $1/32$ of the width of model space, which itself is normalised to 1. The radius is arbitrary, as it can be calibrated to agent number and density cost to produce the morphological outcomes: it is chosen to allow a clear graphical view of outcomes. For most density-cost-based models (section 5.4) the multiplication factor is set to 10 as this allows a range of spatial outcomes where People can find trade-offs between density cost and distance. By increasing the multiplication factor, section 5.4.7 shows what happens when agents are pushed outside of this trade-off range: utility equilibria start breaking down.

4.5 Elements of the model

4.5.1 Key table

Table 4.1 lists a series of graphical components used in figures below to help illustrate how the model framework is put together and how it functions.

Element	**Description**
Actor object	People or firms
Non-actor object	Any other object used in the framework
Variable	Grey boxes indicate variables.
Object array	An array, or set, of objects, such as the set of all firms
Reference to object	For example, a Person may want to keep a reference to their preferred employer
interface	A way to provide shared methods to different objects; not important for understanding model logic, but worth noting where present
method	An object method for achieving a particular model aim
logic	Can be one of three things: a logical decision, a calculation Or an operation on a set (array)

Table 4.1: Key for model diagrams

4.5.2 Decision structures and model timeline

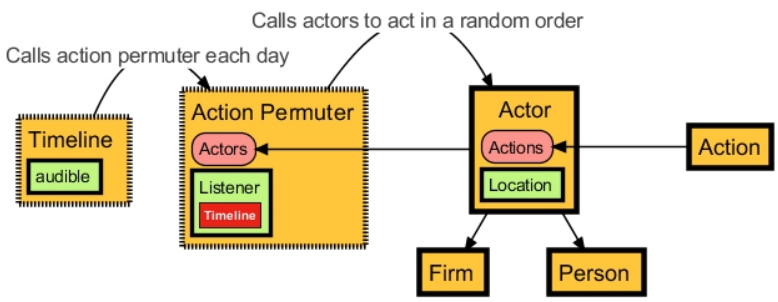


Figure 4.1: Timeline sequence

The timing structure shown in figure 4.1 is very simple: as each day passes on the 'Timeline', an 'Action Permuter' calls Actors in a given order.[1] The issues around exact ordering are

[1]The Timeline is part of an Observer framework used by other model elements like visualisation and data writing; again, see the code appendix for more details.

discussed in section 4.7.4. When an Actor is called, they decide what action to take. Actors' decisions can use any of the following:

- Instant: act on information collected 'now'
- Feedback: Act, wait, respond

As discussed in section 4.7.6, agents take decisions either instantaneously or using feedback from their actions over one or more timesteps. In instant decisions, when agents are called to act, they immediately gather the information they need and optimise based on that. For multi-timestep decisions, agents try an action and then check later whether the outcome was desirable or not before taking their next action. No attempt is made to simulate synchronous decision-making. At the point where an agent is deciding, all others are currently static.

Outside of the Timeline-driven decision structure, there is a third type of 'reflex' decision that Actors can take. Here, an Actor is coded to react to requests from other Actors, rather than the Timeline - for instance when a purchase takes place. These reflex actions are mostly just the exchange of stock or money, but can also involve adjusting prices in a 'granular' way, rather than the Actor waiting for their decision turn. For instance, this is used in the model in section 6.3 where granular price-changes are used to equilibriate stock levels.

4.5.3 Two agent data sources

Actor decisions will be based on one or both of the following sources of information:

- Internal Actor data, changed by other agents since last a decision was taken. (Example: Firm net income.)
- Externally collected data (Example: list of good prices from surrounding Firms.)

Externally collected data can come only from other Actors. This is acquired by searching for a list of all appropriate Actors and forming a set from it.

These three elements - Actor setup, decision structure and Actor data - form the foundation for the model framework. Because all agents face a static world at the point where they decide, their actions can be logically described separately: section 4.6 explains how they are put together.

4.6 Agent setup

4.6.1 How People act

People have 'one day per day' to spend and their fundamental question is, 'what is the maximum utility I can get for my time?' This time can be spent on moving through space or put into Firms in exchange for either goods or money. The description here deals with the most complex of these: People working for money and then spending it on goods.

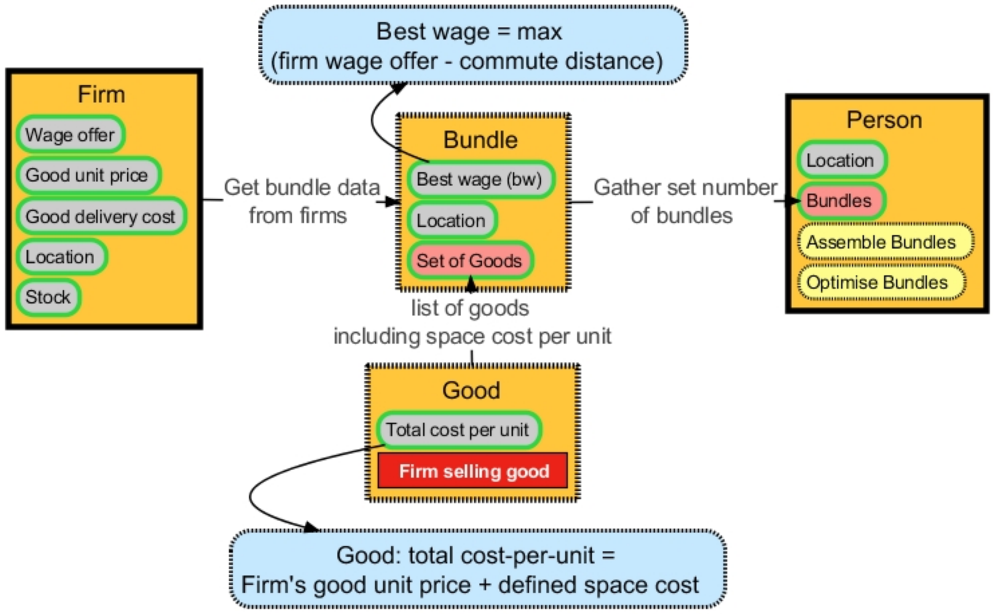


Figure 4.2: Elements for the Person action

Bundles: overview

The Bundle is central to the way People optimise their utility, tying the previous framework sections together. When it is each Person's turn to act, they look to find better economic options in different locations. They do not attempt to exhaustively search all of model space. Instead, they assemble a set number of 'Bundles', each centred at randomly chosen points. They also include a Bundle for their current location to check whether staying put is the best option.

Each of these Bundles contains all the economic information needed to calculate a utility level for that location: the cost of all available goods, including delivery cost; the cost of accessing a wage, including the commute; and the density cost that would be incurred if residing there. Once utility is worked out for each of these Bundles, the best is selected and, if this is not on the current spot, the Person moves there.

Figure 4.3 illustrates this process for the first two iterations of a model run. The particular set-up is that from results section 5.4 as this includes the density cost and provides a clear view of where the agents are located. To help illustrate the Bundle process, ten Firms (indicated with crosses) are located randomly and People use a 'love of variety' utility function, buying a mix of goods (see section 5.4.6 for more details on this function). The figures pick out a typical Person (marked as a triangle). That Person's range of Bundle sample points are shown, with small, black squares being lowest utility value and large, white squares the highest. A thick black circle marks the best utility Bundle; lines indicate where that best Bundle will acquire goods, with thickness of lines in proportion to budget spend on each. Sub-figure 4.3a shows a Person's initial best Bundle; sub-figure 4.3b illustrates their move to this Bundle and sampling again to find the next. The process of assembling Bundles is as follows:

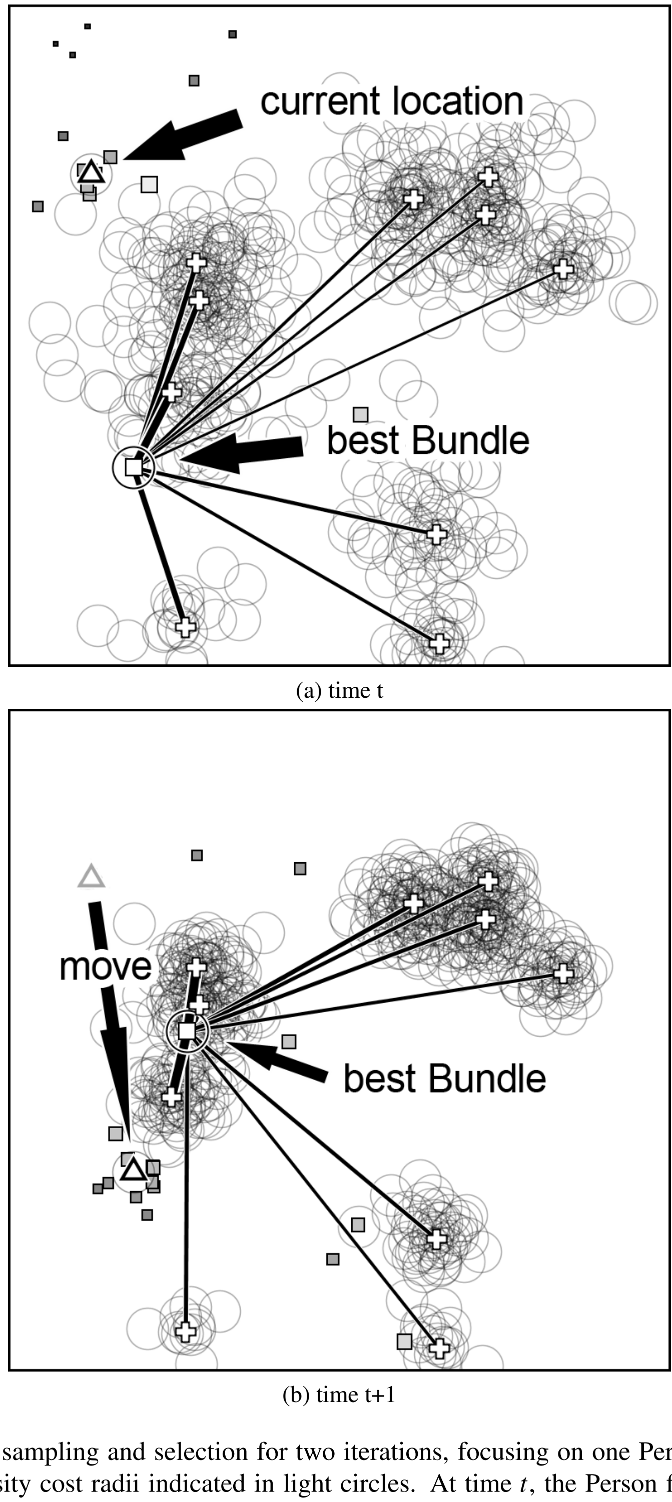
- Create a set number of empty Bundles, each with randomised centroid coordinates in model space.
- For each Bundle, relative to its centre, work out:
  1. What is the optimum wage the Person could get from that location? (What is the most money they can buy with their time, accounting for the base wage, commute distance and density cost?)
  2. What is the optimum utility the Person could get, given the optimum wage, the total price of goods (accounting for distance) and their utility function?
- Select the bundle with the best utility: move to that location, take wage and buy goods. (Note: People do *not* need to 'buy' this movement between points. That is, when moving location, this is free. It costs only to move goods or move themselves to a workplace.)

As mentioned, each Bundle's found utility level is 'objectively' true at the moment it is calculated. However, a Person's set of Bundles is only a very small random sampling of the space around them: a subset of random locations within a given small radius of the Person's current location and another set randomly chosen from the whole space. This allows a balance between finding small extra incremental benefits nearby with having a good chance of finding better options wherever they may be. However, it is almost certain that a small sample of locations will not find the objectively best spot in one iteration. For the method to be successful, this sampling procedure needs to be repeated over many timesteps until no Person is able to unilaterally improve their utility level (a Nash equilibrium). The Bundle sampling process does converge on a stable outcome: People gradually cease moving location as all available options for better utility are exhausted.

Selecting the best wage

The 'transmission belt' model in chapter 6 allows People to spent their time on a range of Firms through a CES function. This provides a way to avoid **Bertrand competition**: People choosing whichever employer is the cheapest, thus usually the nearest. Section 4.7.8 discusses this point further. The rest of this section outlines how, in the rest of the model runs, People choose a Firm to give their time to.

Each Bundle has one unique 'best wage' offer. Selecting this is a two-step process: first, gather a set of all Firms within commute distance. Second, select the Firm offering the best wage, accounting for commute distance from the Bundle's location centre. This best wage can be described as follows. Let $F = \{f_1, f_2, ..., f_n\}$ be the set of Firm objects. Each Firm has a wage offer, which is 'the amount paid for one day's input'. For each $f \in F$, W_f is that wage



(a) time t

(b) time t+1

Figure 4.3: Bundle sampling and selection for two iterations, focusing on one Person (white triangle). Other People's density cost radii indicated in light circles. At time *t*, the Person finds a 'best Bundle' having compared a set of Bundles including density cost and good prices through a CES function.

offer. To describe space cost, the distance in *time* can be used: for each $f \in F$, t_{fa} is the time taken to travel between f and a.

Let l_{af} be the labour available to contribute to f by Person a. The amount of labour they have available for contributing is just their own time (one day, $t_a = 1$) minus commute time. Here, a 'commute' is taken to be twice t_{af}. The upper distance limit on whether a Firm enters into a's considerations has to be less than a's one available day in total - half a day there, half back - otherwise there is no time left at all to contribute to production. So f can only be a maximum of half a day away. Given all this -

$$l_{af} = t_a - 2 \cdot t_{af},\ 0 < l_a < 1$$

If W_f is f's wage for a full day, and a has l_{af} to contribute (a fraction of their full day), this means the wage they will get is just $W_f \cdot l_{af}$. The best wage for Person p can now be defined as:

$$bw_p = max\ \{W_f \cdot l_{af} \mid W_f \in F,\ t_{af} < 0.5\} \tag{4.4}$$

So, the best wage for Person a is the maximum value from the following set: all wage offers proportionate to Person a's labour input from those Firms in F who are less than half a day away.

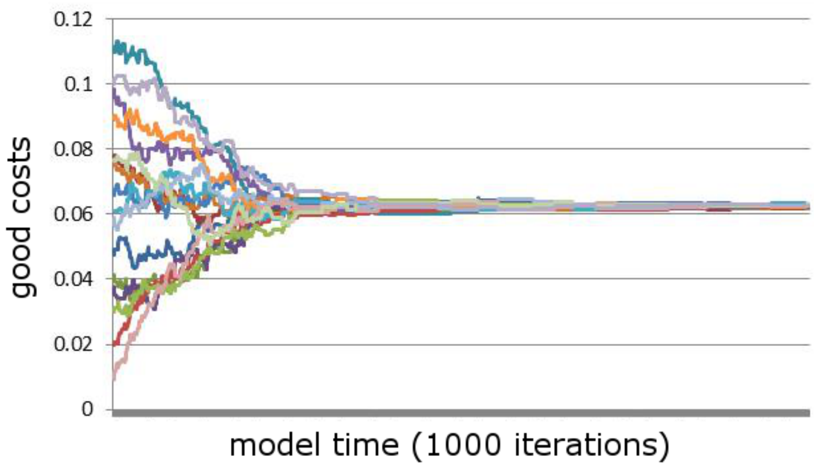
Optimising bundles

Once a Bundle is assembled, it can be passed to the Person's optimisation routine.[2] This produces a utility level from one of two possible utility-maximising methods, and the result is stored in the Bundle. Finally, the Bundle with the maximum utility level is selected: the wage offer is accepted and the optimal quantity of goods is bought.

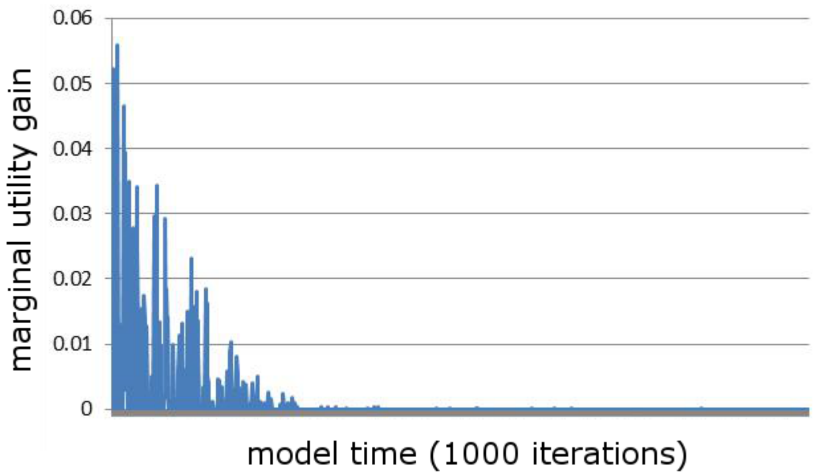
The logical goal is simple: for each Bundle, find out how to spend the best wage on the set of Goods as optimally as possible (without actually spending it) then choose the optimal Bundle. Figure 4.5 is an overview of the process. Assuming the Person has the set of Bundles they will optimise, the sequence is as follows.

- Optimise each Bundle. This happens in one of two ways:
  - (1) A 'hill climber' optimiser uses the following method -
    - The hill climber iterates for a set number
    - The iteration is initialised by randomly slicing the budget (the best wage), one slice for each Good in the bundle, before beginning the iterations, so that each has a starting guess price. The iterations are:
    - 'Jiggle' the slices, changing their proportionate size slightly.

[2] As appendix A explains, there is only one actual, static Optimiser class, but in terms of a 'functional mapping' it can be considered internal to each Person.



(a) 16 initially randomised good costs converge on utility-maximising value



(b) Marginal utility gain (change in utility between hill-climber iterations)

Figure 4.4: Utility-maximising hill-climbing algorithm, optimising a CES function with 16 goods over a thousand iterations, showing both randomised good costs converging on the utility-maximising level and the marginal utility gain for each step over 1000 iterations.

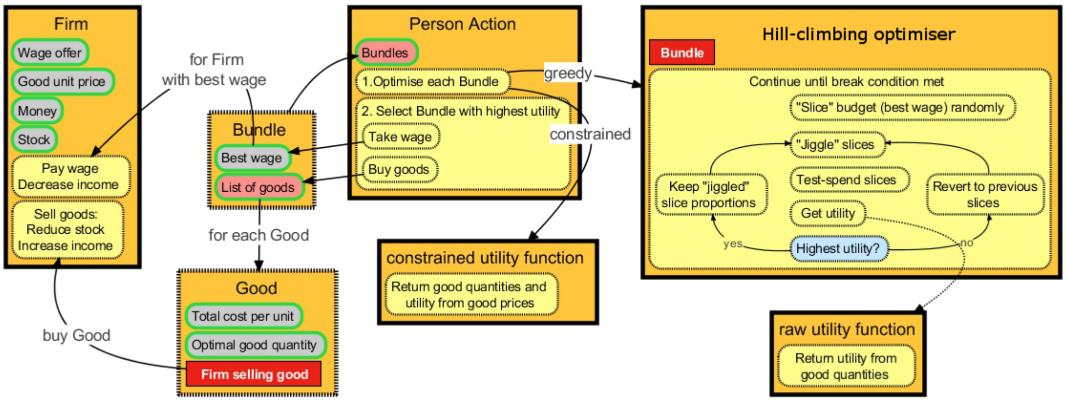


Figure 4.5: Optimisation of Bundles in a Person's action

  * 'Test-spend' those slices on each Good: find out how much could be bought with the current budget slices using the per-unit price for the Good.
  * Get a utility level for these quantities of goods. The Utility function is logically separate: the Optimiser deals only with this process of random slicing and adjusting.
  * If the utility level found is the highest so far, keep the current proportion of slices and return to the 'jiggle' point. Else, revert to the previous slice proportions and try again. This has the effect of ratchetting the budget slices (and thus the spend per Good) towards an optimum.

- (2) 'Constrained' optimiser: a Bundle is passed directly to a code version of the constrained utility optima described in section 2.2.4.
  * The Good prices and budget are used to directly calculate optimal good amounts. The utility level is then 'read back' from those amounts.

• Once each Bundle has an optimum utility, select the Bundle with the highest utility, then:

- Take the best wage: this instructs the Firm to pay the Person.
- Buy goods in the quantities found optimal by this Bundle.

Figure 4.4 graphs data from the hill-climbing algorithm in action. A bundle containing sixteen goods of equal cost is optimised for a CES function over a thousand iterations. (The optimum for a CES function with equal-value goods is already known from section 2.2.3: an even mix.) In this example, the optimum is found quite slowly: this helps to illustrate the process. The first graph shows initially random budget split guesses converging on the optimum mix of 0.06 for each good. The second shows the marginal utility gain from each iteration; most of the gains happen very quickly, after which almost no extra benefit accrues (though there is a very small amount of further convergence.)

Overview of People's utility

The technical detail of optimising utility has just been covered. There are three further points worth covering briefly.

People have utility functions based on those described in section 2.2. In any one model run, all agents have identical tastes: they all use the same utility function. One useful assumption from using identical tastes is that, in one-good models, only the absolute quantity of that good matters. This is because the same quantity of good will always return the same utility for each Person, regardless of the form of utility. For example, if all People have an identical one-good utility function that happens to have increasing returns, the increasing returns will just produce the same level of utility for everyone, given the same good input. In contrast, if tastes were not identical, one Person could gain more utility from the same quantity of good as another. Many of the simplest model tests use the assumption that a single good quantity can stand in for utility.

In all situations where more than one good is available, People use utility functions. The form of utility used must also imply the production structure, though not the actual production function itself. For instance, a two-good utility function means a model where only two goods are made, though many Firms may compete to sell each good. An n-good utility scenario, using the CES function, will have each Firm making one unique good, as they do in monopolistic competition models.

Actors using this method are rational in only the most basic way: they are able to judge when one Bundle is better than another. This rationality reflects basic micro-economic preference assumptions. Bundles of goods are 'complete': any two bundles can be compared. They are also transitive: as Varian puts it, "if the consumer thinks that X is at least as good as Y and that Y is at least as good as Z, then the consumer thinks that X is at least as good as Z" (Varian 2006 p.35).

4.6.2 How Firms act

In many of the partial models, Firms take no actions at all: they serve as geographical points of good supply only. Where they do take actions, there are two types. In chapter 6 the decision process is 'granular': it is explained in-depth in that chapter as it ties very closely to the explanation of the stock targeting mechanism used.

All other Firm decisions come into the DMR category (see section 4.7.7 for more details). The method used is as follows. A Firm's decision process is based on feedback between timesteps. A Firm takes an action, waits, and then responds depending on how its own internal data has changed (see figure 4.6). Because of this, the structure of Firm decision-making is quite a lot easier to describe, since Firms do not collect any data externally themselves.

At time $t-1$, a Firm perturbs its current geographical position[3], altering it slightly in a

[3] The model framework explored perturbing location, good unit cost or wage offer; only location is considered

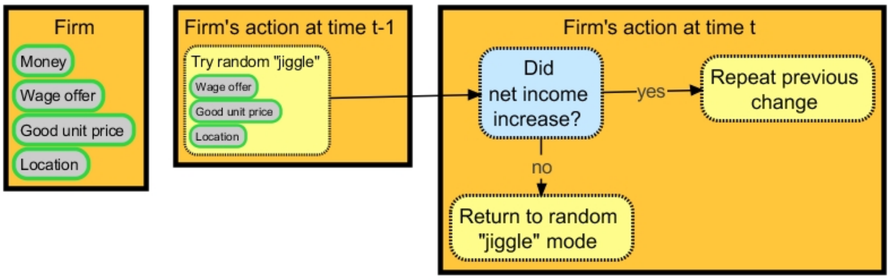


Figure 4.6: A Firm's feedback mechanism: "jiggle" a random value, wait to see effect.

random direction. It remembers these choices. At time t the Firm asks, 'did my net income between timesteps increase?' If yes, it repeats the previous action (hence needing to remember what it was), with the possible addition of increasing the size of the adjustment. If not, the Firm returns to its original location and re-initialises the random search.

The result is that, through blind search, Firms are constantly searching for ways to improve their net income. Note that this method disconnects them even further from any analytic approach to optimising. Where People do use a utility function, the Firm does not directly consider its own input or output requirements.

4.7 Framework discussion

4.7.1 Interdependence of decision structures

It is not easy to meet the OOP ideal of fully encapsulated, autonomous agents in an economic modelling scenario. Indeed, it could be argued that in an economic modelling situation, it is not desirable. Much of the agent literature seems to have absorbed a modelling paradigm required for 'instrumental' agents able to carry out real-world tasks. For example, Wooldridge's 'Introduction to Multi-agent Systems' mentioned in chapter 3 is imbued with this sense of the need for autonomous agents, where "each agent is assumed to have at least one thread of control" (Wooldridge 2009).

Putting aside reasons of computational expense, programming fully autonomous economic agents in this way is a massive task. This becomes particularly apparent when trying to construct the timing framework. The ideal of fully independent objects mediating their interaction with a separately threaded environment is a stark contrast to the reality of the interdependence of different agent decisions. This section looks at a number of examples of this that arose during model development, and includes references to results in chapter 5 that examine these problems more closely.

Any decision which will, in the process of optimising, interact with other agents' decisions can be considered 'online'. In contrast, an agent able to optimise a utility function without

here as the others faced problems. This is discussed in more detail in section 6.2.

affecting the basis for its optimisation is 'offline'. The terminology is Holland's: Mitchell describes its use in the 'two-armed bandit' game, where the player tries to discover each arm's mean payoff and variance:

> "... the goal is not merely to guess which arm has a higher payoff rate, but to maximise payoff in the course of gaining information through samples to the two arms." (Mitchell 1998 p.119 discussing Holland 1975)

The difference between off- and online decision-making is partially blurred with many agents, and depends on timing structures. For the two main types of decision described in section 4.5.2, 'instant' is partly offline, since Actors face a static world and can select the objectively best Bundle. It is also partly online, in that the global optimum is affected by the number of random Bundle locations they can test at any one time; the movement towards a global optimum is thus not instantaneous, and Actor choices interact. This can lead to path-dependence for single Actors, but the global outcomes are statistically identical at the aggregate level. The 'feedback' decision process actually relies on being 'online' since the Actor requires the 'world' to respond to its action to know what to try next. This is discussed in-depth in section 4.7.7.

4.7.2 Communicating costs

Agent models proceed over many iterations. In a spatial economic model involving goods, this presents two immediate problems that need solving. Firstly, in such a world, what moves between economic linkages each timestep: actual resources, where agents actually 'exchange' some excludable, rivalrous quantity? A common way to avoid this is to presume goods perish before the next timestep (e.g. Tesfatsion's 'hash and beans' model; Tesfatsion 2006 p.848). In the thesis model, where good stock is not fixed, the presumption of both persistent and perishable stock are tested to see the effect on the link between production and welfare (see section 6.3).

Things become more complicated with money, however. Actors need some method for communicating costs to others. Again, for the partial models in chapter 5, it is easy enough to assume that, while agents need to use prices, that does not always need to imply money is exchanged. This is a parallel assumption to perishable stock: Firms set prices, and wish to maximise income on a given iteration, but the money they receive 'perishes' immediately. This avoids the huge complications of making money flow function: any simple attempt to implement money exchange as 'actual resources' immediately finds that moneyflow dynamics themselves loom over other model problems.

Figure 4.7 illustrates the basic problem for a set of zero-intelligence agents running for 200,000 iterations. One hundred agents are given an initial one hundred in currency. They then select another agent to give a unit of currency to, if they have any. The consistently stable

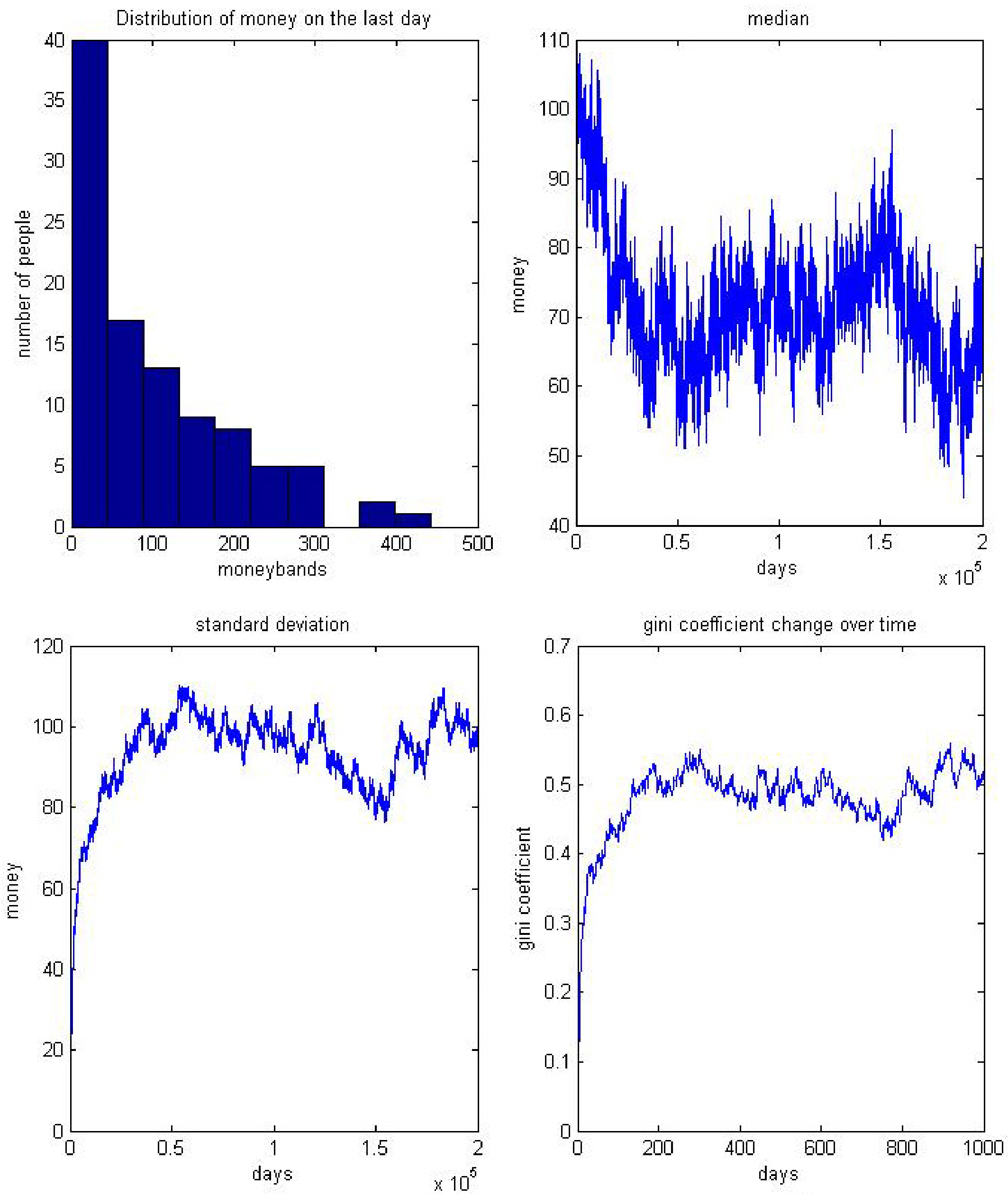


Figure 4.7: Zero-intelligence agents exchange finite money, beginning with equal amounts. Top left figure shows the distribution of money on the last model day: a few very wealthy agents emerge while most are in the lowest $0-50$ band.

result is a bounded random walk resulting in a large inequality[4]. Initial tests using the more economically sophisticated agents of the thesis model suggested that such dynamics could play a large role in outcome.

It is a large assumption to exclude further analysis of this dynamic from the model, since the impact of money flow may be key to interpreting some types of model outcomes. Yet in being strategic about simplifications, this is a useful one to exclude. This is not to say that modelling flows is either unworkable or unimportant: the concluding discussion returns to this point. There are many modelling approaches that could potentially avoid the fate of the zero-intelligence model above, and it is an open question whether more economically sophisticated agents would achieve this, or whether it would require some model implementation of a banking structure forcing agents to 'rent' money rather than keep it. (See Oeffner 2008 for an agent-based macro-economic model that attempts to incorporate this kind of 'monetary circuit'.)

GE (and other economic) models avoid the issue through, firstly, not dealing with flows, and secondly, through using 'price' of one good as the 'numeraire'. In the case of the core model, for example, one good is kept to linear production and that sector's worker number is fixed exogenously. This keeps its value constant, so it can work as a unit of comparison.

In an agent context, if some attempt is made to treat money or goods as 'actual resources' so they can function as a numeraire, problems immediately present themselves. If, for example, two Firms each make a different good, and one is deemed the numeraire, each must pay wages in it. How does the non-numeraire Firm do this? It must acquire the currency, and can only do this by exchanging what it produces - yet there is no guarantee the numeraire-producers want to exchange.

The model in section 6.3 avoids these issues by presuming agents are gifted a quantity of input to spend on each turn, which can be put into either production or movement. It is thus open to interpretation as their time, spent on space or labour. This circumvents the problem of 'real' flows and allows comparisons of change, as the basis for the model becomes agents asking, 'what quantity of utility can I get for my input?'

An overarching question from this is whether Howitt is right to argue (see section 3.3.1) that an agent approach "forces one to make explicit the mechanisms through which individual actions are coordinated, for better or worse" (Howitt 2006 pp.1068). Deciding *which* mechanisms are the focus for a particular model would appear to force quite the opposite: a Friedman-like need to acknowledge which dynamics are excluded or assumed away.

4.7.3 Timing structures

Section 3.4 argued that agent models should be designed with functional mapping in mind, not descriptive mapping. Nowhere is this more true than when constructing the timing inter-

[4]There is a comparison here to Epstein's take on Schelling from section 3.4.7. This model 'grows' an uneven distribution of wealth with a gini coefficient not too far from the reality of many countries. Does it explain it? There is not a large gap between claiming it does to making justifications for wealth inequality as a natural, unavoidable consequence of money flow.

relations of the model. When should an agent take an action? There are a number of approaches. First, is it required that all agents should take an action on every iteration? This would be the case if, for instance, an agent needs to look for employment and spend their wage on every turn. This is the approach used in this thesis: it allows for non-synchronised action, and so meets with the 'no clearing market' condition imposed by space (see section 3.3.1), but forces all agents to take at least one action per turn.

Contrast this to another common approach: to require a specific collective goal to be met on each iteration. This is the method used for market-clearing or auction-based systems, and by the core model's short-run wage equilibrium calculation. The collective goal is to reach a point where no agent is able to improve their welfare further without making someone else worse off: this is Pareto optimality[5]. In an analytic context, this collective goal is achieved through equilibrium, built on a mathematical argument about how agents get there (see section 3.3.1).

In an agent context, stopping conditions could do the same job. In an agent-based clearing market, for example, the iteration may not be over until all Firms have the labour they require, or all People have the wage they want. Note that, in that situation, Firms may end up without having employed all the labour they would prefer, or some People could find themselves unable to work but if either condition is met, the iteration terminates. This scenario would mean agents may act many times before one of those two stopping conditions is met: they are not limited to a single action each per turn.

4.7.4 Randomisation and timing of actions

The model uses a single static 'Randoms' class to provide consistent randomisation globally. A random seed provides both exact repeatability and the ability to change the seed for model runs where aggregate results require a full random spread of outcomes.

Randomisation of agents' timing can be used to avoid artifacts arising as a result of agent order (Gilbert 2007 pp.27) but it raises its own problems. Consider a common scenario where the order of all agents' decision points are randomly permuted on each iteration. The probability of getting any particular point in that order is the same for all agents. However, over *two* days, the time between any one agents' actions will vary. The most common outcome is an exact one day gap, since it has the highest possible number of chances to appear from both days' permutation: a one-day gap can fit into a two-day space beginning at any point on the first day, right up to the last slot. If n is the number of agents being permuted, there are the same number of one-day gaps as n. Each consecutive length of shorter gap has $n-1$ chances, until at both extremes (with a small but equal probability), two actions will either be consecutive (at the very end of one day and the beginning of another) or nearly two full days apart. The result is a

[5] Blaug has a good discussion of this: "the beauty of Pareto's definition of a welfare maximum was precisely that it defined the optimum as one which meets with unanimous approval because it does not involve conflicting welfare changes". This ruled out the possibility that one could "evaluate changes in welfare that do make some people better off but also make other people worse off" (Blaug 1997 p.573-4).

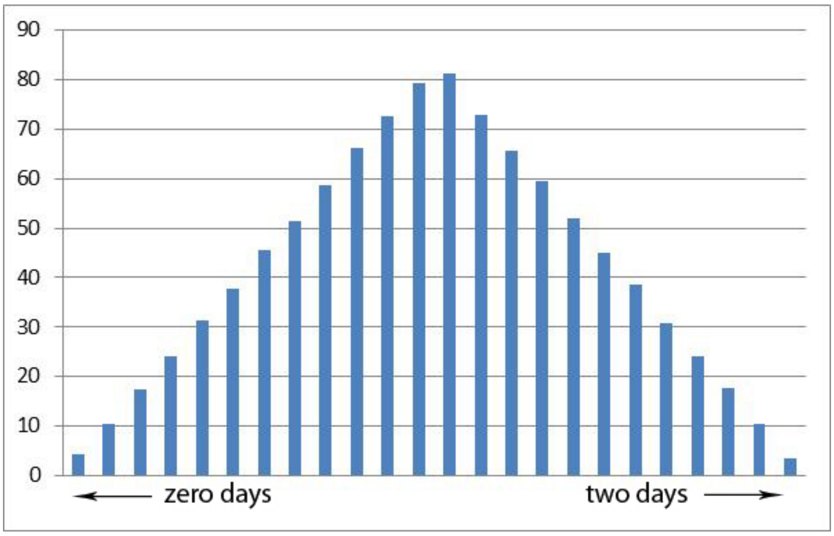


Figure 4.8: Histogram: ten thousand agents over a thousand days, 24 bins, showing how many times on average agents have an action gap between zero and two days, when the order of actions is permuted randomly on each turn.

pyramidal distribution. The histogram in figure 4.8 confirms this from a model run for actions of ten thousand agents with daily randomly permuted actions, over a thousand days.

Using a random permutation, then, the time between actions will never be consistent. However, over many iterations, this averages out, so why should this be a problem? It is important to be aware that the above dynamic exists, as it can explain errors in agent decisions that would otherwise remain mysterious. Any decision process based on feedback between t and $t+1$ may be affected. One of these situations is this: a scenario with both Firms and People, where Firms make decisions based on responding to a change in net income between iterations. Any Firm presuming the gap to be a full day will be using faulty data. There are (at least) three ways around this. Firstly, the actions of People and Firms can be put into two 'bins' and permuted separately. This guarantees that all People will have made their decisions before Firms come to act. Second, Firms can be allowed to know the length of time between decision points and alter their decision-making (for example, by correcting the rate of change of net income accordingly.) Third, some other method of permuting that damps the variability between timesteps could be used, such as a random walk. This thesis uses the first of these.

4.7.5 Economies of scale and avoiding the need for strategy

As section 2.2.5 discussed, considered statically, economies of scale are entirely straightforward: an equation describes the relationship between the level of input and the given output. But if many agents are deciding whether to input time into a production function, how do they decide? The best choice will in part depend on the levels of output available from Firms but this is determined by what all other agents have decided. This leads to a catch-22: should an agent wait to see what others do? If everyone does this, how can production get going at all?

Once they have made a decision, how is the output itself determined or more specifically, when, given that both input and output are flows in a multi-timestep situation?

If agents are deciding simultaneously, this can be framed as a game-theoretic problem, solveable for small numbers of agents, but increasingly problematic with larger numbers, and impossible if agents are not synchronised. The choice agents must make in this situation shares something of the dynamic in the “El Farol Bar problem” (Arthur 1994). Everyone makes a decision about going to the bar, but if more than a certain percentage turn up, it is so crowded that no-one is happy. All agents, however, must decide simultaneously. Edmonds describes it thus:

> “Any model of the problem that is shared by most of the agents is self-defeating. For if most agents predict that the bar will not be too crowded then they will all go and it will be too crowded, and vice versa.” (Edmonds 1999 p.12)

The economies of scale problem is the same dynamic in reverse: all agents gathering in one Firm will give a *higher* payoff, but there is no way to know who will choose where. The thesis model uses a ‘granular’ approach to avoid these pitfalls. While it has its own issues, it does manage to successfully power the ‘transmission belt’ between production and welfare. This method will be described first, followed by a more detailed discussion of its rationale. The following section (4.7.6) also discusses a more detailed ‘pseudo-code’ thought experiment examining the catch-22 problem.

In the thesis model, when agents decide where to contribute their input, they do so then and there, at the point of decision. They then *leave* their input with whoever received it until their next turn to act. It stays there for one whole ‘day’ (on average, given randomisation of agent order, see section 4.7.4), when it is removed by the same agent prior to re-taking the decision. At the same time, the quantity of output produced is *also* worked out. This is done from the producers rolling record of the quantity of contributed input from all agents. Let the producer’s total current contributed input be T_p. When an agent adds their input, t_a, this rolling record can be used to calculate the output from the agent’s contribution, g_a:

$$g_a = f(T_p) \cdot (t_a/T_p) \tag{4.5}$$

Here, $f(T_p)$ is the economies of scale function, giving a total output figure from all current input. The agent’s input divided by the total input, t_a/T_p, gives the proportional contribution of that agent. These two can then be multiplied to give g_a, the ‘granular’ quantity of extra output produced at that point, which goes into the producer’s stock.

What are the problems with this approach? Simply that agents are forced to make optimising decisions given static information: they are robbed of the ability to even attempt to make strategic decisions or develop expectations. Given what has just been said, however, it is the structure of the problem itself that robs them of strategy. The issue here, as Oeffner points out, is the difference between risk and ‘true uncertainty’ (Oeffner 2008 p.25-6). The problems described above represent true uncertainty: probabilistic outcomes are incalculable. As Oeffner points out, agents with 'rational expectations' in these situations will do no better than myopic ones.

Exactly the same issue arises for space-cost decisions in the presence of density costs in section 5.4, though the model in section 5.5.2 looks at a situation where risk is a component. (The ideas in that model are discussed below in section 4.7.7.)

4.7.6 Blind versus rational choice: an illustration of the problem with economies of scale

This section looks at a small concrete example of the interdependency of decision structures and model 'world'. The core model-style of increasing returns (2.12) is used, with the fixed labour requirement set to 3. This means that no production is possible at all unless three units of labour are input: an indivisibility. This is straightforward enough to impose analytically, but how might it work in an agent context?

Consider a scenario where agents choose an employer and goods in the same 'moment' (that is, during one action). Assuming that they are completely myopic, zero-intelligence agents, they are unable to make judgements about future timesteps. As a result, they will not work if no goods are available. There is no point: there is nothing to buy. Equally, if no work is available, they cannot *afford* to buy anything. This can lead to a catch-22: if Firms have no stock, they need labour to make more, but no-one will work if no stock is available. Compare these two possibile approaches agents might take in this situation, both of which have arisen as solutions for getting agents trading:

1. A 'rational choice': agents decide sequentially, looking at what Firms are producing in order to choose whether to work. If no goods are available to buy, there is no point in working.

2. A 'blind choice': At t, agents randomly choose a potential employer. If at $t+1$ this was successful, they repeat the same action.

The first of these, 'rational choice', assumes agents are able to rationally choose a workplace. However, each agent is only able to contribute one unit of labour; a Firm needs three to even begin producing. Less than three, and output is zero. The second, 'blind choice', is not completely blind, since the search requires agents to have the most fundamental attribute required for rationality: to be able to tell on the next turn whether their guess made them better or worse off.

The 'rational choice' result is that no work takes place and nothing is produced: production cannot get started because each single agent considering their options can never get above contributing one unit of time. The blind choice, however, does stand a chance of success. Actors select a random Firm to work for, and see if it improved things on the next day. If it did, they can stick with the same strategy. In the example of a fixed labour requirement of three, even with a set of agents consisting of only four People and n Firms, given enough time they would all randomly find themselves selecting the same Firm and allowing production to take place.

The smooth production function (2.14), while lacking a tidy definition for fixed and marginal labour, works in both 'blind' and 'rational' cases, since any amount of labour will always produce some quantity of output. It is this function that the thesis model uses, since it again offers a route for avoiding some of the complexities inherent in the model's interdependencies. Section 4.7.8 discusses this choice further.

4.7.7 'Data, meaning, response'

A final manifestation of uncertainty for Actors comes from the feedback process of DMR. This idea is taken from Jane Jacobs' final book, 'the Nature of Economies' (Jacobs 2001). In it, Jacobs looks for the common thread that ties supply and demand to other feedback mechanisms in nature. She settles on the importance of integration between 'the data, the meaning of the data and appropriate responses to the data.' (p.109) That framework functions, she argues, for systems as disparate as a termite mound, where supply and demand of soldier termites is precisely controlled through distributed pheromone signals, and price data, which can "carry meaningful information on imbalances of supply and demand and automatically trigger corrective responses" (ibid p.110).

Cybernetician Stafford Beer cites weather vanes as exemplars of perfect 'intrinsic control': "it is the wind that is capricious; the measurement does not make mistakes." (Beer 2002 p.214) In this case, DMR collapses into a single instantaneous feedback. As he points out, "a great deal depends on the speed of the response" (ibid). In situations with slower responses, meaning links data to response and this opens up the possibility that an Actors' interpretation may lead to *in*appropriate responses. An incorrect meaning assigned to data could come from several sources. One of these not used in this thesis is simply that the data may be context-specific. Spedding has an example from cattle markets that illustrates this point nicely:

> "Even when good information is ostensibly available, it may not be precise enough. For example, published statistics would suggest that there is a fortune to be made by transporting top-quality Friesian calves from one market in East Anglia to another no more than twenty or thirty miles away. This is not, in fact, the case - it simply happens to be the practice to sell them rather older and heavier in one market than in the other, and in consequence they fetch a somewhat higher price there." (Spedding 1983 p.458)

In this case, the agents actually need metadata to help formulate its meaning. But even when all agents share a common metadata understanding, as is the case in the thesis model, meaning can be misinterpreted. The model in section 5.5.2 looks at an example that boils

Actors' decision process down to make the DMR dynamic as clear as possible. Firms move and then decide if that action was appropriate based on one datum - revenue change. The simplest interpretation of the data - Firms' take on its meaning - is, 'did my revenue increase? Then the action was a success.'

DMR is a kind of informational 'sonar'; as Vives puts it, firms often "use prices to experiment and learn about demand or costs conditions. Experimentation consists of paying a short-term cost in order to gain information that yields a long-term benefit" (Vives 2001 p.291). This is a very important component of respecting the impositions of space on the market, discussed in chapter 3, where agents must collect information however they can.

Much of the model finds ways to avoid this sort of coordination problem. The ABM tendency toward descriptive mapping makes this sort of problem very prominent: as section 4.7.1 discussed, if autonomous agent-building is the central goal, DMR-like issues will arise. One of the key aims of the 'mobile Firm' models in section 5.5 is to investigate what a large hurdle to full coordination DMR can be.

4.7.8 Examples of descriptive versus functional mapping

This chapter ends with some examples of how the difference between descriptive and functional mapping (section 3.4.1) shows itself in the thesis model. The most obvious example is that a single class takes care of utility (see section 4.6.1 for full details). Each Person asks this centralised class what utility they can get from a given bundle of goods. In reality, of course, the process of deciding what to buy takes place in a person's head: they do not ask some centralised organisation to tell them what they like. While it would be entirely possible to give each Person agent their own private methods for working out their wants (a closer descriptive mapping), in this case there is nothing to be lost by having a single class do the job. Of course, this does not rule out situations where it may be advantageous to internalise these decisions to each Person, as well as good computational reasons. If, for instance, agents were running on several parallel threads, having to calculate utility through a global class would slow the model down.

The method of modelling production used in the thesis is not convincing as a model of actual production and is not meant to be. It attempts to solve much the same issue the core model does: how to link production to welfare via demand? The difference is in requiring disaggregated agents to achieve this. The choice of a smooth over core model-style economies of scale function is, in this light, based on what achieves that goal, rather than on which manages to be 'more realistic'.

The same applies to the model's prevalent use of the CES function. Fujita *et al.* suggested a subtitle for their work might be "games you can play with CES functions" (Fujita *et al*. 2001 p.6), and the same applies here: it is a swiss-army knife of a function, but as with economies of scale, it stands in as a proxy for much more involved processes. For instance, in the thesis models a single Actor can buy any number of goods, each with a separate location and price. No uncertainty is involved in the spatial or price search, no actual travel or coordination issues arise. Equally, the CES can be used to avoid Bertrand competition dynamics where the Firm offering the highest wage (however marginal the difference) gets all the labour. Any more complex, 'realistic' model of actual job markets, able to account for the fact that Bertrand dynamics do not in fact exist in the labour market, may be overkill. A CES approach is used in the model in section 6.3. This is, for the purposes of the model, very useful: functionally mapping the idea of how good cost may affect Actors' acquisition while avoiding detailed dynamic problems. This in no way lessens the importance of understanding those dynamics, of course.

Number of Actors

> Nigel Tufnel: What we do is, if we need that extra push over the cliff, you know what we do?
>
> Marty DiBergi: Put it up to eleven.
>
> Nigel Tufnel: Eleven. Exactly. One louder.
>
> Marty DiBergi: Why don't you just make ten be the top number and make that a little louder?
>
> Nigel Tufnel: [pause] These go to eleven.
>
> *Reiner et al. 1984*

In ABM, agent number is commonly afflicted by descriptive mapping. Each model agent is asked to represent a specific real-world counterpart, and the impact of agent number on outcome is taken to be always relevant. If for computational reasons too few agents can be created, aggregate stand-ins may be used. Approaches include aggregating agents into 'super-individuals' or using a 'technical fix' such as parallelisation and higher processing power (Parry and Evans 2008) or combining an agent element with another aggregate modelling approach like gravity models to create a 'hybrid' (**?**). In a recent example, an Economist article argues for an agent modelling of 'the entire economy' involving 'millions of agents' (Economist 2013). The first thing to consider, however, is what the purpose of agent number is in any particular model.

Many economic arguments analyse away any effects from agent number; as section 2.3.2 outlined, this is how market structure problems are often circumvented. For example, the assumption of price-taking in perfect competition models requires an unspecified 'large' number of agents, so that no single one can influence price. As Blaug says, "perfect competition is, of course, an oxymoron, since it is a case of no competition whatsoever between firms" (Blaug 2009 p.223).

Within ABM there is an implicit assumption that such effects should be accounted for. After all, isn't that the very reason for disaggregating the model in the first place? The answer is, it depends. Just because a number of agents are being used does not mean every number-related effect should be considered. There may be good reasons for excluding them. Most obviously, the modeller may want to ignore some small number effect without having to actually increase agent numbers to do it, simply because of the computational expense. Functional-mapping-wise, this leads to a counter-intuitive point: it is possible to say 'assuming large numbers' while only having a small number of agents, if the dynamic under examination is separate from the assumption. For instance, the modeller may want to assume fixed prices as in perfect competition. This could be justified by assuming that numbers exogenous to the model are large, implying the model is only one region in a much larger world where, for example, fuel prices are set exogenously. But equally, if the goal is to have a number large enough to produce reliable aggregate results (e.g. which regions get what proportion of agents?), it may be just as unnecessary to force a large enough number to avoid sub-perfect-competition dynamics. The price-taking assumption could still hold without invalidating the choice of number.

A related aspect of number is that, for many of the partial models presented in the next chapter, no interaction is taking place between agents. Using partial models is the easiest way to avoid the problems imposed by full coordination. Depending on the problem, any number of model features and variables can be fixed. This is also a very effective method for testing and is used extensively in chapter 5 to do so.

The 'partial' models presented here fix different quantities depending on the problem. Most often, all People get a fixed wage (or a wage determined by the model); this means the overall level of demand is set exogenously. Firms' stock levels do not change: stock is always available to meet demand. All Firms have to do then, is maximise revenue by attempting to capture as much demand as they can. Aside from these two key features, others do vary. Firms may be mobile, optimising their location, while People remain immobile, and vice versa. In partial models using the CES function, the number of Firms each providing a unique good will affect the total possible quantity of utility.

Where People are deciding on location, if wages and prices are fixed, stock is unlimited and no density costs are incurred, each makes decisions entirely independently of the others. As a result, there is no actual difference between running a model with one agent a hundred times, or a hundred agents once (i.e. they are ergodic). There is, however, a reason for running a multi-actor model in this situation: it allows for a quick understanding of the outcome of the underlying stochastic process. Outcomes can be similar to entropy models (Wilson 1970) or can produce 'indifference maps' showing the spread of agent location choices.

In situations where there is interaction, agent number may matter for reasons unique to the problem being modelled. The DMR model in section 5.5.2 discusses a number of ways agent number interacts with noise levels. Finally, the 'transmission belt' model in section 6.3 has its ability to swiftly find equilibrium affected by total agent number.

Agent number, for the thesis model framework, has a somewhat prosaic purpose: to help with its functioning as an 'experiment station', numbers are chosen to allow a balance between good interaction speed and good presentation while also allowing for effectively robust emergent properties.

4.8 Summary

This chapter has detailed the way the thesis model has been put together. It has also discussed the broader issues that arose as a part of the logical journey of model development, and described some ways in which a ‘functional mapping’ has taken precedence over a ‘descriptive mapping’. The next two chapters present results from this framework.

Chapter 5

Framework testing and 'partial' model results

5.1 Introduction

5.1.1 Organisation of the chapter

This chapter presents results from the model framework. It does this in five sections with the following aims:

- Test that the main economic components of the model function correctly, first without any spatial element, and then with space costs included;
- Test the difference between using a 'hill-climber' vs constrained' utility approach;
- Provide People with mobility and a density cost to investigate partial economic outcomes when People optimise;
- Give Firms mobility to investigate their ability to maximise revenue and to examine the DMR issue.

5.1.2 Presentation of results

In this and the following chapter, results and analysis are presented graphically in four ways:

- Model run results include:
  - (1) A visualisation from the model thesis framework. These are screenshots from the 'live' code; versions of the same results can be downloaded via the link provided in the code appendix (A), where parameters can be varied;
  - (2) Graphed data output from a model run

- (3) Analytic results, where particular features are analysed mathematically to understand how a key model feature functions;
- (4) Spreadsheet versions of model dynamics are used in two cases. These boil down model dynamics to their most essential, simplified components in order to understand what is driving the emergent behaviour seen in the model framework. One appears in section 5.4.4, explaining how the optimal choices of People drives spatial outcomes in the ‘mobile People’ models. The other is used in the transmission belt chapter (section 6.3.3) to understand how People’s demand can determine the transition between production regimes.

Each visualisation is designed to make a specific point about the model it shows; a full description of them is given in the sections in which they are used. Figure captions repeat this information in a compacted form to give an accessible summary where they appear.

Correlations from the model framework are used extensively throughout the results. While the basic principle applies to all of these (showing how two quantities are dependent), the data points themselves vary depending on the purpose of the correlation being shown. Again, these are explained in detail in the captions for each, but it is worth mentioning the different ways these correlation graphs will be used:

- Datapoints will show separate agents, correlating two variables for each of them at a given point in model time.
- Single variables (e.g. the distance cost for all agents) will be correlated against other single variables. Often, one of thes will be the mean taken from all agents for a given variable.
- Datapoints will be taken at specific ‘trigger’ points in model time, for example in the ‘mobile People’ models, when Person movement drops below a set value and ‘spatial equilibrium’ is assumed to have been reached.
- In one example only (section 6.4.2), datapoints are taken on each model timestep for a timespan of interest. Here, the direction of model time is indicated in the correlation by using arrows for datapoints.

A more mathematical analysis is used in two types of situation. Firstly, many of the basic dynamics of the model are best communicated using an analytic approach. Secondly, when the model framework produces results that require further explanation. When the agent model outputs a particular outcome, is it due to Actors acting rationally? If so, what factors are at work? The framework proposes these questions, but a full answer may require an analytic comparison. Where results of such an comparison are informative, they have been included. As discussed in the conclusion, this is a happy synergy: the agent model helping to highlight possible lines of inquiry, and an analytic approach shedding further light on the issues. It will be made clear in the text where a model output or analytic approach is being discussed.

The use of stochastics varies depending on the presentation. Some results rely on stochastics to illustrate a point; others present only one model run outcome without showing all stochastic outcomes. For example, a number of the transmission belt presentations show a specific outcome, with emergent producers in a particular region. The specific outcomes are random, but for the purposes of clear presentation, one is shown. Other model presentations (e.g. in section 5.5) present stochastic outcomes; these are explained where they are presented.

5.2 Testing utility functions

This section demonstrates that the model framework successfully replicates the basic microeconomic ideas outlined in chapter 2. This is done by initially excluding any spatial elements to keep a clear focus on the central economic functions themselves. The aim is to demonstrate the robustness of the underlying utility concepts before introducing further model complexities.

5.2.1 CES function

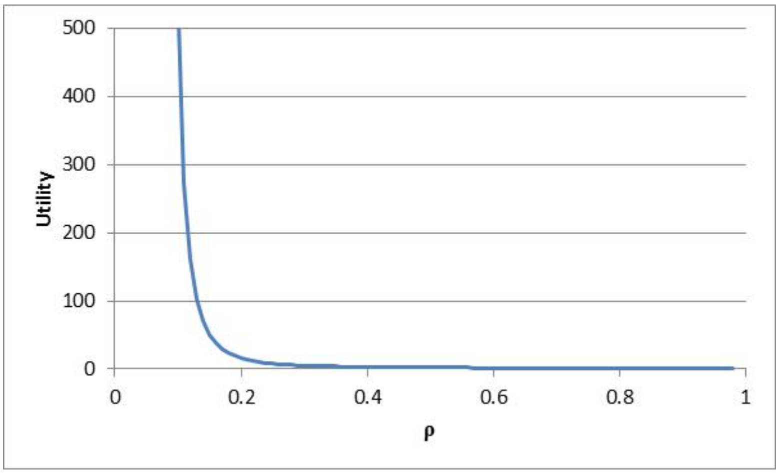
The discrete-good CES function (2.8) is most commonly used in the models that follow. Section 5.3.1 explores the interaction of the CES function with space; the function's basic output is tested here.

The following uses a model run test (see figure 5.1) that removes all spatial costs and normalises wage to 1 and all good costs also to 1. A single Person then optimises: what optimal mix of goods can they buy? Firstly, this is done for $0 < \rho < 1$, in a situation with two goods. (Recall that as $\rho \to 1$, goods become substitutes and love of variety drops.) Two views of that data are given, since for $\rho < 0.2$, utility becomes very large. Sub-figure 5.1b illustrates that, as $\rho \to 1$, utility from a wage of 1 for two goods moves towards 1. When $\rho = 1$ all goods are subtitutable, so any mix would produce the same utility value.

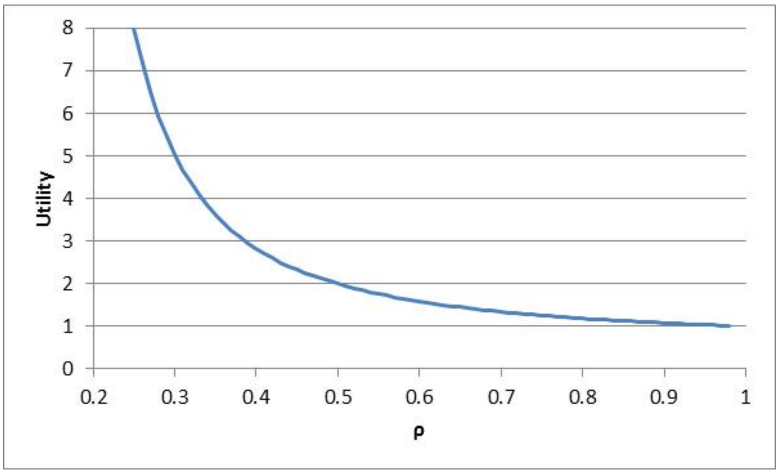
Secondly, in sub-figure 5.1c, ρ is held at 0.75 and the number of goods (in the model, the number of supplying Firms) is changed, up to a thousand. Each good has the same price, so the optimising Person buys a symmetric mix of all available varieties. While the budget constraint does stop utility from growing exponentially with varieties, it can be seen that utility increases (at a decreasing rate) as the number of goods consumed increases.

5.2.2 Cobb Douglas

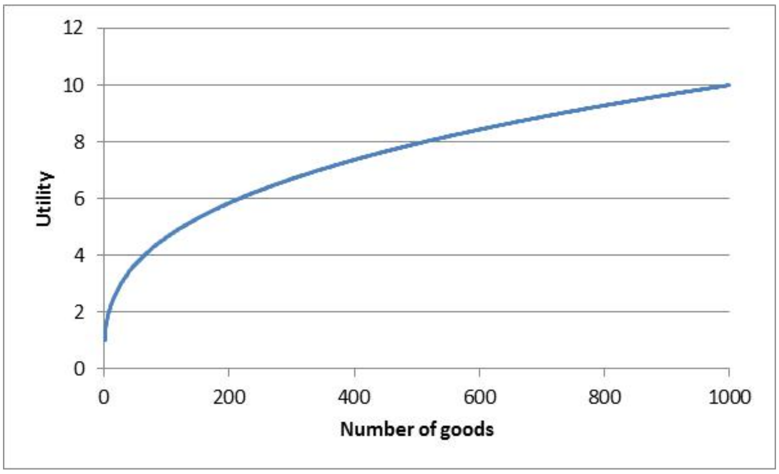
Actors with a Cobb Douglas utility function (2.2) want to buy a mix of two goods F and M, where p_F and p_M are the price of F and M respectively. The constrained optima (5.1) and (5.2) were described in section 2.2.4 and are repeated here:



(a) 2 goods, ρ varied



(b) 2 goods, ρ varied, from $\rho = 0.2$



(c) $\rho = 0.75$, good number varied

Figure 5.1: Testing CES function works correctly. 5.1a and 5.1b look at the same data: two goods, changing ρ. 5.1c keeps ρ fixed at 0.75 and changes the good number.

$$F = \frac{\delta Y}{p_F} \tag{5.1}$$

$$M = \frac{(1-\delta)Y}{p_M} \tag{5.2}$$

The Cobb Douglas share parameter δ exactly describes the income for each sector producing F and M. Price-setting does not affect the income of the seller at all: each always gets exactly a δ share. In the core model, this feature of the Cobb Douglas is useful, since it rules out any elasticity between the two sectors of manufacturing and food, keeping the analysis much simpler.

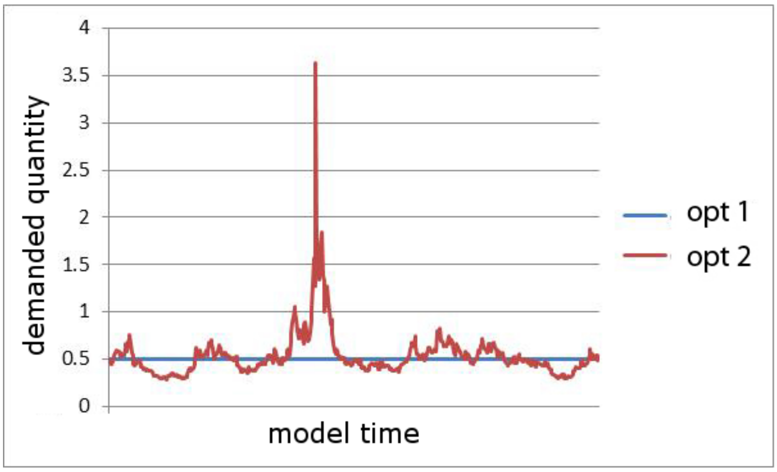
The fact that the share parameter δ dictates each sector's income can be seen analytically from (5.1) and (5.2): the price of each is internal to its own optimum and does not appear in the other. Therefore, changing the price of one *cannot* affect demand for the other. Each sector is independently unit elastic: if the price is dropped, demand will rise - but such that the total revenue remains the same.

Putting this into an agent context, however, if Firms are being asked to take price decisions, they will find they have no effect, removing their ability to make optimising decisions. This can have a spatial consequence. In a situation where only two firms exist and space costs are internal to them, each producing one of F and M, they can gain nothing by changing their location: their income will stay precisely the same. Any reduction in space costs will lower costs, which will lower prices, but this will be exactly balanced by raised demand. In agent model runs where two firms set random prices for F and M, then, both firm's income levels stay the same regardless of price. This is as should be expected from Cobb Douglas utility: price elasticity of demand is unit elastic: any change in price is matched by a change in demand that keeps revenue levels static.

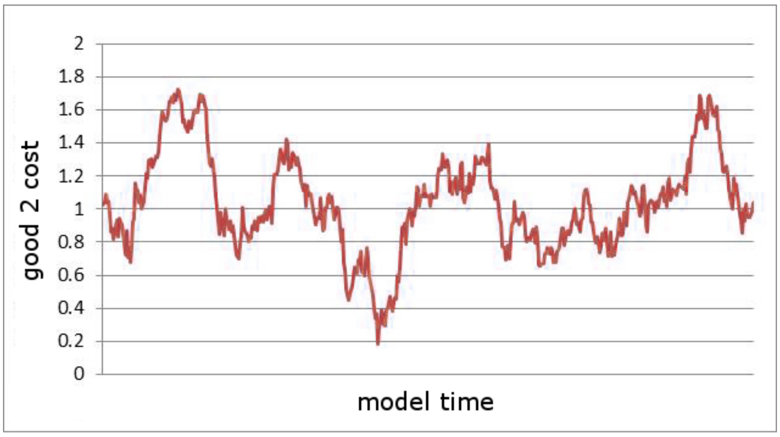
Figure 5.2 makes the same point with data from a model run: one firm's prices are fixed while the other attempts to find a revenue-maximising price. As mentioned above, this fails because they can affect no change on their revenue in this situation. For the fixed-price firm, actors consume a constant amount - and that is not altered by changes in the price of the other good. This shows that if People have a Cobb Douglas utility function in the model framework, demand for each good is independent of the other: there are no elasticity effects between them.

Note that, in situations where distance costs 'melt away' or total model revenue has any other leak-points, the above would not hold: Firms could, for instance, improve their revenue by increasing proximity to their market, if expenditure on distance costs went elsewhere.

An interesting point arises from implementing the Cobb Douglas in agent form. Imagine a situation where only the agent-based approach to testing the utility function existed. Eventually, the independence of the two sectors in the Cobb Douglas would become apparent. But how would one be able to 'explain' it? It is hard to imagine any answer that would not eventually reduce to the analytic result from the Lagrangian optimisation. By clearly showing, in a



(a) Optimal demand for good 1 ('opt1') and optimal demand for good 2 ('opt2'); Firm 2's optimal demand affected by its price changes. Firm 1's in unaffected: there are no elasticity effects.



(b) Price changes of Firm 2, underpinning changes in demand for good 2 ('opt2') in the previous figure.

Figure 5.2: Two Firms selling goods 1 and 2: Firm 1's price fixed, 2 attempting to optimise price. Actors with Cobb Douglas utility: changing price of one has no effect on demand for the other.

mathematical form, that p_F and p_M only appear in the demand equations for the same good, Equations (5.1) and (5.2) demonstrate that Cobb Douglas demand must always be entirely separate for the two goods. This is perhaps the best possible explanation; a cleanly analytic solution that the murkier agent approach makes more difficult to discern. Conversely, it is clear that Cobb Douglas utility *imposes* this strict unit elasticity on sectors. As with the core model this may be desirable, but it underscores the interdependence between Firms and People discussed in section 4.7.1. This is not necessarily an argument against ABM, since - as the rest of the chapter discusses - the flexibility of the method has other advantages. But it does highlight that, for many problems, neither method has a unique claim to explanatory power.

5.2.3 Hill-climber versus constrained utility

Chapter 3 argued that utility is a useful concept for describing how people react to cost changes. In the model framework, two different approaches are tested for modelling People's utility: a 'hill-climber' algorithm (see section 4.6.1) and a more economically traditional constrained utility function, of the type outlined in section 2.2.4.

What might be the problem in using constrained utility in an agent model? Section 3.5.1 pointed out that constrained utility is, as Leijonhufvud noted, "a statement about actual performance, not just motivation". The constrained function takes in a budget amount and goods prices and outputs an amount of demanded good. But what if the demanded good is not available? An 'objective function' of this sort appears unable to deal with this situation. Consider the equation (2.11), the constrained optima for an n-good CES function described in section 2.2.4. For a Person with a bundle of detected goods, the optimal amount for each good c_j is given by:

$$c_j = \frac{p_j^{\frac{1}{\rho-1}} Y}{\sum_{i=1}^{n} p_i^{\frac{\rho}{\rho-1}}} \tag{5.3}$$

- where p_j is the price of good c_j, Y is the budget, and the denominator sums the price of each good raised to $\rho/\rho-1$. What happens in a situation where, for example, $c_1 = 1$ but only 0.5 is available? There is then a different problem to solve: what is the optimal amount for the other goods, given that we know the available stock of $c_1 = 0.5$. The constrained optima from equation (2.11) provides only demand as an output. There is no way to include a set demand amount for one or any of the goods; the function is simply not set up that way, and cannot be due to its roots in constrained optimisation. Neither is it quite as easy as removing c_1 from the list and recalculating, since demand for each affects all the others.

On the surface, the hill-climber algorithm approach appears a more plausible approach to utility: People weigh up a given mix of goods, and compare until they find what makes them 'happiest' (that is, what gives them the highest utility). They do not have a superhuman constrained function informing them of their optimal demand. If any good is not available in the quantities they want, budget slices can be reassigned to the pot for the next iteration. In practice, however, this literal approach to stock consumption is complex to implement - not from the demand side, but because Firms then need to find some way to produce the *correct* amount of stock given demand signals.

Fowler's approach is to make a tactical modelling choice that avoids the need to consider rationing of limited goods: "workers select their optimal bundle of goods without reference to the actual supply of the good" (Fowler 2011 p.10). Thus, they still have an 'objective' demand level. He notes a key informational advantage: this sends "clear signals to firms about the actual level of demand for their goods at the current price" (ibid.) As he says, it also avoids having to work out any explicit rationing strategy if several actors are competing for limited goods.

While this is an effective simplification, it may be problematic depending on what the goal is. For instance, what exactly does it mean for Firms to receive demand signals for goods that People could not actually acquire? This is saying, 'I wanted more than was available', and so in theory a Firm can use that information to inform their production decisions. This is a feasible approach if stock is theorised in a particular way; chapter 6 presents a model that does this.

On a more mundane level, however, it is useful to compare the hill-climber algorithm to its constrained counterpart, to show that the two produce almost identical results. They are only different to the extent that the hill-climber algorithm cannot avoid a certain quantity of noise. Noise in utility is also put to good use in the model presented in section 5.5.2 to examine how it impacts on Firms' ability to make appropriate decisions. The next section looks at the two approaches to utility maximisation in more detail.

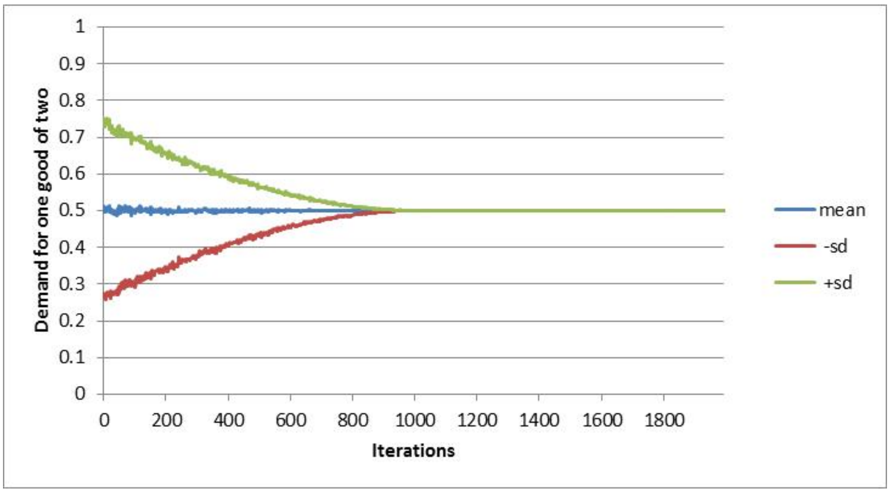
Comparing the two

How do the two approaches compare when stock levels are not an issue? The exact level of difference between the constrained optima and the hill-climber algorithm depends on the number of iterations the latter is given to ratchet towards utility-improving budget slices. Figure 5.3 shows what difference the iteration number makes for two scenarios. Both show aggregate statistics for a thousand optimisations at each iteration-number point, up to 2000. Again, it is a CES function at $\rho = 0.5$, all good costs are even and the budget is one. This means the 'correct' optima for each separate good is known. For the first example (5.3a and 5.3b) optimising two goods, that is 0.5 for each, and as expected the constrained method finds this. Very low iteration numbers do not, on average, manage to get very close to this optimum, but by the time 1000 iterations are used, it is missing the perfect optima only minimally. Sub-figure 5.3b, however, shows that further improvements - though tiny, note the y-axis values - do continue to be made for iteration increases up to 2000 (in this example). There is always a little noise left in comparison to the constrained method.

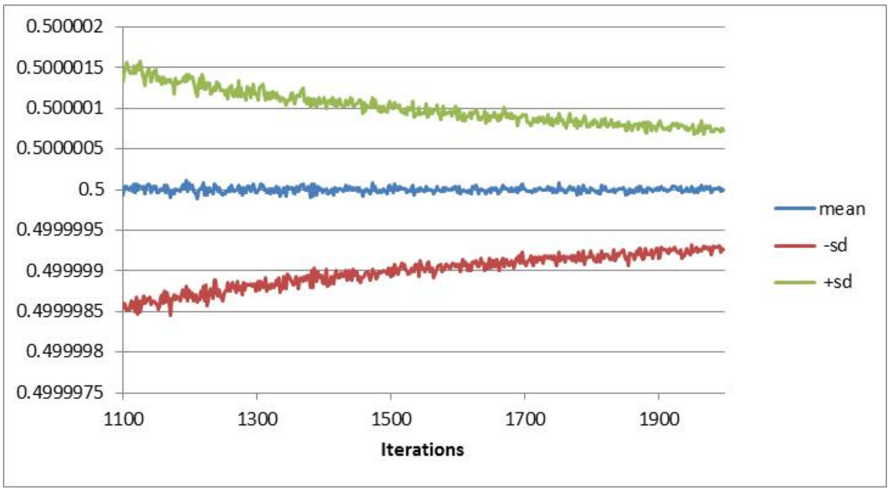
The last sub-figure (5.3c) shows the same outcome if eight goods are used - again showing results for one of those goods, whose analytic optima is 0.125. It is a quirk of the method that a higher number of goods can be optimised more quickly. The basis for this is an effect from good number: budget slices are all individually smaller: there is simply less scope for large variation due to the randomisation method used.

The difference between the hill-climber and constrained algorithm is a useful way for thinking about the issue of realism and descriptive mapping in model structures, and the use of analytic optimisation more generally. Chapter 3 discussed the centrality of equilibrium to neoclassical economic analysis, and the criticisms of agent modellers. Equilibrium skips to the end-point of a process: what about the process itself?

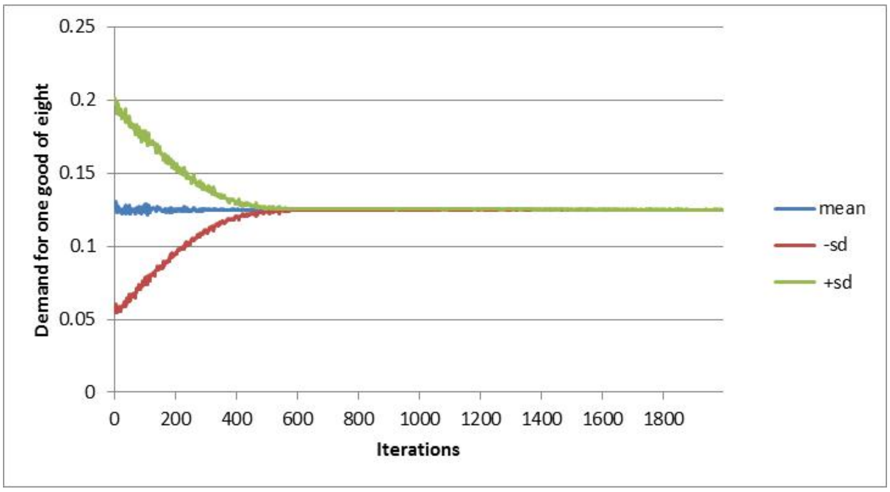
Compare the underlying method of constrained versus hill-climber approaches to utility. Section 2.2.4 explained that the process of acquiring the constrained utility equations is to put them through a Lagrangian optimisation, where an algebraic budget is used to constrain the raw



(a) 1 of 2 goods, zero to 2000 iterations



(b) 1 of 2 goods from 1100 iterations



(c) 1 of 8 goods, zero to 2000 iterations

Figure 5.3: Hill-climber algorithm optimising 2 goods (5.3a and 5.3b) and 8 goods (5.3c). Optimum for one good is shown. 1000 optimisations for each iteration point, mean and one standard deviation.

utility function. This effectively ditches the original utility and budget functions and replaces them with a third set of equations able to give optimal good quantities from an actual budget value and a set of good prices. In the model framework, as section 4.6.1 explained, those constrained equations work well since individual Good costs and People's budgets can be passed in, and optimal quantities of good returned.

Once those values are found, however, the Person's utility level is still unknown. In terms of realism, this is a somewhat 'spooky' outcome: the Person has their optimal quantities but does not know how happy it makes them. To get that information out of the model, the optimal good quantities must be fed *back through* the original utility function. Causally, that makes no sense: a Person uses some function to find the good quantities they want, and then works out precisely how well off it makes them *after* they have identified their optimal bundle.

In contrast, the hill-climber algorithm seems more 'plausible', and certainly coheres more to the real-world process of decision-making. People must 'hold up' bundles of goods and test the utility of each. Yet the hill-climber and constrained approach are clearly producing the same result, both succeeding at homing in on the point where the function can no longer change. However, the analytic constrained utility does the job better at a fraction of the computational expense - despite being causally muddled. This does not mean the hill-climber approach has no merit, but rather that the analytic method applied in this agent model served its purpose very well.

In summary, both algorithms can find optimal utility levels, but the hill-climber algorithm has a processing overhead and noise. For most purposes in this thesis, the constrained approach was better. The difference between the two is no guide in that respect: as Friedman says, 'everything depends on the problem'.

However, the hill-climber algorithm does provide potentially useful flexibility for testing utility. For a start, it does not care about the form of utility it deals with. It takes a sample of good amounts as input: any and all object-oriented and algorithmic methods could then be applied to return a utility level. This could, for instance, allow interaction of tastes between actors as well as evolutionary approaches. (Though note, the particular hill-climb approach used here may not be suitable for situations with many local optima.)

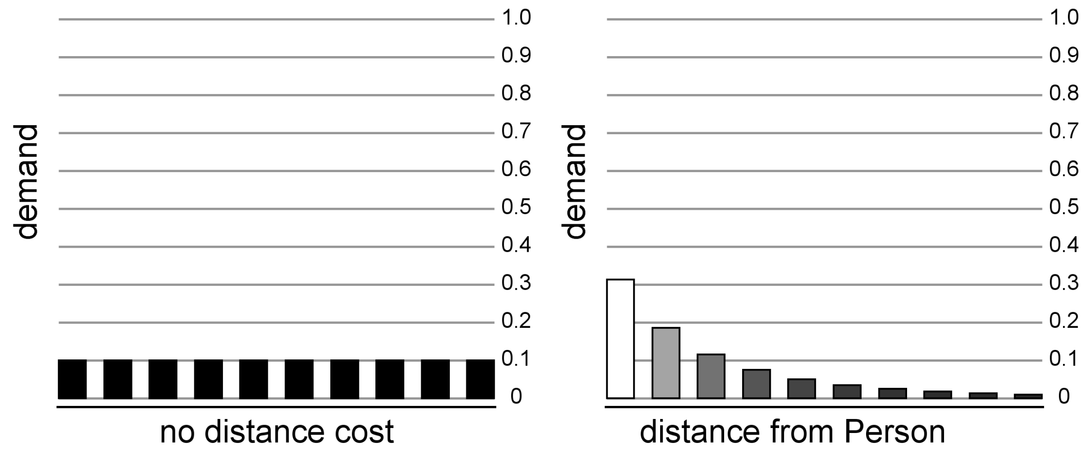
The next section introduces a spatial cost element into the utility functions tested above.

5.3 Two basic distance tests

5.3.1 Love of variety's interaction with distance

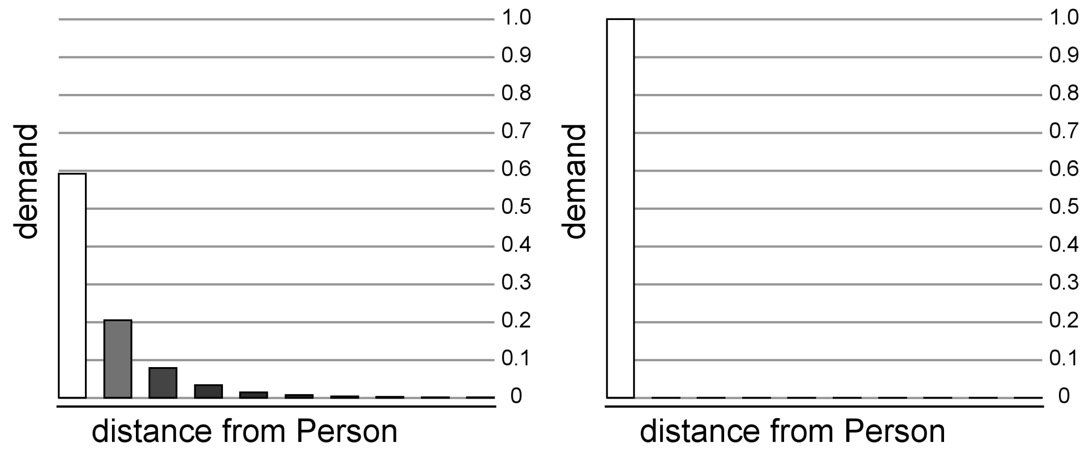
Section 2.2.3 explained that, if good prices are all equal, they will each be demanded in equal quantities when People are using a CES function. This even split does not change for $0 < \rho < 1$ (where ρ is the love of variety parameter). The utility level *derived* from the consumption of that even split increases inversely to ρ, but not the amount of each good demanded.

If prices vary across goods, however, so does demand across goods as ρ changes. This



(a) No distance costs

(b) $\rho = 0.8$



(c) $\rho = 0.9$

(d) $\rho = 1$

Figure 5.4: Effect on demand of changing ρ in the presence of distance costs. Ten Firms are spread along the x axis evenly in distances of zero to one from a single Person. The Person is at the left origin of the x axis. Good cost and wage are = 1. No distance cost in first graph, so demand is even; for others, distance cost is 1 and ρ is varied.

is important spatially: even if the base good prices are the same, distance costs will have a different effect on demand depending on ρ. (This is the basis for the use of the CES function as a spatial interaction tool, mentioned in section 2.8.2.)

There are two separate effects to consider: that of love of variety itself, controlled by ρ, and how changing the cost of moving goods impacts on demand. The following examples consider a single Person with a fixed location and fixed budget, spending it on a number of Firms at different, evenly spaced distances along a line. This can only be an illustrative way of conveying the effect of variable changes for one partial situation. For instance, the exact number and location of Firms will be vital to outcomes; here only nine or ten are considered.

Figure 5.4 visualises the dynamic in a model run scenario. In this model run, a Person is fixed in one position at the far left and ten Firms charging the same base price are evenly spaced in fixed positions from it. In sub-figure 5.4a there are no distance costs: thus all goods cost the same. In this situation, ρ will have no effect on the difference in demand between goods: the budget is evenly split between them. Sub-figures 5.4b to 5.4d all have a positive distance cost and so the further a Firm is from the Person, the higher the cost. The three different values of ρ give a different spatial demand outcome. As ρ approaches one, demand shifts to nearer goods, until at $\rho = 1$, goods are completely substitutable and - stock permitting - demand will be met by the nearest provider.

Figure 5.5 shows the same thing analytically for ten Firms at distances between 0 and 1. Regardless of the distance of the nearest, it will end up getting the whole spend as $\rho \to 1$.

If the cost of moving goods across space is changed, the impact is not necessarily even for all Firms. This is due to the fact that, as those costs increase, demand for more distant goods is re-allocated to closer providers. This can lead to a 'double dip' structure to demand: when space costs are low, demand will be even and nearer goods will sell equally to those more distant. As space costs increase, demand can increase for a time for the more proximate Firms. As costs rise, however, the overall cost effect will drive demand down again. Note, this is only the case for Firms *nearby*: if a Person is at a distance of zero, demand will simply continue to increase as long as it is re-allocated from other goods.

To illustrate, consider the following analytic example. The CES constrained optimisation equation (2.11) can be used in a simplified form. Let $\rho = 0.5$ and assume a budget of one. It then reduces to -

$$g_j = \left[P_j^2 \sum_{i=1}^{n} 1/P_i \right]^{-1} \tag{5.4}$$

- Where g_j is demand for good j, P_j is its full price, and P_i is the full price of goods i through to n. The full price of each good is still just $p + (c \cdot d)$. Assuming the base good price to be the same for all goods ($p = 1$), if no distance costs are present, a perfectly even mix will be the optimum. A series of goods at different distances from the buying Person will differ only in values for d and how that multiplies the current value of per-unit space cost, c. In the following

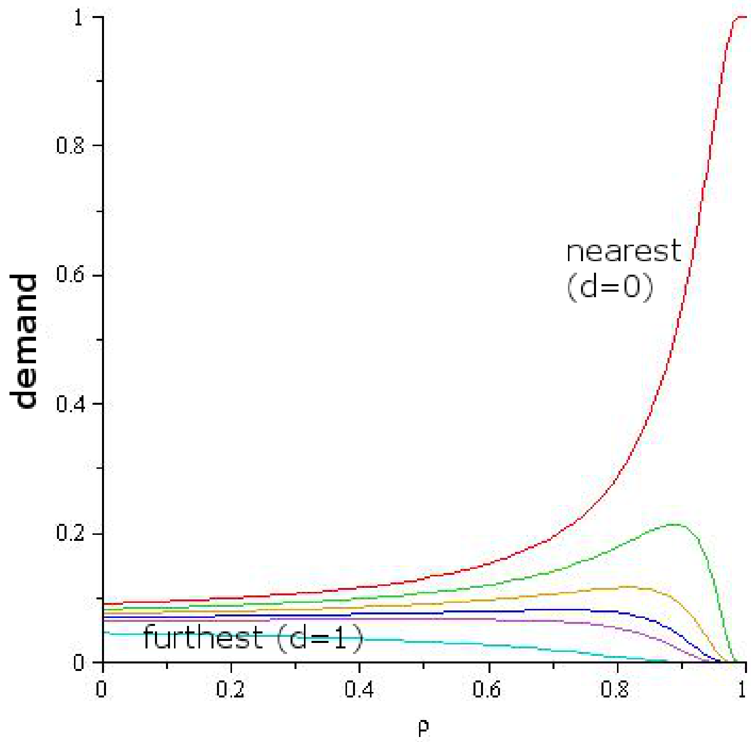
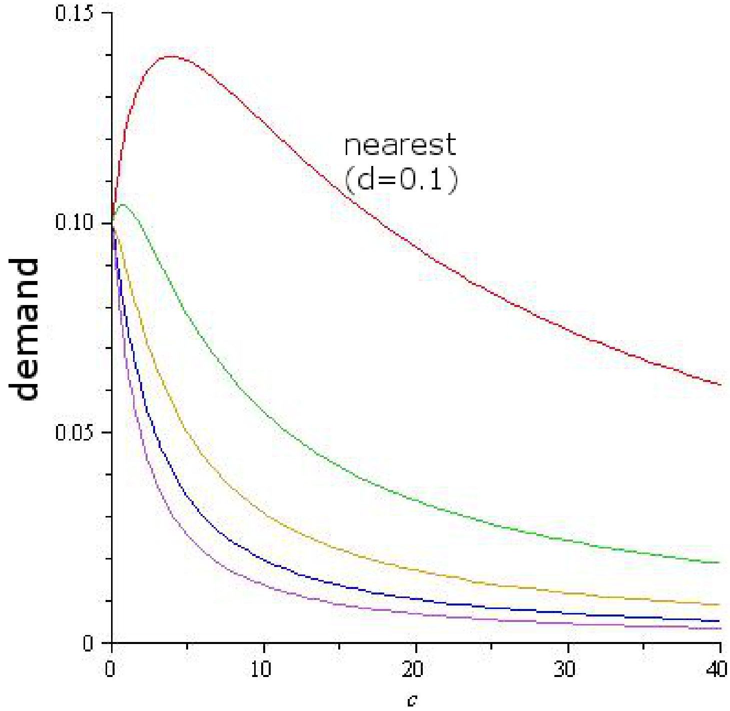


Figure 5.5: Demand levels for ten goods spaced between a distance of zero and one, for $0 < \rho < 1$. Distance cost is 1. One Person, $p = 1$ and $c = 1$, five nearest and one most distant shown here. As $\rho \rightarrow 1$, the nearest good ends up getting the whole spend.

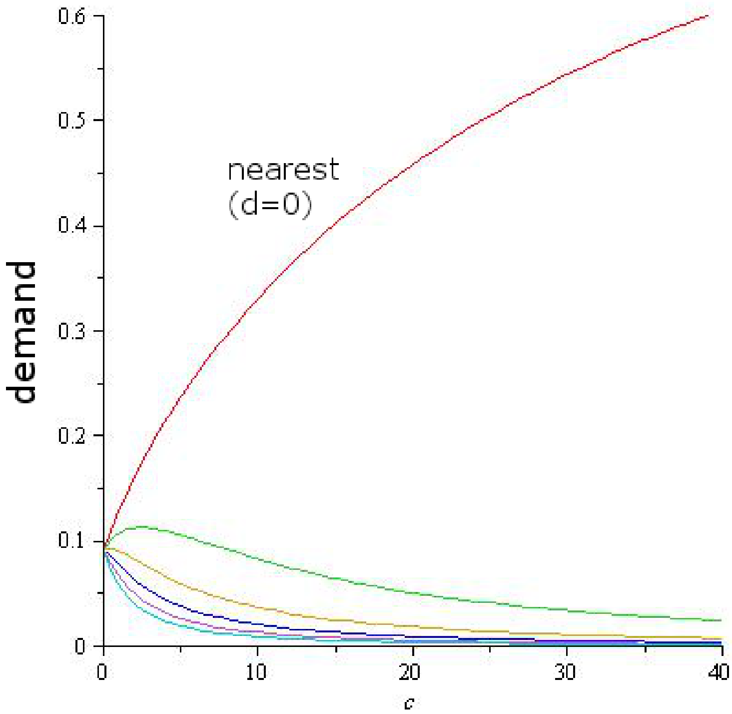
example, a number of goods are spaced at distances from either 0 or 0.1, up to a maximum distance of 1. It is then possible to see the impact of changing the distance cost c. Using all these assumptions, the above equation becomes:

$$g_j = \left[(1 + c \cdot d_j)^2 \sum_{d=0\ or\ 1}^{10} \frac{1}{1 + c \cdot (d/10)} \right]^{-1} \tag{5.5}$$

Here, a number of goods are spread out at different distances up to 1. So: how do changes in the space-cost c affect how demand is divided between the goods? Figure 5.6 shows the above equations analytically for $0 < c < 40$. Ten goods in total enter into the CES, of which only the nearest five are shown here. Where the nearest is at 0.1 (sub-figure 5.6a) the double-dip structure is apparent: re-allocated spend increases demand for that good before larger values of c dominate. As mentioned above, whether there is a good available at distance 0 makes a large difference. Sub-figure 5.6b shows that, if the nearest Firm is at $d = 0$ - that is, if there is zero distance between consumer and producer - they receive all reallocated demand as distance costs increase for other Firms.



(a) Nearest Firm at $d = 0.1$

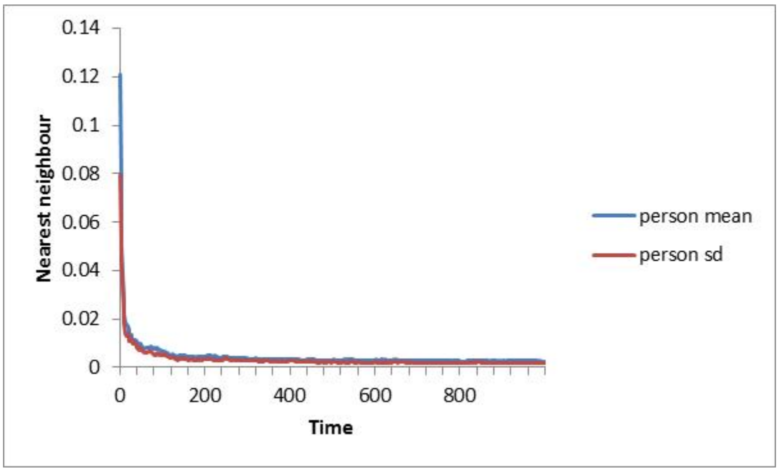


(b) Nearest Firm at $d = 0$

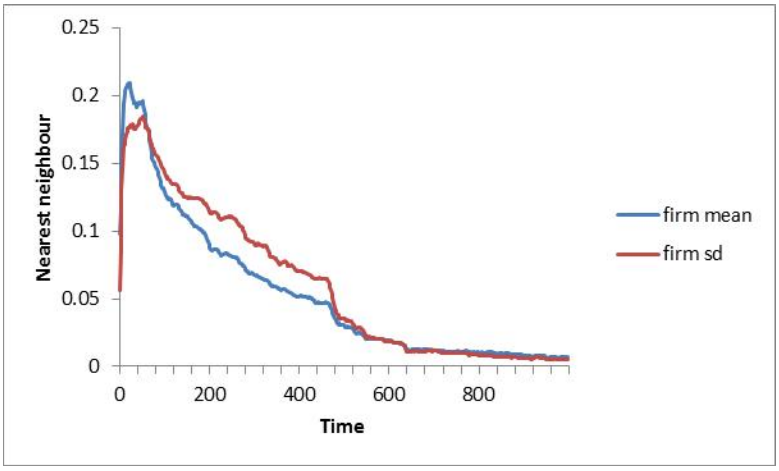
Figure 5.6: Impact of changing good space cost: one Person, ten Firms. In figure 5.6a the nearest Firm is at a distance of 0.1 from the Person; in figure 5.6b the nearest Firm is at 0. Firms are spaced between zero or 0.1 and one, with the five nearest shown here. Value of c (cost of moving good a unit of distance) on x axis. Note: demand axes have different values. 5.6b shows Firm at $d = 0$ receiving all re-allocated demand as distance costs increase, unaffected by distance cost change.

5.3.2 The 'black hole' outcome

Space imposes costs and, like any other costs, actors will minimise them if they can. The most obvious spatial outcome is that all actors will simply remove those costs, collapsing to a point in the process, unless other forces mitigate against this. Such forces exist; one outcome of using them is discussed in the next section. But without them, does the model reproduce the basic 'black hole' finding? In this model run, with all other costs fixed, Firms and People are allowed to find their preferred location using the methods described in chapter 4 - People comparing different Bundle locations and Firms using a feedback that crawls towards revenue-improving locations.



(a) People



(b) Firms

Figure 5.7: Means and standard deviations of nearest neighbour distance, randomised starting locations, averaged from 20 model runs.

Figure 5.7 shows the outcome of this model run over a thousand days for twenty People and four Firms. The model was run twenty times with different random starting conditions; the data shows the average outcome. All actors converge on a point, incrementally seeking positions closer to each other to reduce space costs or, in Firms' case, increase revenue. People quickly converge, while Firms take much longer: this reflects the fact that Firms must use a feedback that cannot always be relied upon - section 5.5.2 discusses this point in detail. Ultimately, all actors do reduce their space cost as close to zero as they can.

5.4 Mobile People

This section investigates the results of allowing People in the model to choose their location. For all but the last model in section 5.4.6, a Firm or Firms are fixed at a single point and People organise themselves around that point, incurring density costs as they do so. The set-up for the models in this section is as follows:

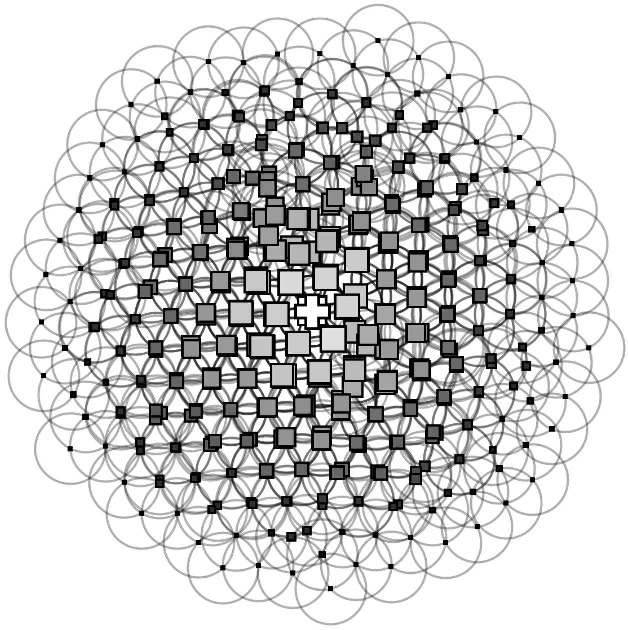
- An R^2 plane of distance 1 per axis.
- 500 People (unless otherwise stated) are mobile, able to select the Bundle with the optimal location as outlined in section 4.6.1.
- Firms have a fixed location. The number of Firms vary depending on the job the model is doing; see each section below for specifics. Base good costs are set to one; distance costs are varied for some model runs.
- People incur a 'density cost' as outlined in section 4.4.

As explained in section 4.4 the multiplication factor for the density cost is set to 10 for all but one set of model runs, because it provides an incentive for People to trade density cost off against distance. Section 5.4.7 discusses how increasing density cost higher impacts on spatial equilibrium assumptions.

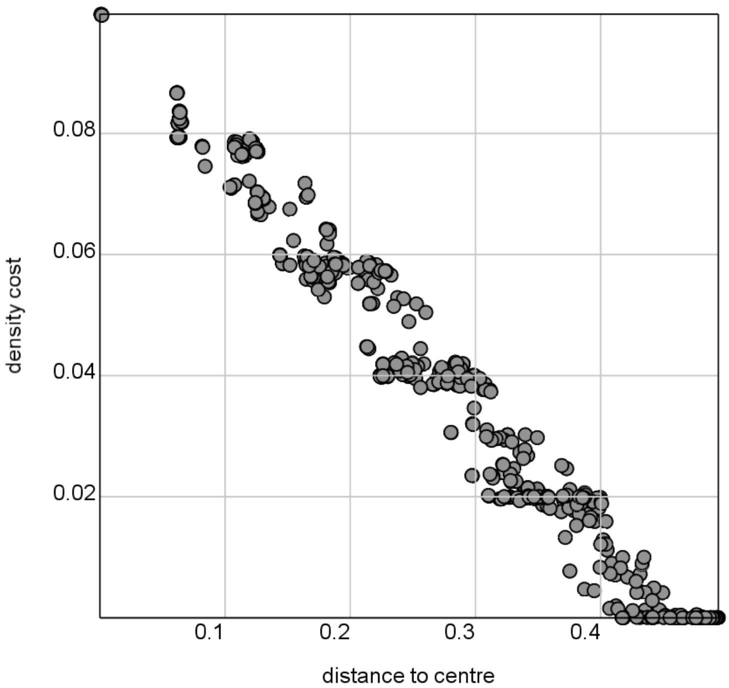
5.4.1 Basic monocentric result: density cost traded off against distance

Figure 5.8 provides an overview of the model's outputs from a typical monocentric setup. Sub-figure 5.10a is a visualisation of People locating in an R^2 space; this style is used throughout this section. These visualisations use size and shade of squares to indicate the **relative** range of values during the iteration being shown. Small, black squares are the lowest value of the shown variable (in this case density cost); large, white squares are the highest. Sub-figure 5.8b shows the same information in a different way, showing each Person's distance to the central Firm on the x-axis versus the density cost they are incurring closer to the centre.

Figure 5.9 is from the same model run. It is telling one clear story: utility is close to identical for all People (staying close to 0.9) regardless of distance from the central Firm. Utility is equilibriating to an equal value for all People over distance. In figures 5.10a and 5.8b, the density incurred by People is shown to drop off linearly over distance from the centre: a trade-off is being made between density and distance cost. This lends support to the argument for



(a) People in R^2 buying from central Firm. Larger/lighter squares indicate higher incurred density cost, rising towards the centre.



(b) Incurred density cost vs distance to centre for individual People

Figure 5.8: Model run: 500 People; density cost = 10; distance cost = 0.25; no commute cost. Correlation shows datapoints for individual People.

assuming spatial equilibrium (section 2.4). However, there are scenarios where this assumption (that spatial equilibrium is also a utility equilibrium) fails; see section 5.4.7.

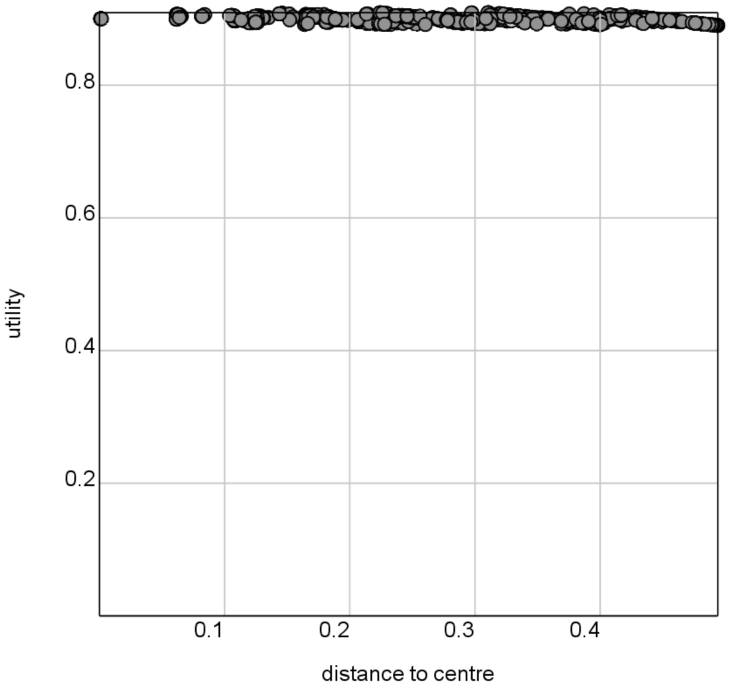
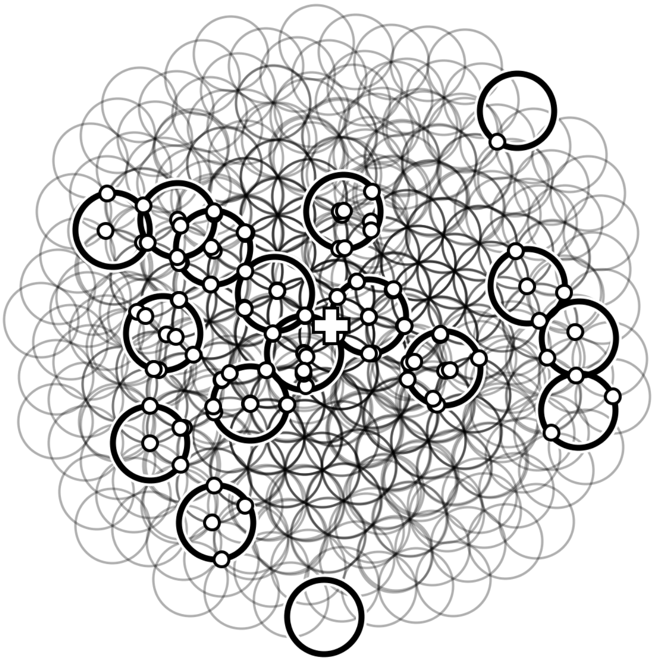


Figure 5.9: Utility vs distance to central Firm: utility equilibriates. 500 People; density cost = 10; distance cost = 0.25; no commute cost.

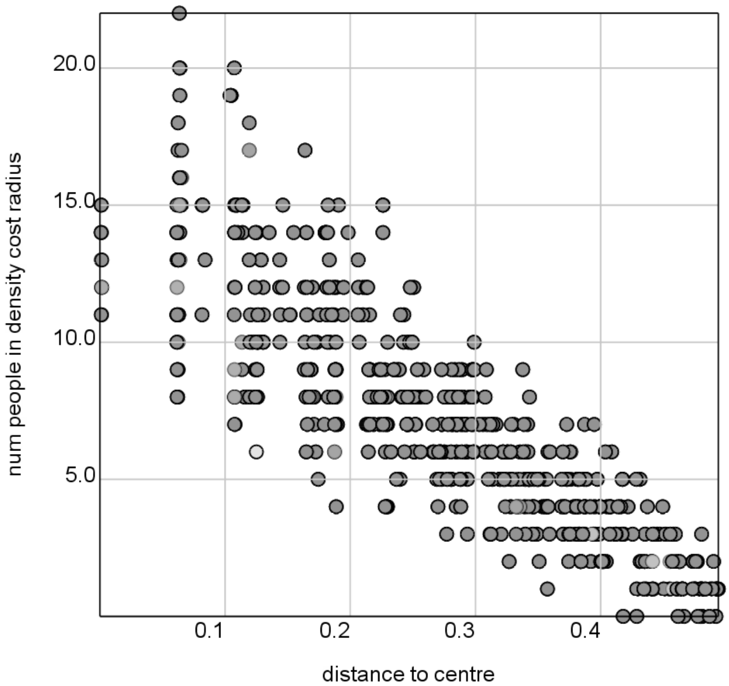
Before section 5.4.7, the following analyses work within the range where density cost sets up a trade-off to explore exactly how it functions from agents’ point of view.

5.4.2 Density cost

To further illustrate the effect of the density cost, figure 5.10 shows a typical result from People making this trade-off in a monocentric model with a single Firm at the centre. Light circles indicate the density cost radius for each Person. Darker circles show a random selection of People, including the position of others they are incurring a density cost from (the smaller white circles). The resulting hexagonal patterning is a striking feature of People’s self-organising behaviour, resulting from minimising density cost by finding interstitial points. However, for the purposes of this thesis - as section 2.4 discussed - it is economically uninteresting. Sub-figure 5.10b shows the discrete number of People entering into density cost calculations correlated to how far each Person is from the central Firm. Each circle represents an individual Person; they are willing to incur higher density costs towards the centre.

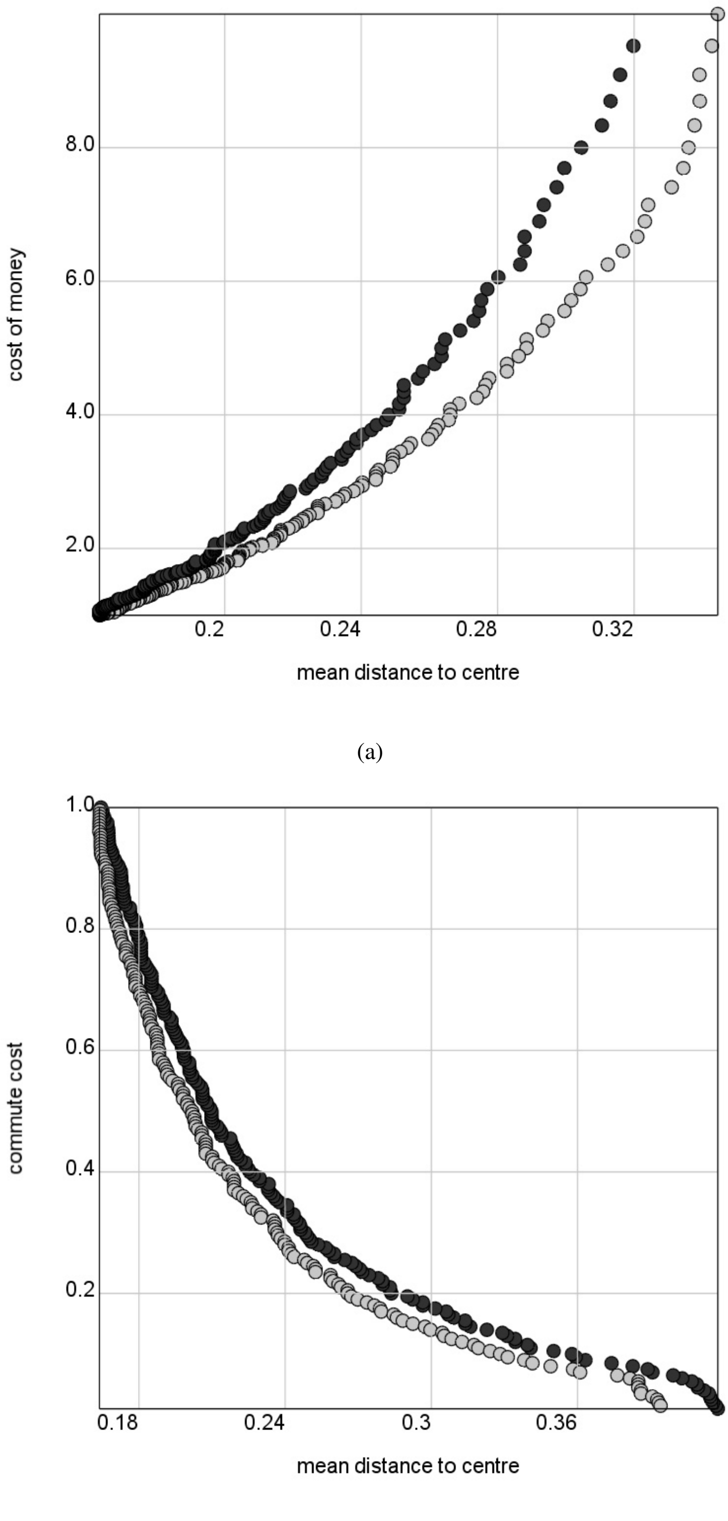


(a) People in R^2 buying from central Firm. People (density cost radii shown) and Firm (white cross in centre). Random selection of People's density cost radii (thicker black circles) showing location of detected People entering into density cost calculation (small white circles).



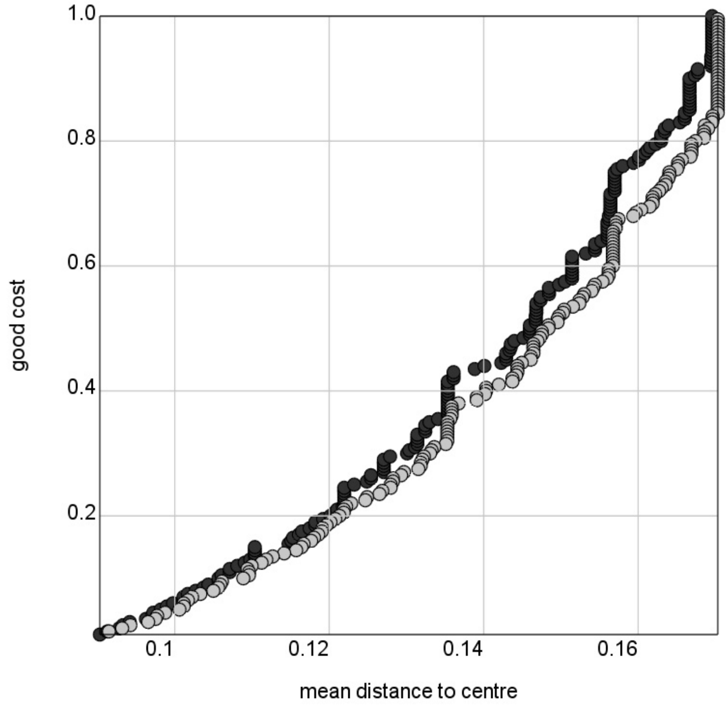
(b) Number of People in each Person's density cost radius vs each Person's distance to central Firm. For each Person, discrete number of People entering into density cost calculation is shown on y axis. This increases with proximity to centre.

Figure 5.10: Effect of density cost in a 'monocentric' model run: 500 People; density cost = 10; distance cost = 0.25; no commute cost. People optimise in response to distance and Density cost.

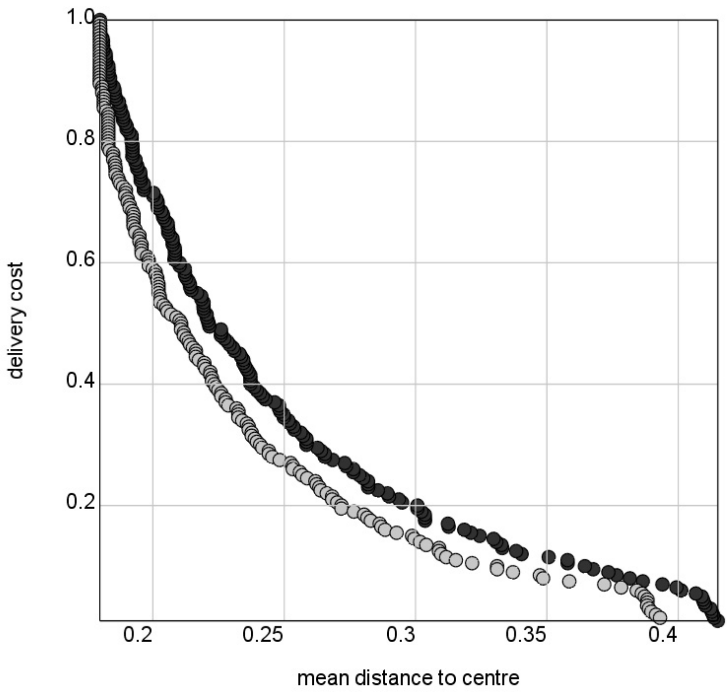


(b)

Figure 5.11: Wage offer ('cost of money') and commute cost versus mean distance to centre; single datapoint shown: average for all People, taken at spatial equilibrium. Each is parameter-swept (positive in black, negative in grey) in increments of 0.005. 'Spatial equilibrium' is assumed when mean agent movement < 0.0001 average over 10 iterations.



(a)



(b)

Figure 5.12: Base good cost and distance cost versus mean distance to centre; single datapoint shown: average for all People, taken at spatial equilibrium. Each is parameter-swept (positive in black, negative in grey) in increments of 0.005. 'Spatial equilibrium' is assumed when mean agent movement < 0.0001 average over 10 iterations.

5.4.3 Impact of changing spatial and non-spatial costs on size of settlement

Section 4.3.3 explained that the cost of moving both people and goods can be broken down into spatial and non-spatial components. Figures 5.11 and 5.12 show the result of sweeping each of these four costs: wage (reframed as base ‘cost of money’ in time) and commute cost in figure 5.11 and base good cost versus distance cost in figure 5.12. The model setup is: 300 People buy a Firm’s single good and work for a fixed wage. Each of the four variables is changed from zero to one and from one back to zero in increments of 0.005. Both directions are checked to highlight any hysteresis: darker circles indicate increasing distance, lighter circles decreasing. ‘Spatial equilibrium’ is assumed when mean movement for all People drops below an average of 0.0001 over ten iterations. A datapoint is taken when this condition is met. At least ten iterations are then allowed to pass before checking for stability again. The impact on spatial morphology is measured using the mean distance of all People from the centre firm, which is larger as People spread further out from the centre.

Sub-figures 5.11a and 5.12a show the effect of varying the non-spatial costs. Changing both the ‘cost of money’ and the good cost effectively alter the real wage, and both have a spatial impact due to their interaction with density cost. Increasing the real wage makes People more able to offset density cost: they agglomerate. Conversely, sub-figures 5.11b and 5.12b show the effect of varying the spatial costs, changing the cost of moving goods and wage across space differently depending on a Person’s location. Decreasing distance cost and commute cost have the same effect: goods and wages can get further for cheaper. While overall this increases Peoples’ real wages, the spatial effect is opposite: they disperse.

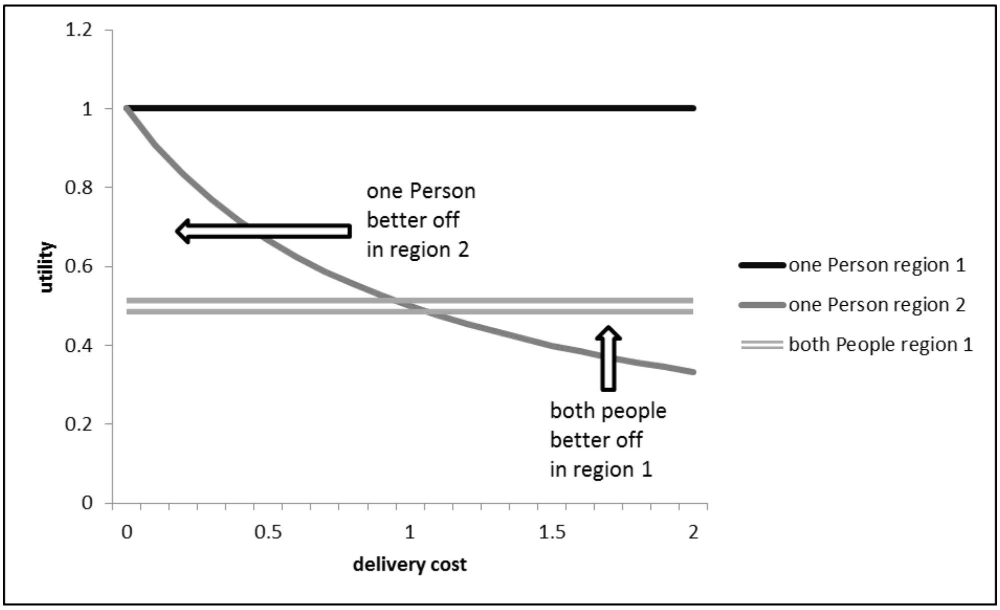
This equivalence of effect in the model framework is useful for conceptualising People buying goods with time directly in the ‘transmission belt’ model (chapter 6).

5.4.4 What drives agent location choice in the monocentric model?

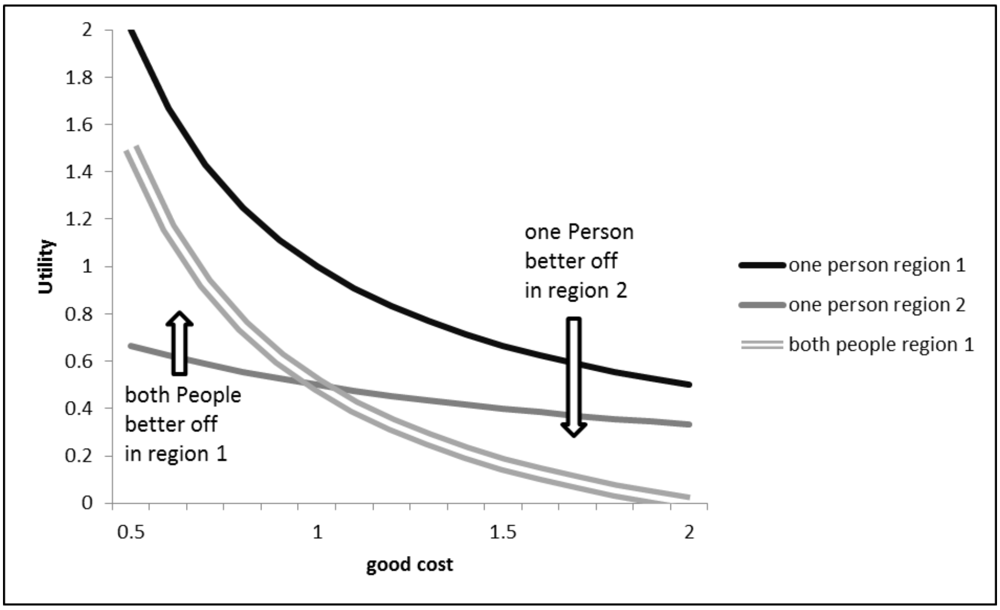
What causes this opposite reaction? This section presents a small thought experiment based on the model framework using the simplest possible assumptions. There are two discrete regions and two People. Region one has no distance costs, but in region two it costs 1 to deliver a unit of good. There is a single Firm in region one paying a wage of 1. It produces G at a base cost of 1. As there is a distance charge in region two, it costs 2 to buy one unit of good there.

The two People must decide where to locate to maximise their utility. They can avoid distance costs by staying in region one, but if they both reside there, density costs are incurred. Let the density cost be zero if only one Person is present, and 0.5 if both are in the same region. If so, subtract the density cost from the wage so it becomes 0.5 for each. Utility is linear: $G = U$.

Exclude the obviously poor choice of both residing in region two, where distance and density costs would be high: at least one Person would choose to minimise distance costs. This leaves only two possible states: both People in region one; or one Person in each region. Figure 5.13 shows the utility for these two states, changing base good cost and distance cost



(a) utility vs distance cost



(b) utility vs base good cost

Figure 5.13: Two People, two regions. Wages good cost = 1. Density cost if two People in same region: 0.5, subtracted from wage. No distance cost in region 1, distance cost = 1 in region 2.

(the spatial and non-spatial components of the total good cost).

It is better for one Person to move from region one to two where the single grey line is higher than the double. Increasing distance cost (sub-figure 5.13a) is agglomerating: while it is better for one person to move to region two when distance costs are below 1, as they go beyond this, sharing region one with the other Person becomes utility-maximising. Conversely, increasing the base good cost (sub-figure 5.13b) is centrifugal: as it increases past 1, one Person will find moving away from the centre the best choice.

If a Person does move from a high-density spot, they leave another Person behind with higher utility than them: as soon as they decide to move to region two (single grey line), the other Person gains a permanent utility advantage in region one (black line). The reverse is also true: if moving to region one is best, this will drag the utility of the existing resident down. These are both obvious externalities: moving from one to two increases both People’s utility - but the Person staying put actually benefits more. Moving from two to one increases only the mover’s utility; the other pays a price in increased density.

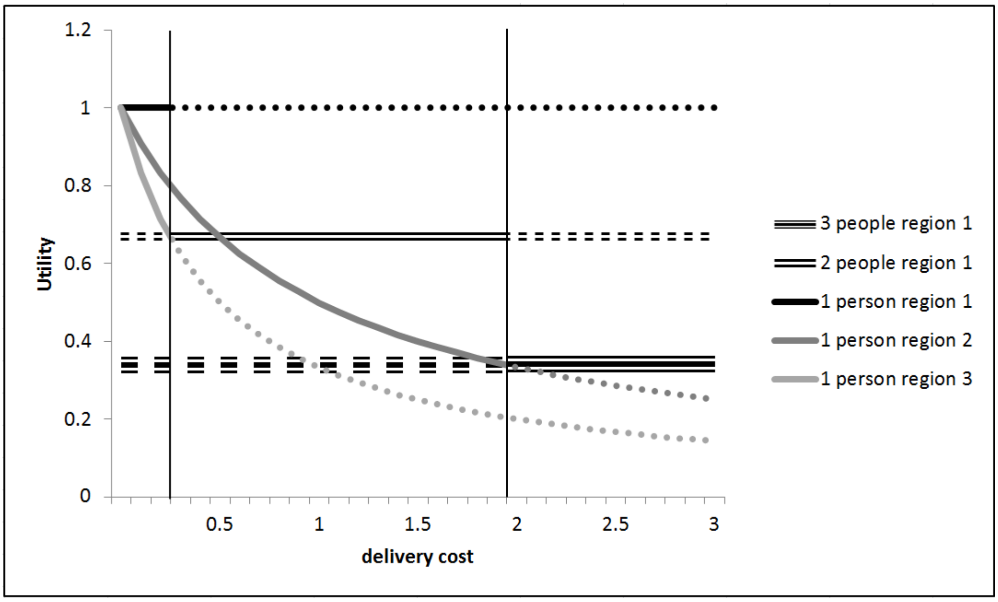
The presence of these externalities as a mechanism for equilibrium is in contrast to their role in neoclassical models. Skitovsky’s characterisation of them as “the villain of the piece” (section 2.7.1) contrasts with how they function in this ABM context, where market interactions are nothing *but* externalities. Each agents’ buying decision in one moment change the prices others face in the next. The spatial outcome of the results presented here are externality-based in the same way.

Extending the thought experiment to three People and three regions, with the third region a distance of 2 from the first - twice as distant again as region 2 (figure 5.14) - produces more complex outcomes, but indicates where the pressure for spatial equilibrium comes from. Wage and good cost are again fixed to 1. Density cost is $1/3$ if two People are present in one region and $2/3$ of all three People are present in one region.

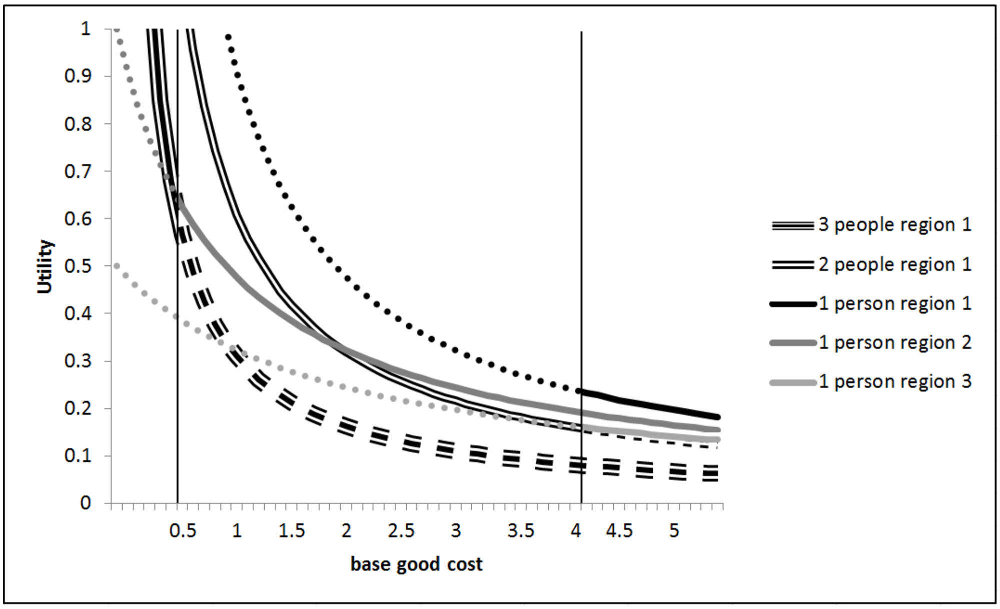
It is made easier to understand be there being only three rational states for People to take: all three in region one; two in region one with another in region two; or one in each of the three regions. Looking at the case of changing distance cost and base good cost again (figures 5.14a and 5.14b) shows where the change between these three options takes place.

Each graph is split into three sections at two key cross-over points where People are incentivised to move. Unbroken lines indicate the utility-maximising choice in each case. Increasing distance costs is agglomerating: starting at lower values, People’s best option is to spread out, one per region. Region three’s resident has no incentive to move to region two: if two people reside there, its utility drops. They should stay put until ‘two people in region one’ becomes the better option. That has a knock-on effect: the existing resident is made worse off. At the second cross-over, region two’s new resident now finds region one better, increasing their utility but lowering that of the other two in the process.

For base good cost, the direction of agglomeration is different, but the outcome is the same: the pressure for spatial equilibrium comes from the external effect of one Person’s optimal



(a) utility vs distance cost



(b) utility vs base good cost

Figure 5.14: Three People, three regions. Wages good cost = 1. Density cost if two People in same region: 1/3; if three People are in the same region: 2/3; subtracted from wage. No distance cost in region 1, distance cost = 1 in region 2; = 2 in region 3. Unbroken lines indicate optimal choices.

choice. Note, it is also possible to read off who will be incentivised to move. In the case of increasing base good cost, the utility of '2 in 1' - higher for cheaper goods - drops until a Person in region one finds moving right out to the edge the best option.

As noted in section 2.4, the assumption of spatial equilibrium means that (as Glaeser put it) "there must be no potential for arbitrage across space" (Glaeser 2008 p.4). The spatial equilibrium resulting from agent interaction in the models presented here do not allow agents to barter in the sense that 'arbitrage' might imply, yet they are able to collectively produce near equilibrium spatial outcomes through a series of interactions, each causing an externality that pushes the model towards that equilibrium.

5.4.5 Giving People a spread of costs

This section presents a dramatic illustration of the opposite effect of spatial vs non spatial costs. This is a way to very clearly demonstrate the implications of spatial and non-spatial costs having the opposite effect as well as showing how von-Thunen like outcomes can emerge through agent interaction using the density cost approach.

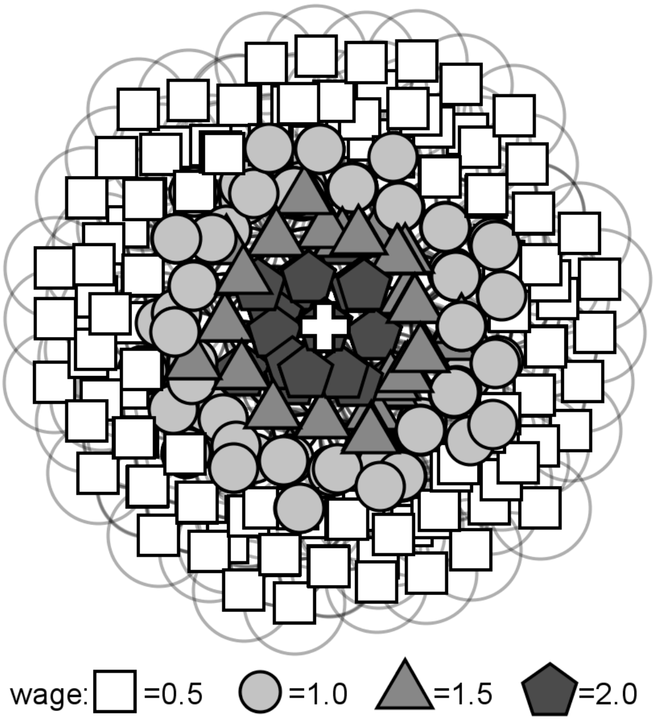
In figure 5.15, four Firms are placed at the centre of the monocentric model. Each Firm has variable costs. Each Person is assigned permanently to a single Firm and cannot change. A quarter of People work for each of the four; the shaded shapes indicate which. Sub-figures 5.15a and 5.15b show the spatial result if wages are varied between 0.5 and 2. Sub-figures 5.16a and 5.16b show the effect if commute cost is varied (from 0.25 to 1). Note that a higher commute cost is equivalent to a lower real wage. This variation is thus applied in reverse to make the diagrams comparable: in both cases, white squares have the lowest real wage and dark pentagons the highest.

As dark-pentagon People are given a higher wage, unsurprisingly they also have the highest utility - but the place where they can find this depends on whether the cost is spatial or not. Being wealthier in a non-space cost allows a Person to buy a denser location and gain from proximity by reducing their commute cost. Having a cheaper commute allows more distance to be bought: the outskirts become the utility-maximising location.

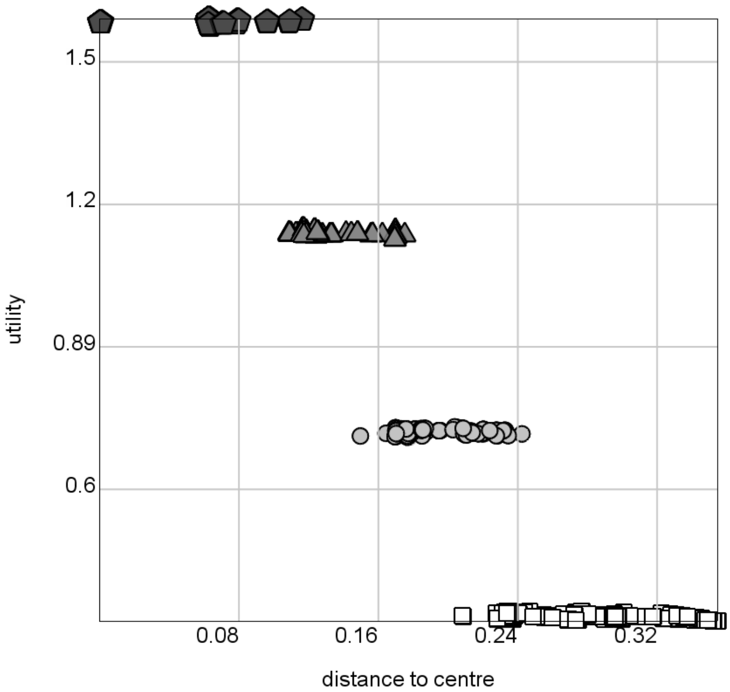
Where the von Thunen approach uses a web of assumptions that include competition via landlords and assumed bidding curves, this example illustrates that a basic proximity cost and a parsimonious set of model assumptions, acting through agents' 'second nature' decision-making externalised to the landscape, can produce a similarly clear emergent outcome. This is done with no market interaction. The discussion of whether this lack of a market is good or bad is taken up in the conclusion, in the context of the equilibrium outcomes of chapter 6.

5.4.6 Love of Variety in a simple two-Firm scenario

The CES function has a very particular spatial effect in the presence of density costs. As section 5.3.1 explained, if all goods have the same base price but vary in their distance, the closest will

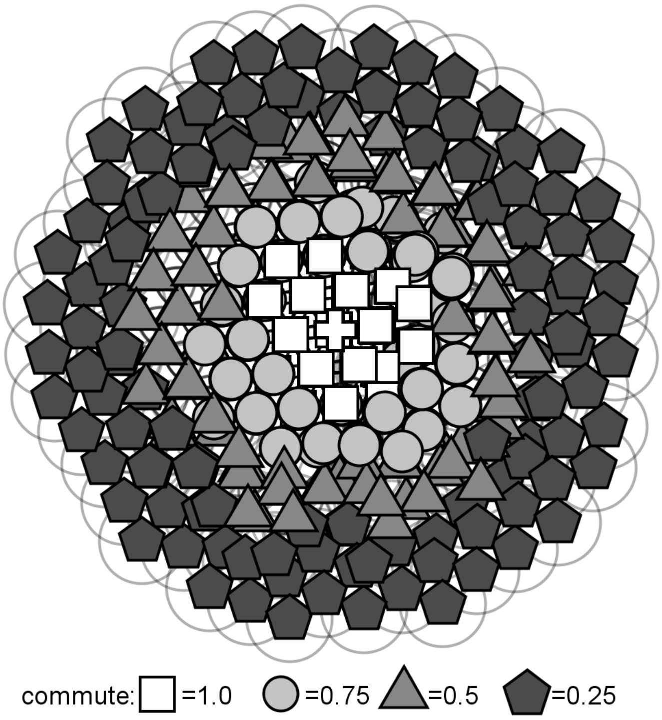


(a) People in R^2 buying from central Firm

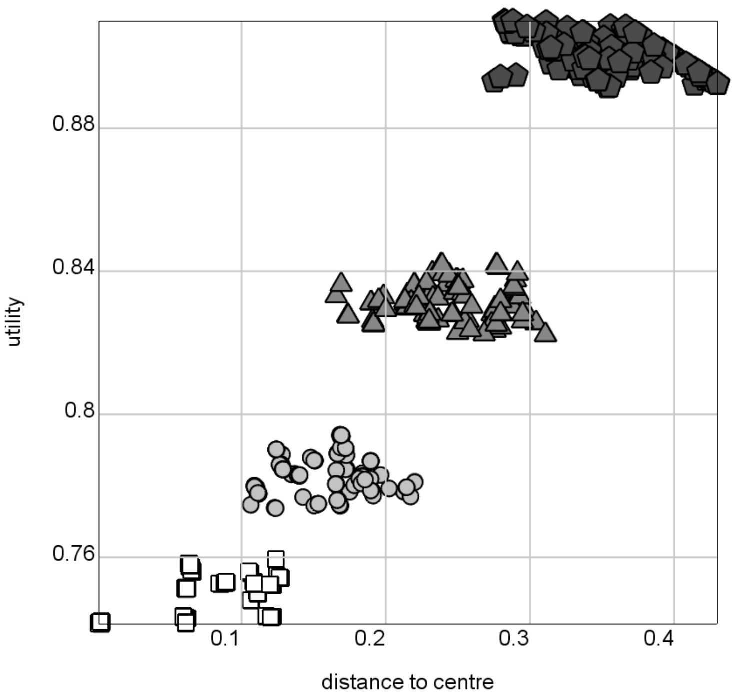


(b) Utility versus distance to centre for each Person.

Figure 5.15: People set with variable wage, a quarter each set to values shown in the key. This key is used in the visualisation and correlation. Wealthier buy proximity.

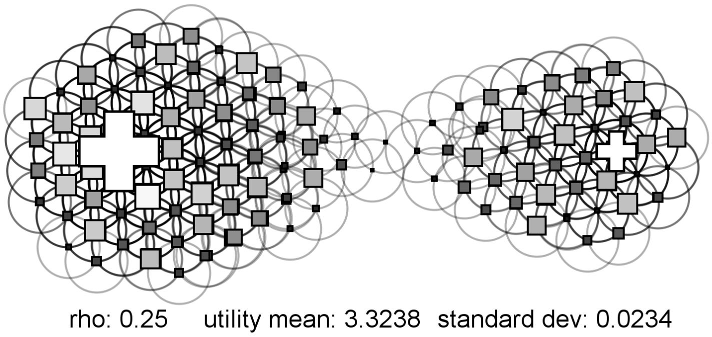


(a) People in R^2 buying from central Firm

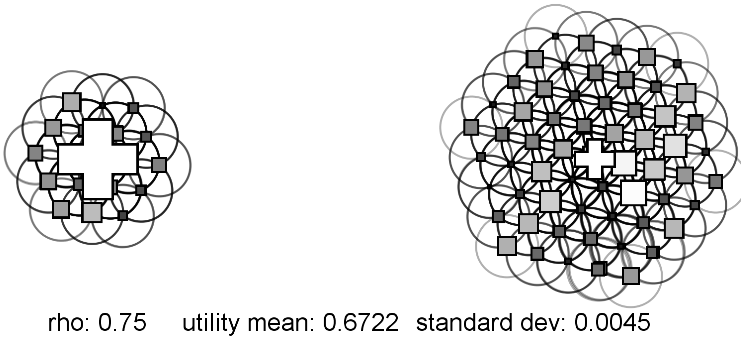


(b) Utility versus distance to centre for each Person

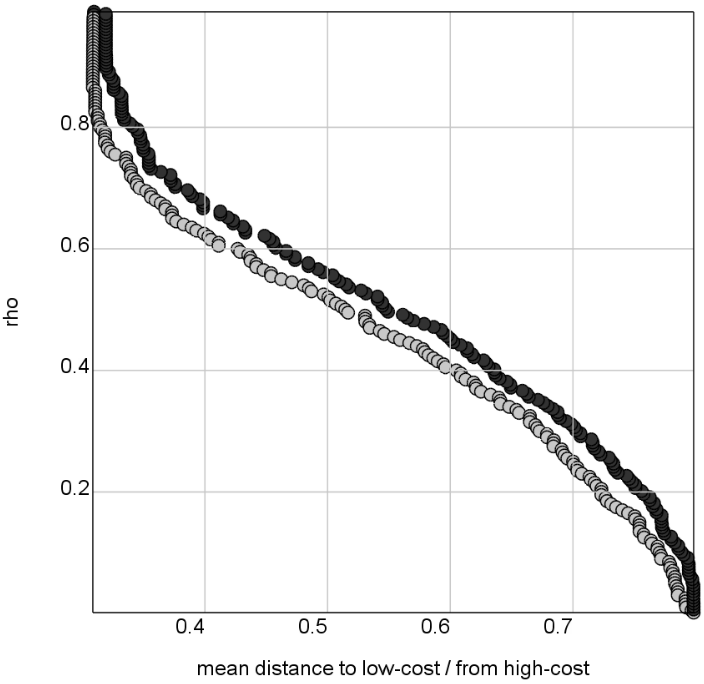
Figure 5.16: People set with variable commute cost, a quarter each set to values shown in the key. This key is used in the visualisation and correlation. Wealthier (paying less commute costs) buy distance.



(a)



(b)



(c)

Figure 5.17: Spatial effect of ρ: firm on right (small cross) has distance cost = 2, the other = 4. ρ is parameter-swept (positive in black, negative in grey) in increments of 0.005. Single datapoint shown: average for all People, taken at spatial equilibrium. 'Spatial equilibrium' is assumed when mean agent movement < 0.0001 average over 10 iterations.

be the cheapest. So when $\rho = 1$ (and assuming demand can be met), only the closest good will be consumed. As $\rho \to 0$, however, more distant goods increasingly enter into a Person's optimal mix. In the example here, all other parameters are set and only 'love of variety' is changed, to demonstrate what impact this can have upon spatial morphology in the presence of density costs.

Two Firms are 1 unit of distance apart. Distance costs for the two goods vary: one firm charges 2 per unit distance, the other 4 (the larger of the two crosses in figures 5.17a and 5.17b). 500 people are free to choose who they buy from and, in proportion to their 'love of variety', they will buy a mix of the two available goods. There is no commute cost: they incur only the distance cost of the good and density cost.

In figure 5.17c, the 'love of variety' parameter (ρ) is swept from 0.99 to near zero, and back again to show any hysteresis. The same stability conditions used in section 5.4.3 are also applied. It is correlated against all People's mean distance from the low-cost Firm (so the inverse is their distance from the other Firm). The higher the 'love of variety', the more People prefer to be near the more expensive of the two.

Figures for the mean and standard deviation of utility are provided in the visualisations. Utility is, unsurprisingly, higher for larger 'love of variety' (as $\rho \to 0$) but there is very little difference across all People: it is close to a spatial equilibrium outcome.

The role the presence of density costs is vital in this example. These set up a trade-off with proximity to Firms. Without a density cost, People could simply avoid distance costs by locating directly next to one of the two Firms (the 'black hole' result). They would choose the Firm with the highest distance cost for all values of ρ. In the presence of density costs, however, two distinct 'cores' emerge is ρ is varied. For higher love of variety, more People locate near to the higher distance-cost Firm. For lower love of variety, when goods are more substitutable, more locate near to the low distance-cost Firm.

Why are more People choosing to locate closer to a higher distance cost source of goods when love of variety is high (sub-figure 5.17a)? Two things combine to cause the shift in morphology: density cost (imposing a distance cost) interacts with 'love of variety' as it increases. In order to maximise the benefit of the mix of two goods, it becomes better to minimise delivery for the expensive good and buy the other from further away. If love of variety drops and People gain less utility from a mix, they increasingly prefer locating near to the cheapest of the two. Though note, in sub-figure 5.17b, some People are still centred around the more expensive Firm; this illustrates the trade-off that density cost causes.

There are points of indifference between the two 'settlements' that develop around the two Firms, with density cost acting as the equilibriating force. This example illustrates that changing direct cost is not the sole determinant of morphology: changing utility alone can make the difference.

This model set-up is also the basis for the final result presented in section 6.4.2.

5.4.7 Spatial equilibria are not always utility equilibria

As mentioned above, there are a range of density cost values that allow People to make a trade off between density cost and distance. If they are able to do this, utility equilibriates: People self-organise into a spatial arrangement that allows them all to be equally well-off. However, this is not the case for all density cost values.

The process begins on the outskirts of the settlement and works its way in to the centre as density costs are increased. Beginning at the outer edge, at a certain density cost, People find avoiding them altogether is their best choice: they locate on the edge where the available free space allows them to have a radius to themselves. At this level of density cost, spatial equilbrium begins to break down. If a Person, on their turn, finds moving right to the edge is best, the external effect is to lower density for others, thus relieving the pressure. The Person who moved first will be worse off than at least one Person they left behind.

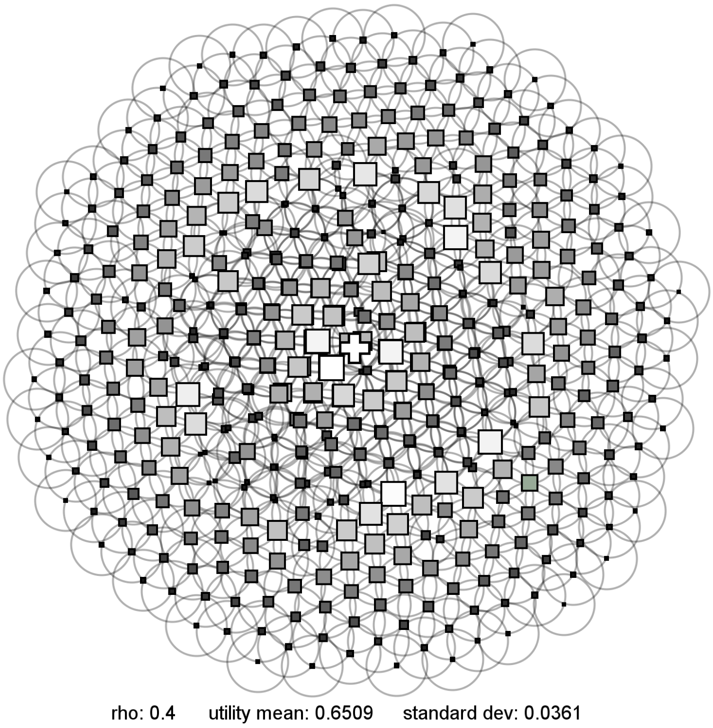
Increasing density cost has two effects. Firstly, it is dispersive as it pushes People out towards the farthest edge. Secondly, the boundary of where density cost avoidance begins creeps in towards the centre as it goes up - until eventually every Person finds avoiding density cost is their best option.

Figure 5.18 illustrates this for a density cost of 100; the boundary of complete density cost avoidance is not yet right at the centre. In sub-figure 5.19b, there is a characteristic dropping off of utility on the outskirts, indicated by the negatively sloped line of People's utility with those on the edge, with the furthest at a distance of approximately 0.7 from the centre. The visualisation of the agents' $R2$ spatial position in sub-figure 5.18a shows the spatial patterning of utility, with many in the centre still incurring density cost.

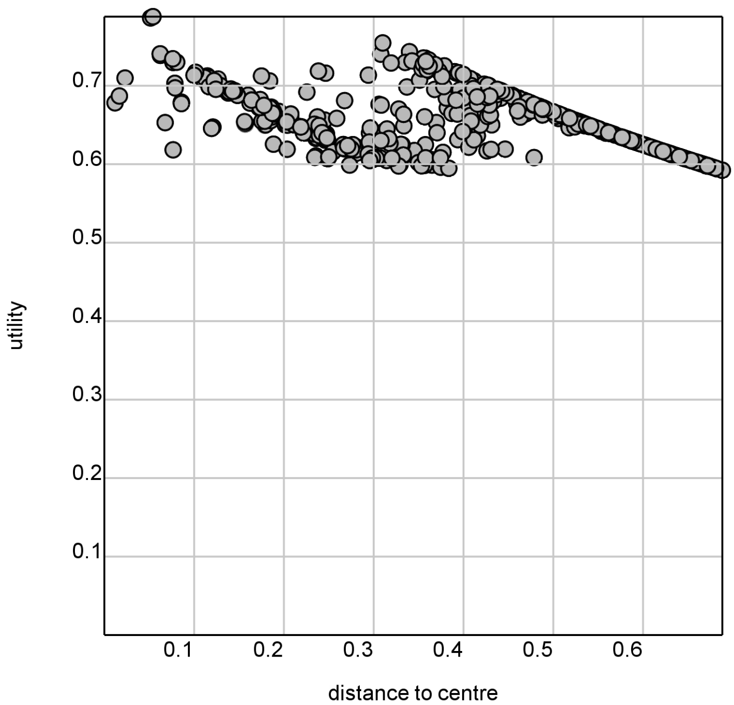
There is a clear second inner band in both figures where utility follows a negative slope. This is the result of two things. Firstly, the interaction of density cost with distance. Secondly, a path dependency in the model: gradually increasing the density cost reduces the utility of those who remain in place. They will only move if a higher-utility option presents itself. As sub-figure 5.19b indicates, there are distinct areas of the settlement with matching utility levels; People are indifferent between them. But if density cost increases, this will push those on the margin in the central area further out.

Figure 5.19 shows the result of all People avoiding density cost (at 500) altogether. People's utility drops the further away they are from the centre (sub-figure 5.19b). At this extreme, spatial utility equilibrium no longer exists. The remaining settlement is stable, however: no Person is able to find a better option. As the visualisation of their R^2 spatial organisation shows, those towards the centre have higher utility and have no incentive to give up their superior location. Those on the edge face the problem that moving further in will always incur the density cost. This is still a Nash equilibrium: no Person can take a unilateral decision that will make them better off.

The implications of the above sections regarding mobile People are discussed in the concluding summary for this chapter (section 5.6). Before that, the following section examines a

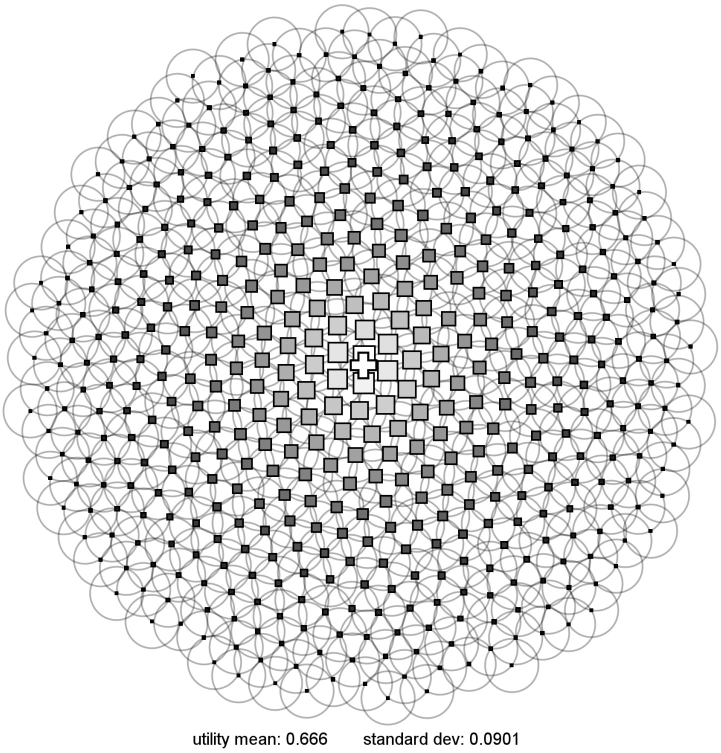


(a) R^2 space visualisation: density cost = 100

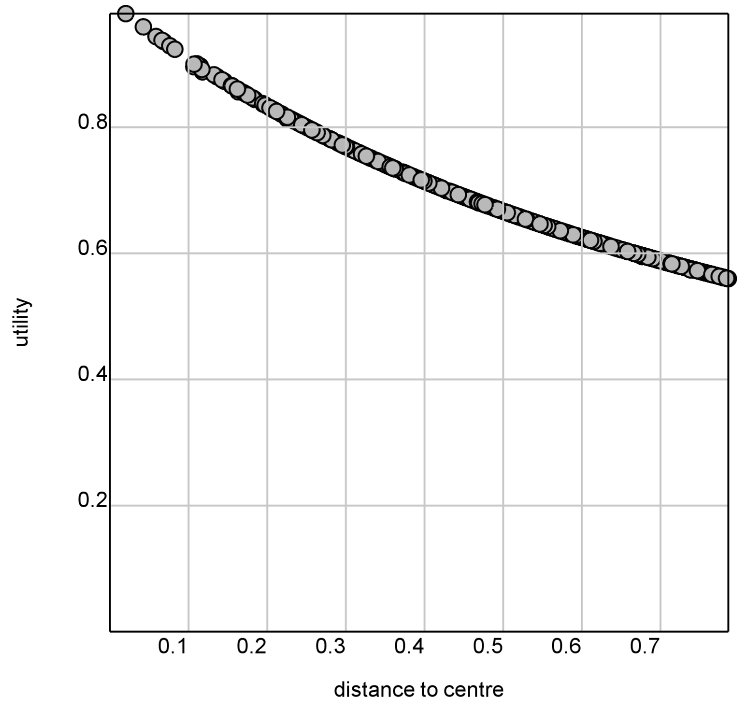


(b) Utility versus distance to centre for each Person

Figure 5.18: Density cost set to 100: some People begin to prefer avoiding them altogether and locating on outskirts. People within settlement holding on to better locations before moving produces path-dependent morphology. Datapoint for each Person shown in graph.



(a) R^2 space visualisation: density cost = 500



(b) Utility versus distance to centre for each Person

Figure 5.19: Density cost set to 500: no Person finds incurring any density cost is better. Settlement fully spread; utility drops off for more distant agents. Datapoint for each Person shown in graph.

scenario where People’s location is fixed and two Firms attempt to optimise.

5.5 Mobile Firms

In order to explore an example of Firms varying their location, the following two sections use a similar set-up. In both, all actors are on an R^1 line of length one. Two Firms can vary their location **only**: they have no control over price or wage, which are both set to one. Firms attempt to optimise their location using the method outlined in section 4.6.2: they take an action and use a feedback between timesteps to judge if the action made them better or worse off, in terms of the amount of goods they sold. If successful, they will pursue the same action (in this case, continue in the same direction) until it stops making them better off. If they become worse off between time steps, they interpret this to mean ‘my action was a failure’ and revert to their previous position.

First, section 5.5.1 spreads fifty people evenly over the line, mimicking in discrete terms the spread of demand seen in Hotelling-style spatial models. Firms are able to optimise outcomes based on different settings for ρ and distance costs, but their behaviour is somewhat erratic. Section 5.5.2 explores the reasons for this in more depth, looking specifically at the problem of DMR discussed in section 4.7.7.

This last section, by introduction coordination problems, is also sets the context for the final model chapter looking at the ‘transmission belt’ model approach to making spatial agent coordination work.

5.5.1 Two Firms optimising location on a line

In this section, Firms attempt to optimise their position as just described. Demand from People is spread evenly over the line: if only a single Firm inhabited the model, it would want to locate exactly at the centre. The model asks: can Firms find optimal positions through their own feedback actions, given the presence of another Firm?

The model diagrams in this section have the following set-up:

- Values shown in the diagrams are means taken over the preceding ten thousand model iterations, in order to display the emergent outcomes rather than the more erratic ‘day-to-day’ behaviour.
- The horizontal axis represents the distance of the R^1 line of length one.
- The average amount (over the previous ten thousand time steps) of each Firm’s good demanded by Each Person at their fixed location along the line is indicated above and below the central portion of the diagrams, and labelled for Firm 0 and Firm 1. The bars are shaded dark for lower values through to white for the highest.

- Each Firm's mean position is shown by the point of the arrow, Firm 0 pointing up and Firm 1 pointing down into their respective demand graphs.

- The position arrows have bars indicating one standard deviation from the mean on either side.

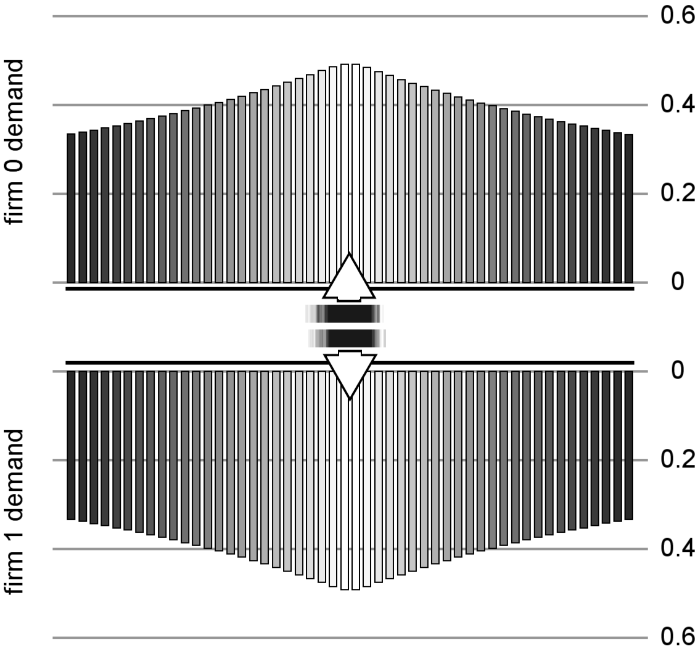
- To give a better sense of the range of location options the Firms attempt, their actual locations over the last ten thousand time steps are indicated impressionistically below the mean position arrows. A single line is drawn per location choice using an 'alpha' value so that more common choices build up and stand out in darker shades.

Note that, with no limitation on stock, each Person is able to have their full demand met (minus distance costs). So if both Firms are co-located (as in sub-figure 5.20a) the Person nearest to each Firm, when they are both central, is able to acquire a mix of 0.5 of each good.

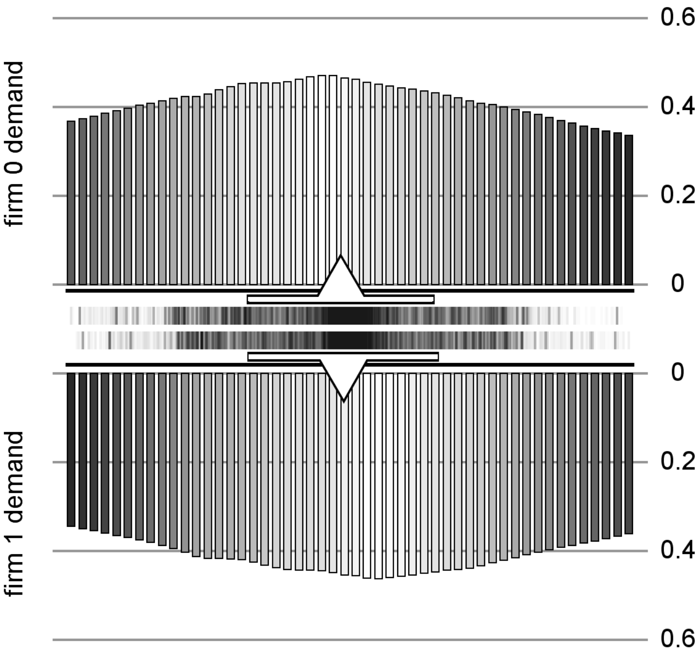
Figure 5.20 shows model outcomes for low distance costs (set to 1). In sub-figure 5.20a, love of variety is high ($\rho = 0.1$): the outcome is that Firms compete over a very small central spot, as indicated by a small standard deviation and a solid block of their actual location choices over the preceding ten thousand time steps. They are locked into a dynamic co-existence as their revenue-maximising spot is to provide a near-perfect mix of the two goods for all People. While sub-figure 5.20b shows much the same average outcome when $\rho = 1$ and goods are entirely subsitutable, the standard deviation and spread of location choices shows much more lively behaviour from the two Firms, spread out over the entire distance line. When ρ is higher, competition for individual demand is much more binary: the cheaper Firm receives the payment. Nevertheless, the average outcome is still central.

5.22 shows a single time-point snapshot to illustrate what demand does from moment to moment. In sub-figure 5.22a, where love of variety is high, demand overlaps for the two Firms. In contrast, sub-figure 5.22b shows how demand is binary when goods are fully substitutable: People buy only the cheapest available to them, not a mix of the two. This cut-off does not show up in the averaged data in figure 5.21.

In figure 5.21, distance costs are raised to 20. This increases the impact of distance on utilty, given the line consists only of distance 1. Again, the sub-figures vary love of variety. In sub-figure 5.21a, when love of variety is high, there is a tendency for the Firms towards the centre again but their location choices are considerably more spread out as they react to distance costs. In sub-figure 5.21b, when goods are subtitutable, a Hotelling-like situation emerges: Firms (on average) find their revenue-maximising position at different ends of the line - though on average over ten thousand timesteps, there is an overlap in demand between them as they try new revenue-maximising locations. In the presence of higher distance costs, the Firms' spread of locations is actually higher when ρ is lower, the reverse of the situation seen for lower distance costs. This is due to Firms more successfully 'locking' into a spread of revenue-maximising locations in the high-substitutability scenario.

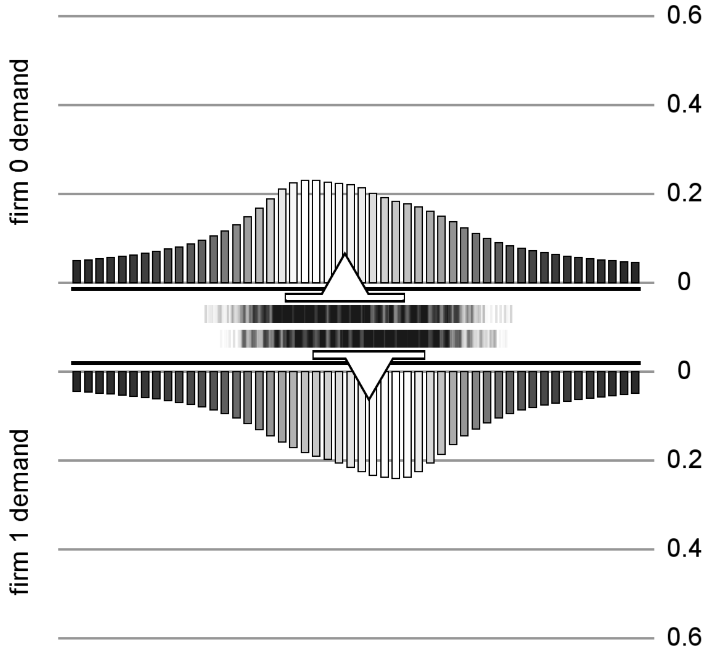


(a) High love of variety (ρ = 0.1); distance cost = 1

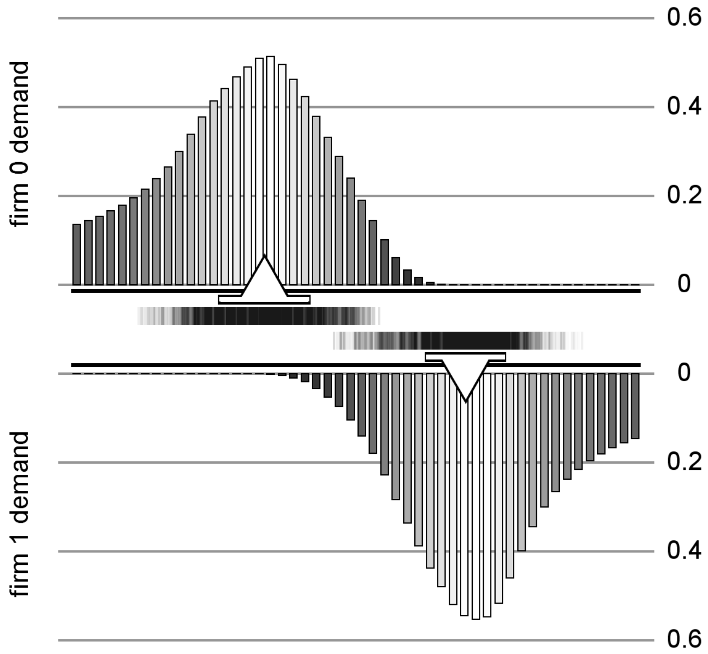


(b) High substitutability (ρ = 1); distance cost = 1

Figure 5.20: Two Firms optimise position on an R^1 line of length one. Fifty People spread evenly along the line from 0 to 1. ρ varied, distance cost fixed at 1. Means taken over 10000 timesteps for Firm position and People’s demand. White arrows shows Firms’ mean location; white bars show one standard deviation each side of this; thin black vertical ‘barcode’ lines between Firm arrows show actual positions taken by Firm over previous 10000 timesteps.

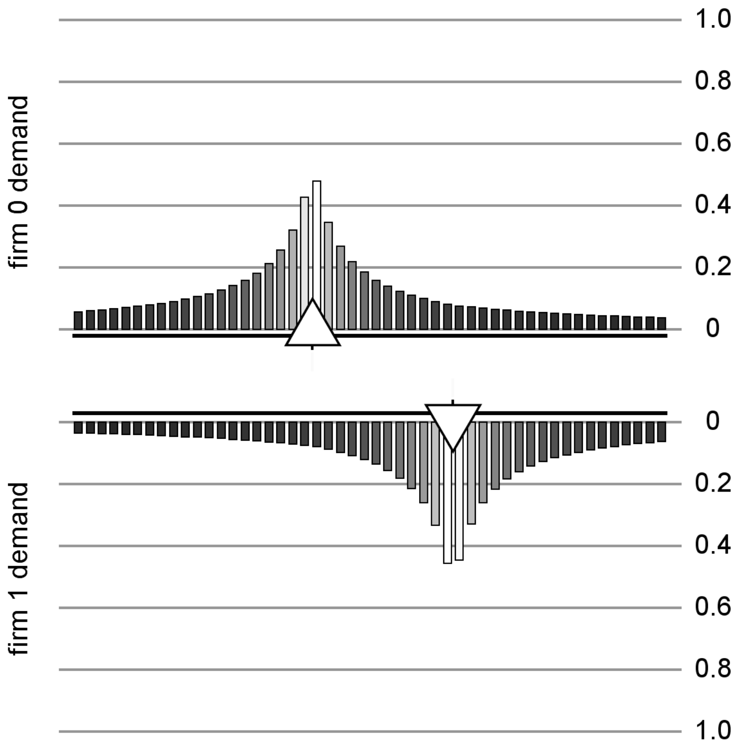


(a) High love of variety (ρ = 0.1); distance cost = 20

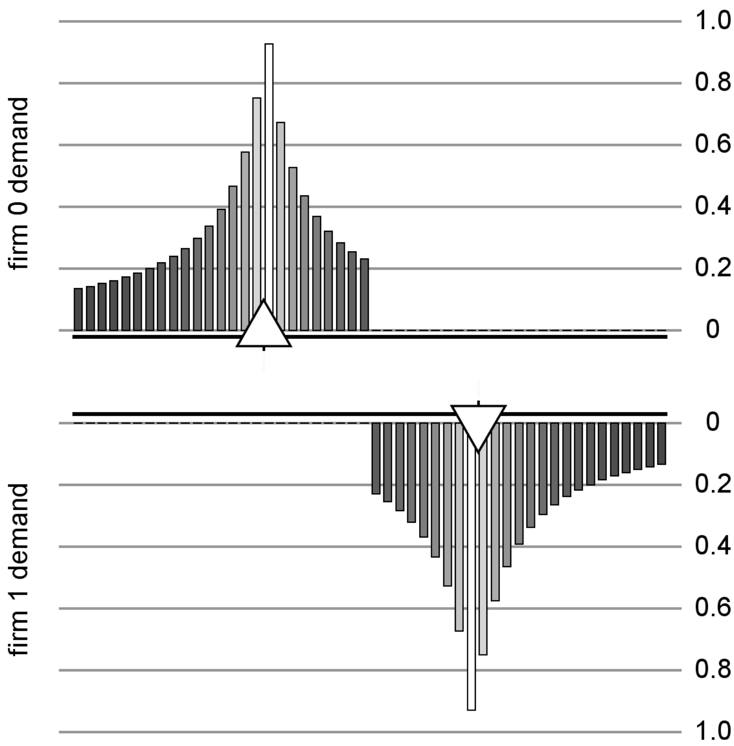


(b) High substitutability (ρ = 1); distance cost = 20

Figure 5.21: Two Firms optimise position on an R^1 line of length one. Fifty People spread evenly along the line from 0 to 1. ρ varied, distance cost fixed at 20. Means taken over 10000 timesteps for Firm position and People's demand. White arrows shows Firms' mean location; white bars show one standard deviation each side of this; thin black vertical 'barcode' lines between Firm arrows show actual positions taken by Firm over previous 10000 timesteps.



(a) High love of variety (ρ = 0.1); distance cost = 20



(b) High substitutability (ρ = 1); distance cost = 20

Figure 5.22: Single time-point snapshot: two Firms optimise position on an R^1 line of length one. Fifty People spread evenly along the line from 0 to 1. ρ varied, distance cost fixed at 20.

Underlying these statistical outcomes, Firms ability to optimise their location using a location/revenue feedback mechanism appears quite weak. Why? The next section explores this problem in more depth.

5.5.2 'Data, meaning, response' with two Firms

The aim of this section is to investigate the interaction of Firms' signals to understand the DMR problem (section 4.7.7). It attempts to boil down the problem to the simplest possible elements to show that, even in a the most basic of examples, Firm agents attempting to optimise revenue through 'sensed' market information alone can struggle to do so. It is an important example for understanding the limitations of Firm optimisation in the transmission belt model in chapter 6.

The model set-up is the same as in the previous section, apart from one key change: People are located only on a small subset of the R^1 line, between points 0.25 and 0.75. Each of the two Firms begins at opposite edges of the line. This model asks the Firms to play a 'game'. A game terminates either when both Firms have succeeded in getting inside the range of space occupied by the People or ten thousand days has passed.

The Firms can attempt to improve their revenue only by moving. Without competition from each other, their revenue-maximising position is near the centre of the People, where distance costs are minimised. Firms are making only this one decision: where to move? Getting closer to their market should, in theory, increase income - and indeed, if only a single Firm exists, it has no problem finding an income-maximising path. The rest of the world is static, so any change must be a consequence of its own actions. As soon as a second Firm is added, however, that is not the case. Confounding factors then mean they may fail to interpret the meaning of their cost signals correctly.

In order to collect the data for figure 5.23, Firms play a large number of these games: utility noise levels are parameter-swept (see below) and for each value the random seed is altered a given number of times to then produce an average for all games with that value.

What is the rationale behind this? To put it into DMR terms (see section 4.7.7) a Firm has only one datum to act on, a change in revenue. The meaning of an increase is taken to be, 'my previous action was successful'. Conversely, a drop in revenue is taken to mean that the action was counter-productive. This leads to the subsequent response: either repeat the action or reverse it. The broader meaning for the Firm in this situation is: 'my actions can determine my revenue levels'. Problems arise when this meaning is incorrect: a Firm may take an action, but if its revenue levels are not determined solely by that action, there is scope for misinterpretation, and thus a faulty response.

The aim is to provide a 'god's eye view' of the Firms' goal in as simple a cost-signal setting as possible. The optimum they are aiming for is clear. Firms aiming between a subset line of People is, from the modeller's point of view, an entirely objective and clearly defined goal. At the same time, it is opaque to the Firms, since their 'view' is restricted to revenue data. Although this looks like a Firm location model, here it is being used to examine the more general problem of conflicting signals involved in DMR-type problems. Its purpose is to think about how Firms make decisions given (possibly noisy) information. Even putting aside that firms obviously do not 'grope' their way to more productive locations, they would in reality be able to make much more informed location decisions. The question for this section is: in as simple an example as possible, what gets in the way of a Firm making an appropriate response to correctly interpreted data?

In this example there are five key factors that impact on revenue. The first is a Firm's own actions - just moving to a new spot. The second comes indirectly from the actions of other Firms: they may move to a location that leads People to shift demand to them, and this in turn may make other Firms' revenue-flow negative where it would otherwise have been positive. The other three factors all relate to the connection between People's demand levels and noise. If People are using pure constrained utility - thus providing a crystal-clear demand signal - Firms have no problems making correct decisions about changes in revenue. However, if demand is less clear than this from turn to turn, revenue signals become less reliable. More than this, as noise increases, the value of ρ - the level of love of variety in the model - begins to determine Firms' success rate.

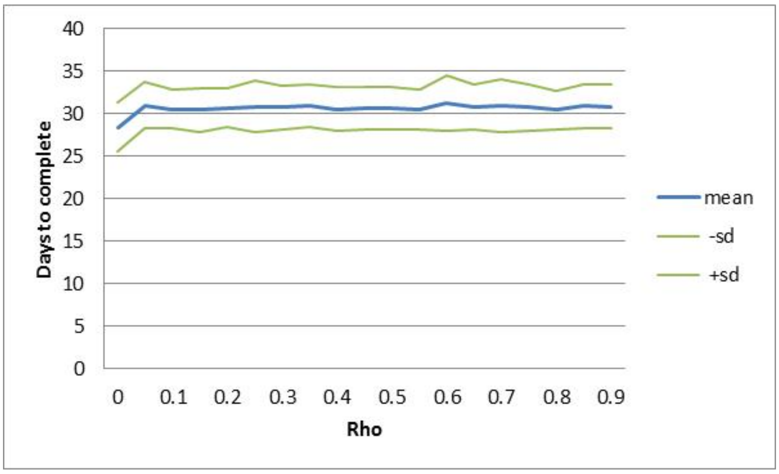
In the model framework, noise naturally appears when using the hill-climber algorithm, and it is noisier the less iterations are used. As section 4.6.1 explained, the algorithm starts by guessing budget slices for all available goods. With fewer iterations, these guesses are further from the 'perfect' optimisation, and thus demand fluctuates. It is in running hill-climber algorithm models that the impact of noise on Firm optimisation becomes apparent. However, for the sake of computational expense and feasibility of large parameter sweeps, a noise term has been added to the constrained utility methods.[1]

The noise term just adds a random fluctuation to the optimal good quantities the constrained algorithm finds. For instance, if $noise = 0.5$, the optimal quantity is randomised in a range of plus or minus half its total value.

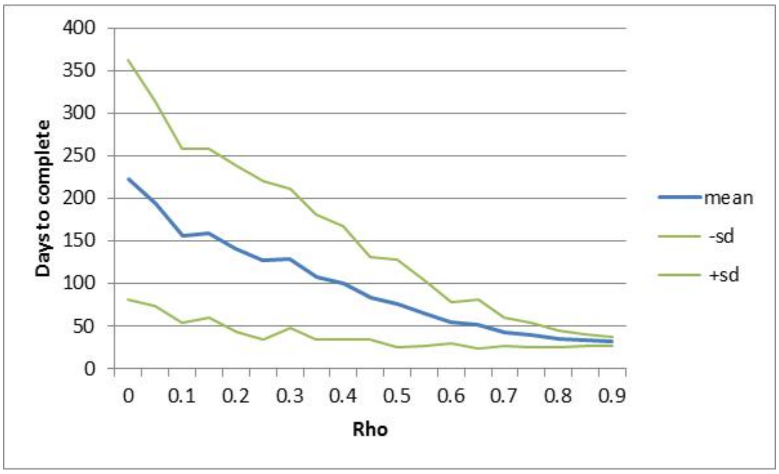
Firm and Person numbers are the final two factors, both impacting on demand noise level, if any is present. Both have a similar effect, for different reasons. Increasing the number of People means a larger source of demand for each Firm's goods, thus tending to average any noise out as these rise. Rather less intuitively, increasing Firm numbers reduces noise when using the hill-climber algorithm simply because the variance of size of budget slices drops. A budget randomly split between two Firms will have a much higher variance than one split between ten or a hundred. Constrained utility noise is equally lower, as it is applied proportionally to the quantity of optimised good amount.

Figure 5.23 graphs the average number of days it took both Firms to get inside the line of People, for a range of values of ρ up to 1, in R^1. A hundred randomisations were averaged for

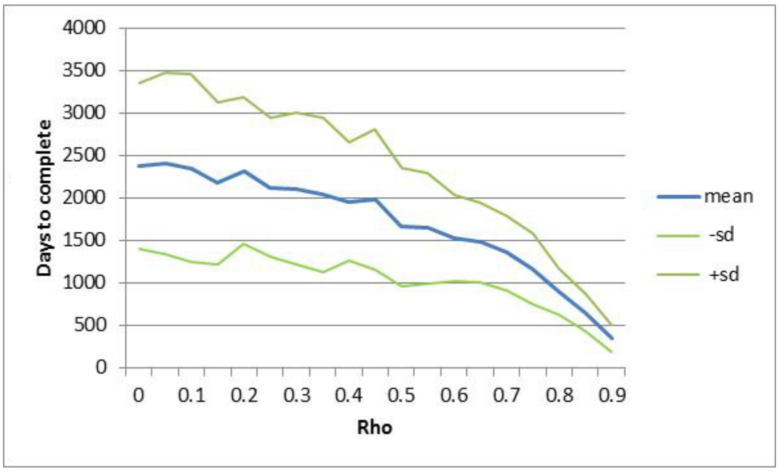
[1] While a noise term can be added to the constrained algorithm, it is quite hard to make it economically sensible. This requires staying in budget: the hill-climber algorithm does this by design. Adding noise to constrained utility, however, may mean People appearing to spend slightly too much or too little. Avoiding this is actually a large optimisation problem in itself, since there is no way to 'read back' from constrained good quantities.



(a) $noise = 0$



(b) $noise = 0.1$



(c) $noise = 0.5$

Figure 5.23: Firms compete to reach centre of market. Number of days to success on vertical axis, value of ρ on horizontal. Mean and S.D. for a hundred randomisations of each value point. Three different noise levels. Note for figure 5.23c, days to complete are an order of magnitude higher than for 5.23b.

each value of ρ. Three different noise levels were tested: it can be seen that when no noise is present, ρ makes no difference to the success of Firms’ decisions. As the noise level increases, however, ρ begins to be important. High levels ($\rho \rightarrow 1$) present no problem, regardless of noise level. However, as love of variety increases ($\rho \rightarrow 0$) noise increasingly interferes in a Firms’ ability to correctly identify revenue-improving positions.

Why does ρ have this effect? Because if love of variety is higher, any change in a Firm’s location has a higher external effect on the demand the other Firm faces. This makes the signalling less clear and, in the presence of noise, more likely to cause a Firm to interpret information from the market falsely.

5.6 Summary

A transparent location example is used as a way to think about other possible DMR decisions involving price and wage setting. There are other reasons why varying good price and wages present problems, discussed further in section 6.2, but the DMR model suggests that noise issues exist in any market where Firms require some form of ‘sonar’ signal back from their actions. While Firms in section 5.5 are able to optimise their location, the outcome is far from the cleanly deduced stable competition outcome examined in Hotelling’s original two-firm line model. As section 5.5.2 explored, in sticking to a DMR approach, Firms have only a narrow view of the effect of their actions.

One option is to allow each Firm to separately pursue its optimal outcome until some maximum is reached while holding all other actions static. This would allow them to fully optimise having removed any possible conflicting signals. As mentioned, single Firms certainly have no problem identifying their optimal position: it is only in the presence of other Firms that it becomes challenging. This method of fully separate decisions is used in Plummer *et al.* (2012), taking their idea from Schendel and Balestra (1969). They suggest “retailers temporarily alter their prices, observe the effect on their profit objective before competitors respond, and then use this information to institute a price change” (p.543), thus avoiding conflicting signals.

However, the goal of the thesis is to examine how to let agents make decisions independently. The DMR example is a simple case, but it presents even more difficulties for the transmission belt model in chapter 6 where constant interacting signals are an unavoidable element of the model’s construction.

In the ‘mobile People’ models, an agent approach makes it relatively easy to equip agents with a ‘second nature’ reaction to proximity (section 2.7.4). The density cost produces a beguiling outcome, able to produce spatial equilibrium with no market arbitrage. Urban economic models built on the spatial equilibrium assumption also rely on the ‘first nature’ properties of land - though land cost itself is ‘second nature’: it is a property of demand. The density cost approach goes some way to capturing this, in a scenario where land is otherwise homogenous (having no intrinsic value that might make one spot more valuable than another).

The underlying message is that, if agents are provided with an ability to find utility opportunities better than their own, in a situation where their actions interact spatially, equilibrium can result without the need for any specific market mechanism. As section 5.4.4, the density cost produces this outcome through the interaction of People's optimising choices and the externalities of those choices. The outcome works precisely *because* of its sequential nature: the choice set of the next Person is effected by the previous Person's decision. As with the way in which strategy considerations are avoided when using economies of scale in chapter 6, actually ruling out game-theoretical considerations is suited to spatial analysis and Oeffner's point about 'irreducible uncertainty'.

As section 2.4 explained, Fujita *et al.* note that the 'monocentric' assumption cannot help directly if one wants to understand endogenous forces leading to settlement location, number and size. But Blaug makes an important point: the power of the monocentric model is precisely in its clarity. Von Thunen knew the 'isolate state' was abstract:

> "in reality, Thunen observes, differences in fertility of the soil which are not themselves related to location will give rise to ground rent in the same manner as do differences in proximity to the central town." (Blaug 1997 p.598)

From this point of view, the 'monocentric' approach is not about a specific settlement pattern, but understanding the polarity and magnitude of spatial forces. The conclusion returns in detail to the models in this chapter, placing them in the context of the following chapter's findings as well as the aims and objectives of the thesis.

Chapter 6

‘Transmission belt’ model results

6.1 Introduction

This chapter presents a model of the ‘transmission belt’ between production and consumers able to produce equilibrium outcomes. These are ‘general’ in the sense that they link supply and demand. It developed through the logical journey of building the model framework, with the models presented in chapter 5 representing some of the important stepping stones revealing which choices could possibly work and which were less promising.

The chapter is organised as follows:

- Section 6.2 discusses elements that were rejected as possible routes to creating a ‘general’ model capable of fully linking supply and demand in a completely open geography where agents were able to make decisions separately.
- Section 6.3 introduces the structure of the transmission belt model. The following subsections develop each stage of the transmission belt model systematically, looking at the following: the stock target mechanism used by Firms to achieve an ‘efficient’ production level; Firms’ scope for using ‘markup’ as a profit mechanism and the problems with implementing this fully into a competition dynamic; a non-spatial explanation of the model’s key ability to transition between two different production ‘regimes’, with one Firm emerging as a more-efficient producer.
- Section 6.4 introduces distance into the transmission belt model. This is done initially with a two region model before examining the dynamic as more agents are added onto a continuous line. This section ends by combining the transmission belt model with the density cost approach, in order to illustrate a way in which the production regime transition can determine spatial morphology.
- The chapter ends with a brief summary; much of the analysis of the transmission belt model takes place in chapter 7. For example, while comparisons to GE are discussed

throughout the chapter, there is a direct comparison to the core model made in section 7.1.

6.2 Moving from 'partial' to 'general'

Producing partial models is a relatively straightforward task compared to getting agents to fully coordinate economic activity. Of the various approaches attempted within this research, many were not able to recreate stable outcomes. Here are two quick examples. One: if the goal is to connect People's labour to production, and thus to stock levels, but also to let Firms set good and wage prices, it is very difficult to get sensible stock levels. This chapter presents a model that solves this stock problem, but it does it by avoiding the DMR problems identified in section 5.5. Firms using price-setting behaviour have a hard time producing stable economic outcomes: stock tends either to crash to zero or to increase indefinitely, without prices finding an efficient stock level. It is also difficult to find any compromise position between partial and general models. For instance, Firms can be given unlimited stock to attempt to avoid the stock problems just discussed.

Second, the problem of price signalling is thorny. The simple random money flow model in section 4.7.2 demonstrated a key part of the problem: treating money literally as a rivalrous good hits up against the tendency for the flow among many agents to 'random walk' into an uneven distribution. The model result even with agents making more sophisticated maximising choices is that some Firms can end up with the entire money supply - more than they need - while others find themselves in the catch-22 of not being able to pay the wages to produce stock to sell to get money to pay wages. Any economic agent model faces being stuck between this naıve approach to money and finding some other method capable of signalling between agents.

The approach of this chapter is to cut out one layer of complexity by making People exchange 'input' (their time) for productive output directly, where the price for the latter is the former. This avoids the need for an intermediary to signal price. This builds on the idea from section 5.4 that the important spatial economic distinction is between the spatial and non-spatial component of the costs People face, not whether they are goods-based or wage-based.

The conclusion goes into more comparative detail about the whole set of dependencies in the thesis model framework and puts them in the context of the role of assumptions in model building.

6.3 The 'transmission belt' between production and welfare

This section presents a model that connects supply and demand resulting in stable equilibria. Production with economies of scale are linked to consumers' welfare. It builds on concepts developed in the preceding analysis, finding a solution that allows efficiencies from economies of scale to feed through to improvements in utility. The key features of the model are:

- People get goods directly as their wage. They buy goods directly with their time (a flow of ‘one day per day’) which then enters into Firms’ production functions. Goods are sold using the ‘goods plus distance cost’ equation from section 2.6.1, $P_g = p_g + (c \cdot d)$. Goods and wage decisions are thus combined into a single utility decision.

- Firms charge base price p_g (the cost in a Person’s time to buy one unit of good). Goods are exchanged directly for People’s time, which then goes directly into the production function. Any good-movement costs ‘melt away’.

- Production has economies of scale: higher input produces a higher per-unit output. A ‘smooth’ economies of scale function is used where any increase in input translates into a higher-ratio output; no fixed labour component is required (section 2.2.5).

- Economies of scale work as described in section 4.7.5: each Person ‘leaves’ their time with each Firm they contribute to until their next decision point. The production of stock is based on the amount of contributed time at each decision point.

- People use a CES function to buy goods and distribute their time to Firms. This circumvents Bertrand competition-like problems with wage-setting, where a wage offer only minutely higher than others will automatically be selected. It means that each Person buys at least a little from each Firm - and thus contributes time to each - unless $\rho = 1$, in which case goods are fully substitutable and the cheapest will be bought.

- Demand for goods is always met: from the point of view of consumers, there are no stock limitations. The key advantage of this for the model is that a constrained utility function can be used safely: a consumer buys a given amount based on the set of prices they observe. They face no risk of stock running out.

- Firms must then make sure that stock levels keep close to the amount being produced (or more than this amount if they are seeking to make a profit; see below). They do this by setting good costs to target a stock level - net zero stock if they are not attempting to skim a profit.

In contrast to the two-timestep feedback decisions of Firms (used in section 5.5) stock levels are controlled through ongoing ‘granular’ price adjustments. These are, as Leijonhufvud puts it, ‘laws of motion’ (section 3.3.1) describing how Firms respond to changes in stock. This adjustment occurs every time a Person buys. It is a ‘market mechanism’ in that price signals are used by Firms to optimise a particular quantity though, as section 6.3.2 discusses, while prices are used and interact to produce the equilibrium outcome, Firms are not able to directly compete. The ‘granular’ process that occurs each time a Person buys a good is as follows:

- A Person chooses their best Bundle as described in section 4.6.1 using their unit of time as the budget. Since it is a CES function, all Firms will receive some share of this (see section 5.3.1) minus the distance cost that 'melts away'; that is, any time a Person spends on distance does not enter into production.

- When each Firm receives a signal that a Person is buying, it pays with the appropriate quantity of good based on its current price and stock is reduced accordingly.

- The quantity of input the Firm receives in return goes straight into production. The production amount is worked out as described in section 4.7.5: economies of scale come from all input from recent interactions, and the actual productive output is calculated when the Person buys. It is proportional to the time that Person contributes. The method of achieving this is to use the smooth economies of scale function described in section 2.2.5.

- There is now a net stock level from this one interaction, which the Firm uses to guide their price adjustment. If stock is more than zero, demand is lower than production, and so the price is lowered to try and sell more. If stock is less than zero, demand was higher than production, and so the price is increased to sell less.

- The change in price is proportional to the amount that stock levels deviate from the target amount: $\triangle p_g = f(stock)/x$, where x controls the magnitude of the price change. See section 6.3.1 for an explanation of the chosen function for stock targeting.

Stock can go below zero for a time; this is vital for the price signal to target it from both sides of the zero line, though the net stock target over time is ideally zero. This zero target can be made physically 'plausible' by assuming Firms have stock buffers. Zero then becomes their target production level rather than a physical lower limit. Indeed, the model can be set up to reach a given arbitrary level of production and then work from there, leaving a positive buffer. For the sake of ease of discussion, though, a target of zero is assumed (except for section 6.3.2 which uses the choice of stock target as a way for Firms to seek profit).

Secondly, in contrast to the 'blind versus rational' problem described in section 4.7.6, the transmission belt model *can* work with the fixed-cost, core model style of production function - but only if total input is able to exceed the fixed labour requirement. Otherwise, People buy goods that the Firm then cannot possibly make, since the production function becomes negative. Because of the fact that People 'leave' their time with Firms over an iteration, the fixed labour requirement can be met, but it leads to some nonsensical situations like the one just described. The smooth economies of scale function produces more reliable outcomes for the transmission belt model.

Lastly, a deeper explanation of the role of stock targeting in the model and its relation to efficiency (details of the actual mechanism are discussed below in section 6.3.1). The motivation was to find a way of linking labour input to good output and consumption. This requires all goods to be priced correctly. Many of the coordination problems involved in doing this have been discussed, especially the issue of getting time input to feed through to more efficient good production via economies of scale (section 4.7.5).

What does ‘efficient’ mean in this context? A given amount of time input into a Firm equates to a given level of good output via its production equation. The most efficient level of production from the point of view of People is where net stock from *all* Firms remains as close to zero as possible. Less than zero stock is both nonsensical in the long run and indicative that economies of scale are not being exploited as fully as they could be. Positive stock levels indicate that People are not getting all the available output that their labour is creating. The zero stock target is thus the point at which the benefits from production are maximised and feed through to utility levels.

This approach described above succeeds in finding efficient equilibria. In situations with many Firms and People, each Firm is able to find a unique price that keeps its own stock levels efficient while buying and selling to each Person. Their price signals find a balance point between production and People’s utility demands as each Firms’ price decisions interact with each other. (As explained below, the external effect of price changes are vital.)

6.3.1 The stock targeting mechanism

This section demonstrates how the transmission belt model and the stock target mechanism work. It compares two model runs with no distance cost: a ‘one Firm, one Person’ model and another with fifty of each. A third model run adds a distance cost and randomly locates fifty of each agent type to illustrate the model’s ability to equilibriate when agents face heterogenous costs imposed by location differences.

For the examples used in many of the following sections, the smooth economies of scale equation is kept in its normalised form of $t^2 + t$, with neither the curve or straight line parameter changing its scale. This allows for some useful simplifications that aid understanding of what the model is doing.

In the case of only one Firm, and assuming there are no distance costs, there is a very straightforward deduction of the correct price for ‘clearing the market’ and this can be compared with the model outcome. If there is only one Firm, all input goes into that single Firm: People have nowhere else to buy from and will spend everything they have there. So (assuming no space costs) the Firm’s output is directly proportion to the number of People, since each Person contributes a single unit of input. A Firm wants to clear its stock - to keep it as close to zero as possible. What price can it set to do this? Price is the output a Firm gives for a unit of input. People pay with a unit of input each so, in this simple case, price is always just the number of people divided by total output. Using the smooth economies of scale function and setting both the curve and magnitude parameters to one, output is just $t^2 + t$ (where t is the total input from all People). In cases where only one Person is contributing one unit of t to each Firm, the price equation reduces to $P = 1/(t+1)$. In the case of one Firm and one Person, then, at equilibrium $P = 0.5$.

In model runs with no distance cost, deducing the equilibrium price is straightforward for other mixes of Firm and Person number, as long as demand is either spread evenly or being input into a discrete subset of Firms, as happens when one Firm emerges as a selected site for large-scale efficient production (see section 6.3.3 below). If demand is spread evenly, output and price is the same as above. If, for example, two hundred people input into fifty Firms evenly, that is four People's time per Firm. Equally, if one efficient producer emerges, all fifty People input their time into that Firm's smooth economies of scale equation.

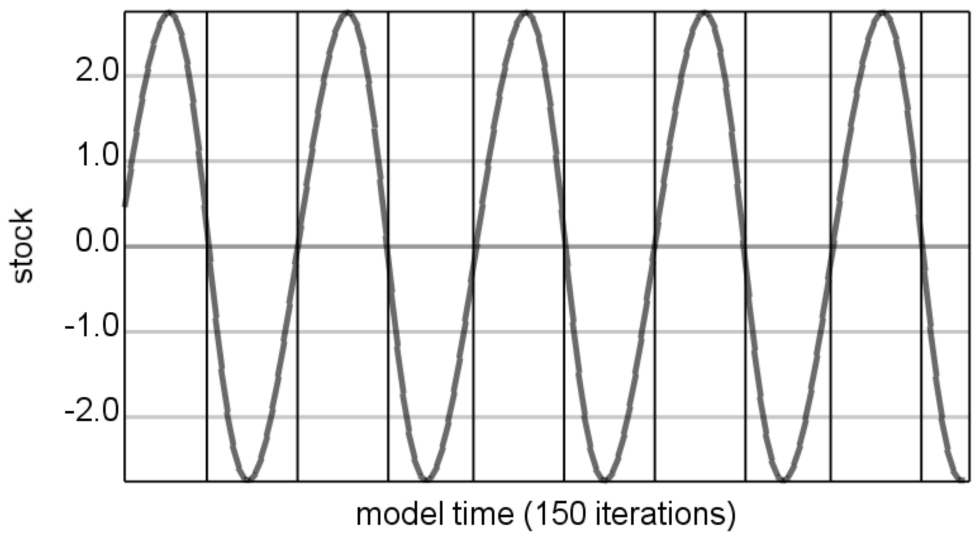
As explained, in the transmission belt model, each Firm adjusts their price in order to hit a set stock target. This adjustment is 'granular': it takes place at the exact moment any Person buys goods. There are problems with the crudest form of this mechanism, however: a mismatch between rates of change for price and stock levels. Using the simplest single-Actor example, this section also illustrates this problem and a solution that uses damping to more effectively equilibriate stock levels using prices.

Earlier iterations of the model used Firms that targeted zero stock only. This works under many circumstances, particularly with many agents as interaction effects act as a dampener on the control mechanism. However, using price as a control to target only zero stock means that, when stock hits zero, price will probably be in the wrong position to keep it there. Stock may well still be on a trajectory beyond zero - or, to frame this more usefully, if the first derivative of stock is not also zero (if it is still in the process of changing) price will have to swing back again. Figure 6.1 illustrates this using a single Firm and Person. As there is only a single producer and good, love of variety has no effect. There are also no distance costs. In this example, a perfectly stable oscillation results as price adjustment and stock levels are always exactly out of sync.

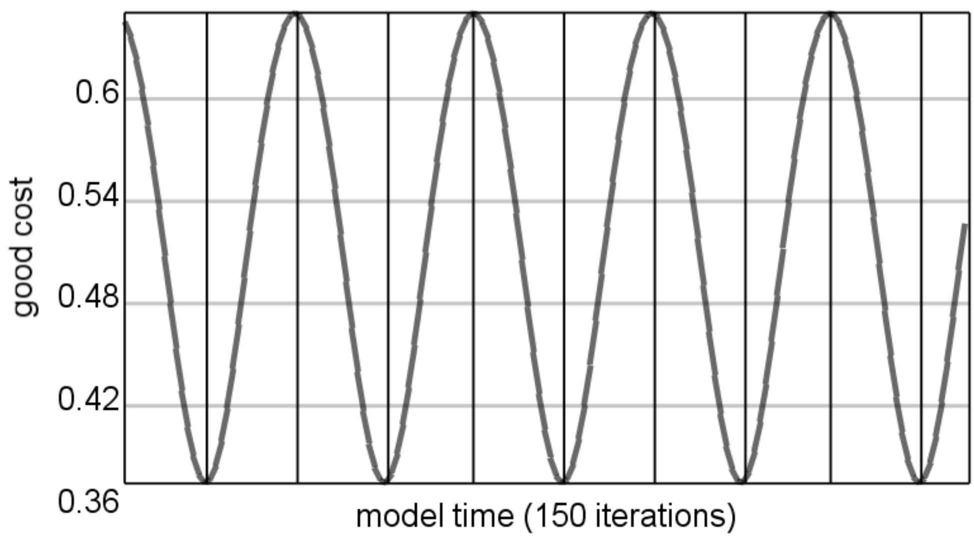
This is a version of a classic problem of harmonic oscillators that can be addressed by using damping (see e.g. Taylor 2005 pp.172). The goal is not only to target zero stock, but to aim for both stock and price rates of change to be zero at the same point. The solution is relatively straightforward: target the first derivative of stock as well as raw stock levels. If price is made to aim for the point where both are zero simultaneously, equilibrium should be reached. In fact, adding a second derivative zero target can improve the targeting even further.

Figure 6.2 shows three model runs, this time using these extra first and second order targets. The 'derivatives' are discrete differences in the model between timesteps. The first order derivative is the difference between stock at time $t-1$ and at t and the second order is the difference between this first order result between timesteps. The target mechanism itself does not change: if $s+s'+s'' \neq 0$, price is changed in proportion to how far from zero it is, as described above: $\triangle p_g = (s+s'+s'')/x$ where x controls the price change magnitude. If $s+s'+s'' > 0$, price is decreased and more stock sells. If $s+s'+s'' < 0$, price is increased and less is sold.

Sub-figure 6.2a shows good cost for a single Firm / single Person run. Sub-figure 6.2b shows good cost for fifty Firms with fifty People, with initial good costs randomised between 0.1 and 1.5. Sub-figure 6.3a adds a distance cost to those hundred agents. It gives them random,



(a) Stock



(b) Good cost

Figure 6.1: Transmission belt model: a single Firm changing price to target zero stock. Vertical lines indicate where stock crosses the y axis. As stock hits zero, however, price is far from its equilibrium value and both stock and prices oscillate.

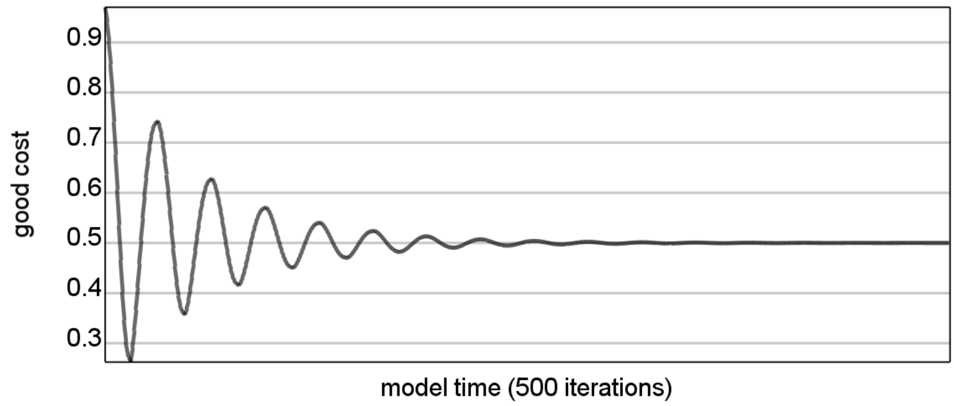
static locations in R^2 so that People buy from Firms at a range of distances, with each position being unique. In the two runs without distance cost, the equilibrium price of 0.5 is reached. It appears faster for sub-figure 6.2b (despite showing fewer iterations) because of the larger number of People: each time one of those People buys, it triggers a Firm's price response, so there are actually fifty such responses per Firm per timestep.

Figure 6.3 introduces an element of distance: agents' locations are randomised on an R^2 plane of one unit distance per side and goods are given a distance cost of 0.25. The result illustrates a basic point about the transmission model: a stable equilibrium can be found with agents facing heterogenous costs from different spatial locations. The result is a spread of equilibrium good prices, as each Firm receives a different quantity of inputted time, and thus the equilibrium production level to maintain the zero stock target is different for each. Sub-figure 6.3b shows this result in more detail by correlating each Firm's good cost to the amount of time being contributed to it. This graph shows data from day 600, which is the very right hand point of sub-figure 6.3a, at a point of stability. For a Firm with a production function exhibiting economies of scale, higher input levels increase per-unit output levels. In seeking to maintain a net-zero level of stock, lower prices are found to clear this higher-efficiency level of production - with the effect also of increasing Peoples' overall utility levels.

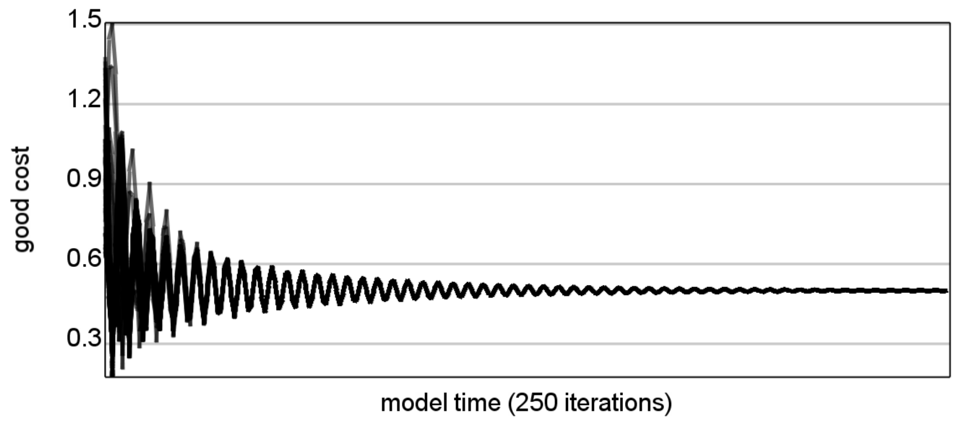
The classic equation for oscillation damping includes a term for angular frequency (ibid). Combined with a scaling parameter, in theory, the level of damping can be exactly targeted. The transmission model is 'under-damped' - it oscillates before reaching equilibrium. Addressing this, however, would require detecting the angular frequency of those oscillations. While it is relatively straightforward to detect inflection points for many model runs, other runs produce complex interaction effects that make this detection process much more challenging. The solution used here is simply to manually set the scale of price reaction to stock changes for each model run using x, but it should be noted that this can impact on the exact point where Firms lose their grip on effective price-setting, as it determines the size of price swings - and consequently the scope for entering a positive feedback loop where stock targets are missed (see the next section).

This example draws attention to another important feature of the transmission belt model. For model runs with a **single** Firm (and assuming space costs stay constant or at zero) production levels cannot change. People will always spend their entire budget on that one Firm. As this budget is also the total time available for production, this also does not change. With a single Firm, this means its production level is fixed: varying price only affects the amount of input that is bought. With more than one Firm, changing price *can* effect production levels - but this is due to an externality. This externality also interacts with economies of scale, as the following section explains.

There is a final point that is useful to make while examining the basic stock-targeting mechanism: in the transmission belt model, People's utility equilibriates. In the presence of distance costs, if People are fixed at different locations relative to Firms (as they are for all presented

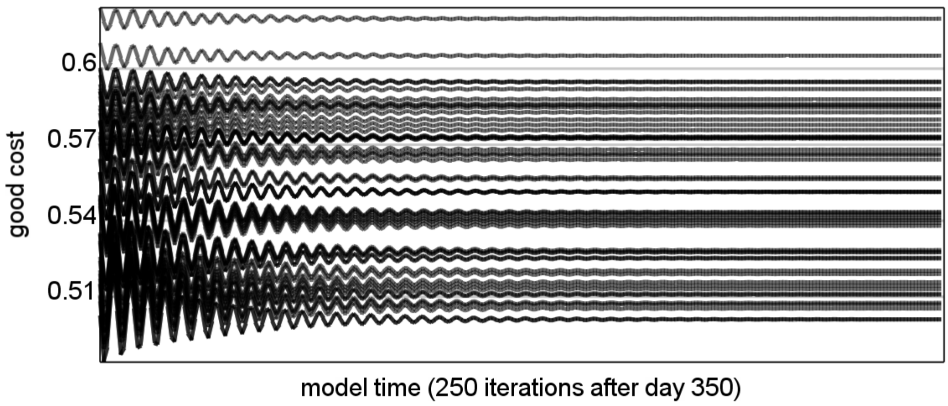


(a) Good cost for One Firm, one Person buying

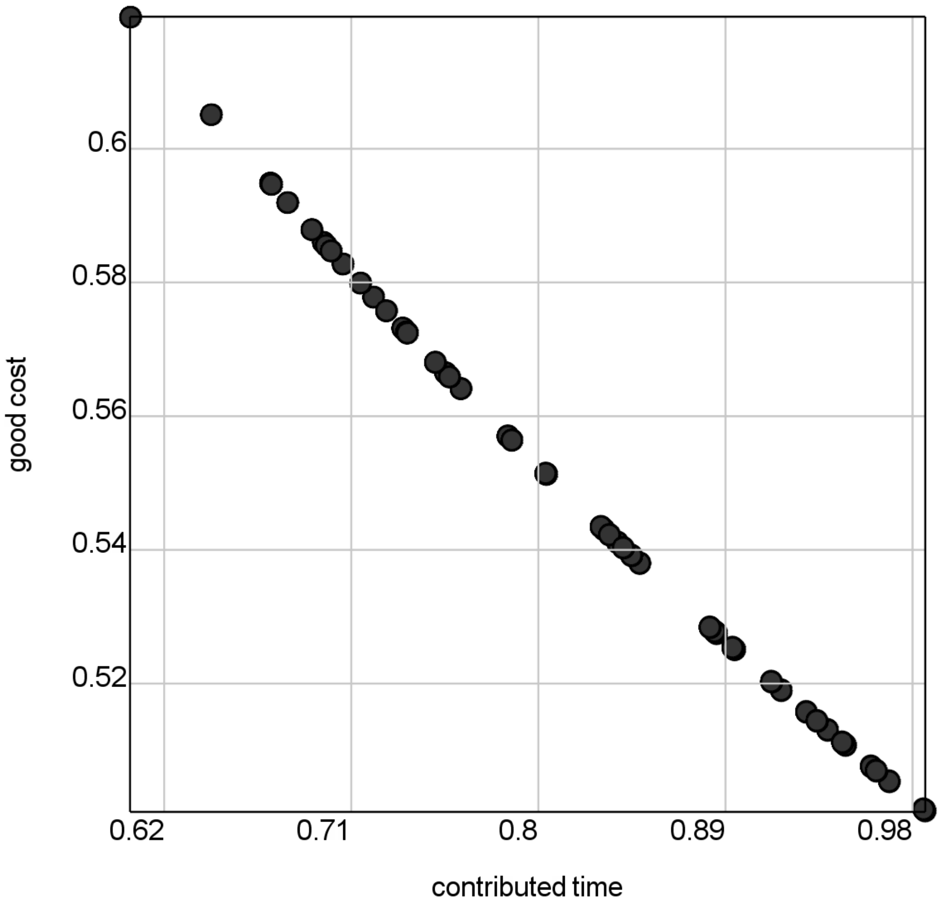


(b) Good costs for fifty Firms, fifty People buying

Figure 6.2: Transmission belt model: targeting the first and second derivative of stock as well as zero stock itself, ‘damping’ the oscillations and achieving equilibrium. Example of single Firm and Person and Fifty Firms and People.

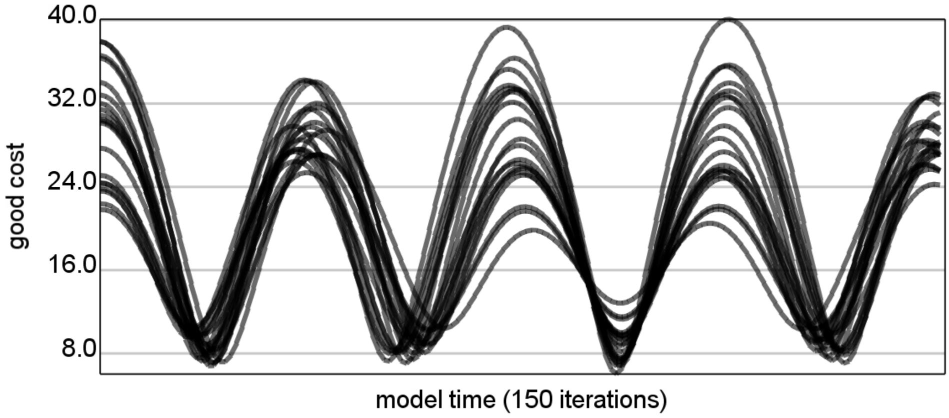


(a) Good costs for fifty Firms, fifty People buying, distance cost = 0.25, random locations in an R^2 plane of 1 unit distance to a side. Showing damping of oscillations over model time.

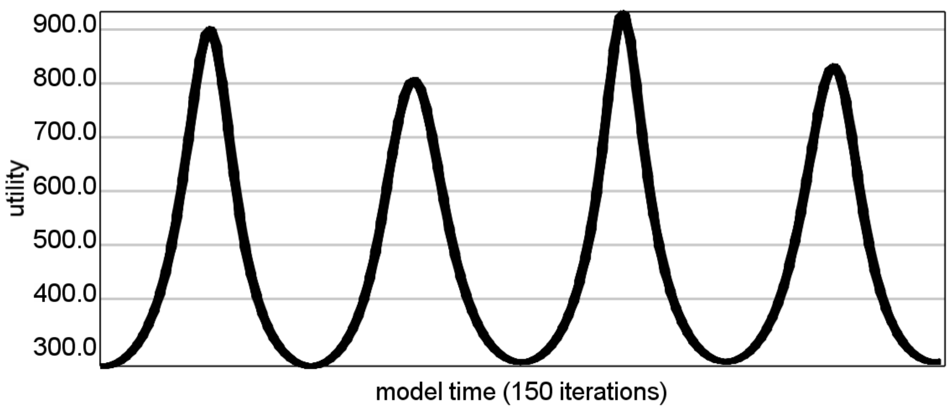


(b) From same model run, correlation of time contributed to each Firm versus their good cost at equilibrium, on day 600 (very right of sub-figure 6.3a).

Figure 6.3: Two graphs from the same transmission belt model run: fifty Firms, fifty People at random locations in R^2; distance cost = 0.25.



(a) Good cost, 20 Firms



(b) Utility, 100 People

Figure 6.4: Transmission belt model, 20 Firms, 100 People, no distance costs. Firms are given randomised stock targeting magnitude response; result demonstrates People’s utility equilibriates to a common value at each timestep, though it varies across model time.

models in this section) their utility is heterogenous simply because some agents can access goods more cheaply. All People's utility is umbilically linked; in a spaceless scenario where no intrinsic utility differences exist, this can clearly be seen, as People's utility remains identical. This is illustrated in figure 6.4, which takes a 150 timestep snapshot from the beginning of a model run with 20 Firms and 100 People. In order to contrast to People's utility, Firm's scaling parameter for the magnitude of their stock-targetting response to demand is randomised (sub-figure 6.4a). The hundred People's utility levels are all shown in sub-figure 6.4b: they are in lockstep, so appear as one line.

Utility does vary between timesteps - but People are following precisely the same path through time, having identical utilities at any one moment. While utility varies over model time, then, it stays in a dynamic equilibrium. Any possible improvement in utility is capitalised on by each Person at their point of decision, keeping the values umbically linked. Section 6.4.1 looks at People's utility in a spatial setting.

6.3.2 Markup and problems with competition

In the core model, a Firm's profit (or 'markup') does not exist as a variable in the economy. It is part of the scaffold that drops away once the regional equations are reached. Its implied role is the signal to other firms about the level of demand: it is argued that, if profits are above or below zero, firms enter and exit the market until profits are driven back to zero again. A change in firm number is the *only* way that production structure changes in the core model and even then, the actual number plays no direct role in deducing welfare outcomes.

This chapter does not examine Firm entry and exit: the number of Firms is set exogenously. It is potentially a fruitful avenue of investigation as it offers another way to examine changes in production structure. However, there are good reasons for leaving it out at this stage. Firstly, the models within this thesis have required agents to utilise only information they can collect from their environment. For a Firm to enter the market, having spied others making excess profits, it would need to be able to access that data directly. To know from a given price what profit another Firm is making requires also knowing their production output and markup. Starting to give agents this level of access to information takes the models into a different realm of knowledge and DMR. It would not necessarily require a 'master agent' watching global variables and adding or removing Firms appropriately; as section 3.5.2 discussed, an evolutionary approach might add Firms randomly and allow them to attempt survival, while also allowing failed Firms (see section 6.3.2) to 'die'. But the mechanics of entry and exit are complex, particularly for the transmission belt model where any change can cause perturbations large enough to knock other Firms over a 'tipping point' (see below), causing runaway problems for the model dynamics.

In the transmission belt model, Firms *are* able to gain a profit - thus the first ingredient for an explicit Firm-level economic micromotive is present. But for this to provide Firms with complete economic motivation, they would need to be able to maximise that profit. In the core model, profit maximisation involves simply finding the first order zero-point of a profit equation.

In the transmission belt model, some explicit mechanism is required. This section explains how Firms in the transmission belt model can gain a markup and how that causes external effects. But it also outlines that, as the model stands, creating an effective competition mechanism is problematic.

The mechanism that a Firm can use for gaining markup is as follows. ‘Markup’ is actually an amount of output they keep for themselves. A Firm sets a target amount of output they want to skim from each transaction and then uses price as a secondary targeting mechanism to reach it. This is a rather unique way to make a profit, so it bears re-stating: Firms’ good price is *dependent* on their markup target. The production target itself is set; the stock targeting mechanism attempts to achieve it, adjusting prices in the process.

Achieving this is slightly awkward, as attaining a markup means that extra output must be produced *constantly*. If a Firm simply aims for a higher stock target level and stops there, it is still net zero change over time. In order to get a markup, it needs stock to be increasing over time, and to take that increase for itself as profit. To do this, the Firm needs a markup level it aims for. This is then added to a moving markup target so that the Firm aims to skim a constant amount in a given time period. The stock target equation is altered to: $s + s' + s'' = mt$ where mt is the moving markup target.

Equilibrium good prices for Firms seeking markup are higher than the zero-stock clearing price. This higher price emerges as it targets the higher level of stock: it ends up charging extra and keeping a cut of the output. The external effect of this raised price is to drive demand to the second Firm. The way other Firms respond to this external effect depends on whether they have economies of scale. Without them, the extra demand they receive as input is linear with output change and the stock clearing price is the same. It is only in the presence of economies of scale that the external effect (beyond shifting demand) will be a change in stock-clearing price: it will become lower as a more-efficient production point is reached.

Without economies of scale, any change in markup by one Firm will not change the absolute price of the other (dependent) Firm. It does change *relative* prices as one Firm (the one making independent markup decisions) targetting higher markup leads to it finding a higher equilibrium price - but it does not change the dependent Firm’s price. When production scale is linear, demand does go up for their good but that is exactly matched by output level, so supply and demand for that Firm’s good does not change. Hence, the stock target remains the same and prices do not move. With economies of scale, however, changes in markup cause absolute price changes in both, as changes in demand cause both to shift up and down the economies of scale equation. This changes per-unit output, which changes the stock target level and then feeds through to equilibrium prices.

This is illustrated for a typical set-up in figure 6.5. One Person buys from two Firms, with Markup for Firm 0 fixed at 0.1. ρ is fixed at 0.5. The parameter controlling the scale of the curve in the economies of scale equation is swept from zero to one (shown on the x axis). The y axis shows the external effect that Firm 0’s markup has on Firm 1’s equilibrium price.

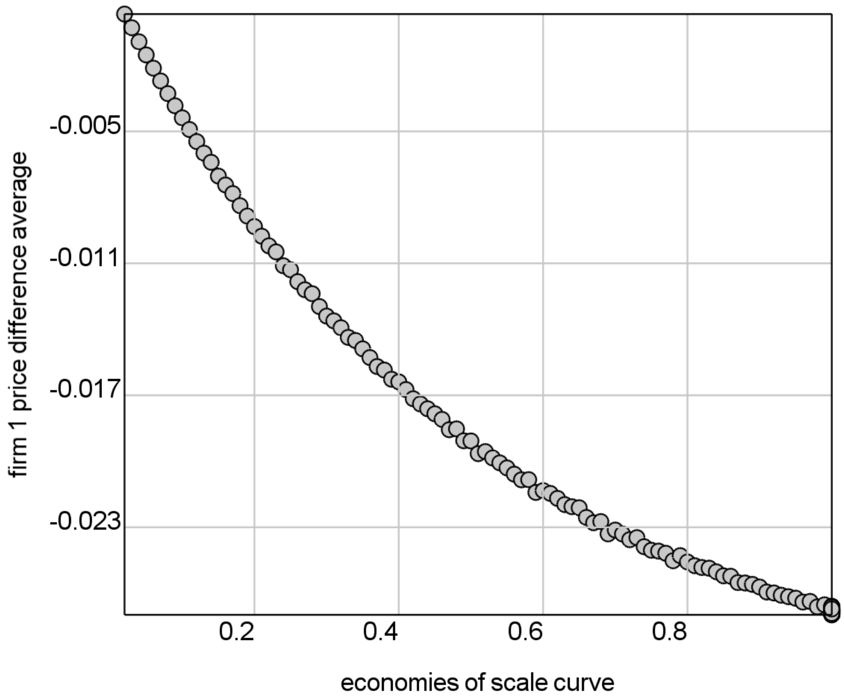


Figure 6.5: Two Firm / one person transmission belt model. Firm 0 has a markup of 0.1. $\rho = 0.5$. The x axis shows ‘economies of scale’ curve swept from 0 to 1. The y axis shows, for the externally affected Firm, the difference between the normal equilibrium price (without markup) at that production scale and the price it finds. Average over previous 500 iterations is used as the value fluctuates at short timescales.

For each value of the curve parameter, if there was no markup to consider, the equilibrium prices are straightforward: one Person buys 0.5 from each Firm. The values on the y axis show the difference between this ‘no markup’ equilibrium price and the actual price the externally affected Firm equilibriates to, as they get more demand. Average price over the previous 500 iterations are used, as the price fluctuates around its stable price daily due to the markup-seeking Firm constantly shifting its stock target. At zero, there is no curve: production scale is linear, so there is no change in the externally affected Firm’s price. As economies of scale increases, this price gap becomes larger. (Note that ρ is set at the level that keeps production from changing regime - see the next section for more on this.)

Figure 6.6 shows the effect of one Firm increasing its markup value (before hitting the ‘tipping point’; see below). It correlates the equilibrium prices of two Firms, with the economies of scale curve parameter fixed at 0.5. Markup is swept between zero and 0.15 for Firm 0. This shows up as a higher equilibrium price for its good as it aims for a higher stock target to keep a cut for itself. The external effect on the dependent Firm is shown on the y axis: its equilibrium price drops. It gets the demand driven away by Firm 0’s higher price, which feeds into a more efficient production scale and thus a lower clearing price. In this case, the ‘raw’ price is used, not an average over previous timesteps; this is to illustrate the volality of the price’s oscillations,

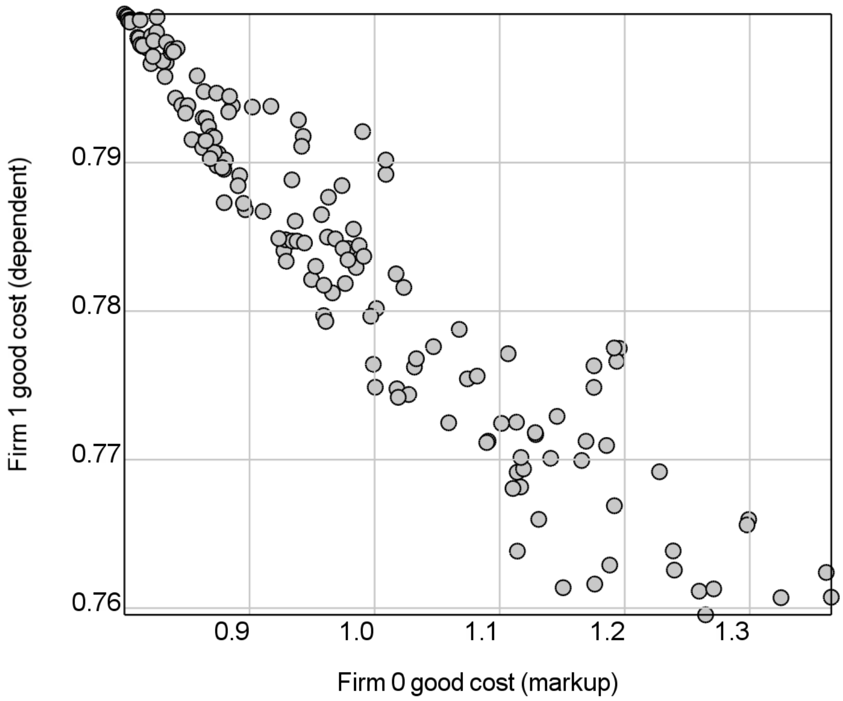


Figure 6.6: Two Firm / one person transmission belt model. Firm 0 varies markup between 0 and 0.15. $\rho = 0.5$, economies of scale curve $= 0.5$. The Firm changing its markup (x axis) finds higher clearing prices, drives demand to Firm 1 (the dependent Firm), whose equilibrium price drops.

though the negative correlation is clear.

These two figures, then, show that economies of scale and profit-seeking interact and cause effects external to Firms making decisions. This is a quite different dynamic to Firms entering and exiting a market, having seen the presence of profits. In the core model, keeping the actual level of production itself static is an important simplification. Here, it is seen that in the presence of economies of scale one Firm changing its markup target pushes the two Firm’s prices in opposite directions. The production structure of the whole economy is affected.

Firms can easily succeed in achieving a stable markup over time, then, for given markup target values. So there is a mechanism for achieving stable profits. But there are a number of problems with building on this foundation towards a fully competitive model. The first issue is DMR related. In order to gain a profit, a firm needs to detect whether a change in markup target means ‘my action improved my profits’ - but the transmission belt model is replete with conflicting signals. The model’s oscillitory behaviour has been discussed; at present, angular frequency makes DMR unmanageable. So this section concentrates on exploring the parameter space of price interaction through blind Firm behaviour - changing markup values directly and observing the effects both for the Firm changing its markup and the externalities on other agents.

The second issue is a problem with Firms attempting to raise their markup: a ‘tipping point’ exists that is disastrous for their stability and they are unable to pre-empt it. What causes the tipping point? One way into understanding the problem is to consider the role that love of variety plays. Assume that $\rho = 1$ in a situation with two Firms, so both of their goods are completely substitutable. If their prices are equal, there is nothing to choose between them. However, if one Firm attempts to gain a markup, once their price is pushed up even marginally, demand for their good will drop to zero.

So the ability of Firms to gain a markup is dependent on the value of ρ. In the model itself, as $\rho \to 1$, a Firm attempting markup reaches a price where it can no longer approach its stock target. This gap between their stock target and what they are able to achieve in production produces a constant pressure to change their price in order to try to get back to the target. The inability to move towards it only reinforces its price changes, and it enters a positive feedback loop: further price increases lock in its unaffordability and drive production to the other Firm. Figure 6.7 makes clear an important feature of this tipping point: it correlates to ρ. Why is this? ρ controls elasticity of substitution, thus the effect of any change on price. For higher values of ρ, as goods become more substitutable, People are less reliant on love of variety: a change in price becomes a stronger signal for the other Firm. In this model run, ρ is swept from $0.1 < \rho < 1$. One of two Firms slowly increases the markup it aims for (this markup is shown on the x-axis). At a certain price, it reaches the tipping point. Each increment of ρ is triggered when this happens: the model is reset and run again, and one Firm again attempts to incrementally seek profit. (Two People are used in this run to indicate the basic dynamic functions the same as for one.)

There appears to be a more fundamental reason why any competitive outcome, in the current model set-up, is not possible. All Firms are incentivised to seek a stable markup but, while the external effect of one doing so is to move demand to others, that demand has no impact on other Firms' profits: each Firm has its own stock target and moves prices to meet it. So, each Firm must separately attempt to raise its markup target if it wants a profit. The dynamic created is a ratchet effect: each individual Firm is able to take extra slices of production for themselves, right up to the point where People are no longer able to afford to buy the remaining output that Firms do not keep as profit - at which point, there is no incentive to work and the system collapses. Note again that People have no exogenous options for spending their time and thus no effective sanction that could produce competitive behaviour in Firms. Even if goods' cost is exhorbitant, they will pay it, as no other options are available to their utility functions and a sliver of utility is better than none.

To summarise: there are only two possible outcomes for Firms attempting to seek profit, as the model is currently structured. If Firms collude, or succeed in following a collusion-like pattern through independent actions (implicit collusion), they can squeeze profit out of People right up to the point where People can no longer afford to buy anything at all. Or, if one or more Firms attempts to get markup, this is sustainable up to a point, but then tips over into a positive feedback loop between the external effect of lowered prices for others and the profit-seeking Firm's inability to use price to target their new stock level due to having lost too much demand.

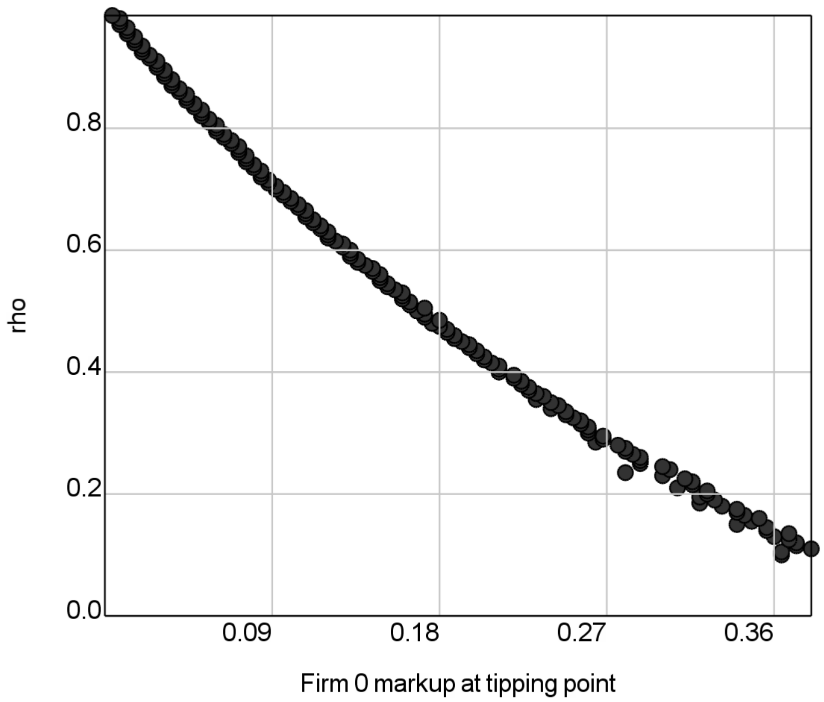


Figure 6.7: Two Firm / two person transmission belt model. The graph shows Firm zero’s markup target correlated to ρ , with datapoints taken when the Firm hits a ‘tipping point’; there are no economies of scale.

Missing their target causes them to continually increase their price, permanently compounding their error.

As the transmission belt model stands then, there is no stable competitive outcome that can emerge through Firm interaction alone that allows for Firms seeking profit. Firms' price signals *can* be used for targeting stock level and can function as an effective distributed mechanism for finding equilibria across all agents.

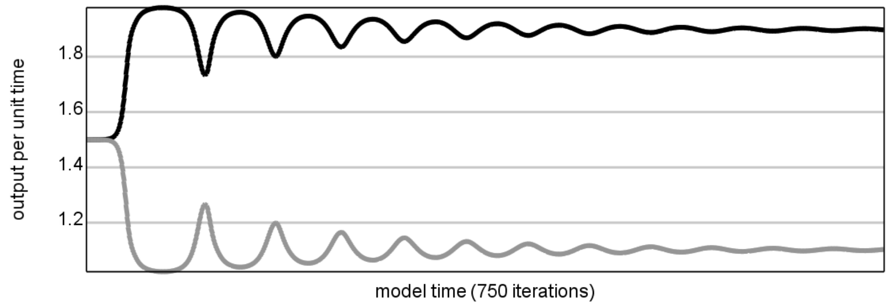
6.3.3 Interaction between economies of scale and love of variety

Continuing with a model with no distance costs, this section outlines how economies of scale and love of variety interact in the transmission belt model. The model uses explicit economies of scale: Firms have a production function where output per unit input increases with input. Consumers have an entirely separate love of variety utility function. This allows interaction between them to be investigated. This section demonstrates that, in the transmission model, the interaction of economies of scale and love of variety can determine the whole model's production structure. In contrast to the core model, distance is not required for a shift between model regimes to occur. Within the parameter space defined by economies of scale and love of variety, agents can find two distinct production regimes. In one, love of variety dominates and production is spread out among all producers symmetrically. In the other, the economies of scale effect dominates, and that symmetry is broken. While the transition is 'catastrophic', changes to relative production scale exist beyond that point, determined by love of variety: the state of the model is not simply binary. If the smooth economies of scale function has zero curve and production is linear, any change in ρ has no effect on price, for the reason mentioned in the previous section.

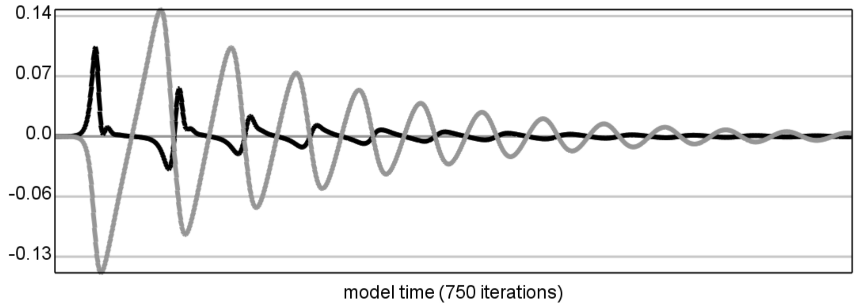
This model example reduces complexity even further by only using one Person buying from two Firms. This makes it possible to compare to a very simple deductive version with the same output values and results, enabling a clear view of how the transmission belt model's agents move between production regimes.

Section 2.5.4 explained that, in the Dixit Stiglitz model, economies of scale at production level are subsumed into two things: the love of variety parameter (or its mirror, the elasticity of substitution parameter ε) and a market structure argument that reduces economies of scale to the entry and exit of firms. It is a very powerful set of deductions that allow everything else that follows a solid mathematical foundation. As Brakman *et al.* (2009 p.106) note, "internal economies of scale are not absent... but show up in only in a rather special way." The core model does contain parameters for production level economies of scale in its 'scaffold', but the final model is not sensitive to their value in any way. Indeed, with the normalisations mentioned in section 2.5.4, they are analysed away.

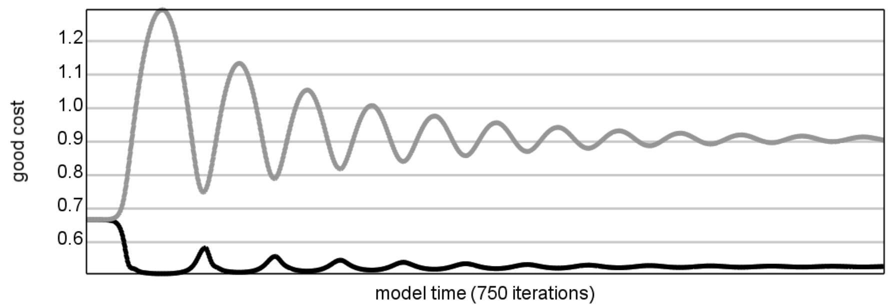
To begin with the model output itself, using all the features described in the sections above, figure 6.8 shows what happens for the two Firms and one Person at the point where love of variety drops low enough to tip the model over into one efficient producer. The change in value



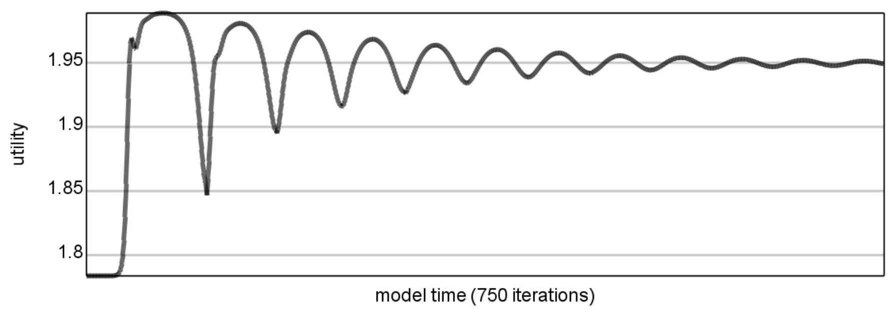
(a) Two Firms: output per unit of input time



(b) Two Firms: stock levels



(c) Two Firms: good cost



(d) Person’s utility

Figure 6.8: Two Firm / one Person transmission belt model moves between production regimes as ρ crosses substitutability threshold. Black and grey identify the two Firms in the first three graphs. 750 Model ‘days’ shown after transition from $\rho = 0.73$ to $\rho = 0.8$.

of ρ that causes this (from 0.73 to 0.8) is not visible - see below for more on this - but rather, the story of what happens is shown through the four sub-figures. The first (6.8a) shows each Firm's output per unit of input time. As more time feeds into one of the two Firms (identified by the black line), it produces at a more efficient scale: output per unit input grows.

Underpinning this is People's demand separating out from a symmetric split between the two and entering a feedback between price and efficiency (this is the essential part of the model and is explained in more detail below). As a consequence (6.8b), this Firm sees its stock level increase initially, while the second firm loses stock as it is now producing at a less efficient scale. It takes some time for both Firms' price reaction to take effect (6.8c): the end result is a more efficient producer able to charge a much lower price to maintain net-zero stock, effectively locking it in as the favoured site of production. The final sub-figure (6.8d) shows the driver for this change: the Person finding that shifting their demand to the higher-efficiency producer increases their utility.

6.3.4 Tracking down the 'transmission belt'

What is causing the shift between these two stable states? The problem can be boiled down into its component mathematical parts, allowing an examination of the 'transmission belt' between production and utility in its simplest form. There are four pairs of variables to consider (with one set each for the two Firms):

- The price each Firm is charging for a unit of output (p_0 and p_1).
- The optimal amount of each good demanded by the Person (c_0 and c_1) based on those prices being input into their CES function.
- The quantity of stock produced by the Firms (f_0 and f_1) based on how much time the Person exchanged for their optimal amount. Their time input is just their optimal amount multiplied by the price they paid for it. (The 'price' is the amount of time they must pay for that optimal amount.)
- The difference between the demanded stock (c) and produced stock (f), $c_0 - f_0$ and $c_1 - f_1$.

To understand how the dynamic works, a useful question to ask is: what makes the model move away from its symmetric equilibrium? Equally, what can hold it there if the variables are perturbed away from symmetry? The coded model suggests that, for values of ρ below 0.75, demand remains symmetric across the two Firms, whereas if $\rho > 0.75$ (and if there is a slight perturbation in initial values) one more-efficient producer emerges.

The model's values can be deduced around this transition point. When production is symmetric, time is split evenly between the two Firms: 0.5 for each. So stock output using the smooth economies of scale function is $f = t^2 + t = 0.75$. For a production equilibrium, that

0.75 needs to be 'cleared': the stock must remain at net zero, so two lots of 0.75 must be sold. If a Person is paying 0.5, it has to cost $p = 0.66'$. (If it costs $0.66'$ for a unit of good, paying 0.5 gives the person $0.5/0.66' = 0.75$.)

So $p = 0.66$ is the equilibrium price at symmetric equilibrium. The goal is to discover what happens when prices are perturbed by a tiny amount from that equilibrium: do they return back or do they move towards another stable equilibrium? This is where the last pair of variables come in: the difference between the optimal demanded amount and the output produced. With slight differences from symmetric equilibrium prices, these two values do not quite match - thus the market is now out of equilibrium as it cannot completely clear at those prices.

This mismatch is a logical outcome of the smooth economies of scale function and the clearing price. If price is slightly lower than the stock clearing amount for a Firm, demand will be higher - and because economies of scale can be gained smoothly, any tiny amount of difference in input leads to a larger input/output ratio, however marginal. Due to this mismatch, Firms must find a new clearing price. Section 6.3.1 has already explained how this is accomplished in the transmission belt model, in a situation where demand may be coming from many different People and places. In this one-Person case, the job is more transparent: there is a simple difference between the single Person's demand and the amount produced and this allows the two Firms to make a price adjustment to seek a stock-clearing value.

This example can be input into a spreadsheet to demonstrate what role love of variety plays. This removes all coordination and agent timing concerns to show the underlying dynamic as clearly as possible. The slight difference between c and f produced when prices move a little from the symmetric equilibrium implies a price change polarity for Firms to use. This can be described as a simple differential equation: $p_n' = f_n - c_n$. As explained in section 3.3.1, this is a Leijonhufvud-like 'law of motion': if demanded stock is more than produced stock, price is going up to reduce demand; if demanded stock is less than produced stock, price is going down to reduce demand. In the spreadsheet, this change is just a very small constant at each step.

The result is to show whether or not the Firms' price adjustments converge back to symmetry or diverge to a new equilibrium, testing whether the symmetric equilibrium is stable at given values of ρ. Figure 6.9 shows the output from the spreadsheet for two different values of ρ on either side of the symmetric/divergence point[1]. Where $\rho < 0.75$ in sub-figure 6.9a, the fixed price adjustment quickly converges back to stable symmetric prices, oscillating around that point due to the fixed adjustment constant used in the spreadsheet. Sub-figure 6.9b shows one producer emerging as the more-efficient, stabilising at a new pair of equilibrium prices.

Looking at only one Person and two Firms in this simple example helps make clear exactly where the 'externality' is doing its work: through the constrained CES utility function. The differential equation describing Firms' price changing behaviour, $p_n' = f_n - c_n$, contains the constrained optimum good amount in c_n. The CES function contains the price of the good being

[1]This spreadsheet is included in the downloadable model 'zip' file via the link provided in the code appendix (A).

bought *and* the price of all others - just a single other good, in this case, but it still means changes in the price of one always impact demand for the other, as there is a finite pot of demand. The scale of the elasticity is controlled by ρ. In the spreadsheet example, the 'external' effect is not a matter of timing or interaction between Firms, as both Firms make their price decisions simultaneously. It is the impact those price decisions have within the Person's CES function on the next step that encapsulate the externality. If one Firm changes its price unilaterally, this will still impact on demand for the other Firm's good - but via the action of the Person's utility function.

Though the transmission model is designed to handle complex spatial demand situations, this simple example illustrates how the outcomes are driven by the underlying economic behaviour and the interaction between Firms' production functions and an 'externality' coming from elasticity of substitution between Peoples' good demand as ρ changes. The tiny price difference required in this non-spatial version is, at root, also responsible for outcomes in the spatial examples used in the following sections.

6.4 Introducing space into the transmission model

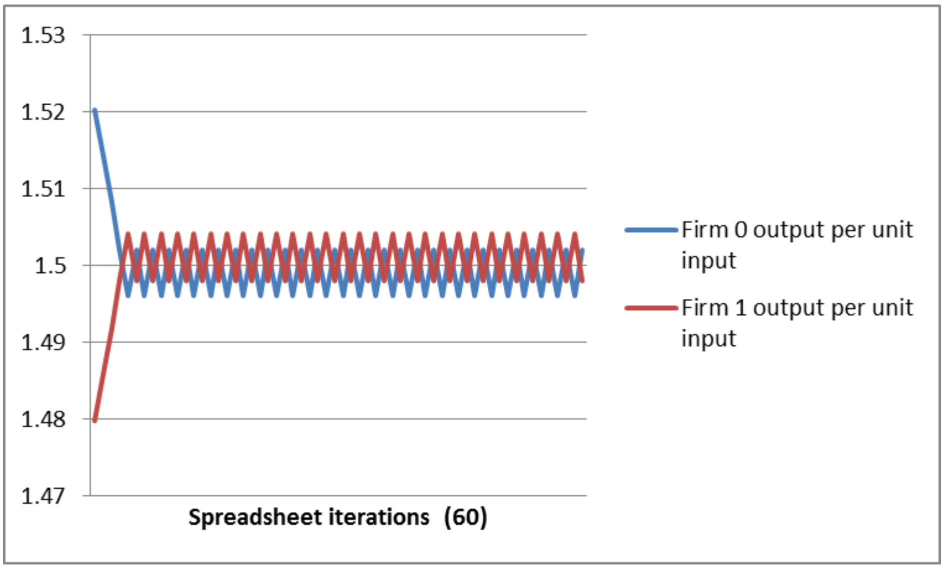
Section 6.3.1 used an example involving random spatial locations. It showed that, if agents have heterogenous location (and thus each Person faces a different set of good prices and each Firm a different quantity of demand) the transmission model is able to find a set of stable equilibrium prices for all Firms. But what role does distance play in the interaction of model variables and the transition between production equilibria? This section examines this question by following the example of the core model, using the simplest possible representation of distance between two discrete regions. There are three model examples in this section illustrating essential points about the effect of distance in the transmission belt model:

1. Keeping People's location static, with one Person in each of the two regions, two production regimes are determined by love of variety, as in section 6.3.3, but distance cost can also determine the transition between regimes.

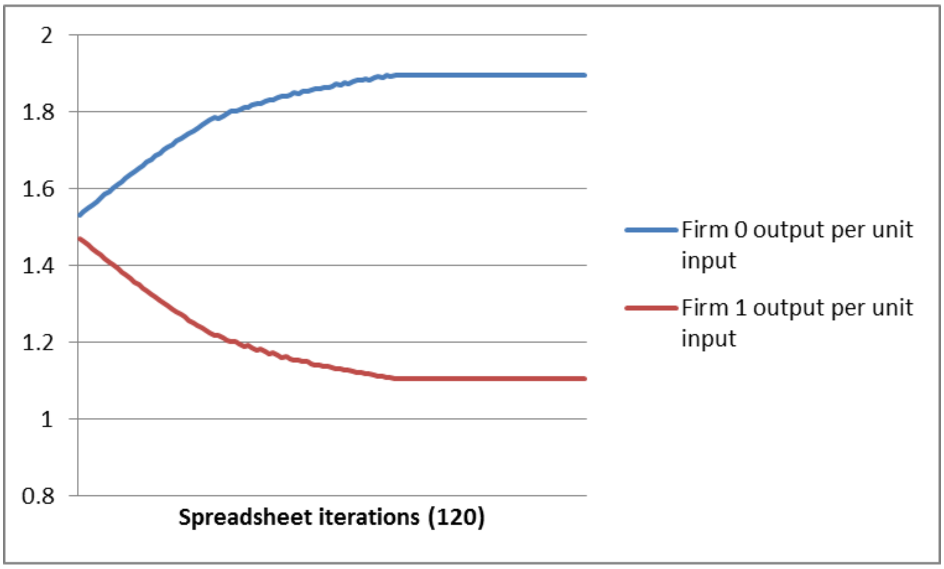
2. Allowing People to move, the interaction with economies of scale locks the resulting emergent region with no way back: a 'black hole' outcome.

3. Two Firms per region, with two People static again, one per region. Four different kinds of production regime are possible by varying ρ and distance cost, demonstrating heterogeneity within regions as well as between them.

Varying rho alone in the presence of distance costs has a direct effect on outcomes (where in the spaceless version in section 6.3.3, goods would remain divided evenly between the two Firms). As section 5.3.1 explained, higher love of variety causes demand to spread out over distance as the higher utility from a mix of goods outweighs the distance cost in acquiring it. As



(a) $\rho = 0.73$



(b) $\rho = 0.8$

Figure 6.9: Spreadsheet output of Two Firm / One Person example. If $\rho > 0.75$ the symmetric equilibrium is not stable: a slight difference in price leads to feedback to one, higher efficiency Firm and one lower. Note: 60 iterations are shown in (a) as convergence happens quickly.

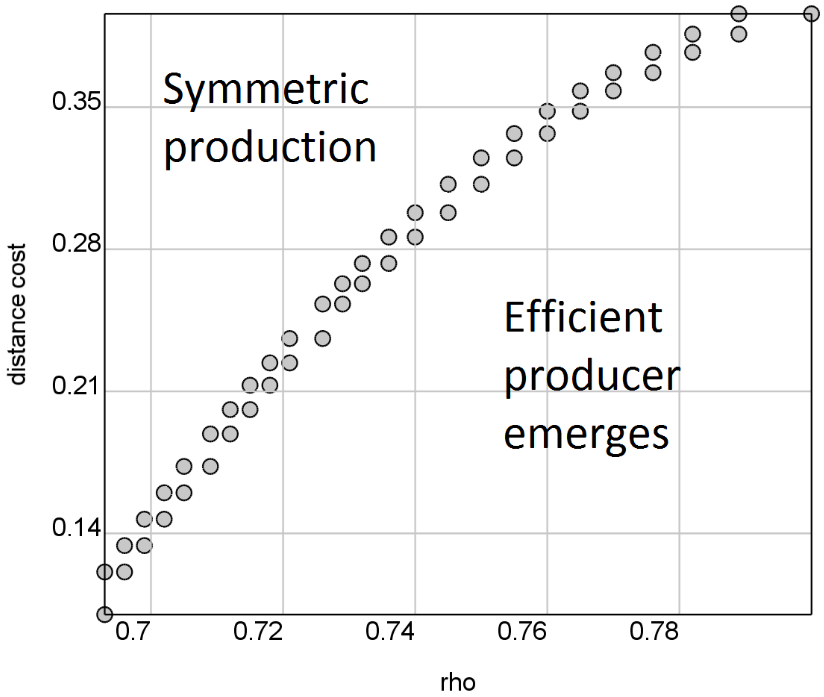


Figure 6.10: Two region transmission belt model: one Person, one Firm per region, starting off in symmetric production. ρ and distance costs start low. Each is raised in turn. When an efficient producer dominates through economies of scale as one is raised, the model shifts to raise the other until equilibrium moves back to symmetry again. A data point is taken at each shift as the parameter sweep threads through equilibria.

love of variety increases, demand for more distant goods increases while that for closer goods will drop as constrained buying is re-allocated. In this two-region transmission belt example, as love of variety drops and demand is relocated to the home region, two things happen. First, less time is lost to the distance cost itself, increasing the scope for production. Second, more time is input into home region producers, increasing their efficiency. This interaction of love of variety and distance costs happens even if no economies of scale are present, as section 5.3.1 already outlined.

Between the two regions, the ratio of time spent on the nearest Firm increases both with ρ and distance cost at all points. Though as $\rho \to 1$ and closes in on perfect substitutes (and People are buying more from their home region) that ratio is affected less and less by distance costs, for the obvious reason that, as People are having their utility needs met more from their home region, changing distance costs has no effect on their local buying.

Figure 6.10 presents a series of data points where the transmission belt model moves between production equilibria. The process for creating this graph was as follows. Distance costs begin at 0.1 and ρ at 0.1. At these values, the model is in a symmetric equilibrium as described in section 6.3.3 and lowering those values further does not alter that state. Rather than just sample points in the parameter space, a thread is woven back and forth across the equilibrium transition point, in order to demonstrate that the model is moving between production regimes. Each variable takes a turn to increase slowly, in increments of 0.001 spaced apart by a gap of three hundred days to let any change settle down, starting with ρ. The model tips over out of symmetry when goods become substitutable enough for efficiency of production to win over love of variety. At this transition point, the model is made to pause again for a set number of days to allow oscillations to dampen. Distance costs between the two regions are then increased until gains from efficiency are lost and the model reverts back to the symmetric equilibrium. The two variables are alternately threaded across the transition point in this way to map the line along which the equilibria are divided.

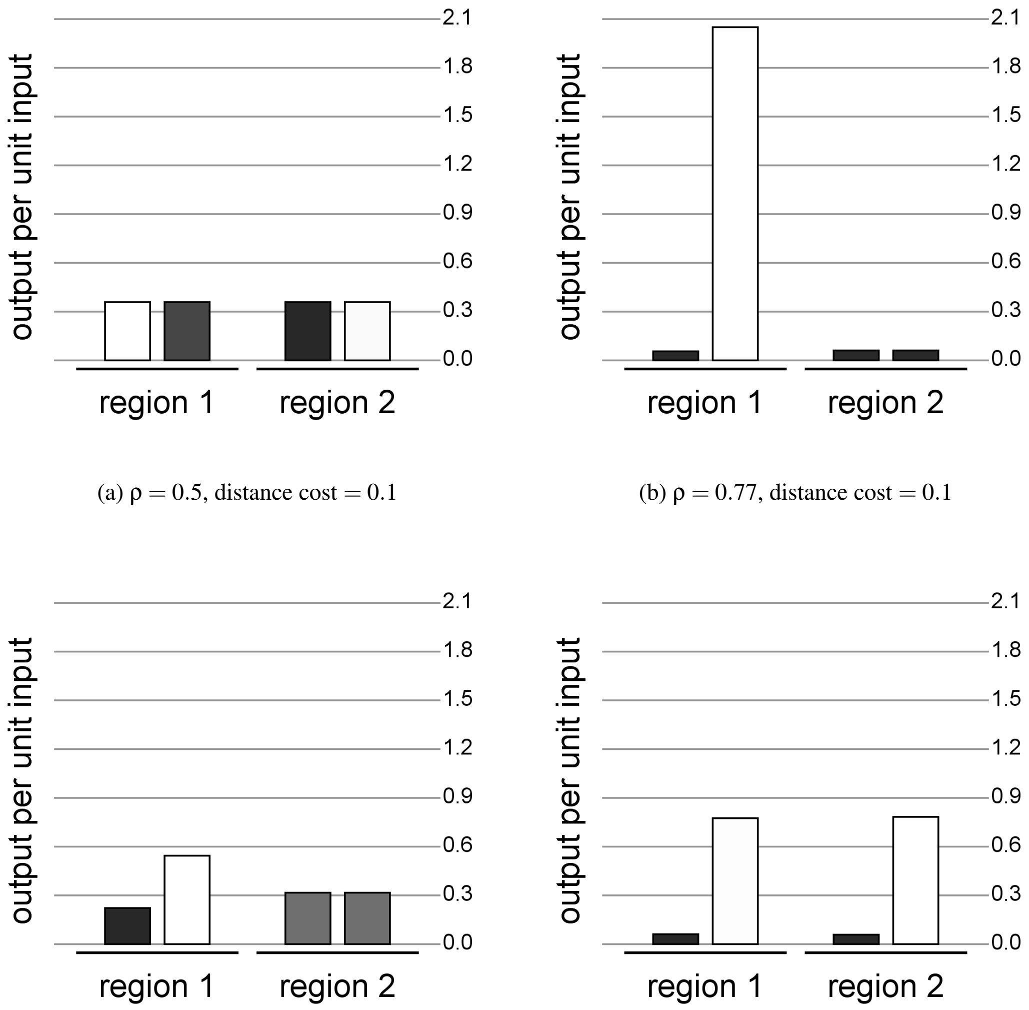
In the above example, when one Firm becomes the more efficient producer, utility is higher for the Person in their region: they face no distance costs to buy from the cheapest, most efficient source and thus have an advantage over the more distant Person. In the core model this difference in utility would be an indication that agents want to move to the better region. In the transmission belt model, if People are allowed to move, this does happen - but the result is different to the core model, in that once People have moved to the more efficient region, this outcome is 'locked in'. If People have chosen one optimal region, the amount of time input they bring with them locks the Firm in their region into efficient production. This is not quite a complete 'black hole' outcome as, in the transmission belt model, Firms are not making separate location decisions and so one remains in the now-deserted region, and receives demand. But it is a black hole for the People: there is no incentive for them to leave the region they have moved to.

Note that providing a permanent level of fixed demand in both regions (as the core model does with its agricultural sector) makes no difference to this dynamic. In the core model, this fixed demand provides a path between equilibria where people and firms can migrate in both directions; the symmetric outcome can be a stable equilibrium at high distance costs. The transmission belt model, however, only requires a single Person out of many to be mobile for 'lock-in' to occur in the more efficient region.

Two Firms per region

Adding a second Firm to each region (taking the model total to four) reveals another dynamic resulting from the interaction of love of variety, economies of scale and distance. The internal production structure of each region can vary, as well as the production regime of the whole model. Figure 6.11 shows the efficiency of the two firms each Region, in terms of 'output per unit input'. The higher this is, the more efficiently the Firm is producing, due to having more time input from People buying its good. (The economies of scale curve and magnitude parameters are both set to 0.5 as this keeps the model stable when two People are buying.) The firms are shaded white to dark grey, where white is the highest efficiency and dark grey the lowest.

Figure 6.11 shows the four states the model takes at differing values of ρ and distance cost.



(a) ρ = 0.5, distance cost = 0.1

(b) ρ = 0.77, distance cost = 0.1

(c) ρ = 0.77, distance cost = 0.35

(d) ρ = 0.85, distance cost = 0.55

Figure 6.11: Two regions, two Firms in each region. Varying ρ and distance costs determines production regime, which can vary within each region. Y-axis shows output per unit input: higher values indicate more efficient production scale.

At the lowest values, when love of variety is high and distance costs low, production spreads evenly across the model. A large proportion of People's time is lost to distance, but this is outweighed by the utility gains from a higher love of variety. As love of variety decreases, a single more-effficient producer emerges. This is possible due to lower distance costs; as these are increased, the regime changes again through two more states. In sub-figure 6.11c, the regions diverge: one reverts to its own internal symmetry while the other splits between more and less efficient producers. In the final sub-figure, as distance costs become higher and goods more substitutable, a more-efficient producer emerges in each region. This contrasts to the core model: regions can develop internal structure, and this development between and within regions is dependent on model-wide interaction.

6.4.1 Space and welfare outcomes in a many-agent model

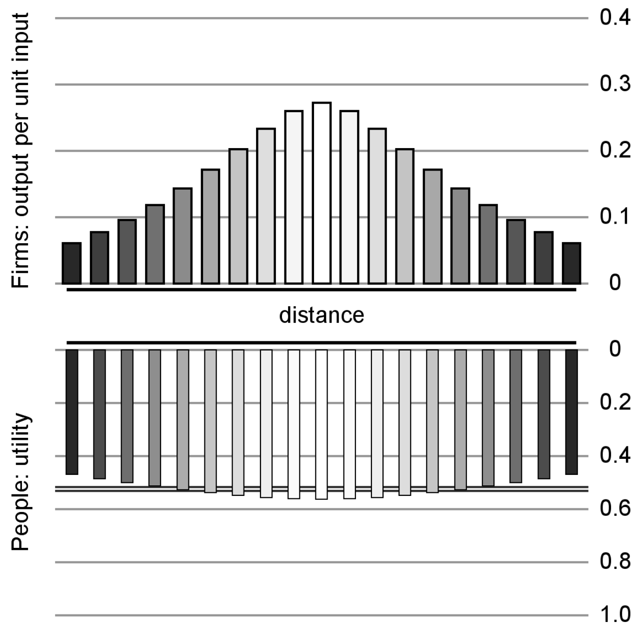
Increasing the transmission model's number of agents produces similar dynamics as in the previous section. In this section's example, a larger number of Firms and People interact to produce equilibrium outcomes. Nineteen Firms and Nineteen people have evenly spaced locations along a line of distance 1; every location containing a Firm also contains a Person. An odd number is chosen so that one Firm/Person pair is directly in the centre of the line. With many agents on a line, the transmission model is very sensitive to parameter settings: many combinations lead to the kind of positive feedback described in section 6.3.2. This section selects sets of parameter values able to maintain stability between production regime changes.

Figure 6.12 shows these pairs of 19 People and Firms with distance represented on the horizontal axis. Above the 'distance' line, each Firms' efficiency levels are shown: their output per unit input. Below the 'distance' line, each Person's utility level is shown. A double line marks the mean utility for all People. White-to-dark colouring marks the relative spread of high to low values for the iteration on which the graph was produced. Sub-figure 6.12a shows symmetric production - no more-efficient producers have emerged. Those nearest the centre of the line have the highest utility and the highest input of time. People closer to the centre are better off simply due to being closer to a wider range of Firms. Those nearer to edge can access less quantity of nearby goods. Equally, the central Firm gets the most input due to its position.

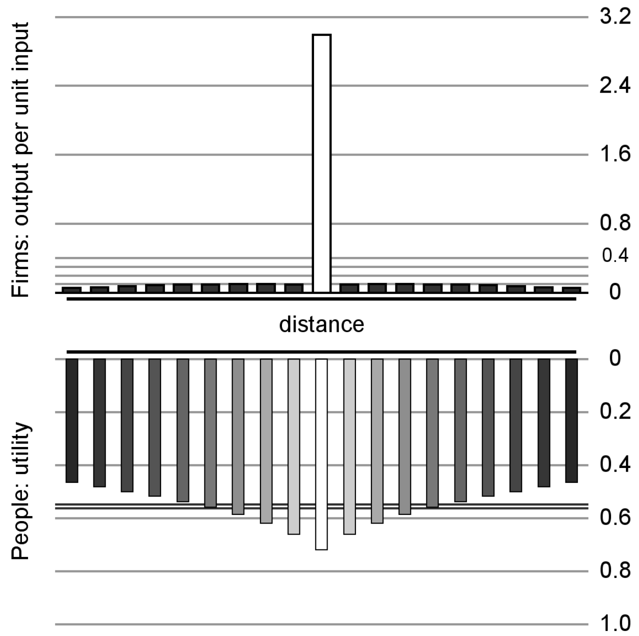
In sub-figure 6.12, distance cost has been reduced from 4.75 to 4.5, resulting in the central Firm emerging as an efficient producer. Note the difference in scales for Firm efficiency: the bottom sub-figure marks the value of 0.4, which is the maximum for the top. The efficiency gain is large.

Figure 6.13 shows two results where more-efficient production regimes have emerged. By increasing distance costs in each case, they illustrate that it can determine the specific equilibrium production morphology. These are each pushed up the same path of distance cost and love of variety outlined in section 6.4. In this case, positions further up the distance cost / love of variety curve produce a larger number of efficient producers.

Transitions to more efficient production *never* make any single Person worse off. It is

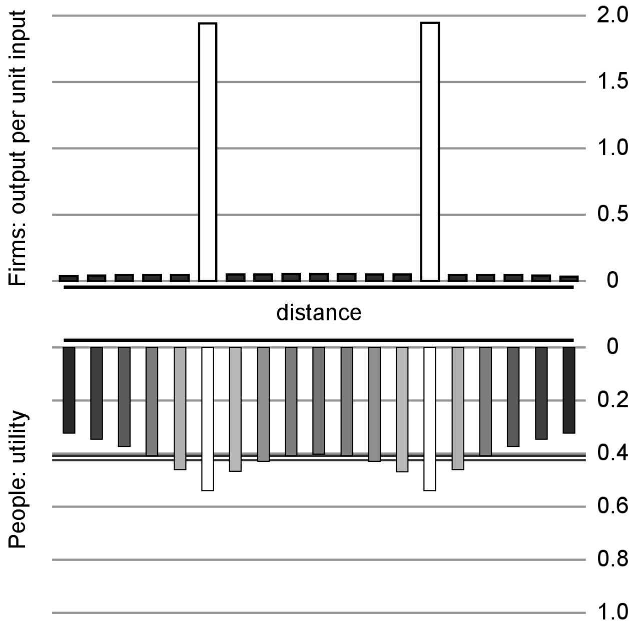


(a) Demand spread out: ρ = 0.7, distance cost = 4.75. Firm y-axis scale to 0.4. For People, double line marks utility mean.

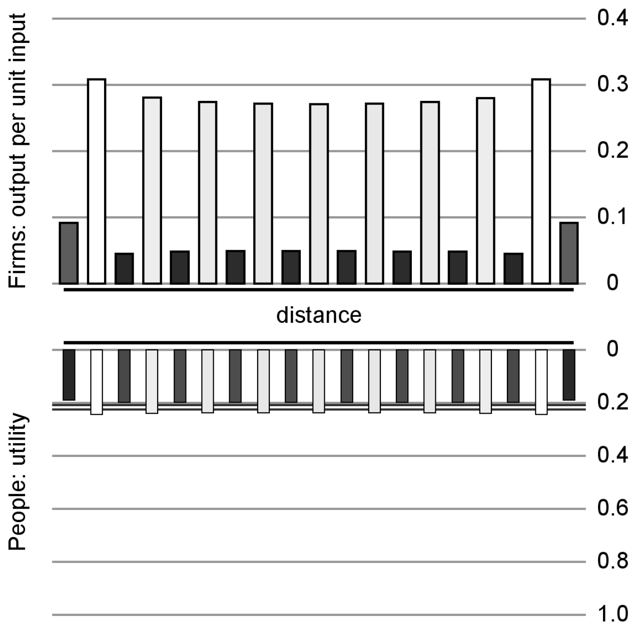


(b) Efficient producer emerges: ρ = 0.7, distance cost = 4.5. Firm y-axis scale to 3.1, with 0.4 marked for comparison.

Figure 6.12: 19 People, 19 Firms, spaced between 0 and 1 on a line. Distance cost determines transition between production regimes. Double black line across People’s utility marks the mean utility value. Different scales used for Firm efficiency y-axis; the second figure marks the lower values used in the first.



(a) Two more-efficient producers: ρ = 0.73, distance cost = 8.5. Firm y-axis scale to 2.



(b) Many more-efficient producers emerge: ρ = 0.79, distance cost = 45.5. Firm y-axis scale to 0.4

Figure 6.13: 19 People, 19 Firms, spaced between 0 and 1 on a line. Two examples of different more-efficient production regimes emerging for differing values of ρ and distance cost.

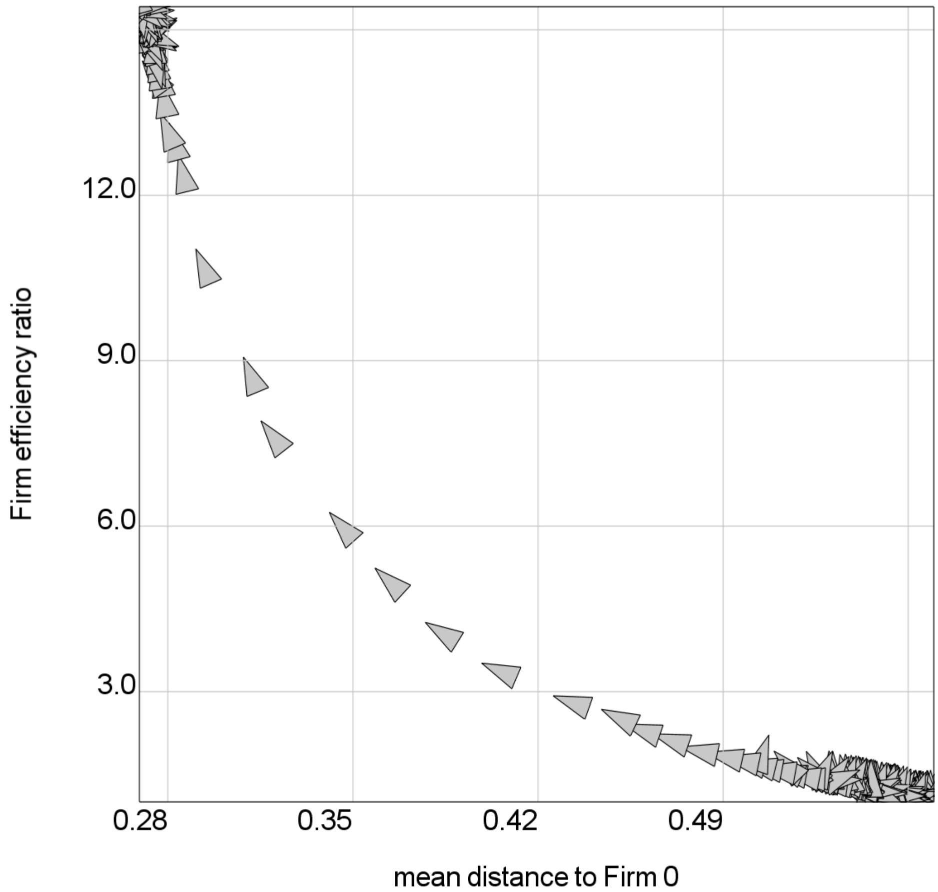
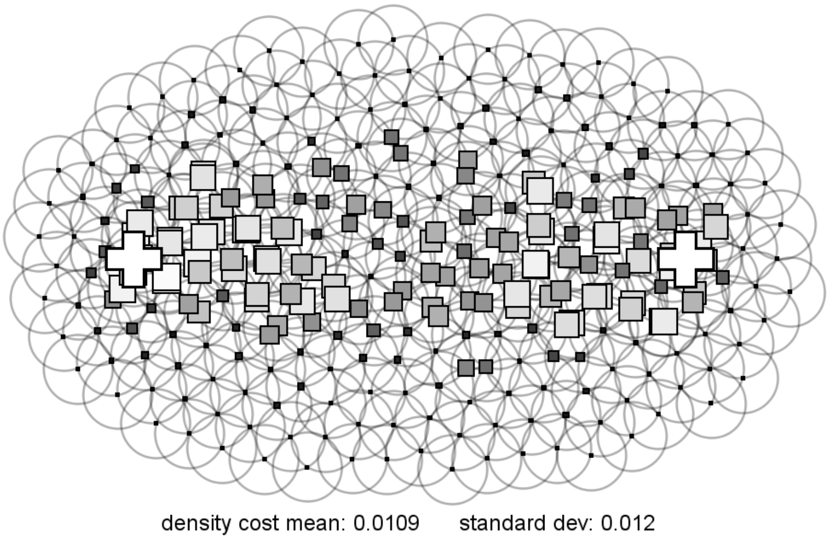


Figure 6.14: Two Firms fixed a distance of 1 apart; a more-efficient producer emerges as $\rho \to 0.9$. Mean distance to Firm 0 (thus mean distance from Firm 1) vs Firm production ratio; arrows point in direction of change between model iterations.

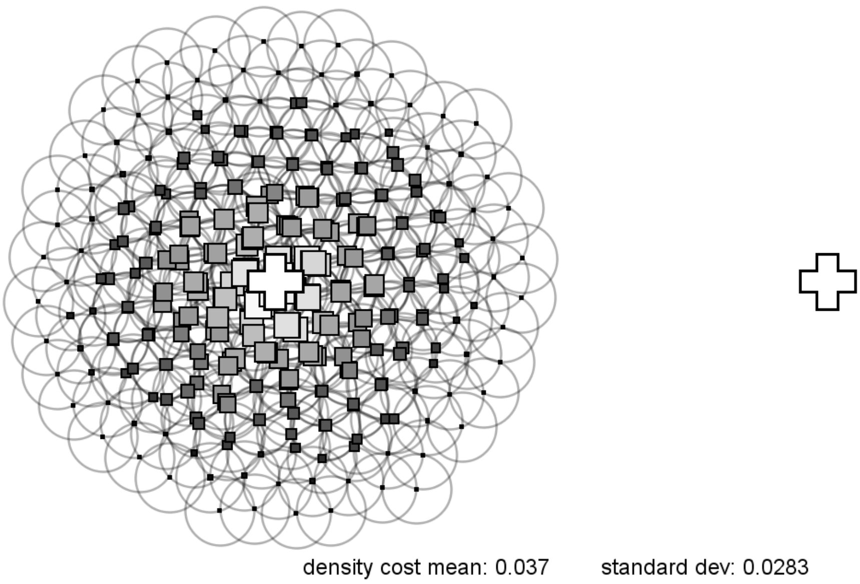
trivially true that, if People's locations are fixed, increasing distance costs will make those further away from the more-efficient producers worse off than others able to minimise those costs. But the transitions to more efficient production regimes always produce both higher aggregate welfare at the same time as making each individual better off, or at the very least, no Person worse off. To the extent that larger demand from more proximate People is responsible for tipping the transmission belt model over into more efficient production, this is a positive externality for others able to benefit from cheaper goods produced at the larger scale.

6.4.2 Combining the transmission belt model with density cost

The final result presented in this section builds on the two-Firm love of variety example presented in section 5.4.6. In that scenario, two Firms at fixed positions in an R^2 space were given exogenously fixed distance costs, with one Firm having double the distance cost of the other.



(a) R^2 space visualisation, $\rho = 0.5$



(b) R^2 space visualisation, $\rho = 0.9$

Figure 6.15: Two Firms fixed a distance of 1 apart; a more-efficient producer emerges as $\rho \to 0.9$.

The aim was to show the effect of changing love of variety in this scenario and how it interacted with density cost: as love of variety increases, it is more valuable to be closer to the more expensive good and 'buy the other in' from further away.

This example demonstrates that a fully linked supply/demand model is able to endogenise something - in this case, the particular region that emerges as the 'core' - that previously had to be exogenously imposed. There are no initial differences between the Firms. Some parameters are set differently to section 5.4.6 to allow the model within the correct range for transitioning between production equilibria. Distance costs are 0.25 and the number of People is 300. The economies of scale curve parameter is set to 0.001 to stop Firms from 'tipping over' as those 300 people send demand signals every time step. Otherwise, the outcome as regards which region emerges as the more-efficient is down to chance. This section illustrates the result for Firm 0 becoming the more-efficient producer.

As ρ is increased (and love of variety drops), more-efficient production starts to kick in - producing a cheaper good. Figure 6.14 shows this transition as ρ moves from 0.5 through to 0.9. As in section 5.4.6, the x axis shows the mean distance to Firm 0 (and thus, because there are only two firms, this is the inverse of the mean distance from Firm 1). The y axis is the ratio between production efficiency for the two Firms, always placing the higher-efficiency Firm in the numerator, thus making sure the ratio is $\geqslant 1$.

This correlation graph introduces a different way to view model time: arrows point in the direction that values changed between iterations. A vector is found between $t-1$ and t and its angle used. The transition to a single more-efficient firm is thus clearly visible: the arrows indicate mean distance to Firm 0 dropping as the Firm efficiency ratio climbs to 13. Visualisations of the agents in this R^2 space are given in figure 6.15 for the two values of ρ, each side of the transition. People's squares indicate, in this case, the relative range of incurred density cost. As the higher-efficiency producer emerges, they move nearer to that production sight. As this is also in the same direction of travel indicated by section 5.4.6 as love of variety drops, the result is doubly locked in.

6.5 Summary

This chapter has introduced an agent-based spatial economic model linking supply and demand to produce a 'general' spatial equilibrium able to transition between production equilibria while allowing agents heterogenous locations. It could be argued the model has four key 'general' components out of a possible five. One: the demand side works effectively, allowing People to make separate optimal buying choices that account for location and love of variety. Two: the supply side has a functioning production system, able to turn that demand into an equilibriating, stable economy that can shift between production regimes as key parameters change. Three: firms are able to gain a profit from the market. Four: it has an effective price system able to tie these separate components together into a 'general' whole. What it lacks is a fifth key element:

competition between Firms.

Section 6.3.2 demonstrated that it is possible for Firms to gain a 'profit' and that doing so has external effects on other Firms (as well as People). However, it does not find a way to create a competitive market structure through emergent Firm interaction. The specific route the transmission belt model found to achieving a 'general' model appears to make this kind of competitive dynamic challenging. In comparison, the core model contains a detailed argument about market structure and competition. The Dixit Stiglitz model approach it is built on is a careful web of reasoning that attempts to deal with competition effects in an increasing returns setting. It is difficult to directly compare approaches on this point, as the market structure of the core model involves assuming equilibrium has been reached.

Nevertheless, the transmission belt model presented in this chapter has some advantages. It is very much in the ABM tradition of examining the *process* of moving between equilibria. The ability to transition between production equilibria is something the Dixit Stiglitz model approach rules out. All goods are assumed to be produced symmetrically, at the same scale. Increasing returns are still present but, as section 2.5 explained, a series of stages transform production-level economies of scale into a change in firm number and thus in utility levels via love of variety.

More fundamentally, the model's equilibrium states come about with no central directing mechanism. This happens entirely through a distributed spatial price interaction between individual Firms with economies of scale and individual People's demand for goods. It can allow for those agents to have heterogenous locations and for People to optimise those locations.

This comparison to GE is taken up further in the conclusion.

Chapter 7

Discussion and conclusions

The modelling paradigm of ABM has held out the promise of freedom from all earlier model-building restrictions, able to create 'virtual worlds' (section 3.4.4) with objects mapping perfectly to real-world structures. If this promise were straightforwardly true, ABM would seem perfectly suited to the 'big questions' of spatial economics: why does spatial economic activity happen in the way it does? How does it lead to the economic landscape seen in reality? How does it change? How do these all happen in a spatial economy made up of agents separated by time and space making individual, but ultimately massively interconnected, decisions?

This conclusion argues that any attempt to answer those questions using ABM must take the limitations of ABM more seriously. This is not a negative conclusion: it reveals a great deal of scope for further work usually hidden under layers of complexity. ABM has huge promise as a tool for digging into the most difficult spatial economic problems; understanding its limitations is a vital step in living up to that promise.

A set of specific dependencies and assumptions have been revealed through the model development process; the upcoming discussion uses these as the focus for highlighting both the limitations of ABM and new opportunities these limitations reveal.

The first aim of the thesis was to:

- **Develop an agent-based spatial economic model framework in which agents with heterogenous locations can optimise welfare given changing costs**

This has the following objectives connected to it:

- Develop a spatial agent model framework and explain its workings in detail in the thesis;
- Compare agent-based and economic approaches to model-building (with specific emphasis on GE) and use this analysis to inform the thesis' model development.

Chapters 4 through 6 presented the framework developed to do this and key results along the way. These described the route to producing the transmission belt model, which successfully enables agents with heterogenous locations to optimise welfare given changing costs, using 'laws of motion' (section 3.3.1) plus damping, allowing Firms to target a stock level and People to maximise utility. The feedback between these two drives a transition between production equilibria where one or more efficient producers emerge.

As well as a series of testing models, chapter 5 focused on two modelling scenarios, each examining a separate issue. The 'mobile People' models (section 5.4) looked at spatial equilibrium by using a 'density cost' proxy that avoided coordination issues. The 'mobile Firms' models (section 5.5) concentrated on the coordination issue, examining how decisions in a simple two-Firm scenario interact.

GE has been examined in detail: both the 'nuts and bolts' of its modelling approach and its overall modelling philosophy (as well as that of economics more broadly). This thesis has taken seriously the arguments about spatial economics put forward by GE. In particular, the need for a link between supply (through some theorisation of production) and demand has been at the heart of the modelling approach.

The benefits of closely examining GE have been two-fold. Firstly, many concrete ideas have been used, not least the CES function throughout the modelling framework. Secondly, and just as importantly, it has revealed a set of challenging problems outside the norm of concerns for ABM. Where the default ABM question has tended to be 'Hayekian' (section 3.3.1) ('how can it come about through decentralised actions?'), keeping the focus on GE concerns about spatial supply and demand has pushed the thesis model framework to find a route that works in an ABM setting.

Doing this has revealed a clear set of dependencies and assumptions. The result is far from negative, as this final chapter hopes to show. Rather than striving for an ABM ideal that supersedes all previous modelling paradigms, turning round and looking the other way, towards the tangle of dependencies and assumptions, reveals a rich seam of possibilities and allows a more open comparison to other, equally valid, modelling approaches.

The second aim of the thesis is:

- **Use the development process of the model framework to answer the following: what obstacles are in the way of ABM exploiting its potential as a tool for spatial economic analysis?**

The following objectives are connected to the second aim:

- Identify obstacles revealed in modelling development at the coding level and a route through them;
- Identify obstacles revealed by the comparison of different approaches to model-building in ABM and economics.

These two objectives are interlinked. Solving problems in the process of building the thesis models meant identifying and overcoming a set of obstacles. In combination with examining the theory of modelling, this process has been used to think about obstacles for agent-based spatial economics and why ABM seems to have avoided the 'big questions' of spatial economics.

Obstacles discovered during the development process have ranged from highly specific coding choices through to broad theoretical ideas. An example coding choice would be working out the need for a 'smooth' economies of scale function rather than the core model's fixed/marginal function (sections 2.2.5 and 4.7.5). The effect of considering theory has been broader. Examining the roots of ABM and GE has been, to paraphrase Hoover (section 1.1.1) 'a quotidian practice, not abstract epistemology', directly applicable at various model turning points. The main effect has been to provide 'permission' to explore a space of modelling choices not usually considered 'valid' in ABM - not least freely using ideas from neoclassical economics, given that many ABM theorists believe the field to be anachronistic.

Through this process of developing the thesis' ABM spatial economic framework, the following two obstacles have been identified:

1. A lack of research directly tackling the space of dependencies in ABM model development. Rather than studying dependencies abstractly, further work should naturally result from model-building while carefully considering the algorithmic 'nuts and bolts' of the process. This is a promising route for ABM *because* of its flexibility: the possibilities for trying out new code structures are huge. GE must rely on the 'engine' of the Dixit Stiglitz model; this ties it to a specific set of dependencies that, for example, rule out considering a situation where firms make decisions over a much longer timescale than people. ABM would in principle have no problem with this.

2. A broader sense that ABM has broken all ties with history and can learn nothing from past modelling efforts. By exploring economic ideas, the thesis suggests that in the endeavour of modelling economic systems, pre-computational thinking is still highly relevant. As the introduction discussed, while computation has radically altered the way modelling is done, it has not changed its fundamental nature.

The rest of this chapter aims to illustrate the fruitfulness of taking dependencies seriously, as well as treating ABM as fully part of a much longer history of modelling thought. This ties together the two main components of the preceding thesis: the presentation of modelling choices, ideas and theory in chapters 2 and 3 and the thesis model framework results presented in chapters 5 and 6. As Hoover points out (section 3.4.1), models cannot be 'cleanly idealised' if they are to be 'operable or realisable in a manipulable form'. Dependencies are unavoidable. Being explicit about this helps track down what they look like and what possible new routes may be open for exploration. The rest of the conclusion looks in detail at these assumptions and dependencies, in the following sections:

- A direct comparison between the core model and the thesis model in order to highlight some of the key assumptions and differences;

- Model framework dependencies identified during the building process;
- The different types of agent interaction the framework has examined and the assumptions involved. These are tied back to considering obstacles facing a spatial economic ABM;
- A more detailed examination of assumptions in both the model framework and in model development more generally;
- The conclusion ends with a discussion of possible further work and some final comments.

7.1 Comparison with the core model

While the thesis model framework is not meant to be a direct comparator to the core model, it is informative to compare them on a number of categories to bring out the differences in both outcome and modelling philosophy of GE and ABM, as well as setting up discussion of model dependencies in section 7.2. Fowler found during the process of developing his agent model that analytic choices made to gain tractability can "actually make the problem less tractable" in an agent setting (Fowler 2007 p.283). This has certainly been the experience throughout the process of model development; this section highlights where many of these differences arose.

Di Paolo *et al.* make the point that no one methodology is inherently superior (section 3.4.4), but each does have different strengths and weaknesses. For example, the approach used by GE versus the agent models of the thesis have quite different areas where they manage to be simple and where they fall into complication (rather than complexity). The detailed presentation of the core model in its original, simplest possible incarnation, illustrates plainly that it is far from simple in its components and the web of assumptions required for it to function, even while the dynamics it reveals are simple to understand. This thesis' agent approach offers a simple way to think about agents, using quite basic functions to make decisions over distance. But the process of making them interact is not simple.

From a top-level view, the difference might be described as follows. The core model creates a tension between worker's income based on a wage, which is worked out from a price index, which is built on a complex market structure argument. The final thesis model has these features: People have time to spend. They exchange it for goods over a distance. Their time goes into a range of Firms each with an economies of scale function. Firms adjust price to keep the market cleared. Doing this allows efficient producers to emerge through cumulative causation. The core model is built on a scaffold: simplifications and useful assumptions that allow the model to build up to its final structure, losing the complexities it used to get there. The thesis models are an attempt to build on the strengths of ABM by keeping the key interactions present.

Timing structure

The core model is broken into a short and long run, both with distinct dynamics - though the long run is by far the simpler of the two. Each is mathematically consistent but the bridge between them is essentially arbitrary, as long as workers move where the real wage is highest. The short-run component of the core model uses an iterative process to find its equilibrium values, but this is simply a method of finding a stable set of values for the region equations; it has no economic meaning.

The thesis models run over many time steps. This fits with Hicks' description of supply and demand functioning "in a recursive manner" (Hicks 1989 p.11). This contrast to the core model shows up well in both the 'mobile People' models and the transmission belt model. The 'mobile People' model's use of the density cost (section 5) sees agents producing emergent morphological outcomes over model iterations with every single Person's choice having external effects. It is this recursive, external-effect interaction that produces this model's spatial equilibrium outcomes.

Similarly, the transmission belt model relies on step-by-step feedbacks to transition between production equilibria, utilising the smooth economies of scale function to do so. It does not produce the migration outcomes of the core model; cumulative causation occurs through People deciding where to exchange their time, and this translating into production equilibria (see below for more on the production structure).

Use of agents and the nature of firms and people

In the core model, the number of agents actually represented by specific mathematical objects or code are small: as section 3.3.3 says, each region has a representation of manufacturing and agricultural workers, making four 'agents' in total (though these reduce to summed region income equations). These are deemed 'representative agents', standing in for all economic actors. Demand from each of these types of agent (manufacturing and agricultural) uses the Cobb Douglas' ability to hermetically seal them into separate compartments; this is very useful for the core model as it allows the manufacturing side of the market to be treated separately. For an agent-based model, there may be situations where such a clean separation could be useful, but not if agents are required to have interactive demand.

The number of firms in the core model can be represented by a continuous variable, deduced from the models' equilibrium structure, though the number itself plays no direct role in outcomes. The number is 'assumed to be large' in order for some key macroeconomic simplifications to be made, allowing vital elements of the mathematical argument to be set to constants. This is a feature of the Dixit Stiglitz model approach the core model uses to good effect, building on it to produce its whole equilibrium structure. It is presumed that when workers move to find higher utility, firms follow. The equilibrium number of firms also follows instantaneously. Firm entry and exit is able to adjust at least as fast as workers are able to migrate.

Firm and Person agents represent each side of supply and demand. While in theory, each have encapsulated decision-making processes, in practice, the structure of these decision processes is determined by the model type and highly dependent on given model choices; this point is discussed in detail in section 7.2. In the ‘mobile People’ models, Firms are static sites that provide goods at fixed prices to People with fixed wages. Many People agents optimise their locations, trading off distance costs, base costs and an incurred ‘density cost’ that allows a spatial morphology to emerge. The ‘mobile Firms’ model uses only two Firms, competing over location on a line of People. The medium of competition (location) and the number of Firms are kept as basic as possible to illustrate a key issue of signalling in the model framework (using DMR).

All of these, including the transmission belt model, use a relatively small number of agents (maximum 500). Agent modelling relies on the agents having some kind of market signal to work with. This does not mean that ‘very large market’ arguments may not be useful, but if reaction to market signals is being investigated, this becomes problematic the larger the number of agents and the noisier the environment. The relatively small number of Firms in the thesis models allows external interaction between price-setting behaviour, but is not ‘oligopolistic’ as Firms do not directly compete.

Production structure

The core model’s use of a fixed and marginal labour production structure is the basis for its argument about how firms enter and exit the market. The mathematical simplifications involved allow the model to remove any elasticity of substitution between goods: the production scale itself in each firm stays fixed and changes in market structure happen *solely* through the implied entry and exit of firms. It is a key simplification that allows the Dixit Stiglitz model ‘engine’ to function.

The transmission belt model works through changing production scale. It does not allow firm entry or exit, but each Firm agent is able to use price signals to target an individually ‘efficient’ level of stock production. At certain values for love of variety, economies of scale and distance, the model transitions between production equilibria, where one or more Firms emerge as sites of much more efficient production.

Fixed and marginal requirements for production make sense in an algebraic setting, where a variable for each can play a role. In the core model, the marginal requirement is used both in the market structure argument and in key normalisations allowing for much simpler regional equations. Getting this production function to work in an agent model is not necessarily an insurmountable problem, but as section 4.7.6 explored, creating a functioning model with economies of scale as a feature presents problems of strategy and timing that must be overcome. In the thesis, this is done by assuming People leave their time with Firms between decision points, thus allowing a flow of production to take place.

While the smooth production function was chosen because it fitted with this ‘leaving time’ approach to producing economies of scale, it also transpired it could be analysed in spreadsheet form to identify the central driving dynamic responsible for producing shifts between stable production equilibria (section 6.3.3; see below also).

Cumulative causation (or feedback) have a role in both the core model and the transmission belt model. It is present in two ways in the core model. Firstly, though an argument about the feedback between firms and workers. Since firms follow workers, there will always be more firms where more workers are; through love of variety (played out via changes in firm number) this locks in a higher real wage (Brakman *et al*. 2009 p.127). This creates the short-run difference in wages between regions. In the long-run, feedback of migrating workers is argued to occur. Section 2.5.5 showed how this dynamic is backed up by reference to the polarity of short-run real wage differences for any given migration change. In short, migration is guaranteed to occur until a stable equilibrium is reached; this leads to either a spread of workers or the emergence of a core in one of the regions.

Cumulative causation in the transmission belt model is due to a feedback between the difference in demand and production. A minimum of two Firms are required for this dynamic (a single Firm would get all market demand and thus its production scale is fixed). A two-Firm model can begin in symmetric equilibrium with both producing at the same scale. If demand passes a substitutability threshold (controlled by a balance between love of variety, economies of scale and, if present, distance costs), the magnitude of the difference between demand and production at a given price can keep both Firms away from their stock target, causing a feedback that ends in a new production equilibrium with the two Firms producing at radically different scales.

The ability to develop non-symmetric demand structures is a key contrast to the core model, in which goods are guaranteed to be demanded symmetrically. No changes in production structure at the firm level are possible; increasing returns occur through firm entry and exit only. The transmission belt model has the opposite dynamic: the number of Firms is fixed and production structure changes play out through production-level economies of scale. In section 6.4, for instance, with two Firms per region, four permutations of production structure emerge for different values of ρ and distance cost.

This method, with the addition of the ‘avoid economies of scale strategy’ code structure (section 4.7.5), allows for more heterogenous agent models with the same dynamic to reach stable equilibria, although stability is dependent on parameter settings and is affected by an interaction of love of variety and the damping response time of Firms targeting a stock level.

Relating back to the ‘ecological criticism’ discussion in section 2.3.5, the ability to explicitly separate economies of scale from love of variety potentially provides a way to examine spatial economic issues involving ‘Weberian’ weights and density. In combination with the ability to produce non-symmetric production structures, the transmission belt model may be suitable for considering production networks; this point is returned to below in relation to dependencies, as well as in section 7.5 discussing further work.

Market structure

Section 2.3 outlined how market structure has determined the divergent paths of non-spatial and spatial economics, up until GE attempted to bridge the gap. ‘Market structure’ determines the way production and consumption connect. In the thesis model framework, this has been labelled a ‘transmission belt’ to emphasise the explicit feedback between production output and demand.

At root, production and consumption are conceptualised in much the same way in both the core model and the thesis models, in that simple output and optimisation functions are used. The thesis has deliberately avoided any more complex structures that ABM may in theory make possible. The two approaches link production and consumption very differently, however.

In the core model, the equivalent of the ‘transmission belt’ between production and consumption happens in a mathematical argument leading to the deduction of regional real wages. Through a series of simplifications, key equilibrium points are found. These are where the argument can state: *if* supply equals demand *then* the equations can be transformed thus. This happens in two places in the core model leading up to the short-run equilibrium. Firstly, an optimal, demanded quantity of good (deduced through assuming some key macroeconomic quantities are constant) is plugged into firms’ profit equation, which is assumed to be maximised. This is plugged into a regional price index describing a proxy for the real wage, or ‘the cost of a unit of utility’. Second, equations for the optimal amount of demanded good and ‘the optimal amount produced in equilibrium’ are combined. Again, the next step is a series of re-arrangements to arrive at an equation in terms of the regional nominal wage. This is combined with the price index to produce a regional real wage.

While the presence of firms is implicit in the core model, the structures it uses develop an argument for a monopolistic competition market structure. At equilibrium, the number of firms (and thus varieties) change; this is responsible for changes in the real wage and thus ultimately determines where workers in the model go.

In the thesis, ‘partial’ models examine first People’s movement, then Firms’ movement. Both models are driven by People’s demand. There is only direct competition in these two when two Firms compete over location. The transmission belt model makes the ‘general’ link between supply and demand by using Firms that target zero stock, where stock is produced using a smooth economies of scale function. The resulting equilibrium comes about through a robust link between price signals, production and demand, but introducing either a profit motive or an ability to genuinely compete has proved difficult (section 6.3.2). The core model (and other static equilibrium models) avoid this problem by assuming competition has already taken place.

A more fundamental question: what does finding this simple ‘transmission belt’ example, using a smooth economies of scale function and a separate love of variety function, say about increasing returns, if anything? In the simplest spreadsheet example, there is no space, two Firms and one Person. Love of variety alone can determine whether one of those Firms emerges as the more-efficient producer. This happens without any of the assumptions involved in the core model's complex, murky and in places somewhat odd market structure argument.

However, without an explicit competition dynamic, it is difficult to directly compare to a GE approach where, in the absence of transport costs, "free trade ensures that the prices of all varieties are equalized between countries" (Brakman *et al*. 2009 p.130). As section 6.3.2 discussed, it may even be impossible for the transmission belt model to produce any stable competition equilibrium in its present form. The 'transmission belt' does, at least, show that it is possible to produce an increasing returns dynamic with some very simple assumptions, and that both production-level economies of scale and love of variety can interact in a feedback process to cause it.

There is a market structure in the transmission belt model, then, but not a fully competitive one. The final section returns to this point as one of the major areas for further work.

7.2 Dependencies

7.2.1 Overview

The concept of dependencies provides both a useful way to think about how model components connect and draws attention to a possible flaw in much of the way ABM is pursued: it is too reliant on code-mapping (section 3.4) and thus to the ideal of 'loosely coupled' agents interacting quite separately in their environment. This section summarises the dependencies found in the process of developing the results in chapters 5 and 6. The introduction suggested that the OOP language of dependencies was useful not only for discussing agent models, but for thinking about how model assumptions in general link together or cause certain approaches to be ruled out. In this light, the original spatial market structure problems of increasing returns and distributed market activity have determined the dependencies of economic theory (section 2.3).

Dependencies are entirely internal to a model: they create 'tightly coupled' structures that dictate the form of a model's assumptions. So, for instance, pre-GE general equilibrium models using only linear functions create a dependency requiring perfect competition. This in turn means it cannot deal with the fundamental spatial economic requirement for increasing returns (section 2.3).

Examining dependencies also produces some very positive outcomes: it helps clearly reveal other possible modelling avenues. This is aided by the transparency and simplicity of the economics concepts used; dependencies can be hard to see under layers of model complication. Modelling avenues identified in this thesis point in a rather different direction to a lot of existing ABM work, toward 'messy' coding issues where meaning and assumptions cannot be 'cleanly idealised'. This is, however, still an area where ABM's flexibility offers a great advantage:

a huge abundance of possible ways to tackle any modelling problem. The choices leading to the transmission belt model presented many branches; some of these specific dependencies are discussed below.

7.2.2 Model framework dependencies

A key determinant of the models' dependencies is the decision to use a CES function for utility. Two parallel methods were tested (section 5.2.3). First, a constrained optimisation that umbilically links good prices to People's demand. A given set of prices, through the constrained utility equations, tell People what amount of each good is optimal. For the model to link demand to supply, that demand needs to actually be met by Firms. Second, a 'hill-climber' approach which converges on the mathematically perfect constrained optima the more iterations it is given, starting from a random guess at a mix of good amounts and 'climbing'.

Both of these utility methods are capable of optimising a given set of goods using a CES function. The 'hill-climber' is more flexible, but this is paid for: it is computationally expensive. It is, however, potentially able to allow People to deal with scenarios involving a lack of optimal stock amount. Zero stock is no problem for either utility approach: a good is simply not entered into the utility function. But the hill-climber is able to re-allocate any spending from goods that have partially enough stock, so it can be spent on others where more stock is available - thus deviating from the 'objective' amount demanded in the constrained utility function.

Using a constrained utility function means People assessing their utility 'now' and thus needing their demand met at the same point in model time. This causes a dependency in the way Firms must be structured: they cannot be given their own 'instant' assessment procedure. This would be possible, for example, if each time a Firm acted, it was guaranteed to be able to clear its stock. It would then be up to People agents to adjust their price-setting over time instead. Both cannot be instantaneous. There are workarounds, but these involve 'time-machine' firms, able to reset the world and re-run each action to compare different outcomes. This is possible, in terms of coding: state can be kept for the whole model and each Firm given the chance to rerun each action, with stochastics dictating the range of other Firms' responses, in order to 'climb' to the best action. But this is a long way from the goal of allowing agents to make independent decisions: this gives them 'spooky' knowledge of all the paths a market could take.

If Firms were equipped to make optimal decisions 'now', the hill-climber approach may also offer another route out of a key dependency that results from using the constrained utility function: the hill-climber iterations could be interpreted as model timesteps, suggesting a way for good buying to be something other than optimal 'now'. Random good 'slices' could continually adapt as supply changed over time without ever reaching a 'perfect' optimum. The constrained utility approach could not do this.

Price competition over good cost and wages does not work without some limit on stock. If stock is potentially infinite, Firms have no restriction on the amount they can sell. If Firms attempt to optimise price as they do with location in section 5.5, prices repeatedly bounce along next to zero or wages enter a spiral as Firms attempt to out-compete each other. Introducing a stock limit, however, introduces new complications: the model is now moving from 'partial' to 'general' territory. It is difficult to find any compromise between these two, as the problem of unlimited stock illustrates. For the introduction of stock to work effectively, coordination needs to take place between wages, good prices and demand. Attempting to solve this by allowing Firms a 'sonar' feedback between timesteps finds itself beset by DMR problems. Section 5.5 explored this issue by allowing two Firms to attempt optimising their location only, along a line of People with fixed, equally distanced positions. The ability of Firms to correctly respond to market signals was shown to be affected by other events within that market: it was possible for them to interpret meaning incorrectly and thus make inapproriate responses.

The exact way that costs are actually communicated is also key. In an agent economic model, what do agents pay with? Firms buying labour and selling goods require some medium of exchange. As section 4.7.2 discussed, many agents using some cost signalling currency may find themselves producing unintended entropy-like dynamics where the currency becomes the limiting factor and distracts from the analysis. While other cost-communication methods may be possible, the solution in this thesis has been to make People's time the currency and to subsume wage and goods into a single exchange, thus massively reducing the complexity. But that does rule out considering labour and good markets separately: another dependency.

The transmission belt model introduces its own dependencies. As section 4.7.5 explained, 'economies of scale' dynamics have the potential to be a thorny coordination problem of an 'El Farol bar' type, caused by many agents choosing what to buy and where to work. This is solved through two assumptions. First, People are guaranteed demand is met at the price they find (as well as combining the wage/good decision into one). Second, they 'leave' their time with the Firms they buy from until their next decision, when they remove it and decide anew. This allows a flow of production to take place that assumes People invest time between decision points. Economies of scale thus become possible and, as a result, stock targeting also becomes possible. This is a particularly tightly bound set of dependencies that the transmission belt model needs to work in a many-agent situation, though the spreadsheet example (section 6.3.3) illustrates how its dynamic functions without these complications.

Where the DMR two-Firm model enabled Firms to optimise location (albeit through some noise), the price volatility inherent in the transmission belt model makes DMR much more problematic. As section 5.6 discussed, there are workarounds that involve making sure Firms can make decisions in an environment where the signal-to-noise ratio is workable for them (which only requires them to take correct actions more often than incorrect ones on average). Given the volality of the transmission belt model's stock targeting mechanism, this would require oscillations to damp enough between actions for any action to produce a readable 'sonar'. However, this would entail the entire model pausing and Firms 'telling' it when to restart. While the current structure is some way from the ideal of agents making perfectly independent decisions, this approach would firmly rule that out.

This noise issue rules out even Firm location optimisation which is, as section 5.5 discussed, in principle no different from the process of searching other price spaces. Equally, as discussed in section 6.3.2, Firms' ability to optimise their profit must also rely on knowing if their own action (changing the markup target) has led to an increase in revenue.

By subsuming exchanges into one transaction, the transmission belt model builds in two other dependencies. One has been mentioned: there is no way to model a separate labour market, which may be key to allowing a stable competition dynamic, since at present People have no 'sanction' against Firms skimming very nearly all production output. Relatedly, it gives the wage-labour exchange a love of variety dynamic: People buy a mix of goods and in doing so also give a mix of time to Firms. Tendentiously, there are real-world parallels to be made to people working for more than one employer or in managing 'portfolios' of food production sites (Rose and Tikhomirov 1993), but this still nevertheless places a restriction on the model.

As mentioned above, the size of the market also matters for price signals, especially for those of the DMR type where noise is an issue. In an agent model, agents need to be able to receive *some* information to have a basis for action. Where the core model *relies* on an argument that the 'market is large' enough to avoid any direct price interaction, in the thesis model framework, a market this large would make signals increasingly difficult for agents to use.

The transmission belt model represents one possible path through these dependencies to a working model. Highlighting dependencies in this way suggests other modelling ideas to test, though many may involve radically different root modelling assumptions. The final section of this chapter discusses some possible further research ideas based on this.

7.3 Different types of interaction

Agent interaction is the main motivation for using ABM. For 'interaction' to take place, outcomes need to not be simply ergodic: the presence of multiple agents must be a factor, not just producing stochastic outcomes where, for example, a model with one agent run many times produces the same result as many agents run once. The dependencies described in section 7.2 underpin four distinct types of interaction in the thesis' models. This section examines how and why they differ in their source and effect, as well as discussing their relationship to assumptions.

Two interaction types avoid coordination: the density cost and the approach to economies of scale. Two others - Firms' DMR location feedback and the transmission belt model's stock targeting - require some form of coordination to occur, meaning that a set of interdependent decisions determine model outcomes.

The economies of scale dynamics of the transmission belt model, as mentioned in the previous section, avoid coordination problems by allowing demand to be met while providing a method for production to change scale as the 'flow' of time input changes. It is, technically, a point of interaction: the change of production scale is essential for the transmission belt mechanism and this comes about through many agents moving more demand to a subset of Firms. It is the most straightforward of the interactions discussed here. It illustrates that there are effective options for avoiding potentially insurmountable coordination problems.

The density cost provides a way to think about how individual agent actions in space can aggregate, through external effects, into spatial morphology. Disadvantages arise if the target of interest is more specific land market dynamics; the density cost has no market exchange. But this is also its advantage: People responding to it are not capable of making a 'wrong' decision, as the DMR Firms are. Whatever outcome they choose is guaranteed optimal *at their point of decision*. The emergent outcome, which may be sub-optimal for some, is externalised to the landscape - it does not happen inside the agents themselves but through an accretion of location decisions and external effects. As section 5.4.7 says, the result shows a path-dependent outcome: a spatial equilibrium will result, but some agents may have hold of a better location than others. On each agents' turn, however, if no better location is available, no change happens: a Nash equilibrium has been reached.

At the other extreme, the DMR scenario in section 5.5.2 makes Firms act, then wait to see how their action affected the world. Firms' knowledge is limited to their own revenue changes in order to see how they cope with the need to react to a simple price signal. They have only one action capable of effecting change and must work out if it has helped or hindered their aim to maximise revenue. Each Firm faces a world with one other Firm in it, both of them competing for a finite pot of demand. In the space between turns, their data may have been affected by factors other than their own action.

It was shown that their ability to interpret the data correctly is dependent on the effect other agents have on the datum they are using, as well as the way in which love of variety can affect the magnitude of those signals because of elasticity between goods. Interpretation of these signals was also shown to be dependent on noise levels. The DMR model provides a coordination mechanism but illustrates the challenges of programming agents able to manage market signals. This weakness does allow for a close study of the problem, however: the DMR issue can be examined in detail by testing in what way Firms succeed or fail.

Why can agents so successfully self-organise into spatial equilibrium arrangements when driven by density costs (section 5.4), while even a simple DMR problem (section 5.5.2) proves so challenging? Both are using minimal rationality, able to know only 'better' from 'worse'. Similarly, why can the transmission belt model successfully use price signals to achieve a stable equilibrium, despite large-scale agent interaction, when in the DMR model, just two Firms struggle? The difference is in the density cost's combination with a demand function to rule out 'wrong' decisions. This does not mean the density cost approach is necessarily better, if DMR is the topic of interest.

What distinguishes the transmission belt mechanism from a DMR approach? There is a clear difference in the speed of signalling; it is closer to the cybernetic ideal of 'intrinsic control' mentioned in section 4.7.7, adjusting prices with each and every demand signal received from the market. Each time a Person makes an exchange, using a CES function, they exchange with every Firm in the model - so each Firm gets to respond using its stock targeting mechanism. A Firm's response is direct, in proportion to the change and guaranteed to be the correct polarity; DMR Firms can make exactly the wrong interpretation and thus do exactly the wrong thing.

This overview suggests a diversity of approaches to getting agents interacting, each suitable for different purposes. In-keeping with the theme of assessing ABM obstacles, Friedman's simple phrase seems apt: "everything depends on the problem". This certainly applies to the forms of interaction that may be relevant for a particular agent modelling problem. Spatial economics presents many interaction challenges; addressing them will almost certainly involve a range of interaction types and a willingness to try out new ideas.

7.4 Assumptions

7.4.1 Overview

Section 3.4.3 discussed the common naïve 'assumptions do not matter' interpretation of Friedman. As he says himself, of course they matter. In the case of the core model, for example, one of the driving motivations behind its creation was an unhappiness with the key assumption that factor differences underlay all explanations for trade over distance, as well as problems with the set of assumptions and dependencies that made so much of economics ignore space altogether. Friedman was arguing instead that no perfect recipe exists for assessing the merits of a particular set of assumptions.

As chapter 3 explored, the assumptions of neoclassical economics (and thus GE) stand accused of creating "citadels of crystalline mathematical perfection that would shatter if touched by the harsh rays of reality" (Ball 2007 p.647). ABM has sold itself as the obvious answer, able to deal with the 'messy' reality of real human behaviour (ibid) in a more 'realistic' way. This thesis has suggested these criticisms may have been misplaced, and that agent modelling has fallen into just the 'descriptive realism' trap that Friedman outlined so well in the 1950s. It has also argued this represents a major obstacle to pursuing further work in spatial economic agent modelling, but one that is surmountable with careful consideration the nature of assumptions and dependencies and a commitment to the 'praxis' of model-building.

Avoiding 'descriptive realism' does not mean avoiding reality. A balance needs to be struck between a dogmatic commitment to mapping model structures to real-world counterparts and a complete disregard for the connection between assumptions and the system being modelled.

Bringing models to reality is often done through a basic requirement for structural parameters that have real-world counterparts, enabling direct model testing from data. So, for example, Bosker *et al.* (2010 p.811) seek out empirical counterparts for the core model's main variables. But the importance of theoretical work is in part as a source of new ideas on what sort of structural parameters might be important. This is the ultimate success of a good theory - making 'substantive hypotheses' that propose new 'ontological predictions' (section 3.4.5) that suggest new places to look for real-world data, or new ways to examine existing information.

The density cost approach in section 5.4 identifies a difference between the effect of changing spatial and non-spatial costs, and provides a framework for considering the spatial and non-spatial components of good and people movement as identical in effect. This suggests some places to look for real-world data; this is discussed in section 7.5. Assuming all structural parameters to begin with can dampen this sort of discovery. Assumption of structural parameters is absolutely necessary for applied models, but potentially ossifying for theoretic models.

7.4.2 Model framework assumptions

Previous sections have discussed the thesis model framework assumptions and dependencies in close detail; this section looks at its more broad assumptions. Perhaps the most important is that it treats time as 'just another factor', exchangeable for other factors, and supplied in a free daily flow. Agents need not concern themselves with the 'time' that happens over many model iterations, except in the case of Firms waiting one timestep for a 'sonar' from a previous action. Otherwise, all model effects emerge through actions agents take 'now'. Section 4.3 discussed some of the key elements of real-world time dynamics this bypasses, while section 2.8.3 outlined the importance of risk and uncertainty in real-world time calculations.

The interaction and coordination problems described in section 7.3 are all deeply affected by time assumptions. Section 2.4 explained that Hagerstrandian diffusion dynamics were to be explicitly excluded from the thesis model framework. The simple time structure used has allowed a range of approaches to thinking about agent spatial economics. The issues raised by a very basic DMR problem throw into sharp relief some of the challenges that might await once broader time-space considerations begin to come into play. This relates also to the use of Euclidean space; reality is networks and changing infrastructure (section 2.8.1). Under assumptions of no network change and a high 'route factor' it is possible to use Euclidean distance as a reasonably accurate proxy but it is still clearly unrealistic.

The use of utility and maximisation is a key assumption. Chapter 3 spent some time exploring critiques of utility. It also discussed how there was some confusion over the 'level' that utility should be applied at, whether assumptions like the representative agent could lead to problems, and whether the fact that people are not, actually, utility-maximisers means it should be avoided in agent models. The framework presented has demonstrated at the very least that rational spatial outcomes can result from the interaction of many agents given utility functions. The density cost would not work without this utility foundation allowing People to 'trade off' differing spatial economic options; elasticity of substitution is key for the transmission belt model dynamic. It has been able to deal with the imposition of heterogeneity that spatial location introduces: agents in different places will face differing costs, and this will in part determine their optimal location choices. It has been argued that there is no need to appeal to psychological plausibility: to the extent that people will react to space cost changes consistently, treating agents as utility-maximisers is useful.

The particular utility approach used, the CES function, proves the value of combining methods. It has been an indispensable 'Swiss-army knife' for simplifying model dynamics: an 'as if' proxy for much more complex processes. The consumption of n goods through a CES utility function is, of course, unrealistic. It subsumes a whole collection of processes that would, in reality, have to take place for such diverse bundles to be assembled. In the thesis model framework, People using the CES do not need to travel to 'fetch' each good; because of this, all route complexities are bypassed. Nor are People required to find a marketplace to buy them. They face no uncertainty in their space search: prices of all goods are known, regardless of distance. But in Feynman's terms, being in a framework that allows the CES approach to be true does not mean being ignorant of the messy reality it is being asked to summarise. As a method for approaching the problem of space-costs using a relatively simple set of assumptions, it has proved invaluable.

The density cost approach was asked to stand in for the messy complexities of price-based land market interactions, as well as more intangible 'congestion costs'. It meshes naturally with treating agents as separate decision-making units, but it has many features that make it unlike any real-world land market, reducing all complexities to a 'second nature' interaction (section 2.7.4) with other proximate agents. It actually does away with any market at all. All assumptions, however, will have some similarities to the phenonema being examined, and some contradictions. In this case, by removing several layers of complexity, a density cost allows the model to focus on its underlying rationale: examining how agents' spatial economic choices are affected by proximity costs; where they work, what they buy and where they choose to locate.

The transmission belt model is also replete with 'heroic assumptions'. In particular, no firm in reality targets a stock level like a hydraulic shock absorber, but this is the method that worked best for the model. It also illustrates the dangers of confusing emergence with explanation. As chapter 3 argued, ABM has tended to privilege the 'Hayekian' research agenda: to "show how a solution is produced by the interactions of people each of whom possesses only partial knowledge" (Hayek 1945). The transmission belt model appears to do this. Can many agents, through price signals, coordinate to produce an optimal production level that maximises welfare? Yes. But this emergence by itself clearly does not explain what is causing it; this took a more detailed, simplified spreadsheet approach to track down a difference in magnitude of change (section 6.3.3).

In an agent context, the coordination question arises due to the nature of the modelling method - but that does not make it the default question to be asked. If there are ways round having to solve it, there may be good reasons to do so. It could be argued this is one of ABM's built-in 'streetlight effect' limitations: precisely that coordination needs so explicitly addressing (or avoiding, as with the density cost and strategy-free economies of scale method) before any research questions can be addressed.

7.5 Future work and concluding comments

Section 1.4 framed some of the ‘big questions’ of spatial economics in terms of contemporary arguments about relocalisation and changing costs. It ended by asking, “is ABM well-positioned to deal with understanding how the ‘gestalt whole’ spatial economy changes?”

The preceding analysis of dependencies and assumptions is chastening and promising in equal measure. Some narrow modelling problems have been solved, most notably the link between supply and demand in a spatial setting, including a dynamic production equilibrium. Keeping the core elements of the model framework simple has made it possible to clearly see some of the other obstacles in the way of progress.

The most important lesson is perhaps this: it is highly unlikely that simply recreating spatial economic actors using an OOP framework will solve the problem. The challenges go much deeper than this. This thesis has argued that the lure of OOP-built virtual worlds able to furnish hermetic explanations has distracted from some hard-earned lessons earlier modellers discovered. The ideal of perfect ‘encapsulation of reality’ has been tried before; arguments that ABM represents a ‘year zero’ able to escape the problems this faced first time around do not stack up.

That is the chastening part. There is promise, however: agent modelling is a powerful and flexible approach. As demonstrated, when compared to GE, it has its own idiosyncratic problems, but it also provides a profound flexibility and is very amenable to a problem-solving, pragmatic approach.

This final section considers some possible future routes that spatial economic ABM might take. Many of these have already been considered during the discussion of dependencies and assumptions; a lot of scope lies here before adding any further complexity. A first problem might be: how to make an agent market structure that can improve on competition dynamics in a spatial setting? This thesis has - like GE - tactically avoided issues of collusion and strategy. The transmission belt model, however, appeared to show genuine competition to be problematic. Is it something about the model’s dependencies that causes this, or can this be addressed within the same framework?

Competition in a market stucture is a fiendish problem from an ABM point of view: many things have to be explicitly specified. While the thesis has argued against always attempting to make every dynamic explicit, ABM does still encourage that approach - and nowhere more strongly than the birth and death of the agents themselves, if entering and exiting the market is vital to understanding competition. What kind of mechanism can be used?

As mentioned, all neoclassical economic theory (including GE) bypasses the problem by assuming competition has taken place. As the core model shows, this can still leave a great deal of complex algebraic organisation to be able to produce a workable model, but it still builds on static assumptions. Perhaps there are simplifications suited to ABM, as the density cost effectively simplified land and congestion costs. Or perhaps competition requires an ABM approach that tackles all its coordination problems head-on.

Some other problems arising from the listed dependencies include: is there a way of separating out Peoples' demand for wages from their good demands? Would that add anything useful to the analysis? Separating out different demand sectors requires some way to communicate costs; this was one of the key difficulties avoided by subsuming wages and goods into one demand function. What other methods for communicating costs can be used that would allow more flexibility while avoiding the 'rivalrous money' / 'currency entropy' problem (section 4.7.2)? Can the hill-climbing approach to utility keep demand economically sensible while allowing Firms a different approach to production that does not involve assuming demand is always met? Is solving DMR-type coordination problems even necessary for understanding spatial economics from an ABM point of view - might it be fine to use Firms with uncanny knowledge abilities?

How do the spatial equilibrium outcomes of the density cost approach compare to a more explicit land market model? Is the 'opposite' reaction of spatial and non-spatial costs (section 5.4.3) something that only makes sense in the framework presented here or is it confirmed by other approaches? Is it detectable in reality? An increase in wages would lead to a relative decrease in commute costs, so the two costs are not as cleanly separable as in the density cost model, but the magnitude of change might be detectable.

Outside of the range of dependencies discussed above, more complex problems await. By far the most challenging of these is time. This is at the heart of the problem of modelling Isard's 'gestalt whole' of agents who may be taking actions very far apart in time and space. It could be argued it is also vital for addressing the 'problems' described in section 1.4. Keeping it simple has been very useful for the thesis models. But a rich array of possible modelling ideas exist beyond this approach to agents optimising 'now' or at most between two timesteps. Any logical journey to make a model capable of examining this would need to make strategic simplifications of its own, but it may well be some approach using a CES function could achieve that goal in a generic way.

As well as demand being distributed over 'timespace', the issue of changes in production structure is deeply time-dependent. The flexibility demonstrated by the transmission belt model hints that perhaps ABM could be effective here. As costs change, this alters forward and backward production linkages (to use Marshall's language). The full problem of modelling this Jacobs-like 'tangled bank' (Ellerman 2002 p.6) of connections between productive units can be imagined as a mesh of differing input and output elasticities. Over distances, these processes overlap with diffusion issues as well as the network modelling concerns discussed in section 2.8.1.

Are there simplifications that can be made to keep this problem tractable when moving towards a more 'tangled bank' model, or are more complex timing relations required? Is it possible to use input-output frameworks to model changes in morphology in this 'tangled bank?' Is it possible to understand its relation to fundamental 'primary sector' production and each following stage? The problem is in combining the fact that "all products are simultaneously inputs to some organisations and outputs from others" (Birkin and Wilson 1986 p.176) with how cost changes may affect the long-term location choices of agents, and the infrastructure development they must work with.

Speaking of infrastructure, does any model of the 'gestalt whole' have to take account of Haggett *et al.*'s challenge: the need for a 'comprehensive model of route development' (Haggett *et al*. 1977 p.95) capable of incorporating agents' response to cost changes, cognisant of the fact that it is 'equally plausible to regard transport investment as a result of a need for movement or as a generator of movement" (Chapman 1979 p.230, from section 2.8.1)?

Chapter 1 mentioned the work of Kerschner and Hubacek (2009; section 1.4) examining the effect of oil cost change in an input-output framework but such approaches are bound both by rigid categories and an inability to conceptualise geographical change. This is something Jacobs was particularly critical of: in theorising how production morphology developed, she argued that:

> "The point is that when new work is added to older work, the addition often cuts ruthlessly across categories of work, no matter how one may analyse the categories. Only in stagnant economies does work stay docilely within given categories" (Jacobs 1970 p.62).

If she is right, the kind of categorisation methods used to collect data for input-output models will struggle with modelling change. While Jacobs' absolutism would appear to rule out data collection altogether, the kind of analysis carried out by Kerschner and Hubacek clearly helps to develop understanding of how different economic sectors are stressed by space cost change. But can generic methods be developed, parallel to the flexibility of the CES function, able to model changes in input-output in a fluid way? Such an approach could look towards other generic approaches to product differentiation such as that developed by Lancaster (1975), where goods are distinguished by their distance in something akin to a 'characteristics space', rather than the generic, across-the-board love of variety used in this thesis.

Food production and consumption is a fascinating case study in this respect: the shift in agricultural production is a quintessential 'Engel's law' dynamic (Zimmerman 1932; Murata 2002; Murata 2008), where an increase in overall wealth leads to a drop in the percentage of revenue going into food production. In terms of production morphology, that dynamic masks a complex process where the changing opportunity costs of time have moved food processing out of the 'productive unit' of the household (see e.g. Taylor and Adelman 2003), slowly 'outsourcing' more elements of food processing into the marketplace. The morphology of food production, then, has changed 'from field to fork'. The problem is a difficult one, however, precisely because those changes are not isolated within any one, neatly definable sector.

The 'transition' literature mentioned in the introduction has a large bias towards food: a burgeoning literature on 'alternative food networks' has developed, in which by and large 'alternative equals good / conventional equals bad' (Penker 2006 p.377). The underlying rationale for recommending localising these networks is a common-sense view that as increase in fuel prices (either due to peak oil or a carbon price) will make the primary need for food (and the localisation of other key non-food activities as far as possible) over-ride all other economic considerations. The economic case for this common-sense view is unproven; most extant theories say it is wrong. But equally, this thesis suggests there is no clear refutation of it either, if one considers space cost change in the round.

Though many approaches exist able to say important things about aspects of the problem, given what has been discussed in this thesis, it is understandable that many economic geography researchers are skeptical of both economic and agent-based methods, if spatial economic approaches simply ignore problems beyond their 'streetlight'. It is unlikely any one theoretical approach will offer definitive answers and as Isard says, "formal theory, in and of itself, is highly unsatisfactory, too general and accordingly too sterile" (Isard 1956 p.36). The tendency to disappear into virtual worlds, rather than find ways to use theory as an 'engine of analysis' (section 3.4.4) must be fought.

Agent modelling could be ideal: as this thesis has demonstrated, spatial economic agents are capable of producing robust rational outcomes in a flexible way. They could perhaps be particularly well-suited to the kind of change just described - not in creating complex, descriptive simulacra, but in using generic micro-economic ideas to study shifts in morphology.

The transmission belt models are examples of how economic change can produce a jump between discrete outcomes, mirroring the 'catastrophic' shifts the core model describes. This naturally presents the question: what sort of radical changes in morphology potentially exist in reality? The 'transition' literature's take suggests cheap energy is a prerequisite for the existing industrial food network morphology. The extreme version of this is that we 'eat fossil fuels' and as the cost of those goes up, that industrial system will not survive. Kerschner and Hubacek's examination of the data says exactly the opposite: the agricultural sector is one of the *least* vulnerable to oil cost change. This lack of agreement suggests there is some way to go before a full understanding of changes in morphology can exist, but it also hints that modelling and data-driven analyses can offer answers quite at odds with common-sense empirical understandings.

A last note on modelling philosophy and how this informs research choices. As mentioned in section 2.3.3, the neoclassical philosophy of 'endogenise everything' may be faulty. The rationale for this philosophy is clear enough: the more can be shown to emerge from a model's internal dynamics, the more one can claim the model 'explains' the dynamic it addresses. In the case of the transmission belt model, the 'transmission belt plus density' model presented in section 6.4.2 utilised its 'general' ability to link supply and demand to endogenise the emergence of a difference in good price that had been externally imposed in section 5.4.6.

But the 'endogenise everything' approach is predicated on the morphology of the spatial economy being entirely economic in origin. Section 2.3.3 noted Fujita musing that "economic development may represent merely a possible result of the more fundamental process of social and cultural development". It certainly may be true that while economic forces may have formed a given settlement pattern, non-economic forces keep it stable, even in the face of strong economic headwinds. There may be mix of dynamics, with economic forces powerfully shaping outcomes but not solely determining them. The difference between those two is crucial: it means pursuing some ultimate economic 'theory of everything' to explain the spatial economy is likely to fail, as it excludes key factors.

ABM has sold itself as a method for pursuing this brave new world of complex social forces, and although this thesis has argued it has got ahead of itself, as the discipline matures it should be well positioned to exploit these possibilities. It has been argued that no model should be asked to stand in as a 'virtual world', only help as an engine of analysis. No modelling analysis will be able to 'solve' the problem of optimal production given changing space costs. What modelling *should* offer is a way to think through assumptions and develop new, informed intuitions. As Marshall put it:

> "Economic laws and reasonings in fact are merely a part of the material of which conscience and common-sense have to make use in solving practical problems, and in laying down rules which may be a guide to life." (Marshall 1895 p.x)

Appendix A

Thesis model code

A.1 Introduction

This appendix describes the models produced specifically for the thesis. A download link to a single 'zip' file containing these can be found via:

- **http://bit.ly/olnerthesis2013**

This zip file contains the following:

- "CoreModel" folder: contains a runnable jar file for the core model version in section 2.5. Instructions are included in a "readme.txt" file in the folder.
- "FrameWorkModelBuilds" folder: contains a series of runnable jar files for the thesis models. "listOfModels.txt" explains them and links to the sections of the thesis they appear in.
- "EoS_vs_LoV_ExcelModel_DanOlner2013web.xlsx": the spreadsheet version of the 'transmission belt' mechanism section from section 6.3.4.
- "exchangetestrandom": Matlab file for the random money exchange example, section 4.7.2.

'Jar' files need to stay in their folders to run as they rely on the libraries and config files there. They should be openable by double-clicking. If this does not work, the command line can be used. Navigate to the 'scenarios' folder, and use:

java -jar "*filename*.jar"

Section A.2 explains the UML overview diagram given at the end of the appendix, and goes into more detail on the structure of those classes and their role in the model framework.

Section A.3 describes the runnable models included in the "FrameWorkModelBuilds" folder and explains the interactive features available in those that were used as a virtual laboratory in the thesis.

A quick final note on the use of floating point numbers: for the vast majority of modelling problems covered in this thesis, the choice of floating point numbers presented no problem, despite their well-documented issues (see e.g. Polhill *et al*. 2006). However, those floating points issues do exist, particularly where many interdependent Actors attempt to equilibriate some value to zero. A way to fix this would be to use very large integer values, but this would require a ground-up recoding.

A.2 UML overview

Figure A.1, at the end of the appendix, provides a UML overview of the model classes. It does not include all classes (see the Javadoc for these) but provides an overview of all the main and abstract classes, such as 'Actor', that have more than one extension. The following list briefly explains each class in the UML overview.

1. Main: runs through the top-level model setup, including:
    - use gl.setUp() to set global variables, number of agents etc.
    - Parameter-sweep loops, call each modelRun() and changes values
    - modelRun(): set up each individual run, includes:
        - ActorMaker.makeActors
        - set up visualisations
        - set initial actor locations
        - set up Action permuters
        - set up data 'Buckets'
        - iterate through single run day number
2. TimeLine: the main 'audible' class (see below) that many others use as a reference to know when to act. TimeLine's day is iterated in modelRun().
3. 'gl': short for Global, shortened version used for quick reference to global variables. 'gl' sets up all of these global variables; other classes use it extensively both for storing and retreiving global variables.
4. ActorMaker: instantiates all Actors, goes their action setup and does any specific jobs like setting variable spreads across Actors.
5. Space: contains all methods related to geography, including methods for Actors to get a list of 'whosInRadius' relative to their location, or the location of a Bundle they are

testing. Also does the following: finds density cost; provides methods for setting Actor location in desired patterns; deals with wrapping if model is a torus or 'racetrack' (wrapping line); implements Drawable, so takes care of carrying out drawing of geographical visualisations

6. Actor: top-level object over-written by sub-types Person, Firm, SingleActorPerson
7. Action: again, overwritten for particular actors by PersonAction, FirmAct and SingleActorAction
8. UtilityShell and ProductionShell: wrappers, each wrapping a 'giveUtility' or 'giveOutput' method for a swappable Utility or Production class.
9. Bundle: a collection of Goods and their sellers and locations, each of which gets its own utility level when maximised.
10. Good: single good, containing details of cost, seller and temporary values when maximising. A Good can be sold by any class implementing the 'GoodSeller' interface, so People can in the single Actor models sell as well as buy.
11. ProcessingDrawer: puts an inner class Processing applet into a frame and has an array of Drawables who are asked to draw when 'heard'.
12. Important interfaces (see below for others in the 'observe' classes).
    - HasLocation: guarantees that Space's methods can work on any class implementing
    - Drawable: any class that implements this and registers with a Drawer will have the Processing applet passed to its 'draw' method when 'heard'. It is then up to the class how and what it draws, within the framework of a particular Drawer.
13. Output classes:
    - DataStore: holds all the 'Buckets', implements listener, calls Buckets to get data when 'heard'. Can be used to store single or multi-run data.
    - Bucket: extended by particular 'Buckets' that will deal with storing the required data.
    - CSVWriter: two versions of this one listens for the end of model runs to write data, the other is static and will write data at any required point.
14. 'Observer' classes
    - Audible (interface): any class that wants to be audible to Listeners. In theory could be used for event-driven model, but in practice, only TimeLine currently implements Audible.

- Listener (interface): any class that wants to know about what day it is needs to implement this and provide a Shouter to an Audible.
- Shouter: keeps a record of its Listener, lets it know when its been asked to shout.
- ListenerWrapper: wraps Listener's 'heard' method, so that Listeners can be exchanged at runtime. Used by the visualisation mouse control system.

A.3 Framework outline

A.3.1 Visualisation and interaction

Main visualisation window

The main visualisation window will display the visualisations as described in the relevant sections of the results chapter. Each scenario is listed below, with a reference to the appropriate section. The window also provides menus for changing the assignment of variables to mouse buttons. These are top-left, just called 'mouse1' and 'mouse2'. When selected, clicking and dragging with the left (mouse1) or right (mouse2) vertically will change values. The current value is displayed bottom left. The variables are repeated in each menu. Those available vary from model to model, but will include the following:

- deliv: sets good delivery cost for all goods
- wage: sets all Firms' wage offer
- price: sets all Firms' good base price
- tech: If People have to 'pay' distance to work, this sets the 'tech' element
- CES: If a CES utility function is used, this sets ρ.
- model: Controls model view angle by dragging
- density: sets density cost level for all People
- uViz: changes magnitude of visualisation of utility levels
- lViz: changes magnitude of visualisation of buying levels

There are also a number of toggle buttons. Some are specific to particular scenarios, but the common ones are:

- peoplemove: toggle whether People are able to change location
- firmsmove: toggle whether Firms are able to change location
- wageTime: toggle whether People must 'pay' the commute time to Firms

Graph visualisation

For those models that include a graph output, note the following:

- Graphs default to showing a moving record of 400 values. ‘f’ will turn on or off the following behaviour.
- The graph will find the lowest and highest value and display between them. It will fix to the highest, so if values drop, the graph can be reset by pressing ‘r’.

All scenarios can be exited by pressing escape while the focus is on the main visualisation window.

A.3.2 List of scenarios

As mentioned, the ‘ModelBuilds’ folder contains a description of each included scenario and the section it relates to this is included here for quick reference:

- MonoCentricDensity (section 5.4.7: “Spatial equilibria are not always utility equilibria”): Starts with density cost = 100, at the point where two bands of utility indifference emerge. Two correlation graphs show density and utility versus distance from centre.
- TwoFirmsLineMove2000dayAverage (section 5.5.1: “Two Firms optimising location on a line”): Two firms on a 1d line of length one compete for demand by moving location only. Fifty People spread evenly along the line from 0 to 1. Means taken over 2000 timesteps for Firm position and People’s demand. White arrows shows Firms’ mean location; white bars show one standard deviation each side of this; thin black vertical lines show actual positions taken by Firm over previous 2000 timesteps. Timestep value to average over can be changed via the config file “Hotelling_Line_DMR.properties” in the ‘configs’ folder.
- ‘SymmetryBreak’ (section 6.4: “Introducing space into the transmission model”) folder contains the following two jars:
  - SymmetryBreakParameterSweep: Greyed out correlation graph will fill with datapoints as the model sweeps up rho and delivery costs in turn. Main visualisation window blocks: Firms’ time input from People.
  - SymmetryBreakParameterSweepFast: after 500 days, model run speeds up to show full range of transitions quickly and the way each Firm reacts. Main visualisation window blocks: Firms’ time input from People.
- ManyAgentsLineTransmissionBelt (section 6.4.1: “Space and welfare outcomes in a many-agent model”): line of 19 firms/people set at point that transition to a single more-efficient producer will emerge. Mouse is linked to rho (left button) and distance costs

(right button). Increase distance costs to return the model to the spread equilibrium. Increasing rho (increasing substitutability of goods) will in turn allow efficient production to re-emerge. At higher distance costs, more than one efficient producer will emerge.

- TransmissionBeltPlusDensity (section 6.4.2: "Combining the transmission belt model with density cost"): Two firms a distance of 1 apart, 300 people. If rho is raised (left-mouse click and drag up) at rho=0.9, transition to single more-efficient producer takes place, People move, create settlement there, buy in from further firm.

- TwoRegionFourFirmsTransmissionBelt (section 6.4: "introducing space into the transmission belt model"): discrete two region model with two Firms per region. For values of rho and distance cost (attached to mousebutton 1 and 2) four different production regime types emerge (see section 6.4).

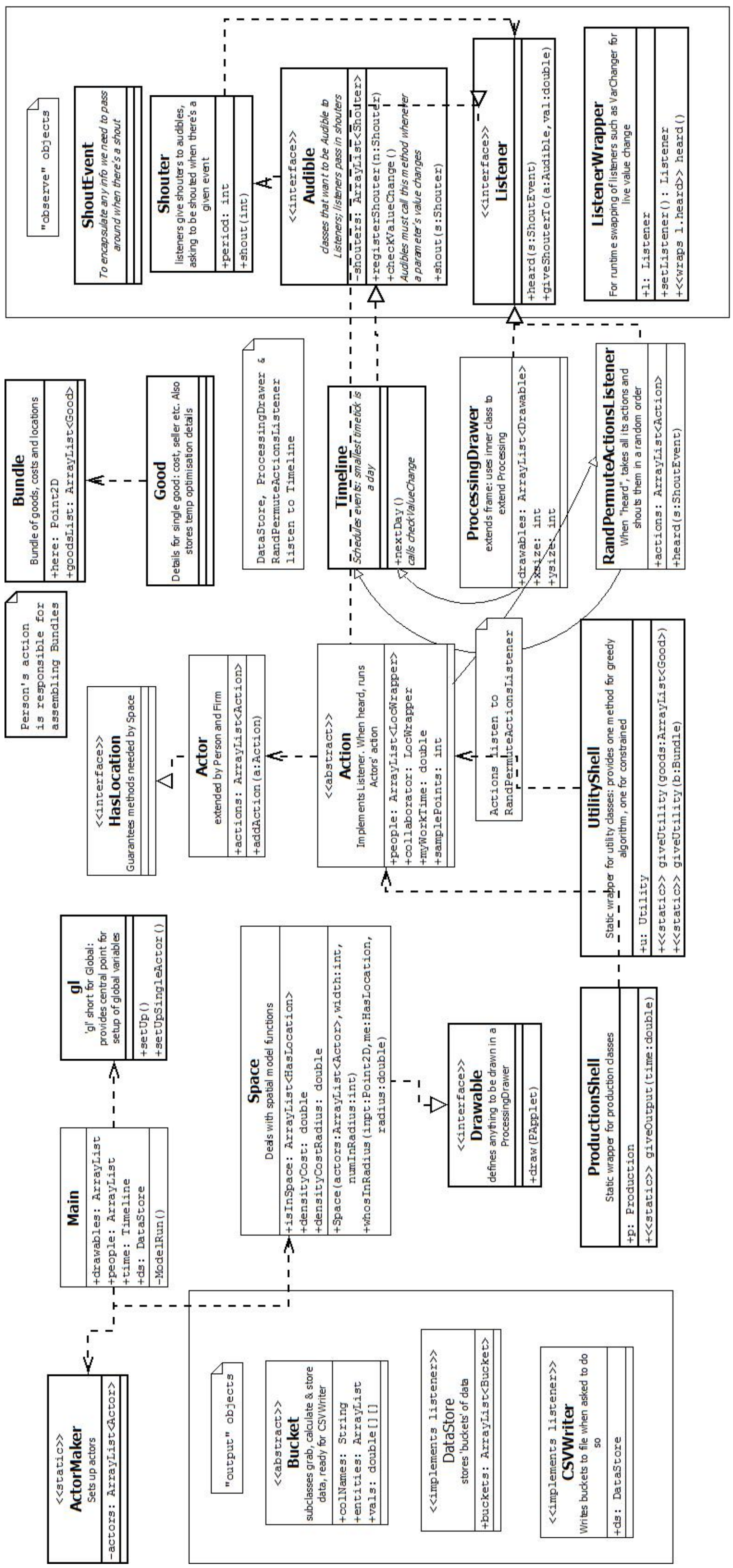


Figure A.1: Model framework UML overview

Bibliography

Abernathy, F. H., Dunlop, J. T., Hammond, J. H., Weil, D. (1999). *A Stitch in Time: Lean Retailing and the Transformation of Manufacturing - Lessons from the Apparel and Textile Industries*. OUP USA.

Almassi, B. (2009). 'Trust in expert testimony: Eddington's 1919 eclipse expedition and the British response to general relativity'. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* **40**(1):57–67.

Alonso, W. (1964). *Location and Land Use: Toward a Theory of Land Rent*. Harvard University Press.

Anderson, J. E., Wincoop, E. v. (2004). 'Trade Costs'. *Journal of Economic Literature* **42**(3):691–751.

Arrow, K. J., Chenery, H. B., Minhas, B. S., Solow, R. M. (1961). 'Capital-Labor Substitution and Economic Efficiency'. *The Review of Economics and Statistics* **43**(3):225–250.

Arthur, W. B. (1994). 'Inductive Reasoning and Bounded Rationality'. *The American Economic Review* **84**(2):406–411.

Bachmeier, L., Li, Q., Liu, D. (2008). 'Should Oil Prices Receive so Much Attention? an Evaluation of the Predictive Power of Oil Prices for the Us Economy'. *Economic Inquiry* **46**(4):528–539.

Ball, P. (2007). 'Social science goes virtual'. *Nature* **448**(7154):647–648.

Ballas, D., Clarke, G. (2001). 'Modelling the local impacts of national social policies: a microsimulation approach'. *Environment and Planning C: Government and Policy* **19**:587–606.

Batten, D. F., Thord, R. (eds.) (1988). *Transportation for the future*. Springer-Verlag, Berlin; New York.

Batty, M. (2012). 'A Generic Framework for Computational Spatial Modelling'. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (eds.), *Agent-Based Models of Geographical Systems*, pp. 19–50. Springer, Netherlands.

Batty, M., Crooks, A. T., See, L. M., Heppenstall, A. J. (2012). ‘Perspectives on Agent-Based Models and Geographical Systems’. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (eds.), *Agent-Based Models of Geographical Systems*, pp. 1–15. Springer, Netherlands.

Baumol, W. J. (1976). *Economic Theory and Operations Analysis*, 3rd Edition. Prentice Hall, London.

Baumol, W. J., Blinder, A. S. (2005). *Economics: Principles and Policy*. South-Western, Div of Thomson Learning.

Becker, G. (1976). *The economic approach to human behavior*. University of Chicago Press.

Beckmann, M. J. (2010). *Lectures on Location Theory*. Springer.

Beer, S. (1994). *Brain of the Firm*, 2nd Edition. John Wiley & Sons.

Beer, S. (2002). ‘What is cybernetics?’. *Kybernetes* **31**(2):209 – 219.

Behrens, K., Lamorgese, A. R., Ottaviano, G. I., Tabuchi, T. (2007). ‘Changes in transport and non-transport costs: Local vs global impacts in a spatial network’. *Regional Science and Urban Economics* **37**(6):625–648.

Betz, G. (2006). *Prediction or Prophecy? The Boundaries of Economic Foreknowledge and Their Socio-Political Consequences*. Deutscher Universittsverlag.

Birkin, M., Wilson, A. G. (1986). ‘Industrial location models 1: a review and an integrating framework’. *Environment and Planning A* **18**(2):175–205.

Black, D., Henderson, V. (1999). ‘Spatial Evolution of Population and Industry in the United States’. *American Economic Review* **89**(2):321–327.

Black, W. R. (2003). *Transportation: a geographical analysis*. Guilford Press.

Blaug, M. (1992). *The Methodology of Economics: Or, How Economists Explain*, 2 Edition. Cambridge University Press.

Blaug, M. (1997). *Economic theory in retrospect*. Cambridge University Press.

Blaug, M. (2009). ‘The Trade-Off between Rigor and Relevance: Sraffian Economics as a Case in Point’. *History of Political Economy* **41**(2):219–247.

Boschma, R., Martin, R. (2007). ‘Editorial: Constructing an evolutionary economic geography’. *Journal of Economic Geography* **7**(5):537–548.

Bosker, M., Brakman, S., Garretsen, H., Schramm, M. (2010). ‘Adding geography to the new economic geography: bridging the gap between theory and empirics’. *Journal of Economic Geography* **10**(6):793–823.

Brakman, S., Garretsen, H., van Marrewijk, C. (2009). *The New Introduction to Geographical Economics*, 2nd Edition. Cambridge University Press.

Brakman, S., Heijdra, B. J. (2001). 'The Monopolistic Competition Revolution in Retrospect'.

Brakman, S., Heijdra, B. J. (2011). 'Introduction'. In S. Brakman & B. J. Heijdra (eds.), *The Monopolistic Competition Revolution in Retrospect*. Cambridge University Press.

Bridge, G. (2010). 'Geographies of peak oil: The other carbon problem'. *Geoforum* **41**(4):523–530.

Brimberg, J. (1995). 'The Fermat-Weber location problem revisited'. *Mathematical Programming* **71**(1):71–76.

Brons, M., Nijkamp, P., Pels, E., Rietveld, P. (2008). 'A meta-analysis of the price elasticity of gasoline demand. A SUR approach'. *Energy Economics* **30**(5):2105–2122.

Brown, D., Xie, Y. (2006). 'Spatial agent-based modelling'. *International Journal of Geographical Information Science* **20**(9):941–943.

Burstein, A., Neves, J., Rebelo, S. (2003). 'Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations'. *Journal of Monetary Economics* **50**(6):1189–1214.

Button, K. (2010). *Transport Economics*, 3rd Edition. Edward Elgar Publishing Ltd.

Cairns, S. (2005). 'Delivering supermarket shopping: more or less traffic?'. *Transport Reviews: A Transnational Transdisciplinary Journal* **25**(1):51–84.

Carrre, C., De Melo, J., Wilson, J. (2009). *The distance effect and the regionalization of the trade of low-income countries*. Centre for Economic Policy Research.

Chalmers, A. (1999). *What Is This Thing Called Science?* University of Queensland Press.

Chamberlain, A. (1998). 'The Socialist Economic Calculation Debate and the Austrian Critique of Central Planning'. URL: http://the-idea-shop.com/article/97/the-socialist-economic-calculation-debate-and-the-austrian-critique-of-central-planning

Chapman, K. (1979). *People, Pattern and Process: An Introduction to Human Geography*. John Wiley & Sons.

Christaller, W. (1933). *Central Places in Southern Germany*. Prentice Hall, New Jersey.

Colander, D., Rothschild, C. (2010). 'Sins of the Sons of Samuelson: Vision, pedagogy, and the zig-zag windings of complex dynamics'. *Journal of Economic Behavior & Organization* **74**(3):277–290.

Conant, R. C., Ashby, W. R. (1970). ‘Every good regulator of a system must be a model of that system’. *International Journal of Systems Science* **1**(2):89–97.

Conlisk, J. (1996). ‘Why bounded rationality?’. *Journal of Economic Literature* **34**(2):669–700.

Consiglio, A. (ed.) (2007). *Artificial markets modeling: methods and applications*. No. 599 in Lecture Notes in Economics and Mathematical Systems. Springer.

Cooper, J., Browne, M., Peters, M. (1990). ‘Logistics performance in Europe: the challenge of 1992’. *International Journal of Logistics Management, The* **1**(1):28–35.

Cooper, J. C. (1983). ‘The Use of Straight Line Distances in Solutions to the Vehicle Scheduling Problem’. *The Journal of the Operational Research Society* **34**(5):419–424.

Craik, K. J. W. (1967). *The nature of explanation*. CUP Archive.

Daly, H. E. (1997). ‘Georgescu-Roegen versus Solow/Stiglitz’. *Ecological Economics* **22**(3):261–266.

Daly, H. E., Farley, J. (2003). *Ecological Economics: Principles and Applications*. Island Press.

Dawid, H. (2006). ‘Agent-based Models of Innovation and Technological Change’. In *Handbook of Computational Economics*, vol. 2, pp. 1235–1272. Elsevier.

Delgado, M., Porter, M. E., Stern, S. (2010). ‘Clusters and entrepreneurship’. *Journal of Economic Geography* **10**(4):495–518.

Desai, M. (1994). ‘Equilibrium, Expectations and Knowledge’. In J. Birner (ed.), *Hayek, Coordination and Evolution: His Legacy in Philosophy, Politics, Economics and the History of Ideas*, p. 400. Routledge.

Di Paolo, E., Noble, J., Bullock, S. (2000). ‘Simulation models as opaque thought experiments’. In *Artificial Life VII: The Seventh International Conference on the Simulation and Synthesis of Living Systems*, p. 497506.

Diamond, D., Spence, N. (1989). *Infrastructure and Industrial Costs in British Industry*. Dept. of Trade & Industry, London.

Dibble, C. (2006). ‘Computational Laboratories for Spatial Agent-Based Models’. In *Handbook of Computational Economics*, vol. 2, pp. 1511–1548. Elsevier.

Disdier, A., Head, K. (2008). ‘The puzzling persistence of the distance effect on bilateral trade’. *Review of Economics and Statistics* **90**(1):37–48.

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

Downes, S. M. (1992). ‘The importance of models in theorizing: A deflationary semantic view’. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 1992, pp. 142–153.

Duranton, G., Puga, D. (2001). ‘Nursery cities: urban diversity, process innovation, and the life cycle of products’. *American Economic Review* p. 14541477.

Duranton, G., Puga, D. (2003). ‘Micro-Foundations of Urban Agglomeration Economies’. *National Bureau of Economic Research Working Paper Series* **No. 9931**.

Duranton, G., Storper, M. (2005). ‘Rising trade costs? Agglomeration and trade with endogenous transaction costs’.

Duranton, G., Storper, M. (2008). ‘Rising trade costs? Agglomeration and trade with endogenous transaction costs’. *Canadian Journal of Economics* **41**(1):292–319.

Durlauf, P. S. N., Blume, P. L. E. (2008). *The New Palgrave Dictionary of Economics*, 2nd edition Edition. Palgrave Macmillan.

Echenique, M. (2002). ‘Separating the Intensity of Transport from Economic Growth’. *Tech. rep.*, Institute for Transport Studies, University of Leeds.

Economist (2013). ‘New model army’. URL: http://www.economist.com/news/finance-and-economics/21569752-efforts-are-under-way-improve-macroeconomic-models-new-model-army

Edmonds, B. (1999). ‘Modelling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models’. In T. Brenner (ed.), *Computational Techniques for Modelling Learning in Economics*. Kluwer Academic Publishers, Dordrecht.

Edmonds, B. (2004). ‘Against the inappropriate use of numerical representation in social simulation’. URL: http://cogprints.org/4262/

Edmonds, B. (2007). ‘Simplicity is not truth-indicative’. In C. Gershenson & D. Aerts (eds.), *Worldviews, Science and Us: Philosophy and Complexity*, pp. 65–80. World Scientific Publishing.

Ellerman, D. (2002). ‘Jane Jacobs as a Development Thinker’. URL: http://www.ellerman.org/Davids-Stuff/Dev-Theory/JJ-on-Dev.pdf

Epstein, J. (2008). ‘Why Model?’. *Journal of Artificial Societies and Social Simulation* **11**(4):12.

Epstein, J. M. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.

Epstein, J. M. (2006). *Generative social science: studies in agent-based computational modeling*. Princeton University Press.

Farmer, J. D., Foley, D. (2009). 'The economy needs agent-based modelling'. *Nature* **460**(7256):685–686.

Fetter, F. A. (1930). 'Review: Alfred Weber's Theory of the Location of Industries'. *The Journal of Political Economy* **38**(2):232–234.

Figueiredo, O., Guimaraes, P., Woodward, D. (2009). 'Localization economies and establishment size: was Marshall right after all?'. *Journal of Economic Geography* **9**(6):853–868.

Fisher, F. M. (1971). 'Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment'. *The Review of Economics and Statistics* **53**(4):305–325.

Fisher, I. (1927). *A statistical method for measuring 'mariginal utility' and testing the justice of a progressive income tax*. Macmillan, London.

Flake, G. W. (1998). *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems and Adaptation*. London, MIT.

Fowler, C. (2006). 'Reexploring transport geography and networks: a case study of container shipments to the West Coast of the United States'. *Environment and Planning A* **38**(8):1429–1448.

Fowler, C. (2007). 'Taking geographical economics out of equilibrium: implications for theory and policy'. *J Econ Geogr* **7**(3):265–284.

Fowler, C. (2011). 'Finding equilibrium: how important is general equilibrium to the results of geographical economics?'. *Journal of Economic Geography* **11**(3):457–480.

Franklin, S. (1997). *Artificial Minds*. MIT Press.

Franklin, S., Graesser, A. (1996). 'Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents'. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*, pp. 21–35. Springer-Verlag.

Friedman, M. (1953a). *Essays in Positive Economics*. University of Chicago Press.

Friedman, M. (1953b). 'Lange on Price Flexibility and Employment'. In *Essays in Positive Economics*, pp. 277–300. University of Chicago Press.

Friedman, M. (1962). *Price theory*. Transaction Publishers.

Fujita, M. (1999). 'Location and Space-Economy at half a century: Revisiting Professor Isard's dream on the general theory'. *The Annals of Regional Science* **33**(4):371–381.

Fujita, M. (2010). 'The Evolution of Spatial Economics: From Thunen to the New Economic Geography'. *Japanese Economic Review* **61**(1):1–32.

Fujita, M., Krugman, P., Venables, A. (2001). *The Spatial Economy: Cities, Regions and International Trade*. MIT Press.

Fujita, M., Mori, T. (2005). 'Frontiers of the new economic geography'. *Papers in Regional Science* **84**(3):377–405.

Fujita, M., Thisse, J.-F. (2002a). 'Agglomeration and Market Interaction'. *Tech. rep.*, C.E.P.R. Discussion Papers. URL: http://ideas.repec.org/p/cpr/ceprdp/3362.html

Fujita, M., Thisse, J.-F. (2002b). *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press.

Gaffikin, F., Morrissey, M. (2011). 'Community Cohesion and Social Inclusion'. *Urban Studies* **48**(6):1089 –1118.

Gastner, M. T., Newman, M. E. J. (2006). 'Shape and efficiency in spatial distribution networks'. *Journal of Statistical Mechanics: Theory and Experiment* **2006**(01):1015–1015.

Georgescu-Roegen, N. (1975). 'Energy and Economic Myths'. *Southern Economic Journal* **41**(3):347–381.

Gilbert, A. (1974). *Latin American development: a geographical perspective*. Penguin Books.

Gilbert, N. (2007). *Agent-Based Models*. Sage.

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A. (1992). 'Growth in Cities'. *Journal of Political Economy* **100**(6):1126–1152.

Glaeser, E., Kohlhase, J. (2004). 'Cities, regions and the decline of transport costs'. *Papers in Regional Science* **83**(1):197–228.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press, Oxford.

Glasmeier, A. (2007). 'Book Review: Worlds of Food'. *Regional Studies* **41**(5):713–718.

Gold, B., Simons, R. A. (2008). *Proof and Other Dilemmas: Mathematics and Philosophy*. MAA.

Goodwin, P., Dargay, J., Hanly, M. (2004). 'Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review'. *Transport Reviews* **24**(3):275–292.

Goyal, S. (2009). *Connections: An Introduction to the Economics of Networks*. Princeton University Press.

Granovetter, M., Swedberg, R. (2001). *The Sociology of Economic Life*, 2nd Edition. Westview Press.

Grauwin, S., Bertin, E., Lemoy, R., Jensen, P. (2009). 'Competition between collective and individual dynamics'. *Proceedings of the National Academy of Sciences of the United States Of America* **106**(49):20622–20626.

Gray, D., Anable, J., Illingworth, L., Graham, W. (2006). 'Decoupling the link between economic growth, transport growth and carbon emissions in Scotland'. *Tech. rep.*, Robert Gordon University, Aberdeen.

Hagerstrand, T. (1953). *Innovation Diffusion as a Spatial Process*, reprinted edition Edition. University of Chicago Press, Chicago.

Haggett, P., Cliff, A., Frey, A. (1977). *Locational Analysis in Human Geography*, 2nd Edition. Arnold Publishers.

Hamill, L., Gilbert, N. (2009). 'Social Circles: A Simple Structure for Agent-Based Social Network Models'. *Journal of Artificial Societies and Social Simulation* **12**(2).

Hamill, L., Gilbert, N. (2010). 'Simulating large social networks in agent-based models: a social circle model'. *Emergence: Complexity and Organization* **12**(4):78–94.

Hamilton, J. (2003). 'What is an oil shock?'. *Journal of Econometrics* **113**(2):363–398.

Hamilton, J. D. (2009). 'Causes and Consequences of the Oil Shock of 2007-08'. *Brookings Papers on Economic Activity* (1):215–283.

Harford, T. (2007). 'Frequent Flier Food'. URL: http://www.forbes.com/2007/11/11/economics-food-local-forbeslife-food07-cx_th_1113local.html

Hayek, F. A. (1945). 'The Use of Knowledge in Society'. *The American Economic Review* **35**(4):519–530.

Heiner, R. A. (1983). 'The Origin of Predictable Behavior'. *The American Economic Review* **73**(4):560–595.

Hicks, S. J. (1989). *A Market Theory of Money*. Clarendon Press.

Hodgson, G. M. (1996). *Economics and Evolution: Bringing Life Back into Economics*. The University of Michigan Press.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press.

Hollis, M. (1994). *The Philosophy of Social Science: An Introduction*. Cambridge University Press.

Hollis, M., Nell, E. (1975). *Rational Economic Man: a philosophical critique of neo-classical economics*. Cambridge University Press.

Holub, H. W. (1977). 'Anti-equilibrium theory vs. equilibrium theory'. *De Economist* **125**(3):393–403.

Hoover, E. (1937). *Location theory and the shoe and Leather Industries*. Harvard University Press, Massachussetts.

Hoover, K. D. (2001). *Causality in macroeconomics*. Cambridge University Press.

Hoover, K. D. (2009). 'Milton Friedman's stance: the methodology of causal realism'. In U. Mki (ed.), *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, pp. 303–320. Cambridge University Press, 1st Edition.

Hoover, K. D. (2010). 'Idealizing Reduction: The Microfoundations of Macroeconomics'. *Erkenntnis* **73**(3):329–347.

Hopkins, R. (2008). *The transition handbook: from oil dependency to local resilience*. Green, Totnes.

Horgan, J. (1998). *The End of Science*. Abacus.

Hotelling, H. (1929). 'Stability in Competition'. *Economic Journal* **39**(153):41–57.

Howitt, P. (2006). 'Coordination Issues in Long-Run Growth'. In *Handbook of Computational Economics*, vol. 2, pp. 1605–1624. Elsevier.

Hume, D. (1739). *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Penguin Classics.

Hummels, D. (2001). 'Time as a Trade Barrier'. *Tech. rep.*, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University. URL: http://ideas.repec.org/p/gta/workpp/1152.html

Hutchins, E. (1996). *Cognition in the Wild*, new edition Edition. MIT Press.

Inc., N. C. (1993). *Object-oriented Programming and the Objective C Language*. Addison-Wesley, Reading, Massachusetts.

Isard, W. (1956). *Location and Space-economy: General Theory Relating to Industrial Location, Market Areas, Land Use, Trade and Urban Structure*. MIT Press.

Isserman, A. M. (1996). ‘It’s Obvious, It’s Wrong, and Anyway They Said It Years Ago? Paul Krugman on Large Cities’. *International Regional Science Review* **19**(1-2):37 –48.

Jacobs, J. (1970). *The Economy of Cities*. Vintage Books, USA.

Jacobs, J. (1986). *Cities and the Wealth of Nations: Principles of Economic Life*. Penguin Books Ltd.

Jacobs, J. (2001). *The Nature of Economies*. Vintage Books USA.

Jana, D. (2005). *Java and Object-Oriented Programming Paradigm*. PHI Learning Pvt. Ltd.

Jefferson, M. (2008). ‘Accelerating the transition to sustainable energy systems’. *Energy Policy* **36**(11):41164125.

Johansson, B., Quigley, J. (2003). ‘Agglomeration and networks in spatial economies’. *Papers in Regional Science* **83**(1):165–176.

Kaldor, N. (1972). ‘Irrelevance of Equilibrium Economics’. *Economic Journal* **82**(328):1237–1255.

Kerschner, C., Hubacek, K. (2009). ‘Assessing the suitability of Input-Output analysis for enhancing our understanding of potential effects of Peak-Oil’. *Energy* **34**(10):1662–1668.

Kirman, A. (1992). ‘Whom or What Does the Representative Individual Represent’. *Journal of Economic Perspectives* **6**(2):117–136.

Kirman, A. P., Vriend, N. J. (2001). ‘Evolving market structure: An ACE model of price dispersion and loyalty’. *Journal of Economic Dynamics & Control* **25**:459–502.

Klamer, A. (1984). *The New Classical Macroeconomics: Conversations with New Classical Economists and Their Opponents*. Prentice-Hall.

Korowicz, D. (2010). ‘Tipping Point: Near-Term Systemic Implications of a Peak in Global Oil Production, An Outline Review’. *Tech. rep.*, Feasta, Dublin.

Kramer, G. J., Haigh, M. (2009). ‘No quick switch to low-carbon energy’. *Nature* **462**(7273):568–569.

Krugman, P. (1991a). ‘History versus Expectations’. *The Quarterly Journal of Economics* **106**(2):651–67.

Krugman, P. (1991b). ‘Increasing Returns and Economic Geography’. *Journal of Political Economy* **99**(3):483–99.

Krugman, P. (1993). ‘The Fall and Rise of Development Economics’. URL: http://web.mit.edu/krugman/www/dishpan.html

Krugman, P. (1996). *The Self Organizing Economy*. WileyBlackwell.

Krugman, P. (1999). ‘Was it all in Ohlin’. URL: http://web.mit.edu/krugman/www/ohlin.html

Krugman, P. (2005). ‘How I work’. URL: http://web.mit.edu/krugman/www/howiwork.html

Krugman, P. (2009). ‘The Increasing Returns Revolution in Trade and Geography’. *American Economic Review* **99**(3):561–571.

Krugman, P. (2010a). ‘Debt, Deleveraging, and the Liquidity Trap’. URL: http://krugman.blogs.nytimes.com/2010/11/18/debt-deleveraging-and-the-liquidity-trap/

Krugman, P. (2010b). ‘The New Economic Geography, Now Middle-Aged’. URL: http://www.princeton.edu/ pkrugman/aag.pdf

Krugman, P. R. (1979). ‘Increasing returns, monopolistic competition, and international trade’. *Journal of International Economics* **9**(4):469–479.

Krugman, P. R. (1991c). *Geography and Trade*. MIT Press.

Ladley, D., Bullock, S. (2007). ‘The effects of market structure on a heterogeneous evolving population of traders’. *Emergent Intelligence of Networked Agents* **56**:83–97.

Ladley, D., Bullock, S. (2008). ‘The Strategic Exploitation of Limited Information and Opportunity in Networked Markets’. *Computational Economics* **32**(3):295–315.

Lancaster, K. (1975). ‘Socially Optimal Product Differentiation’. *American Economic Review* **65**(4):567–585.

Lang, C. (2010). ‘Heterogeneous transport costs and spatial sorting in a model of New Economic Geography’. *Papers in Regional Science* **89**(1):191–202.

Launhardt, W. (1885). *Mathematical Principles of Economics*. Edward Elgar.

Leamer, E., Storper, M. (2001). ‘The economy geography of the Internet age’. *Journal of International Business Studies* **32**(4):641.

Leijonhufvud, A. (2006). ‘Agent-Based Macro’. vol. 2 of *Handbook of Computational Economics*, pp. 1625–1637. Elsevier.

Li, S., Haefen, R. v., Timmins, C. (2008). ‘How Do Gasoline Prices Affect Fleet Fuel Economy?’. *Working Paper 14450*, National Bureau of Economic Research. URL: http://www.nber.org/papers/w14450

Losch, A. (1940). *The Economics Of Location*. Yale University Press, New Haven.

Lovell, A., Saw, R., Stimson, J. (2005). 'Product value-density: managing diversity through supply chain segmentation'. *The International Journal of Logistics Management* **16**(1):142–158.

Lucas, R. E. (1988). 'On the mechanics of economic development'. *Journal of Monetary Economics* **22**(1):3–42.

Malmberg, A., Maskell, P. (2002). 'The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering'. *Environment and Planning A* **34**(3):429–449.

Manson, S. (2001). 'Simplifying complexity: a review of complexity theory'. *Geoforum* **32**(3):405–414.

Marshall, A. (1895). *Principles of Economics*, 3rd Edition. Cambridge University Press.

Martin, R. (1999). 'The new 'geographical turn' in economics: some critical reections'. *Cambridge Journal of Economics* **23**(1):65–91.

Martin, R., Sunley, P. (2007). 'Complexity thinking and evolutionary economic geography'. *Journal of Economic Geography* **7**(5):573–601.

Maslow, A. H. (1966). *The Psychology of Science: A Reconnaissance*. Maurice Bassett.

Mcluhan, M. (1964). *Understanding Media: The Extensions of Man*, new edition Edition. MIT Press.

Melberg, H., Rogeberg, J. (2010). 'Rational Addiction Theory: A Survey of Opinions'. *Journal of Drug Policy Analysis* **3**(1).

Melchior, E. (1972). *The integration of space in Latin America*. Sage.

Miller, G. F. (1995). 'Artificial Life as theoretical biology: How to do real science with computer simulation'.

Miller, J. H., Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.

Mirowski, P. (2007a). 'Markets come to bits: Evolution, computation and markomata in economic science'. *Journal of Economic Behavior & Organization* **63**(2):209–242.

Mirowski, P. (2007b). 'On kicking the habit: A response to the JEBO Symposium on "Markets Come to Bits"'. *Journal of Economic Behavior & Organization* **63**(2):359–371.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*, new edition Edition. MIT Press.

Morgan, K., Marsden, T., Murdoch, J. (2006). *Worlds of Food: Place, Power, and Provenance in the Food Chain*. Oxford University Press.

Moss, S., Edmonds, B. (2005). 'Towards Good Social Science'. *Journal of Artificial Societies and Social Simulation* **8**(4):13.

Murata, Y. (2002). 'Rural-urban interdependence and industrialization'. *Journal of Development Economics* **68**(1):1–34.

Murata, Y. (2008). 'Engel's law, Petty's law, and agglomeration'. *Journal of Development Economics* **87**(1):161–177.

Myrdal, G. (1957). *Economic theory and underdeveloped regions*. Harper & Row, London.

Neary, J. P. (2001). 'Of Hype and Hyperbolas: Introducing the New Economic Geography'. *Journal of Economic Literature* **39**(2):536–561.

Newlyn, W. T. (1950). 'The Phillips/Newlyn Hydraulic Model'. *Bulletin of Economic Research* **2**(2):111–128.

OED (2002). *Oxford English Dictionary*, 3rd Edition. Oxford University Press.

Oeffner, M. (2008). *Agent-based Keynesian Macro-economics: An Evolutionary Model Embedded in an Agent-Based Computer Simulation*. Ph.D. thesis, University of Wurzburg, Wurzburg, Germany.

Ohlin, B. (1933). *Interregional and International Trade*. Harvard University Press.

Ojo, G. J. A. (1973). 'Journey to Agricultural Work in Yorubaland'. *Annals of the Association of American Geographers* **63**(1):85–96.

on Peak Oil, I. T., Security, E. (2010). 'The Oil Crunch: A wake-up call for the UK Economy'. *Tech. rep.*

O'Sullivan, D., Haklay, M. (2000). 'Agent-based models and individualism: is the world agent-based?'. *Environment and Planning A* **32**(8):1409–1425.

O'Sullivan, D., Unwin, D. J. (2002). *Geographic Information Analysis*. John Wiley & Sons.

Otter, H., van der Veen, A., de Vriend, H. (2001). 'ABLOoM: Location behaviour, spatial patterns, and agent-based modelling'. *JASSS - the Journal of Artificial Societies and Social Simulation* **4**(4).

Overman, H. (2004). 'Can we learn anything from economic geography proper?'. *Journal of Economic Geography* **4**(5):501–516.

Palander, T. (1935). *Contributions to Location Theory*. Almquist and Wiksell, Uppsala.

Parry, H. R., Evans, A. J. (2008). 'A comparative analysis of parallel processing and super-individual methods for improving the computational performance of a large individual-based model'. *Ecological Modelling* **214**(2-4):141–152.

Penker, M. (2006). 'Mapping and measuring the ecological embeddedness of food supply chains'. *Geoforum* **37**(3):368–379.

Phillips, D., Karn, V. (1992). 'Race and housing in a property owning democracy'. *Journal of Ethnic and Migration Studies* **18**(3):355–369.

Plummer, P., Sheppard, E., Haining, R. (2012). 'Rationality, stability, and endogenous price formation in spatially interdependent markets'. *Environment and Planning A* **44**(3):538–559.

Polhill, J. G., Izquierdo, L., Gotts, N. M. (2006). 'What every agent-based modeller should know about floating point arithmetic'. *Environmental Modelling & Software* **21**(3):283–309.

Posada, M., Cesareo, H., Lopez-Paredes, A. (2007). 'An Artificial Economics View of the Walrasian and Marshallian Stability'. In A. Consiglio (ed.), *Artificial markets modeling: methods and applications*, no. 599 in Lecture Notes in Economics and Mathematical Systems. Springer.

Quarmby, D. A. (1967). 'Choice of Travel Mode for the Journey to Work: Some Findings'. *Journal of Transport Economics and Policy* **1**(3):273–314.

Reiner, R., Guest, C., McKean, M., Reiner, R., McKean, M., Guest, C., Stringer, K. (1984). 'This Is Spinal Tap'.

Reynolds, C. (1987). 'Flocks, herds and schools: A distributed behavioral model.'. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* p. 2534.

Ricardo, D. (1817). *The principles of political economy and taxation*. Dover Publications, Mineola, N.Y.

Robbins, L. (1935). *Essays on Nature and Significance of Economic Science*, 2nd Edition. Macmillan.

Robinson, H., Sharp, H. (2009). 'The emergence of object-oriented technology: the role of community'. *Behaviour & Information Technology* **28**(3):211–222.

Robinson, J. (1933). *The economics of imperfect competition*. Macmillan, London.

Robinson, J. (1971). 'The Measure of Capital: The End of the Controversy'. *The Economic Journal* **81**(323):597–602.

Rogeberg, O. (2004). 'Taking absurd theories seriously: Economics and the case of rational addiction theories'. *Philosophy of Science* **71**(3):263–285.

Rose, R., Tikhomirov, E. (1993). *Who grows food in Eastern Europe?* No. 209 in Studies in public policy. University of Strathclyde, Glasgow.

Ross, D. (2008). 'Economics, cognitive science and social cognition'. *Cognitive Systems Research* **9**(1-2):125–135.

Rothschild, M., Stiglitz, J. (1976). 'Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information'. *The Quarterly Journal of Economics* **90**(4):629–649.

Safarzynska, K., Bergh, J. (2010). 'Evolutionary models in economics: a survey of methods and building blocks'. *Journal of Evolutionary Economics* **20**(3):329–373.

Sasaki, Y., Box, P. (2003). 'Agent-based verification of von Thunen's location theory'. *Journal of Artificial Societies and Social Simulation* **6**(2).

Schelling, T. C. (1969). 'Models of Segregation'. *The American Economic Review* **59**(2):488–493.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton & Co.

Schelling, T. C. (2006). 'Some Fun, Thirty-Five Years Ago'. In *Handbook of Computational Economics*, vol. 2, pp. 1639–1644. Elsevier.

Schendel, D. E., Balestra, P. (1969). 'Rational Behavior and Gasoline Price Wars'. *Applied Economics* **1**(2):89–101.

Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. Routledge.

Scitovsky, T. (1954). 'Two Concepts of External Economies'. *The Journal of Political Economy* **62**(2):143–151.

Scott, J. C. (1998). *Seeing like a State: How Certain Schemes to Improve the Human Condition have Failed*. Yale University Press.

Seebregts, A., Goldstein, G., Smekens, K. (2001). 'Energy/environmental modeling with the MARKAL family of models'. In *Proc. Int. Conf. on Operations Research (OR 2001)*.

Segal, P. (2011). 'Oil price shocks and the macroeconomy'. *Oxford Review of Economic Policy* **27**(1):169 –185.

Shackle, G. L. S. (1983). *The Years of High Theory: Invention and Tradition in Economic Thought 1926-1939*. Cambridge University Press.

Silver, B. L. (2000). *The Ascent of Science*. Oxford University Press.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Capstone.

Solow, R. M. (1997). 'Georgescu-Roegen versus Solow-Stiglitz'. *Ecological Economics* **22**(3):267–268.

Sorrell, S., Speirs, J., Bentley, R., Brandt, A., Miller, R. (2009). 'An Assessment of the Evidence for a Near-term Peak in Global Oil Production'. *Tech. rep.*, UK Energy Research Centre, London.

Spedding, C. (1983). *Fream's Agriculture*. John Murray.

Stafford, T. (2009). 'What use are computational models of cognitive processes?'. In *Connectionist models of behaviour and cognition II: proceedings of the 11th Neural Computation and Psychology Workshop, University of Oxford, UK, 16-18 July 2008*, p. 265.

Starr, R. M. (1997). *General Equilibrium Theory: An Introduction*. Cambridge University Press.

Steenken, D., Voss, S., Stahlbock, R. (2004). 'Container terminal operation and operations research - a classification and literature review'. *OR Spectrum* **26**(1):3–49.

Storper, M. (2010). 'Agglomeration, Trade, and Spatial Development: Bringing Dynamics Back In'. *Journal of Regional Science* **50**(1):313–342.

Storper, M., Scott, A. (2009). 'Rethinking human capital, creativity and urban growth'. *Journal of Economic Geography* **9**(2):147–167.

Suarez, M. (2003). 'Scientific representation: against similarity and isomorphism'. *International Studies in the Philosophy of Science* **17**(3):225–244.

Summers, L. H. (1991). 'The Scientific Illusion in Empirical Macroeconomics'. *The Scandinavian Journal of Economics* **93**(2):129–148.

Sun (1995). 'What Is a Class?'. URL: http://download.oracle.com/javase/tutorial/java/concepts/class.html

Taylor, J. E., Adelman, I. (2003). 'Agricultural Household Models: Genesis, Evolution, and Extensions'. *Review of Economics of the Household* **1**(1):33–58.

Taylor, J. R. (2005). *Classical Mechanics*. University Science Books.

Tesfatsion, L. (2006). 'Agent-Based Computational Economics: A Constructive Approach to Economic Theory'. In *Handbook of Computational Economics*, vol. 2, pp. 831–880. Elsevier.

Thompson, N., Derr, P. (2009). 'Contra Epstein, Good Explanations Predict'. *Journal of Artificial Societies and Social Simulation* **12**(1).

Thunen, J. v. (1826). *The Isolated State in Relation to Agriculture and Political Economy*. Palgrave Macmillan.

Torrens, P., Benenson, I. (2005). 'Geographic Automata Systems'. *International Journal of Geographical Information Science* **19**(4):385–412.

Torrens, P. M. (2010). ‘Agent-based Models and the Spatial Sciences’. *Geography Compass* **4**(5):428–448.

van der Panne, G. (2004). ‘Agglomeration externalities: Marshall versus Jacobs’. *Journal of Evolutionary Economics* **14**(5):593–604.

Van Dyke Parunak, H., Savit, R., Riolo, R. L. (1998). ‘Agent-based modeling vs. equation-based modeling: A case study and users guide’. In *Multi-Agent Systems and Agent-Based Simulation*, p. 1025.

Varian, H. R. (2006). *Intermediate Microeconomics: A Modern Approach*, 7th Edition. W. W. Norton & Co.

Veblen, T. (1898). ‘Why is Economics not an Evolutionary Science?’. *The Quarterly Journal of Economics* **12**(4):373–397.

Vives, X. (2001). *Oligopoly Pricing: Old Ideas and New Tools*, new ed Edition. MIT Press.

Vriend, N. (1991). ‘On Walrasian models and decentralized economics’. *Tinbergen Institute Research Bulletin* **3**:2537.

Vriend, N. (2002). ‘Was Hayek an ace?’. *Southern Economic Journal* **68**(4):811–840.

Vriend, N. J. (1994). ‘Artificial Intelligence and Economic Theory’. In E. Hillebrand (ed.), *Many-agent Simulation and Artificial Life (Frontiers in Artificial Intelligence & Applications)*, pp. 31–47. IOS Press,US.

Vriend, N. J. (1996). ‘Rational behavior and economic theory’. *Journal of Economic Behavior & Organization* **29**(2):263–285.

Wallace, I. (1974). ‘The Relationship between Freight Transport Organization and Industrial Linkage in Britain’. *Transactions of the Institute of British Geographers* (62):25–43.

Webber, M. J. (1984). *Explanation, Prediction and Planning*. Research in planning and design. Pion, London.

Weber, A. (1909). *Theory of the Location of Industries*. University of Chicago Press.

Webster, C. (1965). ‘The discovery of Boyle’s Law, and the concept of the elasticity of air in the seventeenth century’. *Archive for history of exact sciences* **2**(6):441–502.

Wheeler, C. (2007). ‘Do localization economies derive from human capital externalities?’. *Annals of Regional Science* **41**(1):31–50.

Wilhite, A. (2006). ‘Economic Activity on Fixed Networks’. In *Handbook of Computational Economics*, vol. 2, pp. 1013–1045. Elsevier.

Wilkinson, P. (2008). 'Peak oil: threat, opportunity or phantom?'. *Public Health* **122**(7):664–666; discussion 669–670.

Wilson, A. (1970). *Entropy in urban and regional modelling*. Pion, London.

Wilson, A. (2000). *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*. Prentice Hall.

Winner, L. (1980). 'Do Artifacts Have Politics?'. *Daedalus* **109**(1):121–136.

Wong, S. (1973). 'The 'F-Twist' and the Methodology of Paul Samuelson'. *The American Economic Review* **63**(3):312–325.

Wood, J. (1994). *John Maynard Keynes: Critical Assessments, Second Series*. Routledge.

Wooldridge, M. (2009). *An Introduction to MultiAgent Systems*, 2nd Edition. John Wiley & Sons.

Zimmerman, C. (1932). 'Ernst Engel's Law of Expenditures for Food'. *Quarterly Journal of Economics* **47**:78–101.