5 MODELLING AGGLOMERATION WITH NEW ECONOMIC GEOGRAPHY

5.1 Introduction

The deconcentration issue in the previous chapter is modelled as a simulation, but of course the more fundamental question is why and where concentration occurs in the first place, which is clearly one that i-o/CGE models cannot answer. As the leading role of large agglomerations inside a region is of obvious relevance in this respect, it seems fruitful to shift our focus from regions to a lower level of analysis. Apart from this academic argument, a lower spatial level is also becoming increasingly relevant from the policy point of view. The province level has long been the main focus for Dutch regional policy and regional analysis, but, as Oosterhaven (1996) points out, from 1995 onwards there is a shift towards a more differentiated spatial approach. As we have already touched upon in chapter 4, spatial policy is facing more and more local problems of industrial site availability, traffic congestion and discussions on long-term priorities for land use. Bringing home and workplace closer together can reduce mobility and traffic congestion. Concentration of firms in targeted economic zones can improve productivity and leave land use in other zones for residential development (SNN, 1999; REV, 2002). On top of this, environmental policies and programs for nature conservation areas add growing constraints on land use for nature, agriculture and, last but not least, on residential development in rural areas.

The rapidly growing literature on what in recent years has become known as the New Economic Geography (NEG) seems to present a fruitful new framework for economic modelling at lower spatial levels. NEG finds its origin in the seminal work of Krugman (1991a, 1991b, 1993), who first came upon the idea to apply the path-breaking revival of monopolistic competition theory by Dixit & Stiglitz (1977) to geography. The first NEG models were very basic two-region systems, but the focus in NEG has been shifted quickly to agglomerations and urban structures (Fujita, Krugman & Venables, 1999). NEG models turn out to be capable of analysing how and under which conditions spatial concentration of economic activities emerges as a self organizing system, and – more important – how agglomerations change in size due to changes in external economies of scale and transportation costs. So-called spatial CGE (SCGE) models are based on the NEG idea, but also incorporate multisector modelling and input-output information on technology and interregional trade (Bröcker, 1995).

With the wisdom of hindsight the application of the Dixit-Stiglitz framework to spatial modelling seems now so obvious, that it is surprising that it took more than 10 years after the "monopolistic competition revolution" of 1977 before somebody thought about it. Krugman modestly considers himself just the lucky guy because "nobody else picked up that $100 dollar bill lying on the sidewalk in the interim" (Krugman, 1999b). After his 1993 article on the multiregional NEG model, however, it did not take long for me to realize that Krugman had yet dropped another $100 dollar bill, lying around for me. This was the introduction of two-dimensional economic space into the NEG model, on which I

27 A large part of this chapter was published in Stelder (2002).
first started to work in 1995 (Stelder, 1995c). To the rapidly expanding number of authors in the field concentrating on the economic implications and possible refinements of a spatial Dixit-Stiglitz model with a simple one-dimensional circle or horizontal line was apparently already fascinating enough.

As this and the next chapter will illustrate, however, the empirical form of the geography in question is essential for the model, and an erroneously specified economic space will lead to erroneous economic conclusions. Indeed, as a spatial economist one tends to get the impression that as long as the NEG discipline neglects its geographic dimension its acronym could also stand for Not Enough Geography.

This chapter introduces the basic concepts of NEG modelling and presents a general two-dimensional implementation. The next chapter will move to empirical geography.

5.2 The basic NEG model

As a background to the models discussed here I will not elaborate on all the equations of the basic NEG model in detail but only give its main characteristics 28. Assume an economy of \( R \) locations with labour as the only production factor and a fixed total of labour \( L \) available. In each location the economy is divided in two parts. One sector does not move to other places while the other is footloose with respect to its factors of production and can choose its optimal production location. The former is usually referred to as agriculture and the latter as manufacturing but clearly one can think of activities like mining, harbours, central government etc. belonging to “agriculture” as well. Agriculture produces one and the same product in all locations with constant returns to scale and at a fixed price that serves as a numeraire for manufacturing prices 29.

All manufacturers produce a unique variety of a manufactured good with economies of scale at the level of the individual firm. Based on the monopolistic competition framework of Dixit & Stiglitz (1977) all individuals are assumed to share the same utility function

\[
U = C_m^\mu C_a^{1-\mu}
\]  
(5.1)

with \( 0<\mu<1 \) being the total share of consumption expenditures on manufacturing goods. Consumption of manufactured goods \( C_m \) is given by the CES function

\[
C_m = \left[ \sum_i {C_i}^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}
\]  
(5.2)

28 See Krugman (1995) for a more elaborate discussion.
29 This implies that all prices, including that of labour (wages) are expressed in terms of the numeraire “agricultural good.”
with $\sigma > 1$ being the elasticity of substitution between the product varieties, also called the "love-of-variety".

Total national labour $L$ equals

$$L = L_m + L_a$$

(5.3)

and no substitution between $L_m$ and $L_a$ takes place. The choice of units is taken so that $L_m = \mu$ and $L_a = 1 - \mu^{30}$.

Economies of scale in manufacturing firms are introduced by assuming a linear cost function with a fixed cost:

$$l_{mi} = \alpha + \beta x_i$$

(5.4)

and profit maximisation by each producer leads to the setting of the price $p_{ir}$ of product $i$ in region $r$ as a fixed mark-up over marginal costs:

$$p_{ir} = \left[ \frac{\sigma}{\sigma - 1} \right] \beta w_r$$

(5.5)

where $w_r$ is the manufacturing wage rate in location $r$ relative to the agricultural wage, which is assumed the same in all regions and is used as a numéraire (Krugman, 1991, p.489). Note that in the zero-profit equilibrium $\sigma/(\sigma - 1)$ is the ratio of marginal product of labour to the average product of labour, and, consequently, the degree of economies of scale. In other words, the parameter $\sigma$, originally introduced as a taste for variety in (5.2), is also an inverse index of the equilibrium economies of scale.

With free entry of firms profits must be driven to zero, so for each firm in region $r$ we have

$$(p_r - \beta w_r) x_r = \alpha w_r$$

(5.6)

so $\forall r$:

$$x_r = \frac{\alpha(\sigma - 1)}{\beta}$$

(5.7)

That is, output per firm is the same in each region implying that the number of goods produced in each region is proportional to the manufacturing labour $l_{mi}$ in each region:

$$N_r / N = l_{mi} / L_m$$

(5.8)

30 This choice of units ensures that the manufacturing wage equals the agricultural wage in the long-term equilibrium.
Equation (5.8) is the result of the economies of scale that make it profitable to produce each variety in only one location. Each location therefore produces a different bundle of varieties, the number of which is proportional to local labour supply.

As the free entry of firms lead to no excess profits, labour wages are the only source of household income. Total income by region is thus given by the sum of regional agricultural income and manufacturing income:

\[ y = I_a + w \cdot I_m \]  \hspace{2cm} (5.9)

with

- \( y \) = \( R \times 1 \) vector of income by region
- \( I_a \) = \( R \times 1 \) vector of agricultural labour
- \( w \) = \( R \times 1 \) vector of manufacturing wage
- \( I_m \) = \( R \times 1 \) vector manufacturing labour

Because, as mentioned, total labour \( L \) is scaled to unity, the entries of the vectors \( I_a \) and \( I_m \) add up \((1-\mu)\) and \(\mu\) respectively. Note that because prices and wages are expressed in the numeraire agricultural good, there is no agricultural wage in (5.9).

Transportation costs are assumed to be of the Samuelson 'iceberg' type: if a unit of a good travels a distance \( D \), only a fraction \( e^{-\tau D} \) arrives, with \( 0 < \tau < 1 \) being the transportation cost parameter. The relative price in region \( r \) of manufacturing products produced in region \( s \) is then a symmetric \( R \times R \) matrix \( P \) depending on regional production costs and transportation costs between \( r \) and \( s \):

\[ P = \frac{w}{e^{-\tau D}} \]  \hspace{2cm} (5.10)

Matrix \( D \) is the symmetrical \( R \times R \) interregional distance matrix.

The composite price index \( p \) can be derived to be (Krugman, 1991a, p492):

\[ p = \left[ I_m \cdot P^{[1-\sigma]} \right]^{-1/(\sigma-1)} \]  \hspace{2cm} (5.11)

Consumption in region \( r \) of manufacturing products from region \( s \) is given by the \((R \times R)\) matrix \( C \). It can be shown (Krugman, 1991a, p491) that it is given by

\[ C = \pi \cdot y \cdot I_m \cdot P^{[1-\sigma]} \cdot p^{[\sigma-1]} \]  \hspace{2cm} (5.12)

Finally, income of manufacturing in each region must equal total sales:

\[ w \cdot I_m = \pi C y \]  \hspace{2cm} (5.13)

---

31 From this point onward, a vector notation becomes efficient for the multiregional NEG model. The next equations form a vector generalisation of the original two-region model of Krugman (1991a). The multiregional version in Krugman (1993) did not contain any formal generalisation for \( R \) regions. Each bold written vector \( a \) represents \( R \) regional entries \( a_i \). In vector notation \( a \cdot b \) and \( a/b \) here mean cell-to-cell multiplication and division.
with $C$ being the normalized matrix of $C$ with columns adding up to 1.

Short-term equilibrium of $y$, $p$ and $w$ can be derived from (5.9) - (5.13) but what really matters to the workers is their real wage $\omega$, which is given by (Krugman, 1991a, p492)

$$\omega = w.p^e$$  \hspace{1cm} (5.14)

For long-term equilibrium the regional dynamics assumption is that manufacturing workers respond to regional differences in real wages and migrate to regions with the highest real wage. This migration takes the form:

$$l_m(t+1) = \rho l_m(t) \cdot (\omega / \omega_n)$$  \hspace{1cm} (5.15)

where parameter $\rho$ reflects the sensitivity of labour migration to relative regional real wage differences and $\omega_n$ is the average national real wage. If $\rho$=1 equation (5.15) simply states that a location where real income is 10% above the national average will see its share in national manufacturing labour rise with 10% in the next period.$^{32}$

5.3 An agglomeration model in one-dimensional economic space

The model presented above has a built-in tendency of regional concentration to reinforce itself: profit maximizing behaviour of producers under economies of scale leads them to locate their production where large markets are, i.e. where many workers live. At the same time, larger concentrations attract more workers because their prices are lower and therefore real wages are higher.

Given some initial distribution $l_0$ of manufacturing over the regions, a concentration process takes place towards those regions that can offer lower prices and higher real wages. These equilibrium regions are usually the ones that were large in the first place (in the initial distribution), or the ones that have relatively low transportation costs to the other locations, or a combination of the two. The solution found depends on the initial distribution, the size of manufacturing as a whole ($\pi$), the transportation cost parameter $\tau$, the substitution elasticity $\sigma$ and the distance matrix $D$.

Although the logic of the model sounds simple, the mathematics is not and solving has to be done by numerical iteration. The outcome is achieved numerically by choosing starting values for the distance matrix $D$, the model parameters $\pi$, $\sigma$, $\tau$ and the initial distribution of manufacturing labour $l_0$ and letting the model iterate until long-term equilibrium is reached. This is the case when real wages have become the same in all locations and according to (5.15) migration stops.

$^{32}$ This is not exactly true because after applying (5.15) $l_m(t+1)$ is rescaled to 1. This multiplicative form is different from the original additive formulation by Krugman (1993), which in some cases leads to negative values for some $l_m(t)$. 81
Some simulation results are shown in figure 5.1a-f for a “racetrack economy” with 12 locations located on a circle, i.e. \( D(1,2)=1, D(1,3)=2 \) etc. but also \( D(1,12)=1, D(1,11))=2 \) etc. because trade in both directions along the circle is possible. The reason for using this circle space is to analyse the pure economic effect of combinations of \( \pi, \sigma \) and \( \tau \) in an economy in which no location has any geographical advantage over the other in terms of transportation costs. The model needs, however, some initial distribution of manufacturing \( l_0 \) that is unevenly spread over the locations along the circle, to start with. If we would assume that manufacturing, just as agriculture, would also be evenly distributed over the circle, then there will be no incentive to migrate from any location to somewhere else. Because all locations are of equal size, all prices and wages will also be the same everywhere and no agglomeration will take place. This is an important limitation of spatially neutral economic spaces that will be discussed later on.

Figure 5.1a shows the equilibrium that results from some arbitrary chosen initial distribution of manufacturing labour that has the highest shares in location 3 and 9. In the example of figure 5.1a manufacturing becomes concentrated in two cities of equal size at 3 and 9. The spatial distribution of the locationally bounded agricultural sector does not change by definition. Because no city has any locational advantage over the other, it is the size of the largest cities in the initial distribution that eventually leads them to taking over the smaller ones. In the example given, transportation costs are sufficiently high relative to scale economies, so that it is not profitable to concentrate all production in one city. Instead, each city takes up half of the manufacturing production with an agglomeration reach of about three locations in each direction. Generally speaking, fewer cities are left over when scale economies are increased (lower \( \sigma \)), when the size of the footloose sector (manufacturing) relative to the bounded sector (agriculture) is set higher (higher \( \pi \)) and transportation costs are lower (lower \( \tau \)) \(^{33} \). In the case of figure 5.1b, when \( \tau \) is lowered to 0.3 and \( \sigma \) is lowered to 3, all manufacturing concentrates into one city at location 3.

As the values of the initial distribution vector mentioned at the bottom of figure 5.1a-f show, location 3 does not have the exact highest value in the initial distribution (63 compared to 64 for location 9), but together with its 2, 3 or 4 direct neighbours in both directions it still forms a larger cluster than location 9 (191, 274 and 330 versus 189, 272)

\(^{33}\) Economies of scale are determined by \( \sigma/\sigma-1 \) and therefore increase when \( \sigma \) decreases. The influence of \( \tau \) is less straightforward then mentioned here (Krugman, 1995). Unless mentioned as otherwise, the migration parameter \( \rho \) is set to 1. This parameter only effects the speed of the concentration process but not the final outcome. High values of \( \rho>1 \), however, may cause back-and-forth fluctuations across a long-term equilibrium that is never reached.
Figure 5.1a-f Six simulations of a racetrack economy with 12 locations

5.1a $\pi=0.4$, $\sigma=4$, $\tau=0.4$

5.1b $\pi=0.4$, $\sigma=3$, $\tau=0.3$

5.1c $\pi=0.4$, $\sigma=6$, $\tau=0.4$

5.1d $\pi=0.4$, $\sigma=4$, $\tau=0.4$, uneven distribution agriculture

5.1e $\pi=0.8$, $\sigma=6.5$, $\tau=0.2$

5.1f $\pi=0.85$, $\sigma=6.5$, $\tau=0.2$

Initial distribution manufacturing: 10,42,63,33,43,38,9,20,64,49,47,45

Uneven distribution agriculture: 8,8,8,8,8,40,8,8,8,30,8,8
and 325). If we would also take the 4th neighbour in both directions into account, then the cluster around 9 is slightly larger then the cluster around 3 (400 versus 399). In figure 5.1b, this effect would come about when we would lower the transportation cost parameter $\pi$ further to 0.2. Then also only one manufacturing city emerges, but in 9 instead of 3.

Krugman noted that the outcome of his basic model shows that “there is a systematic tendency toward formation of central places roughly evenly spaced across the landscape” (Krugman, 1993). As figure 5.1c shows, if the agglomeration effect of figure 5.1a is weakened by raising $\sigma$ to 6, four cities of equal size appear at 3, 6, 9, and 12. Because of the spatial neutrality of the racetrack, any long-term equilibrium will consist of cities of equal size (Krugman, 1991b, p20; Brakman et al., 2001, p122).

A necessary condition for this result, however, is the assumption of even distribution of the agricultural sector across space. As figure 5.1d shows, if we assume an uneven initial distribution of agriculture in the simulation of figure 5.1, the long-term manufacturing equilibrium result can be asymmetric. Due to the larger agricultural markets at location 6 and 10, the equilibrium manufacturing agglomeration not only moves from location 3 and 9 to 6 and 10, but also with the largest manufacturing agglomeration in 6. The geographical influence of an uneven distribution of agriculture of course also depends on its total size relative to the manufacturing sector. Figure 5.1e shows another parameter configuration with a much larger manufacturing sector ($\pi=0.8$). There is still an equilibrium of two cities at 6 and 10, but reversed in size. At an even higher value of $\pi=0.85$ the influence of agriculture disappears entirely and leads to only one manufacturing agglomeration at neither 6 nor 10, but at 9.

The need for some uneven distribution of manufacturing can be seen as a virtue because it enables the model to take account of agglomerations that have taken place in the past. It means that - in the words of Krugman - "history matters" because different initial distributions lead to different outcomes. It should, however, also be immediately clear that "geography matters" as well. As soon as the model takes its locations from the real world, the economic space will not be a circle anymore, but a configuration of locations in which some will have locational advantages over others.

Figure 5.2a shows what happens when the original simulation of figure 5.1a is run again, but now with the 12 locations on a horizontal line instead of a circle. To distinguish it from the "racetrack economy" in figure 5.1, the horizontal line economic space is labelled as the "Hotelling economy". Again two cities emerge at 3 and 9, but now the city at 9 is larger. Because the circle is literally "broken" between location 1 and 12, for location 3 there is nothing left beyond the left side of 1 to capture in the agglomeration process. Being one point closer to the centre, location 9 has a more strategic position in this respect. There can now be a two-city equilibrium with two cities

---

34 Hotelling (1929) was one of the first economists who modelled location choice of two firms on a horizontal line.
Figure 5.2a-f  Six simulations of a Hotelling economy with 12 locations

5.2a $\pi=0.4, \sigma=4, \tau=0.4$

5.2b $\pi=0.4, \sigma=3, \tau=0.3$

5.2c $\pi=0.4, \sigma=6, \tau=0.4$

5.2d $\pi=0.4, \sigma=4, \tau=0.4$, uneven distribution agriculture

5.2e $\pi=0.4, \sigma=4, \tau=0.4$, "no history"

5.2f $\pi=0.4, \sigma=4, \tau=0.4$, "no history"; $R=13$

initial distribution manufacturing and uneven distribution agriculture: same as figure 5.1
of unequal size, even when agriculture is evenly distributed. The one-city agglomeration in location 3 of figure 5.1a changes to location 9 in the Hotelling case for the same reason.

Figure 5.2c shows the interesting case of 5 cities in the Hotelling economy, again of unequal size, versus the 4 cities in figure 5.1c. Location 3, 6 and 9 enter the equilibrium again, but the 4th city 12 in figure 5.1c shifts one place to the left to 11, because it cannot attract manufacturing workers from 1 and 2, as is the case in the racetrack. Location 11, therefore has a more competitive position than 12, which it is able to absorb in the agglomeration process. At the other end of the line location 1 can survive as a small city because of its distance of 2 to the nearest larger city 3. Finally, when we compare the case of uneven distribution of agriculture in figure 5.2d with figure 5.1d, location 6 and 10 emerge again as the two agglomerations, but in the Hotelling economy 6 is relatively larger due to its more strategic position in the centre.

So it is clear that indeed geography matters as well. What makes the Hotelling case much more interesting than the racetrack economy, however, is the fact that the properties of the economic space itself can be analysed without any arbitrary initial distribution. If we give each location an equal share of agriculture and manufacturing in the initial distribution, no location will have any initial size advantage over the other and then no "history" will determine the model outcome. This will be called the "no history" or zero agglomeration assumption. In such an application, it will then be only the geographical position of each location, in other words, the geographical characteristics of the economic space itself that will determine the outcome.

Figure 5.2e shows again the Hotelling economy with the same parameter configuration of 5.1a and 5.2a, but now with this no history assumption. The two largest cities are located symmetrically at 2 and 11, with two smaller cities close to each other in the centre positions 6 and 7.

The reason why there is not, as we would expect, one large city in the centre is because there is not one middle point in a row of even numbers. Locations 6 and 7 indeed have a locational advantage over the other locations, but they cannot agglomerate together into one large centre city at the hypothetical point 6.5. This centre does appear in simulation 5.2f, which is identical to 5.2e, except for the fact that the Hotelling economy now consists of 13 instead of 12 points. Location 7 is now the largest centre accompanied by two smaller cities at 2 and 12.

So now we have a simple proto-type of a spatial economy that works as a self-organizing system without the need of assuming any specific initial distribution. The "no history" assumption can also be paraphrased as "in the beginning there were only little villages". What the model can do is to find out where it would be likely - from the economic point of view - for agglomerations to arise in the long run, starting with an

---

35 The result stems from the broken symmetry in the distance matrix $D$ and, consequently, in the price index $P$ and the real wage $\omega$ in (5.14).
economy with low levels of agglomeration. The model also makes it redundant to run series of random initial distributions in order to find out general model behaviour under certain parameter configurations, as is inevitable for spatially neutral models. In a non-neutral space one should mathematically expect the outcome of the no-history assumption to be the typical outcome, to which the average result of  random initial distributions will converge as $n \to \infty$.

Finally, it should be noted that when the number of possible locations increases, there comes more room for cities of intermediate size to enter the equilibrium. This is an issue underestimated by many authors who study the behaviour of the NEG model in its most elementary form using just two regions. The two-region case is way too special and hardly relevant for a multiregional world because "a world of twos is very different from a world of many" (Henderson, 2004, p292). As a mere illustration, figure 5.3 shows the equilibrium distribution of manufacturing labour for a Hotelling economy with 371 locations, with $\pi=0.4$, $\sigma=4$ and $\tau=0.3$. Again, as in figure 5.2e-f, the zero agglomeration assumption is applied, i.e. in the initial distribution all 371 locations have the same share of both agriculture and manufacturing. The distance between two adjacent locations is the same as in 5.2e-f, which means that this economy is in fact more than 30 times as large.

**Figure 5.3 Equilibrium distribution in a Hotelling economy with 371 locations**

\[ \pi=0.4, \sigma=4, \tau=0.3 \]

There is now a much more complicated equilibrium with a total of 48 cities in various sizes. The remaining locations (not shown) only have their equal sized bounded sector left over. The agglomeration structure is divided into the expected two symmetrical left and right half of the horizontal line, but, contrary to the situation in figure 5.2e-f, without a city in the centre location 186. Apparently, transportation costs are too large to make a large centre competitive in an economy of this size. Instead, the largest cities appear somewhat shifted away from the centre.

---

36 In a spatially neutral model one can only find the typical model behaviour for a certain parameter configuration by running many random initial distributions. See for instance Fujita et al. (1999), p84 and Brakman et al. (2001), p122.

37 A large city does appear at 186 with much lower values for $\tau$ and $\sigma$, but then the equilibrium has less different cities of different sizes.
The most striking feature of figure 5.3 is of course its resemblance with a system of central places, but this issue will be discussed after the introduction of the two-dimensional model in the next section.

### 5.4 A two dimensional model

Starting from the Hotelling economy in the previous section, the most logical next step towards the introduction of two-dimensional space is a rectangular grid of equidistant locations. We can think of this grid as a discrete approximation of a closed continuous economic plane in $\mathbb{R}^2$. As on the horizontal line, the model cannot simulate a continuous space because it needs a finite number of possible locations. We will start with the most simple example of a square economy of $n \times n$ possible locations, in which each location $q_{ij}$ has a distance 1 to its four direct horizontal and vertical neighbours $q_{i(j\pm 1)}$ and $q_{(i\pm 1)j}$. The distance between any two arbitrary locations $p$ and $q$ is assumed to be the direct Euclidian distance $\sqrt{(p_i - q_i)^2 + (p_j - q_j)^2}$. In other words, it is assumed that there is a straight "airline connection" from $p$ to $q$ along which goods can be transported.

**Figure 5.4 Experiments with a 3x3 economy**

| 5.4a | $\pi=0.5$, $\sigma=6$, $\tau=0.4$ |
| 5.4b | $\pi=0.1$, $\sigma=5$, $\tau=0.52$ |
| 5.4c | $\pi=0.1$, $\sigma=4.9$, $\tau=0.52$ |
| 5.4d | $\pi=0.1$, $\sigma=4.8$, $\tau=0.52$ |
The smallest version of this model is a 3x3 grid\textsuperscript{38}. Figure 5.4 shows 4 simulations with different parameter configurations. The blue locations have both agriculture and manufacturing and the red locations have only agriculture. The blue dots are plotted proportional to their manufacturing share\textsuperscript{39}. The initial distribution (not shown) again follows the same no history assumption as mentioned before and thus consists of nine blue dots of equal size: each location has an equal 1/9 share of total agriculture and total manufacturing\textsuperscript{40}.

Figure 5.4a shows the most likely outcome: manufacturing agglomerates totally into the centre city. The eight remaining red locations on the grid have lost all their manufacturing activities to the centre and have become true “villages” in the sense that only their unmovable agricultural production has remained. All parameter combinations in the range of 0.3 - 0.6 for $\pi$ and $\tau$, and 3-6 for $\sigma$ lead to the same result. Rare intermediate cases were only found for a very low share of manufacturing ($\pi = 0.1$), combined with small variations around 5 for $\sigma$ and 0.5 for $\tau$. Figure 5.4b – 5.4d show increasing agglomeration towards the centre. The corner cities tend to survive longer than the other four because they are more remote from the centre. In the equilibrium in figure 5.4d, they still have a small share of manufacturing while the other have lost all of their manufacturing.

Figure 5.5 shows the result for 4 grids of increasing geographical size, with an increasing number of locations ranging from 9x9 to 12x12, each with the same parameter configuration of $\pi = 0.5$, $\sigma = 6$ and $\tau = 0.4$. Because of the larger distances and larger number of locations this set of parameters does not lead to the agglomeration into one centre as was the case in the 3x3 example (figure 5.5a). In the 9x9 example (figure 5.5a) there is one central city combined with four second order cities in the corners that stay out of reach from the centre.

The 10x10 model (figure 5.5b) shows an opposite case: the four largest cities are located in the corners, each of them serving as a local capital of their quadrant. The reason is simple: because of the even number of the model dimension there is not one strategic central location but a pair of four, who only survive as small cities. Due to their mutual competition they cannot develop as one large centre that can attract manufacturing from the corners.

In the 11x11 case (figure 5.5c) a diamond pattern emerges with no cities in the centre. The trade-off between scale economies and transportation costs does not lead to a single centre as in the smaller sized uneven 9x9 economy. As one can easily imagine, other 11x11 experiments show that with lower values for $\sigma$ and $\tau$ the diamond becomes smaller and converges to a single centre.

\textsuperscript{38} In a 2x2 grid all 4 locations are perfectly symmetric and have no locational advantage.
\textsuperscript{39} The graphs are only relatively comparable: the plotted size of the largest city is the same in each simulation.
\textsuperscript{40} The no history assumption applies to all simulations in the remaining part of this chapter and will therefore not be mentioned again.
The 12x12 example (figure 5.5d) shows four medium sized cities in the centre at a distance of 3 from each other. This is unlike the 10x10 case, were the four centre cities were direct neighbours. Their competitive advantage, however, is not large enough to prevent the four corner cities from staying the largest local centres. There is also room left for eight small cities somewhere between the markets reach of the eight big cities.

We should be aware that when the grid size is increased, there are in fact two effects possible at the same time. First, as we have seen with the large Hotelling example in figure 5.3, because the number of locations on the grid increases, more and different city patterns become possible. Second, with the transportation costs held constant, some location in the centre that may have a competitive position in a small economy, may not be that competitive any more in a larger economy because its products become too expensive in the more remote regions. Instead, the equilibrium can then move from a dominant single centre city to a set of multiple first-order cities away from the centre.
The best way to isolate these two effects from one another is by adjusting the distance matrix so that the total size of the grid remains the same. If we start again with a 9x9 grid, we can adjust the distance in a comparable 10x10 grid with a factor $8/9^{41}$. The 10x10 grid then has the same size of the 9x9 grid but with a larger resolution. This is done in figure 5.6a – 5.6d with the parameters held constant at $\pi = 0.4$, $\sigma = 6$ and $\tau = 0.4$ for a grid resolution increasing from 9x9 to 12x12. Figure 5.6a-b is about the same as figure 5.5a-b, only the second order cities in 5.6a-b are slightly larger. In 5.6c the diamond pattern of 5.5c has disappeared. Instead, the pattern of 5.6a is repeated, but with more concentration into the centre. The same holds for 5.6d, which repeats the pattern of 5.6b but again with more concentration. The main cause of this result is that due to the higher resolution more locations come within the reach of the main centres.

Figure 5.6 Four equal-sized grids with different resolutions

$\pi = 0.4$, $\sigma = 6$, $\tau = 0.4$

$\pi \sigma \tau$

$5.6a$ 9x9

$5.6b$ 10x10

$5.6c$ 11x11

$5.6d$ 12x12

$41$ The horizontal distance between two corners in an nxn grid is n-1.
5.5 Parameter variations with a large square economy

Just as the number of pixels in a digital picture, the accuracy of modelling an economic space of a particular shape is determined by the resolution of the grid. The computational limits of present desktop computers allow us a maximum number of around 2700 locations, which gives us the opportunity to model more complex city configurations and look at the effects of different parameter values. Figure 5.7a shows the result for a square economy of 51x51=2601 locations and parameter values $\pi=0.5$, $\sigma=5$ and $\tau=0.4$. We have chosen an uneven number for n in order to have one single location in the centre of the grid. The dots show the long-term equilibrium proportional to their share of manufacturing labour. For clarity of the graph, the “villages” that have lost all of their manufacturing labour are not plotted. Note that each location on the grid still has its unchanged share of 1/2601 of the agricultural workforce. Out of the total number of 2601 locations 81 cities emerge in 10 different size classes.

The agglomeration structure is perfectly symmetric because of the square spatial shape of the economy, but its structure is not what we would have predicted intuitively. There is a large city in the centre, but it is not the largest one. The parameter configuration leads to an eight angle of four symmetric pairs of cities of size class 1 in each quadrant.

In the accompanying table it is shown that more than 25% of total manufacturing labour is concentrated in these eight cities. The graph shows the sizes of all 10 classes. In terms of rank-size nr 4, 6 and 8 are about 25-30% smaller than their predecessors 3, 5 and 7 (see the third column in the table). The fall in size when going from one size class to the next is less then 20% for the other combinations so the overall picture is that of a rather smooth continuous size distribution. The smallest cities are along the borders, where they stay out of reach of the larger agglomerations.

Figure 5.7b shows what happens if we reduce the economies of scale by raising the substitution elasticity parameter $\sigma$ from 5 to 5.5. As expected, the weaker centripetal forces now lead to more smaller cities. In this case there is a hierarchy of 104 cities in 16 size classes. Instead of the eight angle in simulation A, in B the four largest cities are placed in a diamond shape, more close to each other and more close to the centre. The allocation of the smaller cities shows a fascinating complex pattern, which makes one wonder whether this is art or economics. The bar chart and the statistics in the table indicate that the size distribution is rather smooth, a smaller size class being around 90-95% of it preceding higher level, but with some exceptions. Cities of level 3, 5 and 12 are 70-80% smaller, and the smallest category 16 is only 41% of the size of nr 15.

In figure 5.7c $\sigma$ is raised further to 6 resulting in 108 cities in 18 size classes. The equilibrium now enables four main centres in the corners, which are larger than the largest size class in simulation B (each of them taking 2.7% versus 2,3% of total manufacturing employment), but the other cities are smaller than in simulation B. The size distribution shows the largest breaks for level 2 (56%) and 18 (54%) and minor breaks at level 11 (80%) and 15 (61%). At first sight it may seem strange that the lower
Figure 5.7 Three simulations with a square grid of 51 x 51 locations

5.7a $\pi = 0.5$, $\sigma = 5.0$, $\tau = 0.4$

Size class in % of total

<table>
<thead>
<tr>
<th>size class</th>
<th>nr cities</th>
<th>% of total</th>
<th>share of class in % of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>3.2</td>
<td>25.4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2.3</td>
<td>18.3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1.6</td>
<td>12.9</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>1.2</td>
<td>19.7</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.8</td>
<td>12.8</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0.7</td>
<td>5.9</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>0.5</td>
<td>5.4</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0.3</td>
<td>1.3</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>0.3</td>
<td>2.3</td>
</tr>
<tr>
<td>total</td>
<td>81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.7b $\pi = 0.5$, $\sigma = 5.5$, $\tau = 0.4$

<table>
<thead>
<tr>
<th>size class</th>
<th>nr cities</th>
<th>% of total</th>
<th>share of class in % of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>2.3</td>
<td>9.0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.5</td>
<td>8.1</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>1.6</td>
<td>13.0</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>1.0</td>
<td>9.0</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.8</td>
<td>6.7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0.8</td>
<td>6.6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.8</td>
<td>6.2</td>
</tr>
<tr>
<td>total</td>
<td>104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.7c $\pi = 0.5$, $\sigma = 6.0$, $\tau = 0.4$

<table>
<thead>
<tr>
<th>size class</th>
<th>nr cities</th>
<th>% of total</th>
<th>share of class in % of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>2.7</td>
<td>21.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.5</td>
<td>11.3</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>1.4</td>
<td>11.3</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.2</td>
<td>4.9</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>1.2</td>
<td>4.9</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>1.0</td>
<td>7.9</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0.9</td>
<td>3.6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.9</td>
<td>7.0</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>0.8</td>
<td>6.2</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>0.6</td>
<td>4.6</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>0.6</td>
<td>8.9</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>0.5</td>
<td>4.6</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>0.5</td>
<td>4.5</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>0.5</td>
<td>3.9</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>0.3</td>
<td>2.4</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0.3</td>
<td>1.2</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>total</td>
<td>120</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
economies of scale in simulation C lead to (some) larger cities than in simulation B but the fact that the centripetal forces for the system as a whole are weaker is illustrated by less concentration in the centre. However, this does not prevent the four large cities in the corners to gain a main position in each of the four quadrants.

5.6 Rank size properties

If we leave the spatial dimension of the grid models aside for a moment, we can concentrate on the size distribution of the model results only. Ever since George Zipf (1949) stated his rank size rule, and, among many other social phenomena, found that many city distributions follow this rule, some spatial and urban economists have spent considerable time and effort on this subject, trying to find a theoretical explanation and estimating rank size regressions. Recent overviews of the literature on the rank size rule can be found in Gabaix (1999) or Brakman et al. (2001). The rank size rule states that most empirical city distributions can be fitted well with OLS according to a rank size regression of the form:

\[
\log (\text{Rank}) = a + b \cdot \log (\text{Size})
\]

If \( b = 1 \), the largest city with rank number 1 is exactly twice as large as the second city with rank number 2, three times as large as city nr 3, etc..., which is know as Zipf’s Law. If \( b > 1 \), the distribution is steeper and if \( b < 1 \), a situation found frequently in reality, the distribution is flatter. The question now arises whether the three simulations A, B and C from Figure 5.7 follow this specification.

Figure 5.8 Rank size distribution of simulation A-C

<table>
<thead>
<tr>
<th>Simulation</th>
<th>b</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation A</td>
<td>-1.08</td>
<td>0.79</td>
</tr>
<tr>
<td>Simulation B</td>
<td>-1.48</td>
<td>0.77</td>
</tr>
<tr>
<td>Simulation C</td>
<td>-1.14</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Figure 5.8 shows the city size classes of the simulations 5.7a-c plotted against their ranking number on a logarithmic scale. Every city of equal size is given an arbitrary subsequent number. In simulation B, for example, the 4 cities in size class 1 are numbered 1 through 4, the 4 cities in size class 2 are numbered 5 through 8 etc. In the table on the right side of the figure the regression results are given. All three estimates for \( b \) are significant at the 5% level. Simulation A comes the closest to the Zipf value of \( b = 1 \).
As discussed above, the presence of multiple cities of equal size in A, B and C is due to the spatial symmetry in the square model grid, and of course this limits the fit to a rank size rule: within each size class every next city gets a higher ranking number while its size remains the same. As will be shown in the next chapter, these classes of cities of identical size disappear when an asymmetric grid is used. Like in reality, then hardly any pair of two cities has exactly the same size.

So the results of the grid model do seem to follow the rank-size rule to a certain extent, but in a spatial context this is of little significance. As will be discussed at length in the next chapter, where we move to empirical geographical grids, a model that predicts a set of cities of the correct size and rank-size distribution, but not on the correct locations, should not be considered to be a good model. When measuring the performance of a spatial agglomeration model, location and size should always be taken into account jointly and not separated from each other.

5.7 Other shapes

It has become clear from the previous sections that the model results are shape dependent. Just as the agglomeration pattern on a large grid is different from that on a small grid, other shapes than squares will give other results. As a mere illustration, Figure 5.9a – 5.9f show six results with different parameter sets for a rectangular 7x17 grid. In 5.9a two city belts along the North and the South side appear with smaller dispersed cities in between. The $\pi$-value is set lower and the $\sigma$-value higher simulating less concentration than in the square examples of figure 5.4. Setting the manufacturing share $\pi$ to 0.4 and lowering $\sigma$ to 5 gives more agglomeration into the two North and South belts only, with on each belt two largest cities away from the centre (5.9b). If each parameter is further changed one point into the direction of more agglomeration ($\pi$ to 0.5, $\sigma$ to 4 and $\tau$ to 0.4), the two belts shrink into 3 cities only, with the largest city in the centre, and both belts closer to the centre (5.9c).

Figure 5.9 Six experiments with a 7x17 grid
When $\tau$ is reduced further to 0.3 only two cities remain, dividing the rectangle into two square parts with one centre each (5.9d). Both cities move towards the centre when $\tau$ and $\sigma$ are reduced together to 3 and 0.25 respectively (5.9e). Only one city remains when $\tau$ is set to 0.2 and $\sigma$ to 2 (5.9f), but just as in the other grids discussed before, this result can be achieved with many other parameter combinations as well.

5.8 NEG versus central place theory

As already mentioned, the city distributions that the grid model produces seem to have some similarity with the patterns of central places known from Christaller (1933) and Lösch (1940). Figure 5.10a illustrates the classical Christaller hexagonal pattern with $K=3$, indicating that each market area of a higher-ordered good $N+1$ is a constant factor of $K$ times larger than the market area of good $N$. Figure 5.10b shows the general hierarchy case, where $K$ can be different for each level (Parr, 2002; Beckmann & McPherson, 1970).

Although the geometry of their city distributions may look the same, there are some important conceptual differences between an NEG model and a central places model. First of all, the concept of a market area does not exist in an NEG framework: each variety of a manufactured good is considered to be unique, only produced at one single location. Consequently, each variety is consumed in all other places in decreasing amounts as distance and transportation costs increase.\footnote{The iceberg specification of the transportation costs implies a continuously increasing amount of the good "melting away" as distance increases. This amount is asymptotically approaching...}
In a central places model, each homogeneous good is produced at more locations (market centres) and only purchased by consumers within their market area. The borders of these market areas are the break-even points, at which it becomes cheaper to buy the same good from somewhere else. In other words, NEG is characterized by the heterogeneity of one single good with cross-hauling, while central places assumes homogeneity of different goods and absence of cross-hauling\textsuperscript{43}.

A second difference is that central place models are defined on an infinite borderless economic surface. As Parr (2002) discusses in detail, the hierarchical central place system can either be derived from a high-level centre downward, or from a low-level structure upward, but this can only be done by assuming the structure of the initial starting level somewhere. The same argument holds for the metropolis model of Lösch (1940), which has to assume a metropolis somewhere to start with. The resulting urban systems, then, can be anywhere on the economic plane. The NEG grid model, however, but never reaching the 100%, which means that even at a very large distance between the producer and the consumer still an infinitesimal small amount of the good is traded and consumed.

\textsuperscript{43} Cross-hauling is trade of an identical good between two regions in both directions. This phenomenon is usually explained by assuming that an identical good produced somewhere else is by definition not identical. This assumption is known as the Armington assumption (Armington, 1969).

Source: Parr (2002)
uses a closed economic space, which enables a prediction of the rise of cities starting from a situation of no history: given an economic space with all population and economic activity equally distributed across this space, where will cities arise?\textsuperscript{44} The fact that size \textit{and} absolute location are simultaneously predicted at the same time without any prior information is therefore a major advantage of the NEG-grid model.

There are, however, also some similarities between the two models. In central-place theory the smallest market area for a product is determined by its “threshold value”: the minimum market needed to start one production unit. The size of a market area is therefore determined by the fixed costs involved, which is a form of economies of scale: products that have higher fixed costs need larger market areas. Second, Isard’s (1956) amendment of central-place theory, that central places will not only be production sites but also places where more population is concentrated, is consistent with the agglomeration effects that Krugman assumes. According to Isard, this will lead to market areas that become larger in less densely populated areas. The exact shape of the market areas, however, does not need to be hexagonal anymore. Not having today’s computers at his disposal he guessed that “the resulting pattern is at best only one of many which can be evolved” and that “non-hexagonal forms are more consistent with the full interplay of location forces”\textsuperscript{45}. The urban structure in our NEG model may not be of the Christaller-hexagonal type, but when a symmetric closed economy is used, like the squares discussed above, it is indeed symmetric as well. Finally, the central places model can be seen as a monopolistic competition model too: two identical goods produced in different market areas are two varieties of that good. In this view it is the production location that defines the variety, and each producer of that good is the monopolist in its own market area.

Theoretically, one can think of the option of using the same grid model in a central places approach. Following the no history assumption, all grid points having the same size in the initial situation can then be thought of as indeed "little villages" to start with, which represent the production centres of the lowest order good. However, to derive the higher-level centres taking account of the fact that low-level centres close to the border have truncated market areas, seems very complicated and requires new algorithms to be developed. It has been shown in this chapter that the NEG approach does not need any new algorithms for this purpose because it already has all the tools we need. The next step is now to move from abstract grids towards real geographical structures. This is done in the next chapter.

\textsuperscript{44} See Ottaviano & Thisse (2003) for more discussion on even distributions in homogenous space. Brakman et al. (2001) discuss various asymmetrical economic spaces in the context of NEG models.

\textsuperscript{45} Isard (1956), p 274.