



**HAL**  
open science

# On the emergence of the urban phenomenon. Part I

Kristian Behrens

► **To cite this version:**

Kristian Behrens. On the emergence of the urban phenomenon. Part I. [Research Report] Laboratoire d'analyse et de techniques économiques(LATEC). 2001, 40 p., Illustration, ref. bib.: 28 ref. hal-01526539

**HAL Id: hal-01526539**

**<https://hal.science/hal-01526539>**

Submitted on 23 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LATEC

## LABORATOIRE D'ANALYSE ET DE TECHNIQUES ÉCONOMIQUES

*UMR 5118 CNRS*

DOCUMENT DE TRAVAIL



**Pôle d'Économie et de Gestion**

2, bd Gabriel - BP 26611 - F-21066 Dijon cedex - Tél. 03 80 39 54 30 - Fax 03 80 39 54 43  
Courrier électronique : [secretariat.latec@u-bourgogne.fr](mailto:secretariat.latec@u-bourgogne.fr)

ISSN : 1260-8556

---

**n° 2001-11**

**On the emergence of the urban phenomenon  
Part I**

**Kristian BEHRENS**

**décembre 2001**

## On the emergence of the urban phenomenon - Part I

K.BEHRENS †

LATEC, UNIVERSITÉ DE BOURGOGNE

December 2001

### Abstract

In this paper we develop a relatively simple but general model describing the formation of urban agglomerations in a pre-industrial setting. Rural agglomeration on a local scale arises from the multiplication and specialization of intermediate goods and labor services, while dispersion is due to the commuting costs of agricultural workers. Urban agglomeration on a more global scale stems from the trade-off between increasing returns to scale and transport costs of goods. We derive a general equilibrium model of the formation of urban structures and show how population growth, strict indivisibilities and structural changes in the production sector can modify quantitatively and qualitatively the resulting urban configuration.

### Résumé

Dans ce papier nous développons un simple modèle d'économie spatiale décrivant la formation d'agglomérations urbaines dans un contexte pré-industriel. Dans un premier temps, au niveau local, l'agglomération rurale résulte de la multiplication et spécialisation des biens et services intermédiaires tandis que la dispersion est due aux coûts de transport des personnes agricoles. L'agglomération urbaine au niveau plus global est le résultat de l'interaction entre les coûts de transport des marchandises et les économies d'échelles. Nous développons un modèle d'équilibre général de formation de structure urbaine et nous montrons comment la croissance démographique, des indivisibilités strictes et les modifications du secteur productif peuvent modifier quantitativement et qualitativement les configurations spatiales.

**Keywords / Mots clés :** pre-industrial cities, city-formation, indivisibilities

**JEL-Classification :** R11, R12, O18

---

† The author is indebted to Jean-Marie Huriot, Christian Michelot and Frédéric Gilli for valuable comments. A short version of this paper has been presented at the XIth SESAM conference, 12–14 September 2001 in Lille, France. Submitted in “Regional Science and Urban Economics”.

## 1 Introduction

Despite the significant advances that the theory of urban formation has made these last few decades, especially with the general equilibrium models developed in the “new economic geography”, several issues have only been sparsely examined. One of them concerns the formation of urban agglomerations in the context of a pre-industrial economy <sup>(1)</sup>. While most current models are able to investigate the formation of agglomerations in a modern context, they are unable to explain in an intellectually satisfying way the formation of urban structures when either differentiated industries are inexistent or no other cities are present. Two facts should be highlighted : first of all, most models are based on hypotheses that do not allow their application to early historical periods. Second, most models start with characteristics that are already inherently urban; even when there are no agglomerations in the beginning we already suppose that there is a highly developed manufacturing or industrial sector which, through technological or pecuniary externalities, generates agglomeration forces. The most ancient form of division of labor, namely between rural-agricultural and urban-manufactured work, so crucial in the explanation of the early phases of urbanization, is simply assumed. As stated by FUJITA M., KRUGMAN P., VENABLES A., [7], “[...] *we find it most natural to think of new cities as emerging, as an economy, that already has an urban structure, grows over time*”. This leaves us with great models capable of explaining how agglomerations form in “modern” economies but gives no hint at how the urban phenomenon can be initialized when there is no industrial activity and urban structure at all in the beginning. We believe that this question is fundamental since there is a large consensus about the fact that the division of labor is intimately related to urbanization and that one can probably not be understood without the other. Even if we will probably never be able to say if the division of labor between agriculture and manufacturing is a source or a consequence of urbanization (it probably is both at the same time), we should jointly consider the emergence of the urban phenomenon and the specialization and diversification of the manufacturing and service activities. This is what we will try to do in this paper.

Throughout this paper we will develop a simple *two stage model* of city formation and evolution. Although several approaches have already been taken, see i.e. DURANTON G., [6], we believe that city formation and evolution in a pre-industrial context has still not been sufficiently examined. The questions we will try to answer are how the urban phenomenon is related to *the emergence* of the division of labor between agriculture and manufacturing, how the urban phenomenon passes from a *local* to a *global scale* <sup>(2)</sup> and how *strict indivisibilities* can restrict adjustment mechanisms and give rise to “bad” agglomeration forces. How does, at a *first stage*, urbanization interact with the formation of a manufacturing sector and how do, at a *second stage*, different agglomerations interact in order to keep this agglomeration mechanism alive ?

As already pointed out by historians a long time ago, the existence of an *exchangeable agricultural surplus* has been a necessary condition for the formation of

---

(1) A major exception is the paper DURANTON G., [6].

(2) One should note that in our context the label “urban” automatically refers to a global phenomenon. See section 5 for an explanation of the terms “local” and “global”.

cities <sup>(3)</sup>. BAIROCH P., [1] brilliantly describes in his first chapter how the “neolithic revolution”, the “invention” of agriculture, led to a significant increase in productivity and made possible the sedentarization of people. This point will be covered by the first stage of our model. We will especially show that city size is highly reactive to gains in agricultural productivity when this productivity is initially at a very low level. This necessity of an agricultural surplus explains partly why certain regions, endowed with especially favorable “first nature factors”, were the first to develop agglomerations. It does however not answer the question why some of those regions (like e.g. the region around Jericho) suffered later decline and why some other regions, also endowed with favorable “first nature factors” were not able to develop similar agglomerations. Therefore, one must appeal to “second nature factors” in order to explain the formation and evolution of agglomerations. We especially would like to show under what conditions the existence of an *agricultural surplus* is not only a **necessary** but also a **sufficient** condition for the formation of urban agglomerations. As we will argue, this can only be the case if the agricultural surplus is *exchangeable* and if there are compensation mechanisms limiting and restricting factor movements. These exchange – compensation – restriction mechanisms, which as we will see are loosely related to what is usually referred to as *local governments*, are at the heart of city-formation and the second stage of our model, which we have termed the global dynamics.

Throughout our analysis, we will adopt the following definition of *urban agglomeration*, inspired by the following remark of BAIROCH P., [1] : “*L’homme, est-il le seul animal à avoir inventé la ville ? Oui, si l’on prend la ville dans son acception la plus spécifique, celle qui suppose une spécialisation des activités qui conduit à cette caractéristique spécifique des concentrations humaines qu’est les villes, à savoir que ces concentrations sont incapables de se nourrir elles-mêmes, donc de survivre sans l’apport des habitants des régions avoisinantes. Et c’est là l’essentiel du fait urbain, la différence qui sépare la ville de la fourmilière ou de la ruche*”. This had already been stated in a slightly different way by PIRENNE (1925) who argued that “*A city group, in fact can live only by importing its food supply from outside*”.

**Definition 1.1 (URBAN AGGLOMERATION)**

*An agglomeration will be called urban if its spatial location is invariant with time, if it has been created on a permanent horizon and if it is unable to provide itself with all the agricultural goods it needs.*

Throughout all of this paper we will adopt the “usual” notations of the literature and denote by *A-goods* the (homogenous) agricultural good(s) and by *M-goods* the manufactured goods. Note that the term manufactured goods should be understood in a very broad sense since we will consider that *M-goods* can be either physical goods or immaterial services. Simply consider that the term *M-goods* refers to all production that “can’t be eaten”. The agricultural workers will be called *A-workers* and the manufacturing workers will be called *M-workers*. Throughout this paper, the prefix *A-* will always refer to the agricultural sector, while the prefix *M-* will always refer to the manufacturing sector.

---

<sup>(3)</sup> We will use the word city in a rather generic way without defining for now what we exactly mean by it. Thus the term city will, for the time being, be synonymous with agglomeration.

An important preliminary remark must be made at this point before starting to develop the model and to analyze its implications in the next Section. When one tries to apply modern economic concepts to historical situations, one always risks to involuntarily create anachronisms and incoherences. We have therefore tried to put ambiguous terms systematically between quotation marks in order to remind the reader that he should not take all terms literally, that is to say in the modern accepted sense. This is especially important in our case since, as already stated, we will analyze the very formation of agglomerations in an “economic context” where there is no real economy up to now. This should nevertheless not directly affect the model’s underlying logic, which is largely independent of modern economic mechanisms <sup>(4)</sup>.

This paper is organized as follows :

In Section 2 we will stipulate some basic axioms and discuss some of our modeling choices. We will try to justify these choices with historical evidence and compare our approach with the standard approaches in the existing literature. Several terms will be defined and notations will be introduced.

In Section 3 we develop our basic general equilibrium model at a local scale. First we will define the constrained consumer problem and derive the demand functions, then we will turn to the conditions of *A-good* and *M-good* production. The equilibrium conditions will be established and the elasticity of city-size with respect to agricultural productivity will be analyzed.

In Section 4 we will very briefly explain some elements concerning the integer problem in (spatial) economic modeling and justify our approach in terms of *critical threshold*.

Section 5 explains the local dynamics of a single isolated agglomeration as its population size increases gradually. We will especially show that optimal agglomeration size is directly related to individual preferences and will discuss the implications. Some numerical examples will help to illustrate the fundamental concept of *potential urban site*.

In Section 6 we will finally drop the isolation assumption of Section 5 and examine how real urban agglomerations form as a result of factor- and goods-movements. There we will establish the link with more general concepts of spatial economics like e.g. economic potential and accessibility and we will show how these considerations narrow down the set of potential configurations.

Section 7 offers some preliminary conclusions and discusses future research directions.

---

<sup>(4)</sup> We believe that basic economic equilibrium- and optimization-mechanisms like e.g. factor movements in response to utility (resp. income) differences and profit (resp. utility) maximization are largely independent of any historical and economic context.

## 2 Basic axioms, historical background and modeling choices

Consider a rudimentary “economy” where there is no real manufacturing activity up to now <sup>(5)</sup>. The agricultural techniques have been adopted, so we are shortly after the “neolithic revolution”. As for every model, we need some simplifying assumptions which we will give now. We will consider an artificial initial configuration in which a relatively small population is evenly spread out over the landscape. Land is supposed to be of the same quality everywhere, even if we can, as we will show later-on, relax this assumption without essentially violating the importance of the *second nature factors*. Land is supposed to be abundant so that the small rural settlements are sufficiently dispersed and their agricultural domains do not overlap.

Without being exhaustive, we will stipulate the following two behavioural axioms for our analysis. A third axiom, concerning the agricultural production, completes this simple axiomatic framework. Although they are axioms and hence do, by their very definition, not need any justification, we feel obliged to say a few words concerning each of them.

### Axiom 2.1 (RATIONALITY)

*All individuals are supposed to be rational in the traditional microeconomic sense.*

This axiom will not be “justified”, so we will not enter again the discussions on its empirical relevance. We refer the interested reader to the abundant literature on this topic.

### Axiom 2.2 (SOCIAL PREFERENCES)

*All things being equal, individuals prefer to live near other individuals. As a corollary, we will assume that the utility of a single agent living in isolation is zero.*

This axiom may seem *ad hoc* at first because it immediately gives rise to an agglomeration force which seems to have no economic justification at all. Let us insist on two points that we believe are essential to keep in mind :

First, it has often been argued that the very formation of agglomerations (which will not be urban according to our definition) is, at least in early periods of human history, not primarily an economic matter <sup>(6)</sup>; considerations concerning religious, social and defensive matters prevail, as so many historians have pointed out. Since the beginning of time, humans have lived in groups. Without going into details let us just mention that humans are essentially social animals which do not naturally live

---

<sup>(5)</sup> That is to say there are no workers which spend their whole work-time on the production of manufactured goods exchanged on a “market”. Subsistence manufacturing like e.g. production and repair of primitive agricultural tools or hunting equipment will not be considered as being a manufacturing activity *stricto sensu*. In order for a worker to be a manufacturing worker, manufacturing must be his principal source of revenue and hence his primary (i.e. full-time) activity.

<sup>(6)</sup> It is even doubtful, as already mentioned, that we can speak of an “economy” at all at this period of history.



a solitary life. On the one hand this is due to the fact that there has always been a vital need to “communicate” and to “exchange ideas” (“*the propensity to interact with others is a fundamental human attribute, as is the pleasure to discuss and to exchange ideas with others*” FUJITA M., THISSE J.-F., [8]), on the other hand it is rather sure that the first humans were not at the top, but rather near the bottom, of the food chain. Even most simple tasks as for example hunting and keeping the fire lit required the joint efforts of several individuals. Protection, if not by physical strength could only be achieved by numbers. All those considerations lead us to reasonably assume that “being together” is better than “being alone”, thus “justifying” partly our axiom.

Second, although our reasoning may be hazardous, we believe that there is nevertheless a “strictly economic” reason able to “justify” this axiom. In a world where each one’s survival depends on the agricultural production, agglomeration can be (up to a certain *surplus-maximizing and variance-minimizing size*) some kind of insurance against famine and death. Illness and age have always been, and still are in most societies, critical factors which can threaten survival. Those two factors usually lead to a strong variability of the food supply, but this variability gets weaker the more individuals share the agricultural production. The lower the variability of the food supply, the higher the chances of survival and reproduction. Although the answer to this question will probably never be known, we believe that initial *rural agglomerations* at early periods of history can be viewed as some kind of informal insurance mechanism <sup>(7)</sup> ; therefore even at this early stage of history some kind of “economic” agglomeration force could have played a role. In all our subsequent developments we will, for simplicity, drop the probabilistic assumption on agricultural surplus and consider a deterministic model.

### Axiom 2.3 (DECREASING RETURNS IN AGRICULTURE)

*For each given and fixed level of technology, the marginal A-good production for a given agglomeration at a location  $s$  is decreasing in limit and tends to zero.*

Two important remarks must be made at this point of our analysis. First of all, we will adopt a slightly different approach than most of the other formal works in spatial economic theory. While most models consider that agriculture is land intensive and that *A-workers* are **therefore necessarily dispersed in space**, we consider that up to a certain point an agglomeration of *A-workers* is sustainable. Agricultural production is still land intensive and hence dispersed, but *A-workers* can form clusters. In that case they simply have to commute to their fields every day in order to produce the agricultural good <sup>(8)</sup> . Hence we will suppose that *A-workers* are not evenly spread

---

<sup>(7)</sup> There is an extensive literature that shows that informal insurance mechanisms still play a major role nowadays in developing countries as e.g. the Phillipines and Malaysia.

<sup>(8)</sup> This fact has already been noticed by VON THÜNEN as the following note attests : “*Relativement aux travaux exécutés sur le champ lui-même, la fraction de temps employée aux allées et venues des hommes et des attelages est complètement perdue; et cette fraction croît également en raison directe de la distance des champs à la ferme. Donc les frais de travaux sont moins forts pour le champs rapproché de la ferme que pour celui qui en est éloigné, et le premier doit donner un plus grand rendement net que le second, en supposant que la fertilité du sol soit égale de part et d’autre*” (VON THÜNEN as quoted in HURIOT J.-M., [13])

out on the “economic line” but are rather concentrated in *rural agglomerations*. This has strong historical evidence, as attested to by BAIROCH P., [1] :

*Le monde rural, un monde de paysans . . . le monde urbain, un monde d'artisans . . . Certes oui. Et cette différenciation des activités [. . .] est l'essence même de l'urbanisation. Mais il convient de ne pas pousser la différenciation urbain – rural trop loin et de considérer qu'industries et services sont des activités uniquement localisées dans les villes et surtout que l'agriculture est une activité purement rurale. En effet, si l'industrie rurale est une réalité bien connue et considérée même parfois, pour certaines périodes et certaines régions, comme une activité très importante, il en est tout autrement de la composante paysanne des villes que l'on tend à ignorer ou à négliger.*

BAIROCH, pp. 38 – 39

These commuting costs are of course synonymous with transportation costs and lead to decreasing returns in agriculture since the time lost in commuting cannot be used to work on the fields. Hence we have, contrary to many other models of that type, commuting costs for *A-workers* but costless intra-city transport of *A-goods* (which are brought back home by the agricultural workers). While the formal results are nearly the same, we believe that this approach is important for empirical reasons : while there are non-negligible inter-urban transport costs on *A-goods*, the intra-urban transport costs are less important than the commuting costs, since commuting costs are incurred on a *daily basis* while transport costs are incurred at most *several times a year* when the *A-good* is harvested and brought to the agglomeration. Hence at early historical stages of the formation of agglomerations, the commuting constraint has probably played a stronger role than the intra-urban transport cost constraint in restricting agglomeration sizes. DURANTON G., [6] states that “*The other two potential sources of dispersion could have been the transport costs of manufactured goods and the commuting costs of people. These two explanations must be ruled out for pre-industrial urbanization. It has already been argued that for small cities commuting costs can be neglected (and most pre-industrial cities were “small”)*”. Although this is true for manufacturing activities in a city, it is no longer true when we consider agglomerations of agricultural workers. As BAIROCH P., [1] depicts so vividly when speaking of antique Rome, “*Si seulement 5% de la population active de Rome s'étaient livrés à l'agriculture, cela aurait impliqué quelque 25000 à 30000 agriculteurs; et en postulant que seulement 10 hectares de terre par actif [. . .] cela signifierait une superficie de 2500 à 3000 kilomètres carrés qui, si elle était uniformément disposée autour de Rome impliquerait un trajet aller-retour de quelque 50 à 60 kilomètres pour les fermes les plus éloignées (ou quelques 10 à 14 heures). En fait le ravitaillement de Rome ne peut même pas se concevoir dans le cadre de la péninsule italienne*”. The transport cost constraint on *A-goods* will play its full role only when there are real *urban agglomerations* which need to import *A-goods* in order to supply their inhabitants. In that case the commuting constraint does no longer play a major role since it is replaced by the inter-urban transport cost constraint as soon as it becomes more profitable to import *A-goods* than to produce them locally.

The commuting of A-workers is one of the two factors responsible for the diminishing returns. The farther away the agricultural land, the less time the workers can spend on their fields. If we assume, which seems reasonable in view of the rationality axiom (2.1), that the most fertile and nearest land is cultivated first, increases in population lead to the cultivation of farther away and (or) less fertile land. To paraphrase NORTH D., THOMAS R., [20], “*the land surrounding any given agglomeration is inevitably limited in extend. Continued population growth eventually forces the people of the agglomeration to take up land of inferior quality, having in the past already cleared the best virgin land available*”. Therefore we can safely assume, as we do in axiom (2.3), that the agricultural returns are non-increasing. Since there is an absolute limit to daily commuting, determined by the level of technology, returns will tend to zero in limit as population continues to grow (when there is no more land available, the “worker per acre” ratio rises and we have diminishing returns à la RICARDO; we will neglect this issue since we work with fixed technological coefficients in the agricultural sector). The only way to escape the “spectre of diminishing returns” is to move out of the agglomeration, as NORTH D., THOMAS R., [20] clearly illustrate for 10th century Europe : “*As new fields were cultivated farther and farther from the village, some families tended to save travel time by moving out and enclosing small farms [ . . . ]*”.

**Remark 2.4** Let us just say some words considering our modeling choices. When modeling economic phenomena one has always to make a choice : general functional forms or particular functional forms. Although we believe that one should use general functional forms over particular ones, the second approach is nearly always adopted for reasons of mathematical tractability. As noted by HENDERSON V., [10], “[ . . . ] *specific functional forms are also chosen over general functional forms, because for certain propositions closed form solutions are highly desirable [ . . . ] before solving  $\partial U / \partial N = 0$  for the efficient  $N$ , we must specify the nature of the external localization economies function [ . . . ]*”. We will show in this chapter that one can derive several interesting results on equilibrium city sizes under general functional forms when introducing non-differentiability into the different functions. Non-differentiability broadens significantly the set of solutions to the optimization problems at hand since the solutions to a non-differentiable problem are mostly to be found at *kinks*, that is to say points of non-differentiability. Therefore, non-differentiable modeling provides us with a *priori* knowledge on the potential solutions to the optimization problems at hand. The essential problem, giving rise to the main critics generally adressed to this technique, is that this kind of modelization should be introduced in a non *ad hoc* way into the economic problem. There are, generally speaking, two different possibilities to introduce non-differentiability : on the agglomeration or on the dispersion side. We will introduce it on the dispersion side and use general functional forms in order to tackle diseconomies of agglomeration. This can easily be done, as we will show in the next Section, by simply constraining the traditional consumer problem. The additional subsistence constraint that we will introduce is both economically pertinent and gives us an additional degree of freedom since we have one more parameter in order to adjust our model. This comes however at a certain price. First, despite being tractable, the analytical expressions of our functions are a bit more complicated; second, we must use the subdifferential calculus tool of convex analysis which is generally not known and used in economics. We believe that the advantages outweigh the additional technical difficulties and that a lot can be learned from that kind of modelization <sup>(9)</sup> .

---

<sup>(9)</sup> We refer the interested reader to our forthcoming paper “Weak heterogeneity in spatial

### 3 The basic model

Assuming from now on that axioms (2.1) – (2.3) apply, we will briefly describe the basic mechanisms at work in the model. Consider, for simplicity, a linear economy stretching out on the real line. We will suppose that land is homogenous everywhere but, as already mentioned, we will show that this hypothesis is neither necessary nor desirable.

#### 3.1. The consumption side

As usual we will suppose that all agents are identical and have identical preferences, given by the following *Cobb-Douglas* <sup>(10)</sup> function

$$u(x_1, x_2) = x_1^\mu x_2^{1-\mu}, \quad 0 < \mu < 1 \quad (3.1)$$

where  $x_1$  is the quantity of the (composite) agricultural good (*A-good*) consumed and  $x_2$  is the quantity of the (composite) manufactured good (*M-good*) consumed <sup>(11)</sup>. Let  $p_1 = 1$  be the normalized unit cost for the *A-good* and  $p$  be the relative price of manufactured good expressed in *A-good*. The resolution of the constrained optimization problem with *agricultural subsistence constraint*

$$(\mathcal{P}) \begin{cases} \max_{x_1, x_2} u(x_1, x_2) \\ x_1 + px_2 = w \\ x_1 \geq c \end{cases}$$

yields the following demand functions <sup>(12)</sup>

$$x_1^*(w) = \begin{cases} \mu w & \text{if } w \geq \frac{c}{\mu} \\ c & \text{if } w \leq \frac{c}{\mu} \end{cases} \quad (3.2)$$

---

economic modeling". In this paper we study the implications of non-differentiability on a general equilibrium framework à la FUJITA M., KRUGMAN P., VENABLES A., [7] and show how this non-differentiability can be justified in an economically pertinent way.

<sup>(10)</sup> We use a Cobb-Douglas specification for consumers' preferences since this approach is the most frequent one in the existing literature. Additionally, the Cobb-Douglas specification allows for a neat treatment of revenue effects which will play an important role in the rest of our model. We have however not investigated the sensibility of the results obtained with respect to the functional forms adopted on the consumption side.

<sup>(11)</sup> When there is no manufactured good at the beginning, we will just consider that the utility level is zero. Note that this is problematic since the quantities consumed are normally supposed to be strictly positive in that type of modelization. Consider that by definition in our model  $u(x_1, 0) = 0$  for all  $x_1 > 0$ ; as one can verify, this corresponds to a continuous extension of the utility function.

<sup>(12)</sup> Unfortunately, the case with a CES-type subutility function for the differentiated manufactured goods yields great technical difficulties in the analytical resolution. The case  $w < c/\mu$ , corresponding to a positive KKT multiplier, does not yield any explicit solution in terms of the  $x_i$ ; nevertheless, several theoretical properties of the demand- and the indirect utility function could possibly be derived. We have not investigated this issue any further.

$$x_2^*(w) = \begin{cases} (1 - \mu)wp^{-1} & \text{if } w \geq \frac{c}{\mu} \\ (w - c)p^{-1} & \text{if } w \leq \frac{c}{\mu} \end{cases} \quad (3.3)$$

for *A-goods* and *M-goods*. Note that these expressions yield the traditional demand functions of the literature when  $c = 0$ , i.e. when there is no subsistence constraint. The parameter  $c$  can be interpreted as the minimum consumption expressed in units of *A-good* necessary for subsistence. This parameter is loosely related to the nutritional value of the composite *A-good*. Lower values of  $c$  are for example directly related to a higher nutritional value of the *A-good* (the now classical example of the comparison of rice and wheat yields a neat illustration of the signification of this parameter; see for example BAIROCH P., [1] or DURANTON G., [6]). Note that we do not treat the case where there is a saturation constraint on the consumption of *A-goods* since this can't be reasonably assumed in our setting on the emergence of urban structures. The *Karush-Kuhn-Tucker (KKT)* multiplier associated with the inequality constraint of the corresponding minimisation problem is given by

$$\alpha(w) = -\frac{c^{\mu-1}}{p} \left( \frac{w - c}{p} \right)^{-\mu} (\mu w - c) \quad (3.4)$$

which is always positive for  $\mu w < c$ . Indirect utility is therefore given by

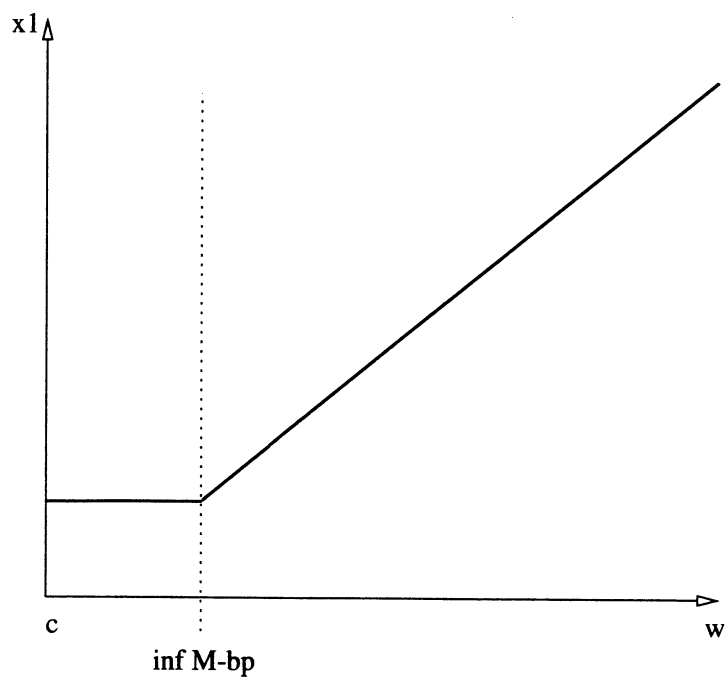
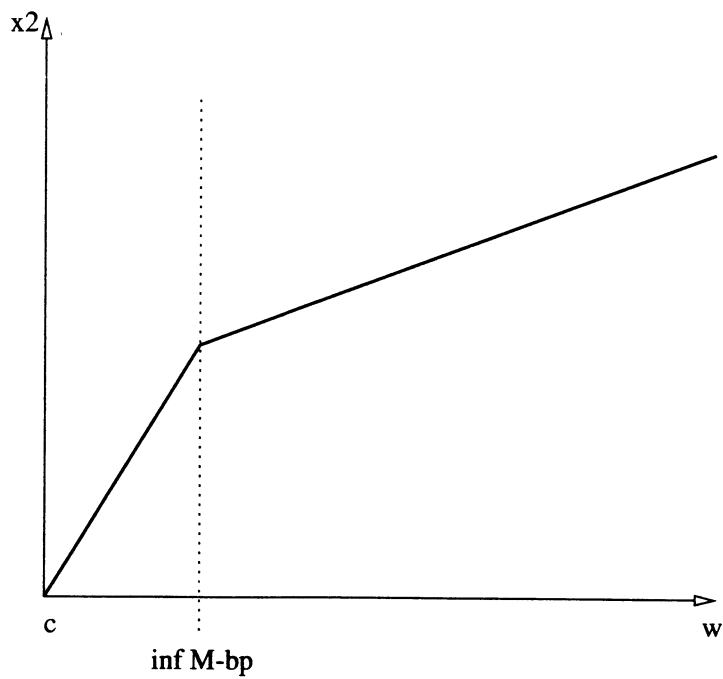
$$u^*(w) = \begin{cases} \mu^\mu (1 - \mu)^{1-\mu} p^{-(1-\mu)} w & \text{if } w \geq \frac{c}{\mu} \\ c^\mu (w - c)^{1-\mu} p^{-(1-\mu)} & \text{if } w \leq \frac{c}{\mu} \end{cases} \quad (3.5)$$

The subsistence revenue is given by  $w = c$ , in which case there is no consumption of *M-goods* possible since the whole revenue is used for the consumption of agricultural goods. Note that revenues below  $c$  are not feasible since this would imply that the agent can't even afford the strict minimum for subsistence. The *inf M-breakpoint* is given by  $w = c/\mu > c$  which is the point from which on any contraction in revenue leads to a stronger contraction in the demands of *M-goods* than before. One should note that any increase in wage below the *inf M-breakpoint* yields a stronger impact on the demand for *M-goods* than an increase in  $w$  above that point. Therefore demand of *M-goods* will be *more reactive* when revenues are low than when revenues are high. Note also that indirect utility is increasing more than proportionally when augmenting low revenues than when augmenting high ones. This seems plausible to us.

Figures [1] and [2] illustrate this for a given set of parameters. The discontinuities of the derivatives of  $x_1^*$  and  $x_2^*$  at the *inf M-breakpoint* are given by

$$\partial x_1^*(c/\mu) = [0; \mu] \quad \text{and} \quad \partial x_2^*(c/\mu) = [p_2^{-1}; (1 - \mu)p_2^{-1}].$$

The larger those two intervals, the stronger the modification of the revenue-elasticity of the demand functions as the revenue crosses the critical value  $c/\mu$ . These two intervals capture a phenomenon of *structural modification in the demands*; we abandon here an old implicit hypothesis in economics, namely the structural constance and the

FIGURE 1 : THE DEMAND FUNCTION FOR *A-goods*FIGURE 2 : THE DEMAND FUNCTION FOR *M-goods*

differentiability of the demand functions (for functions with non zero/infinite revenue-elasticity). The *inf M-breakpoint* will play an important role in the determination of the optimal size of the agglomeration. We will show later that optimality coincides with this point for a rather large choice of dispersion force functions. Hence we have a model in which the optimal size of the agglomeration is directly related to the demand functions of the economic agents (or more precisely to the *changes* in those demand functions). Before we continue our developments, two remarks are in order.

**Remark 3.1** A natural extension of this reasoning leads to the introduction of a *saturation constraint*, placing an upper bound on the consumption of *A-goods*. Although this is certainly pertinent, it yields additional technical difficulties and will not be treated in this paper. Considering the results we obtain in this model, it is reasonable to assume that the *sup M-breakpoint* will also enjoy certain optimality properties and coincide with particular economic configurations.

**Remark 3.2** The dynamics of the demand functions are quite intuitive to understand. When revenue is low (that is to say when we are below the *inf M-breakpoint*), *M-good* demand is constrained by agricultural subsistence considerations. Hence demand for *M-goods* is highly reactive when revenue increases from a very low level (people would like to consume more *M-goods* but their budget constraint hinders them). When revenue is at an intermediate level, any increase in revenue is spent proportionally to personal preferences given by  $\mu$  and  $1-\mu$  (this corresponds to the traditional *Cobb-Douglas* demand without constraints). When revenue gets high and if we consider the case where there is a saturation constraint (that is to say we are above what we could call the *sup M-breakpoint*), physical saturation for *A-good* constrains the consumption. Any increase in revenue will correspond to an increase in *M-good* consumption which, once again, becomes highly reactive to increases in revenue. Note that above the *sup M-breakpoint* only increases in population will lead to an increasing demand of *A-goods*. The first situation corresponds probably to what happens in developing countries or in countries in transition; those countries are typically very reactive to *M-good* consumption as revenues increase (just think of how industrialized countries' large firms dream of the awakening of the Asian markets). The second situation describes "balanced" growth of demand for countries in transition, while the third situation is typical for industrialized countries which need to cope with their agricultural surplus and which can only increase demand for industrial goods and services as revenue increases.

### 3.2. Production of A-goods

Let us suppose that to produce  $\xi$  units of *A-good* per unit of agricultural surface

$$L : s \rightarrow L(s) > 0$$

units of labor are needed at a distance  $s$  from the agglomeration. Note that we do not specify the exact functional form of  $L$  for now. Nevertheless, certain "dispersion properties" will be needed for  $L$  in order for axiom (2.3) to hold; we will suppose in particular that  $L$  is monotonously increasing beyond a certain value of  $s$ . Hence in order to produce  $\xi$  units of agricultural output, more agricultural workers are needed as the fields are farther away. This simply amounts to saying that the productivity of each individual worker gets less as he has to commute longer distances. If we note  $N$  the total *A-worker* population and  $r$  the agricultural fringe distance

$$N = 2 \int_0^r L(s) ds \quad (3.6)$$

must hold if there is full employment in the  $A$ -sector (we suppose that  $L$  is symmetric about the origin and that the agglomeration is located in the middle of the agricultural area). Equation (3.6) has a unique solution in  $r$  which will be noted  $r^*$  and which will be called the *A-worker exhausting agricultural fringe distance*. Let

$$r^* = r(N)$$

be the function giving the *A-worker exhausting agricultural fringe distance* for any given value of  $N$ . Total production of  $A$ -good is therefore given by

$$P_A = 2r^*\xi = 2r(N)\xi$$

while total  $A$ -good surplus is given by

$$S_A = P_A - Nx_1^*(w) = 2r^*\xi - Nx_1^*(w).$$

Of course  $S_A$  must be non-negative for an agglomeration of size  $N$  to be sustainable. We suppose throughout this paper that agriculture is carried out collectively<sup>(13)</sup>, so that the individual revenue in agriculture is given by

$$w(N) = \frac{P_A}{N} = \frac{2\xi r(N)}{N}. \quad (3.7)$$

Note that since revenue depends on  $N$ , the demand functions (3.2) and (3.3) can be rewritten as functions of  $N$  too. Hence we have

$$S_A = 2r(N)\xi - Nx_1^*(N). \quad (3.8)$$

---

(13) This assumption has strong historical evidence and should be pertinent for our analysis. As noted by MOKYR J., [17] when speaking about the introduction of the heavy plow in the european Middle Ages, *"its impact was especially momentous because it required a team of oxen to pull it. Few peasants could afford to own such an expensive capital good, and in part in an attempt to solve the fixed cost problem, medieval society developed a semi-cooperative organization sometimes referred to as the manorial system"*. Or as FUJITA M., THISSE J.-F., [8] put it, *"An alternative institutional system is obtained by assuming that a group of workers decide to form a co-operative producing the final good. They do so because they understand that by joining efforts, they may enjoy the benefits of increasing returns (at least up to a certain employment level). To this end [...] they form a community which then sets a local government. This government is entitled with the objective of maximizing the well-being of the community members by choosing the optimum community size"*. Although agricultural communisms should not be exaggerated, as LATOUCHE R., [16] emphasizes when discussing the historical pertinence of the *Markengenossenschaft hypothesis* concerning the germanic agricultural organization during the high Middle-Ages, we can still reasonably retain it as a proxy for historical organizational structures. As we will see, the community – local government approach is essential in understanding how dynamics are restricted and how certain processes can be kept alive. Finally, one should also note that cooperative production methods in agriculture are still observed to a certain extend in todays developing countries.



### 3.3. Production of M-goods

Suppose that the composite *M-good* is produced under a constant returns to scale technology using  $n$  differentiated intermediate inputs. Production of a final and consumable *M-good* is given by <sup>(14)</sup>

$$Q = \left( \int_0^n q(i)^\rho \right)^{\frac{1}{\rho}}, \quad 0 < \rho < 1$$

where  $q(i)$  is the quantity of the  $i$ th intermediate input used in the production of the consumable final good. The intermediate inputs (which can be goods and/or services or immaterial inputs like e.g. experience, knowledge etc. . .) are assumed to be the result of either manufacturing or service activities. Hence increases in  $n$  are synonymous with a growing diversification of the intermediate input sector. This growing diversification within the non-agricultural sector is a consequence of the primary division of labor between agriculture and manufacture and will be referred to as *secondary division of labor*; it gives rise to agglomeration forces and will permit the emergence of *urban agglomerations* <sup>(15)</sup>. Suppose that each intermediate input is produced by a single “firm” with inverse production function

$$l(i) = f + mq(i). \quad (3.9)$$

Thus each intermediate input is sold at price <sup>(16)</sup>

---

<sup>(14)</sup> Like in BECKER R., HENDERSON V., [2] we assume that “each firm “sells” its own span of products (though the market structure for this is not really specified) and all firms’ local outputs are combined costlessly to form the final city output”.

<sup>(15)</sup> Note that all  $q(i)$  will be used in the same quantity; hence production is done under constant returns to scale for any fixed number of differentiated intermediate input, but under increasing returns to scale as the number of differentiated intermediate input increases. Note also that we suppose that *M-good* production is carried out collectively in the sense that total sales revenue in the production of this composite good is used to pay the wages in the intermediate input firms.

<sup>(16)</sup> In DURANTON G., [6] “labor is paid at its marginal productivity because the production of each task is perfectly contestable”. This is not the case in our model where tasks are not contestable (due to the existence of the fixed costs) and where the intermediate input firms hence use a constant price mark-up. This has strong historical evidence. It is well known that one of the major reasons for the existence of guilds was the fact that urban artisans had to incur large fixed setup costs in both human and physical capital. In order to protect those large investments of their members, guilds precisely constrained the contestable character of the market by controlling the number of manufacturing workers in the city. Thus guilds enabled their members to capture some kind of monopoly rent, which can precisely be viewed in terms of price mark-ups. One should also note that final output will be produced no matter the number of intermediate inputs available (we will drop this hypothesis later on by stipulating the existence of “absolute” indivisibilities). Hence we do not require tasks to span the whole of a predetermined range, which eliminates bargaining problems and potential coordination failures. For a synthesis of a *Dixit-Stiglitz* and *Becker-Murphy* approach, see BECKER R., HENDERSON V., [2].

$$p(i)^* = p^* = \frac{wm}{\rho} \quad \text{in quantity} \quad q(i)^* = q^* = f \frac{\rho}{(1-\rho)m}$$

using the amount of labour

$$l(i)^* = l^* = \frac{f}{1-\rho}.$$

Therefore we have

$$Q^* = n^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \quad (3.10)$$

as total production of composite  $M$ -good. Since  $\rho < 1$  it is easy to see that  $Q^*$  has increasing returns in the number of differentiated intermediate inputs used. This can loosely be interpreted as the gains due to diversification of intermediate inputs and labor specialization. Although we do not explicitly model in a *Becker-Murphy framework* or like DURANTON G., [6] the micro trade-off due to specialization in specific tasks, we believe that the interpretation is finally very analogous to those more explicit models. Take a given service  $i$  and suppose this service is not highly specialized; in that case it can probably be broken down in its more specialized parts, hence giving rise to a wider set of available varieties. Note  $j, j \in \{1, 2, \dots, k\}$  those basic components. Each of those more specialized services requires workers to spend their whole time at producing those service. Hence breaking down  $i$  will automatically require a larger labor force, which shows once more that “specialization and diversification is limited by the extend of the market”. We do not agree with DURANTON G., [6] who claims that in this kind of modelization “[...] the idea that more labor devoted to the production of a given good leads to a higher marginal productivity does not appear”. When only  $i$  is produced, a fraction  $l^*/k$  of labor is devoted to the each basic component of the service. When specialization increases, each of the basic component captures the whole quantity  $l^*$  of labor. Since the passage from  $i$  to  $j, j = \{1, 2, \dots, k\}$  increases the marginal productivity in the final  $M$ -good industry, we see that the fact of using more labor at the production of specialized services (hence specialization) leads to a higher global productivity. Therefore, this kind of formulation captures, in a certain (admittedly special) way, the benefits from labor specialization.

Let  $L^* = n^*l^*$  be the total  $M$ -worker population. Since  $M$ -workers do exclusively produce  $M$ -goods, but do also consume  $A$ -goods, their maximum number is determined by the agglomeration’s agricultural surplus. This is what we will call the *A-good exhausting M-worker size*. Hence

$$n^*l^*x_1^*(N) = S_A \quad \text{which implies that} \quad n^* = \frac{S_A}{l^*x_1^*} = \frac{S_A}{x_1^*} \frac{1-\rho}{f}$$

must hold if the whole agricultural surplus of the agglomeration is used in order to feed the  $M$ -worker population and hence the market for  $A$ -goods is cleared. Zero profit condition in the  $M$ -industry requires that

$$\pi^* = PQ^* - \sum_{i=1}^n p^*q^* = 0.$$

This yields the equilibrium price

$$P^* = n^{1-1/\rho} \frac{m}{\rho} \frac{2\xi r(N)}{N} = \frac{2\xi m}{\rho} \left( \frac{1-\rho}{f} \right)^{1-1/\rho} \left( \frac{S_A(N)}{x_1^*(N)} \right)^{1-1/\rho} \frac{r(N)}{N}$$

which is the price which equilibrates demand and supply for *M-goods* according to Walras' Law. Equality of demand and supply can be checked via the equation

$$n^{*\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} = (N + L^*)x_2^*(N) \quad (3.11)$$

which can be written as

$$\left( \frac{1-\rho}{f} \right)^{1/\rho-1} \frac{\rho}{m} \left( \frac{S_A(N)}{x_1^*(N)} \right)^{1/\rho} = \left( N + \frac{S_A(N)}{x_1^*(N)} \right) x_2^*(N)$$

when  $n$  is replaced by its *A-good* exhausting value. This clearing condition holds for all values of  $N$ . A proof is given in appendix A.

Real revenue <sup>(17)</sup> in the agglomeration is given by

$$\frac{w(N)}{P^*} \quad (3.12)$$

which will behave much the same way as does the total amount of *M-good* production. One can check that

$$P^*Q^* = L^*w$$

which implies that there is equality between the revenues in the *A-* and in the *M-sector* since total sales revenue is equally divided amongst *M-workers* (since by symmetry wages are the same in all intermediate industries).

Note that the analytical expressions of  $Q^*$ ,  $P^*$ ,  $S_A$ ,  $x_1^*$  and  $x_2^*$  all depend on the value of the wage  $w$  for a given  $N$  (all those functions are bi-partite and are not easy to write in a compact analytical way). Their analytical expressions structurally change at the *inf M-breakpoint*.

Note finally that city-size is given by

$$s(N) = N + n^*l^* = N + \frac{S_A(N)}{x_1^*(N)}$$

as long as the agricultural surplus is positive. Therefore city-size is limited by a value  $\tilde{N}$  which corresponds to  $S_A(\tilde{N}) = 0$ .

---

<sup>(17)</sup> Strictly speaking this is a transformation of the real revenue. Since this function has globally the same behaviour, we will use (3.10) as a proxy for real revenue.

**Remark 3.3** By using the definition of  $S_A$ , city size can be rewritten as  $s(N) = \frac{2r(N)}{\alpha_1} \xi$  which, due to the expressions of  $x_1^*$  and  $w(N) = 2\xi r(N)/N$ , yields

$$s(N) = \begin{cases} \frac{N}{\mu} & , \quad w(N) \geq \frac{\epsilon}{\mu} \\ \frac{2r(N)}{c} \xi & , \quad w(N) < \frac{\epsilon}{\mu} \end{cases}$$

This expression shows that for high revenues, city size is a multiple of its “agricultural base”. The multiplier is given by the inverse of the share of revenue spend on agricultural products. This highlights once more the now well-known result that increases in the share of revenue spend on  $M$ -goods is one of the most significant factors leading to large agglomerations. But this expression also shows the fundamental trade-off to city size. As  $\mu$  decreases and  $N$  increases, city size will increase more than proportionally, but in order for revenue to stay sufficiently high,  $c$  has to decrease while  $\xi$  and  $r$  have to increase. Hence growth of city size goes, at least in the period we’re interested in, hand in hand with higher agricultural productivity, larger nutritional values of food-stuffs and lower commuting/transport costs in the agricultural sector. When we turn to the low revenue case, things become even more interesting. Consider the relative increase of city size due to increases in agricultural productivity. We have  $\frac{\partial s/\partial \xi}{s}(N) = \frac{1}{\xi}$ , which shows that growth will be “explosive” when  $\xi$  is sufficiently low. This can explain why we had some spectacular historical examples of city growth in the Ancient World (e.g. in ancient Sumer, city sizes “exploded” after the generalization of irrigation systems; some cities increased, according to SOUTHALL A., [25], by a whole order of magnitude). A similar result is obtained by DURANTON G., [6] who states that “it is no surprise then, that the relative increase in the sizes of pre-industrial cities can be dramatic after a rise in agricultural productivity [...]”.

#### 4 The problem with indivisibilities

Traditionally, most general equilibrium models suppose that  $n \in \mathbb{R}$ , so that adjustment due to factor movements can be handled via differential equations. Although one tries to model indivisible production activities, the continuity assumption  $n \in \mathbb{R}$  reintroduces a certain kind of divisibility into the model. This becomes especially visible when one wants to examine below which level of agricultural surplus no manufacturing activity can take place at all (hence if we introduce a certain *critical population size*). When  $n$  is a continuous variable, one can, *no matter how small the agricultural surplus is*, equilibrate the model by considering that  $n$  is *arbitrarily small* (that is to say, the *mass* of firms goes towards zero). Therefore there is no level beneath which no  $M$ -good production can take place. Nevertheless, we believe that such “discrete” and rigid barrier phenomena play an important role in explaining how initially uneven situations can give rise to even more uneven distributions of population and economic activities. This was already very clear to KALDOR N., [14] as he wrote that “*Allowance for indivisibilities means that for activities involving certain commodities there is a minimum scale of output, and the activity can only be “attained” at integral multiples of that minimum scale [ . . . ] if, in other words, there is a whole hierarchy of activities not all of which are feasible or attainable at any point of time – the choice among “activities” becomes primarily a matter not of prices but of the scale of production*”.

Let us drop the assumption  $n \in \mathbb{R}$  for a moment and consider that  $n \in \mathbb{N}$ <sup>(18)</sup>. This has

(18) Although  $n \in \mathbb{R}$  can be a convenient approximation of the situation where  $n \in \mathbb{N}$  is large,

several implications in terms of the equilibrium conditions derived above and leads to equations that we will handle in a rather sloppy and heuristic way during this section. We will denote the integer part of  $n$  by  $[n]$ .

First, one should note that the wage equations are always verified. We have

$$Q = [n]^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \quad \text{and} \quad P = [n]^{1-1/\rho} \frac{m}{\rho} w$$

so that

$$PQ = [n]l^*w.$$

The major problems due to the integer nature of  $n$  will arise in the demand- and supply-conditions in the  $A$ - and  $M$ -sector. Let us define the *excess A-good surplus* and the *excess M-good demand*

$$XS_A(N) = (n - [n]) \frac{f}{1-\rho} x_1^*(N) \geq 0 \quad (4.1)$$

and

$$XD_M(N) = \left( N + [n] \frac{f}{1-\rho} \right) x_2^*(N) - [n]^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \geq 0. \quad (4.2)$$

The two expressions above show that market clearing will not be achieved if  $n$  is not integer (a condition which will never be spontaneously realized since  $\mathbb{N}$  is of zero-measure in  $\mathbb{R}$ ). Note that due to increasing returns to scale, excess  $M$ -good demand is positive <sup>(19)</sup>. We will show how compensation can take place at a local level and at a global level. We will also show that local compensation is only possible at short term and that long run compensation must be global in order for an equilibrium to be achieved. It is important to realize that disequilibrium situations are both important from a theoretical and pertinent from an empirical point of view. Economists are not at easy with the concept of disequilibrium, especially since “[. . .] *in the rarefied world of Walrasian perfection where markets are continually in equilibrium, the question of how markets respond to “disequilibria” does not arise because all such “disequilibria” are ruled out – all equilibrating adjustments are assumed to be [possible and] instantaneous, either because changes are timeless or because all changes have been perfectly foreseen. However, the markets of the real world are not in continuous equilibrium in this sense; there are, or can be, persistent differences between production and consumption [. . .]*” KALDOR N., [14] (terms in brackets added by the author). The implications of disequilibria in terms of production and consumption will be further developed in Section 5.

---

we believe that this is not applicable to our historical examples. In pre-industrial cities, the degree of specialization and the number of goods and services available was not sufficiently large, so that we cannot reasonably assume that  $n$  goes to infinity (whatever this means concretely anyway).

(19) Since  $n \geq [n]$ , the quantity produced and demanded is lower in the integer case. Since increasing returns are at work, the quantity produced contracts more than the demand so that there is a positive excess demand for  $M$ -goods.

## 5 The model's first stage : local dynamics

We will briefly describe in this section the *local dynamics* of this model. By *local dynamics* we understand the dynamics of a single isolated and autarchic agglomeration. These "first stage dynamics" correspond to the emergence of what we have called *rural agglomerations*. The "second stage dynamics", the *global dynamics*, consist in the passage from rural to *urban agglomerations*.

Introduction of the *inf M-breakpoint* enables us to control the behaviour of the model over large ranges of parameters. Nevertheless this comes at a price : non-differentiability of most of the model's functions. Since the analytical expressions of all functions change when the revenue crosses the *inf M-breakpoint* given by  $c/\mu$ , we will start by using some numerical simulation examples to show how this model works. Afterwards we will derive analytical properties and show that the *inf M-breakpoint* coincides with the agricultural surplus maximizing population size for a large choice of functions  $r$ .

As stated before in our axioms, we assume that returns are decreasing in agriculture. These decreasing returns are needed in order to limit the agglomeration's size and the total amount of available agricultural surplus. We will therefore assume from now on that  $r$  is concave. In the case where returns are constant or increasing in the *A-sector*, we are in the traditional "black hole" situation; agricultural surplus can always be increased and hence there is no limit to the agglomeration's size and concentration of *M-industry*. Since productivity in the *M-sector* rises indefinitely, due to increasing returns, all production will take place in a single gigantic agglomeration. This case is of course of no interest.

Consider a rudimentary "economy" where agriculture provides the only source of "revenue"<sup>(20)</sup>. "Manufacturing" activity needs, in order to be efficient, workers to spend their whole time on non-agricultural tasks. Therefore they must be supplied by the surplus the other workers produce in agriculture. Since the manufacturing activities are subject to increasing returns to scale, several agricultural workers are needed in order to produce sufficient surplus for any *M-activity* to take place (refer to equation (3.9))<sup>(21)</sup>.

We will assume throughout this model that there exists a size  $N$  of *A-workers* such that the generated surplus is sufficiently high in order to develop an *M-industry* (that is to say at least one *M-firm* can be supported so that there is a strictly positive output for  $q$ ). Since the utility level is zero as long as there is no *M-good* production (remember axiom (2.2)), we will have agglomerations, that is to say that small clusters of *A-workers* will form which will support a certain level of manufacturing activity by the agricultural surplus they generate. These clusters can be of different sizes depending on *first nature factors* such as local fertility of land (given for example by

---

(20) Terms will often be enclosed in quotation marks to indicate that their meaning is not to be taken literally in the modern sense.

(21) Throughout this model we will reasonably assume that no single *A-worker* can generate sufficient local surplus on his own to support any manufacturing activity. Therefore, we need to introduce an indivisibility hypothesis since, as argued before,  $n \in \mathbb{R}$  allows even a single worker to support the production of a fraction of *M-goods*.

$\xi$ ) and the form of the function  $l$  (giving the nature of the dispersion forces at work). We will discuss later-on the homogenous versus heterogenous implications.

**Definition 5.1 (CRITICAL POPULATION SIZE)**

We will call  $\bar{N}$  the critical population size. It is the minimum size of  $A$ -worker population needed in order for consumable  $M$ -good production to take place, that is to say  $n^*(\bar{N}) \geq 1$ . Hence we introduce a strict indivisibility by supposing that

$$Q^*(n^*) = \begin{cases} 0 & \text{if } n^* < 1 \text{ (resp. } N < \bar{N}) \\ Q^*(n^*) & \text{if } n^* \geq 1 \text{ (resp. } N \geq \bar{N}) \end{cases}$$

Note that we could choose a different critical value than 1.

Note that the critical population size is increasing with the level of fixed costs and that it is decreasing with  $\rho$ . The lower the fixed costs and the more differentiated the goods (resp. services) are, the earlier the agglomeration will hit its critical population size and develop manufacturing activities. While the first argument is easy to understand, the second one has a more complicated interpretation. In fact, it relies on the implicit assumption that the more differentiated the intermediate goods and services are, the more likely it is that a small number of intermediate goods and services will be able to produce final goods. This amounts to saying that a wood craftsman and a smith will be able to produce more easily jointly a final good than a cooper-, gold- and silver-smith will be able to do. In pre-industrial times, strong diversity in inputs yielded, once combined, more easily usable final goods than very similar inputs would have done.

**Definition 5.2 (POTENTIAL URBAN SITE)**

A site will be called a potential urban site if there exists a value of  $N$  being a critical population size.

Note that a region can be stuck in an “urban underdevelopment trap” if it has no potential urban sites. In that case, agglomerations will grow but they will reach their maximum sustainable size (where all local agricultural surplus is consumed) before they can ever reach a critical population size (note that this result is not true if we don't introduce “real” indivisibilities since  $n \in \mathbb{R}$  always allows an agglomeration to support a fraction of  $M$ -industry; hence in this case all sites would be potential urban sites). No manufacturing activity can be developed and hence no urban agglomerations will emerge <sup>(22)</sup>. This “explains” why agriculturally less fertile regions can sometimes

---

(22) We insist on the fact that the rationality axiom and the definition we have given for an urban agglomeration imply that there must be a manufacturing activity for an agglomeration to be urban. If an agglomeration has an urban character, it is non-autosufficient and hence it must import agricultural goods. If there was no manufacturing activity, this would imply that the agglomeration has  $A$ -workers which absorb more than the whole surplus because they produce much less than they consume. This implies that  $A$ -goods must be imported from another agglomeration. Since transportation involves strictly positive costs, this violates the rationality principle; rationality assumes that the excess workers migrate to other under-sized agglomerations or form new clusters.

not develop an urban system; what is needed is not only an agricultural surplus but a surplus of sufficient magnitude to cover the fixed costs involved in non-agricultural tasks. The *second step dynamics* will explain more thoroughly why several ancient cities like e.g. Jericho could not develop as expected, even though they had an advanced agriculture and were located in fertile regions (see BAIROCH P., [1]).

**Remark 5.3** It is clear that manufacturing activities must already be known to exist before they can be developed. This can be safely assumed in our analysis since we focus on manufacturing that is carried out on a full-time basis. Historically, most manufacturing activities were first developed on a part-time basis in the context of a subsistence economy : each individual, besides cultivating his land, used his spare time in order to produce rudimentary craftsman goods for his own needs. Hence woodcraft, metallurgy, glass-blowing etc. were first developed on a small-scale basis by individuals for their own personal purposes, before being developed as real full-time manufacturing activities.

The “first stage dynamics” consist in selecting one or more *rural agglomerations* and to check their dynamics by assuming that there is no contact between the agglomerations (this is a reasoning in the spirit of VON THÜNEN’S *isolated State*). Depending on the parameters one gets different results which **nevertheless behave qualitatively the same way when one chooses a concave function** for  $r$  (these are the decreasing returns in agriculture of remark (5.1)) <sup>(23)</sup> . An analytical investigation of the agricultural surplus will allow us to derive sufficient conditions for the *inf M-breakpoint* to correspond to the optimal agglomeration size in the first step if  $r$  is globally a concave function of  $N$ . Once the *local dynamics* have been checked, the interesting work consists in dropping the “no-contact assumption”. We will show that the *global dynamics* consist in the emergence of one or more *urban agglomerations*, growing at the expense of their neighbors. Global dynamics are optimal responses and adjustment mechanisms, consisting in redistribution of population and production.

Without any analytical proof for now, we will start by investigating the following numerical example :  $\mu = 0.5$ ,  $\rho = 0.4$ ,  $m = 0.7$ ,  $c = 0.7$ ,  $f = 1$  and  $\xi = 1$ . For this same parameter-set we will investigate two different cases. In the first one we take  $r(N) = \sqrt{N}$  (case 1), in the second one we take  $r(N) = 0.6 + 0.3N$  (case 2). This yields the following figures and results :

Let us start with case 1 and take a look at Figure [3]. The graph of the agricultural surplus is obtained by using the above parameter values and equations (3.8) and (3.2). Note that the critical value for  $N$  is given by  $\bar{N} = 1.667$  in this example. As soon as  $N \geq \bar{N}$  the agricultural surplus is sufficiently high in order to support the production of *M-goods*. Therefore  $N = 1.667 = \bar{N}$  is the *critical population size*. Production of *M-goods* is sustainable over the interval  $N \in [1.667 ; 4.163]$  as can be seen from Figure [3]. Note that the point  $N^* = 2.0408$  plays a special role. It is the value of  $N$  for which we attain the *inf M-breakpoint* and hence have a structural change in the agents’ demand functions. It is the value of  $N$  for which  $w$  just crosses the “critical value”  $c/\mu = 1.4$ . As one can see from the different Figures [3], [4] and [5], this structural change creates cusps in all the important functions. Even better, the size  $N^*$  corresponds to maximum agricultural surplus, maximum *M-good* production,

---

<sup>(23)</sup> One can also have increasing returns for a certain range of  $N$  but one needs decreasing returns in limit.



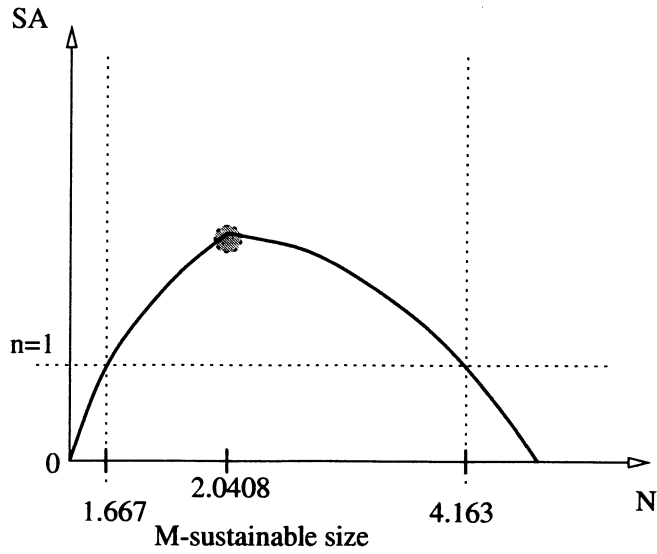


FIGURE 3 : AGRICULTURAL SURPLUS IN CASE 1

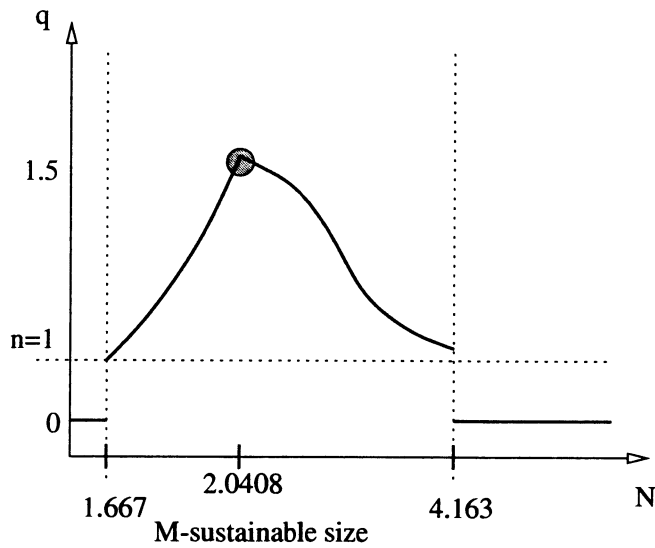


FIGURE 4 : M-GOOD PRODUCTION IN CASE 1

maximum utility, maximum mass of firms and (as we have not shown on the graphics) maximum real revenue. These results are fairly general under this parameter set and do not depend much on the concave function retained in order to measure diseconomies of agglomeration. This is due to the use of non-differentiability in order to control more easily optimal city size without explicit functional forms

The second case yields different results. In this case we have no interval of sustainable sizes for *M-good* production, while  $N^* = 1.5$  is the *inf M-breakpoint* (which also yields optimality in this example). Note that the maximal city size, which

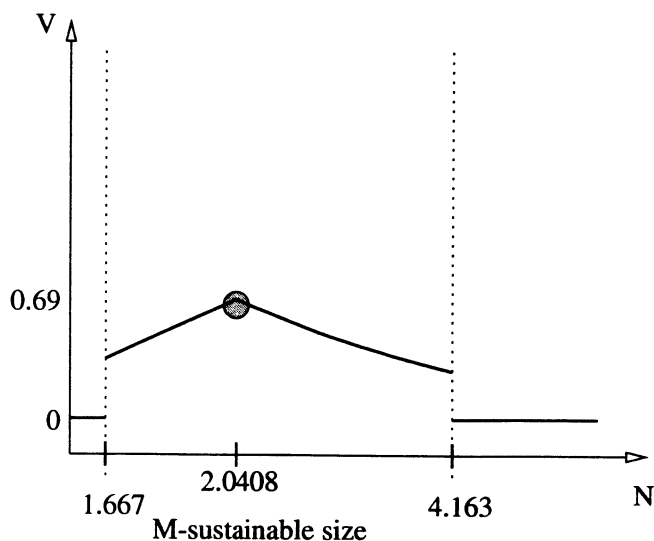


FIGURE 5 : INDIRECT UTILITY IN CASE 1

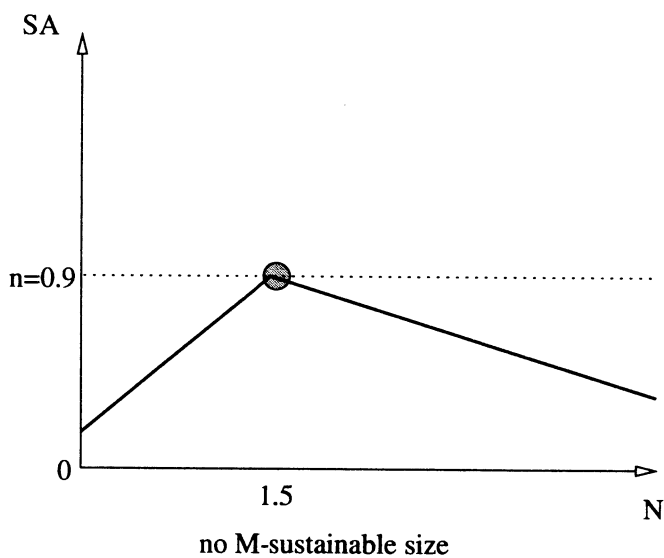


FIGURE 6 : AGRICULTURAL SURPLUS IN CASE 2

we will note  $\bar{N}$ , is larger in the second case than in the first one : 8.163 in case one and 12 in case two. The agglomeration of case 2 is not a *potential urban site*, that is to say no manufacturing activity can be supported by local agricultural surplus and break even (make zero profit while solving its optimization problem). Therefore utility and *M-good* production is zero, no matter how large the agglomeration gets.

As can be seen from our two illustrative examples, different agglomerations are characterized by different optimal and maximal sizes. These differences are due to exogenous "first nature factors"; some sites may develop a manufacturing industry while others may not (and as can be seen from the above examples, sites which

can develop a larger maximal local population than other sites, need not be able to develop a manufacturing industry; this is in accordance with the historical example of BAIROCH P., [1] who insists on the fact that there were big agglomerations in ancient China that were “urban” considering their size but “rural” considering the structure of their economic activities). These spatial heterogeneities can be considered as being the source of historical “accidents”<sup>(24)</sup> explaining partly why an agglomeration has appeared at a certain location. Nevertheless, as we will show in the section concerning the *global dynamics*, these “first nature factors” play only a minor role in the emergence of *urban locations*. They are secondary in the sense that they only explain why we have that particular observed spatial pattern **not why that pattern emerged**. First nature provides initial conditions, while second nature tells the story of how these initial conditions evolve. In the absence of initial conditions, that is to say in the case where space is perfectly homogenous, we have an indetermination that is inherently unstable. Random perturbations will lead to the global dynamics as if space was heterogenous.

### 5.1. On the optimality of the inf M-breakpoint

As seen in the two preceding examples, the *inf M-breakpoint* enjoys certain particular properties over a large range of parameter values. We will especially show in this subsection that this point is often characterized by  $N = N^*$ , that is to say that the *inf M-breakpoint* coincides with the optimal size of the agglomeration. We therefore have a model where *the optimal size of the city is directly related to the consumption behaviour of the individual agents*. In order to check for the optimality of the *inf M-breakpoint*, we can investigate under what conditions the agricultural surplus will be maximal. This surplus is given by

$$S_A(N) = 2r(N)\xi - Nx_1^*(N)$$

Since  $r$  is assumed to be concave and since  $x_1^*$  is convex,  $S_A$  is a concave and “sub-differentiable” function (since it is concave, “super-differentiable” would be a more appropriate term) in the traditional sens of convex analysis. A sufficient condition for a global maximum is given by<sup>(25)</sup>

$$0 \in \partial S_A(N) = \left\{ 2\partial r(N)\xi - N\partial x_1^*(N) - x_1^*(N) \right\}$$

where  $\partial S_A(N)$  is the sub-differential of  $S_A$  at  $N$ . Assuming that  $r$  is differentiable with respect to  $N$  (a hypothesis that could be relaxed in order to use e.g. piecewise affine functions, which are very realistic formulations for transport costs), this condition is equivalent to

<sup>(24)</sup> This terminology, largely used, is strictly speaking inadequate since exploitation of first nature factors is a rational response to observed heterogeneities in geographical space. They are called accidents because they are totally independent of human existence.

<sup>(25)</sup> Refer to HIRRIART-URRUTY J.-B., LEMARÉCHAL C., [11] for sub-differential calculus and its properties. Since all our functions are “well behaved” in the sense of convex analysis, sub-differential calculus in this paper can simply be considered as being a generalized form of traditional differential calculus.

$$d = \frac{2r'(N)\xi - x_1^*(N)}{N} \in \partial x_1^*(N).$$

According to equation (3.2)

$$x_1^*(N) = \begin{cases} \mu \frac{2r(N)\xi}{N} & \text{if } N \leq N^* \\ c & \text{if } N > N^* \end{cases}. \quad (5.1)$$

The function  $x_1^*$  is differentiable everywhere except at the *inf M-breakpoint*  $N = N^*$ . One can easily verify that

$$\partial x_1^*(N^*) = \left[ \frac{\mu}{N^*} (2r'(N^*)\xi - \frac{c}{\mu}) ; 0 \right]$$

since  $(2\xi r(N^*))/N^* = c/\mu$ . Optimality of the *inf M-breakpoint* with respect to the maximization of the agricultural surplus is given by

$$d = \frac{2r'(N^*)\xi - c}{N^*} \in \left[ \frac{\mu}{N^*} (2r'(N^*)\xi - \frac{c}{\mu}) ; 0 \right]. \quad (5.2)$$

Using the above condition, we will show that the *inf M-breakpoint* is quasi-optimal for the first case where  $r(N) = \sqrt{N}$  and optimal for the second case where  $r(N) = 0.6 + 0.3N$ . Consider the first case. Since  $r(N) = \sqrt{N}$ , we have  $r'(N) = 1/(2\sqrt{N})$ . Knowing that  $N^* = 2.0408$  we can, using equation (5.2) and the numerical values for the parameters, compute  $d$  as being equal to 0.000001372. Or  $d \notin [-0.1715 ; 0]$  (use again equation (5.2) and the parameter values). Nevertheless  $d$  is “nearly” in the “sub-differential” and we will therefore consider that the *inf M-breakpoint* is quasi-optimal (we can show that the maximal surplus is attained just after  $N$  grows beyond  $N^*$ ).

Consider the second case. Since  $r(N) = 0.6 + 0.3N$ , we have  $r'(N) = 0.3$ . Knowing that  $N^* = 1.5$  we can compute  $d$  as being equal to  $-0.0666$  which is contained in the “sub-differential” given by  $[-0.2666 ; 0]$ . Hence the *inf M-breakpoint* is indeed the surplus maximizing point.

One can check that  $r(N) = \log(1 + N)$  does not yield optimality for the *inf M-breakpoint* while  $r(N) = \log(2 + N)$  does. Other concave functions yield similar results. One therefore sees that the exact analytical expression of dispersion forces plays no major role. What is important is the strength of those forces, a traditional result in spatial economic theory.

The principal advantage of our approach is that we can easily check whether a specific functional form for  $r$  does yield optimality for the *inf M-breakpoint* or not. This allows easy numerical resolution of the problem for complicated functional forms, without having to derive new equilibrium expressions each time the functional form of  $r$  is changed. Hence we are able to use various different forms for the dispersion forces with minimal analytical effort. Additionally this shows that the model’s basic properties are relatively robust and relatively independent of modelization choices, at least on the dispersion side concerning the function  $r$ .

## 5.2. Optimal vs effective agglomeration size

What happens when agglomerations continue to grow, due to an exogenous increase in population ? As we have seen above, some will drop out as unsustainable urban locations while some others will begin to develop *M-good* production. The question we want to investigate in this section is that of optimal vs observed city size. Will agglomerations grow until they reach at least their “optimal size” ? Can they grow beyond that size and can such a configuration be sustained as an equilibrium ? The problem with the effective size arises right here.

Consider an agglomeration  $A_j$  whose size is larger than  $\bar{N}$  but smaller than  $N^*$ . As long as its size is less than its optimal size, growth of the agglomeration leads to growing utility and is therefore rational for all of the agglomeration's inhabitants. Consider that this optimal size is reached and that it corresponds to the agricultural surplus maximizing size (the whole of this surplus being used to supply the *M-workers*). Since there is no surplus left and since the generated surplus is maximal, it is obvious that an absolute limit to the concentration of *M-workers* has been reached in the agglomeration. No matter if the number of *A-workers* increases or decreases, there is not enough surplus to support the new number of *M-workers*. Hence when a new agent enters the agglomeration, this agent will necessarily be an *A-worker* so that  $N$  will grow beyond  $N^*$ , which means that the available agricultural surplus decreases. This decrease in agricultural surplus reduces the number of *M-workers* that can be supported in the agglomeration and leads to decreasing utilities in the agglomeration. Although this is not desirable for the agents already residing in the agglomeration, it is for the entering agent who would, by axiom (2.2), have a utility level of zero if he was to live in isolation. Is this a “realistic” scenario ? We believe that it isn't in our model for the following reason : we consider a historical period where cities were traditionally small and we suppose that agriculture is carried out collectively. Hence there must be some organizational institution, which we will call for simplicity the *local government*, which will restrict access to the agglomeration in order to keep the agents' utilities at a maximal level. Hence we can reasonably assume that access to the agglomeration will be restricted as soon as the optimal size is reached<sup>(26)</sup>. Can sizes between the optimal and maximal size of the agricultural population (hence in the interval  $[N^* ; \bar{N}]$ ), be sustained as equilibria once established ? Consider that we have an agglomeration of

---

(26) While the historical pertinence of the hypothesis of agricultural communism may seem fragile, that of the collective exclusion mechanism is highly robust and accounted for. There are countless historical examples of communities regulating their size by exclusion and restriction. Ostracism has been a powerful controlling device in mercantile Venice and is also explicitly stated much earlier in the 45th title of the Salic Law : *De migrantibus*. As LATOUCHE R., [16] explains, “*Ses trois articles règlent le sort de l'homme qui change de domicile et qui, accueilli par plusieurs habitants du village où il veut s'installer, se heurte à l'opposition d'un ou deux autres sous prétexte qu'il a voulu s'établir sur la terre d'autrui. Une longue procédure s'engage qui s'achève par l'expulsion de l'indésirable, l'abandon de la terre qu'il a commencé à labourer et sa condamnation à une amende de 30 sous d'or si les habitants hostiles à l'intrus ont persisté dans leur refus*”. MOKYR J., [18] also emphasizes the fact that cities always have found ways to expand their economic base while maintaining their optimal size (mostly through guild restrictions).

size  $s(N_j, n_j) = N_j + l^*n_j^*$ . Consider also that the *inf M-breakpoint* corresponds to the surplus maximizing (optimal) size. Since  $N_j > N_j^*$ , the agricultural surplus is not maximal. A reduction in  $N_j$  will increase the total available surplus, which will allow for an increase in  $n_j$  and the utility of the agents. Hence if we assume that there is perfect mobility between the *A-* and the *M-sector*, sizes of the *A-worker* population beyond the optimal size risk to be unstable (when there is no inter-sector mobility, the size can be sustained as an equilibrium if no *A-worker* exclusions are possible). Since  $s$  is increasing with  $N$ , two cases must be considered; the case where the local government is able to exclude agents and the case where it is not able to exclude agents. Consider an equilibrium with an agglomeration size  $N_j > N_j^*$ . If  $N_j$  is decreased,  $n^*$  increases but the total net effect on the agglomeration size  $s$  is still negative. Hence only a fraction of the agricultural workers can be integrated in the manufacturing industry if the equilibrium conditions are to be kept. Therefore an equilibrated decrease of the *A-worker* population towards the optimum requires the exclusion of some of those workers. If this exclusion is possible, the configuration beyond the optimal size is not an equilibrium. If the exclusion is not possible, the configuration beyond the optimal size can be sustained as an equilibrium. We believe that *ex post* exclusions are much less plausible than *ex ante* restrictions on the access to the agglomeration. Hence we believe that agglomerations will not grow beyond their optimal size. Of course this assertion must not be extrapolated to more modern scenarios; as one knows, restrictions on access are more and more difficult as agglomerations grow and information gets imperfect. And once the optimal size has been crossed, suboptimal equilibria can be sustained as argued above (see the excellent paper by BECKER R., HENDERSON V., [2] for a clear illustration), since individual agents have no incentive to leave the agglomeration and since they can't be excluded.

The following table resumes the sustainability results for different agglomeration sizes once they have been established; it can be seen that the introduction of a critical mass yields somewhat richer results than in more traditional models (e.g. BECKER R., HENDERSON V., [2]). The case of  $[N^*; N^{**}[$  is of special interest since it requires some more hypothesis on how the underlying mechanisms of the model are specified. It can be seen that different historical periods (hence different institutional and "technical mobility" environments) may yield different results. This result especially suggests that in modern developing economies (characterized by low mobility between the *A-* and *M-sectors* and poor exclusion control) oversized cities can be sustained as sub-optimal equilibria.

Size	$[0; \tilde{N}[$	$[\tilde{N}; N^*[$	$[N^*; N^{**}[$	$]N^{**}; \tilde{N}]$
Sustainable	Yes	No	depends on exclusion and mobility hypotheses	Yes

## 6 The model's second stage : global dynamics

If we limit ourselves to the *local dynamics* where each agglomeration is supposed to develop in autarchy, things are pretty simple and rather uninteresting (assumptions of autarchy most times do not fit nicely with economic theory anyway, since economics is mostly about interactions between different components of a system considered as a whole). In this section we will investigate the *global dynamics* that arise when multiple agglomerations interact with each other. As BECKMANN M., [4] has put it neatly, “*To understand urban growth processes [...] one must get away from the single metropolis*”. This is what we will do in the following sections.

Consider a region where the population is (not necessarily evenly) dispersed and lives on agriculture only. Since by assumption no individual *A-worker* is able to generate on his own enough agricultural surplus to support any (even arbitrarily small) level of *M-good* production, no manufacturing activity exists and utility is zero everywhere. This configuration is both suboptimal and highly unstable. Suppose for example that two individuals agglomerate at a certain location and that those two individuals are not able to produce jointly enough agricultural surplus to support any *M-good* production; utility is still equal to zero but in view of axiom (2.2) the two individuals will prefer to live together rather than separated. Therefore individuals will begin to agglomerate at various locations which are determined by *first nature factors* like for example fertility of the soil and natural amenities. The initially dispersed population tends to agglomerate in several locations of rural nature. These agglomerations are characterized by their self-sufficient character and their absence of manufacturing activities. This is what happens in the first stage of our model in the previous section.

Without loss of generality, we will now drop the assumption that land is of homogenous quality <sup>(27)</sup> and we will assume that the different “embryonic” rural agglomerations grow at the same exogenous rate. Differences in first nature factors simply amount to stipulate that each agglomeration is characterized by a different agricultural productivity  $\xi_i$  and dispersion force  $r_i$  (we could also assume that  $c_i$  is different for each agglomeration, but this hypothesis is more difficult to defend since it implies heterogeneous agents). Let us assume that there are  $k$  such agglomerations which will be labeled  $A_i, i = 1, 2, \dots, k$ . Suppose for simplicity that returns are strictly decreasing in agriculture (a more realistic assumption would be to consider that returns are first increasing and then decreasing). As the agricultural population  $N_i$  of each agglomeration grows, per capita revenue decreases which in turn leads to decreasing consumption of *A-goods*. Nevertheless, consumption decreases faster than per-capita production as  $N_i$  stays below  $N_i^*$  and agricultural surplus therefore increases <sup>(28)</sup>. Several agglomerations, located in the less fertile areas of our region, will eventually

---

<sup>(27)</sup> This symmetry breaking assumption is necessary in order to solve an otherwise undetermined situation. It can be relaxed and we could assume that land is perfectly homogenous; in this case we would either have to solve a bargaining problem or to introduce some other exogenous perturbation in order to initialize the process of city formation. The important thing to note is that *first nature factors* are not necessary for the subsequent analysis to work. They “explain” how the process is initialized but they are not a necessary condition for this process to take place.

<sup>(28)</sup> As long as there is no manufacturing activity, the market for *A-goods* does not neces-

hit  $N_i^*$  while they are still not able to support any production of  $M$ -goods (that is to say that  $S_A(N_i^*)(1 - \rho)/x_1^*(N_i^*)f = n^*(N_i^*) < 1$ ). Those agglomerations are not *sustainable urban locations* and will not develop any manufacturing industry (this is the case of the second example we have developed in the preceding section). Since utility is still zero in those locations, they will continue to grow until they eventually hit a *malthusian upper bound* (given by  $S_A = 0$ ), limiting definitively their population expansion (those locations are caught in the “urban underdevelopment trap”).

While several locations may drop out as *unsustainable urban locations*, one or more other locations will eventually hit the critical value  $\bar{N}_j$  with

$$\frac{S_A(\bar{N}_j)(1 - \rho)}{x_1^*(\bar{N}_j)f} = n^*(\bar{N}_j) \geq 1.$$

Those locations are able to produce enough agricultural surplus in order to support a full-time manufacturing production which breaks even, maximizes profit and offers a revenue as high as in the agricultural sector. In order to avoid a bargaining problem, we will suppose that there exists an agglomeration  $A_j$  which hits the critical value first (we therefore need some heterogeneity of fertility since otherwise all agglomerations would hit this value at the same moment). This agglomeration, which will be called the *leader*, will therefore develop an  $M$ -good production before any other agglomeration does. Note that if two or more agglomerations hit the critical value at the same moment, the bargaining problem consists in determining which agglomeration will develop the manufacturing activity (remember that due to increasing returns to scale, and if the agglomerations are “sufficiently near”, only one location will develop this activity) and which one will need to get compensation for not developing a manufacturing activity of its own.

Assume for a moment that  $n$  is an integer. This assumption will enable us to highlight a fundamental local adjustment problem and to introduce the global dynamics examined later on in this section. If  $n \in \mathbb{N}$  it is clear that the probability that  $n^*(N^*) \in \mathbb{N}$  is zero. Therefore  $[n^*] < n^*$ ; nevertheless the  $A$ -worker population will continue to grow until  $N^*$  is reached because of the utility differential. This creates disequilibrium in both the  $A$ - and the  $M$ -markets. We have an excess in the supply of  $A$ -goods and an excess in the demand of  $M$ -goods which are given by equations (4.1) and (4.2). Local compensation requires the relative price of  $M$ -goods to increase. This leads to revenue and hence utility differentials between the  $A$ - and  $M$ -sector. A local adjustment is impossible;  $A$ -workers want to migrate to the  $M$ -industry, which can only support a new firm when a sufficiently large number of  $A$ -workers integrates this new firm. Or this is not possible since in that case the agricultural surplus is no longer able to sustain

---

sarily clear (we have a positive agricultural surplus). This is of course due to the fact that we normalized the price of the only available good, which leaves no possibility for adjustment. We should interpret this situation as follows : in a world where the level of agricultural production should be considered as being a random variable and where daily survival depends on the non-negativity of the agricultural surplus, this surplus will be biased towards strictly positive values. It is better (even indispensable) to have a strictly positive expected surplus rather than a zero expected surplus. Hence the existence of an agricultural surplus can be considered as being some kind of *informal insurance mechanism*.



the *M-worker* population. So one possibility is that a *local governments* restricts further growth of the *A-worker* population when the best value  $[n^*]$  is attained. We will not consider this case but rather turn to the case where agglomerations interact with one another. This gives rise to what we have called the global dynamics of the system. The global dynamics is at the heart of the formation of *urban agglomerations*, thing we have not considered at all when considering only the local dynamics <sup>(29)</sup> .

Suppose that  $A_j$  hits its critical *A-population* size given by  $\bar{N}_j$ . From that point on this agglomeration can support a full-time manufacturing activity and will produce *M-good* as it continues to grow. This *M-good* is consumed locally and utility of the agglomeration's agents becomes strictly positive, that is to say  $U_j > 0$  (refer e.g. to Figure [4] for an example; as one can see the production of *M-goods*, and hence the utility, has a strong discontinuity). Since the other agglomerations are still unable to develop an *M-good* production of their own, we have  $U_i = 0$  for all  $i \neq j$ . Therefore we have a utility differential which, by assumption of interaction between the agglomerations, will lead to several *adjustment dynamics*.

Traditionally, economists consider that factor movements are the main adjustment mechanism in economics. In our case, we will have factor movements but they will not play an equilibrating role, at least in the beginning. Suppose that all agents are perfectly mobile and can costlessly relocate. Since agents can reach a higher utility in  $A_j$  than in any other agglomeration, there will be movement of workers from the agglomerations  $A_i$  towards  $A_j$ . This will further increase the size  $\bar{N}_j$  until we hit  $N_j^*$  (the size which maximizes the utility of residents of agglomeration  $j$ ). Once this value has been reached, agents living in  $A_j$  will restrict access in order to preserve their maximum utility level (this has strong historical evidence, especially with immigration/emmigration restrictions by Guilds and other local authorities like merchant councils on so on). Therefore, factor movements stop when utility in  $A_j$  is highest. Since relocation of workers from  $A_i$  to  $A_j$  imply that  $N_i$  has decreased, we still stay below the critical value  $\bar{N}_i$  in the other agglomerations which keep a utility level of zero. Hence **factor movements have amplified the utility differential**. We will explain below why we think that factor movements can still have even nowadays a disequilibrating influence on the spatial configuration of developing countries.

After a first phase of factor movement, we have maximum utility in  $A_j$  while utility is still equal to zero in each  $A_i$ . Let us suppose first that factor movement restrictions do not work well (that is to say that there is no efficient local government in  $A_j$  capable of limiting access to the agglomeration in order to keep its residents' utility-level at its maximal value). In this case, factor movements will continue as long as  $U_j > 0$ . Utility differentials will disappear as population of  $A_j$  explodes and reaches its malthusian upper bound  $\bar{N}_{\max}$  from which on no *M-good* production is longer possible (this corresponds to  $\bar{N}_{\max} = 4.163$  in our first example and is illustrated by Figure [3]).

---

(29) Our argumentation resembles that of traditional microeconomics when there are indivisibilities. As is well known, the integer problem (or the non-convexities) becomes negligible when the number of agents becomes large. In our case adjustment requires the interaction of several agglomerations, exchanging surpluses and concentrating production in a few locations in order to equilibrate as much as possible the markets and to level economic conditions.

**Remark 6.1** The situation where the population explodes and leads to a utility level of zero in the agglomeration is highly implausible both from a logical and a historical point of view. To our knowledge, there have been no regressions from manufacturing societies back to rural ones. Agents have always been able to defend their “privileges” by restricting access, forming coalitions or more brutal and military means. A more interesting interpretation is from the view of today’s developing countries. In developing countries, insufficiently developed rural areas, bad transportation infrastructures and low revenues usually inhibit the compensation mechanism by good flows. Hence compensation is “tried” by factor movements, consisting essentially in rural exodus towards the higher revenue areas materialized by the larger cities. This excess factor movement consists in what BAIROCH has adequately termed the *third world urban inflation*. Paradoxically this does not lead to an equalization of real revenues and utilities; the compensation mechanisms are unable to work properly since those cities are kept alive by western countries’ agricultural surplus. This leads to what is commonly referred to as “poverty trap” in which those countries are stuck. Rural development is not encouraged (since the large metropolitan areas can be supplied exogenously at low costs with food), real revenues in the metro areas stay higher than in the country-side and will not diminish sufficiently to stop excess factor movements since the city stays attractive as external aids prevent adjustment. This aid in turn is not for free and absorbs most goods that could be used in order to level the inequalities between country-side and urban areas. Therefore we have the situation often observed where overcrowded cities like e.g. Manila offer immigrants from the country-side very low utility levels, which are nevertheless still higher than in the country-side where they are “zero”. Hence cities continue to develop since adjustment cannot take place. This mechanism shows that urban areas are not automatically synonymous with balanced development or simply development; this mitigates the fact that “in its recent *World Development Report, the World Bank (1999) stresses the importance of economic agglomerations and cities for boosting growth and escaping from the poverty trap*” FUJITA M., THISSE J.-F., [8]. What a country needs in order to develop is the development of a *system of cities*; or what we observe in most developing countries is a polarization between some few gigantic megalopolises and underdeveloped country-sides. We refer the interested reader to DURANTON G., [6] for a more thorough discussion on the inefficiencies due to the existence of *primate cities*.

Since we do not think that excess factor movement is the relevant adjustment mechanism to explain the emergence of *urban agglomerations*, we will turn to the second adjustment factor: good flows. Good flows will play a double role; they are both an adjustment mechanism and a rational response to the exploitation of scale economies at a more aggregate level. Trade, as has often been argued, is an “[. . .] *indirect way to trade factors of production and a way to achieve the economies of localization. It will also ordinarily be beneficial to all concerned*” KRUGMAN P., [15]. We will now explore in some more detail the feasibility of trade and we will show that trade can be seen as consisting in the replacement of one agglomeration diseconomy (decreasing returns in the local supply of *A-goods*) with another agglomeration diseconomy (transport costs of goods). Suppose as before that the first relocation of workers from  $A_i$  to  $A_j$  has led to a maximum utility  $U_j^*$ . The total production of *M-goods* at  $A_j$  is of  $q_j^*$ , which gives us a certain *per capita* consumption. As explained above, agglomeration  $j$  cannot grow anymore since increase in  $N_j$  will cause this *per capita* consumption (and hence utility) to decrease. This is problematic since utility could still be increased further if there was expansion in the *M-good* sector (which is subject to increasing returns to

scale). Or this expansion is impossible to realize at a *local level* but not at a *global level*. Therefore the following adjustment dynamics could take place :

Several agglomerations  $A_i$  are below the critical level of agricultural surplus where they can support the production of *M-goods*. Nevertheless, there is an agricultural surplus available in these agglomerations. This local surplus can be shipped to the agglomeration  $A_j$  at a certain positive transport cost. Therefore the total agricultural surplus available in  $A_j$  is given by

$$\psi(A_j) = \sum_{i=1}^k S_{A_i} f_A[d(A_i, A_j)]$$

where  $f_A$  is a standard distance substitution function for *A-goods* depending on the distance  $d$  between  $A_i$  and  $A_j$  and representing the *dissipative effect of space*. This *global surplus*, which by construction is greater than the *local surplus* of the leader, raises the number of *M-good* firms that can be supported in the agglomeration  $A_j$  and increases more than proportionally (depending on the strength of the economies of scale) the output of *M-good*  $q^*$  in  $A_j$ . A fraction of this additional production must of course be exported to the other agglomerations  $A_i$  as compensation for the imported *A-goods*. We will assume that the transport of *M-goods* can be done at no cost : the *A-workers* who supply the leader with *A-goods* simply transport them on their way back home. The *central place*  $A_j$  provides its *subordinate* centers  $A_i$  with *M-goods* while it receives the agricultural supply it needs for subsistence. Note that  $A_j$  is now what we call a *real urban agglomeration*, unable to provide itself with the food it needs. Note also that, even if *M-goods* could be produced in  $A_i$ , they are produced at much lower price in  $A_j$  due to increasing returns to scale. This differential in the cost of production is partly captured by the city  $A_j$ , which can offer higher utility levels than before, **compensating the agglomerations  $A_i$  for not developing an *M-good* production of their own**. The distinction rural – urban does well appear with the specialization and division of labor at a more global level.

### 6.1. Some numerical examples

In this section we will give a brief illustration of how global dynamics work. We will use the two numerical examples developed at the local level in Section 5. As we have seen, the agglomeration of case 2 hits its optimal size  $N^*$  before the agglomeration of case 1 but it is not able to develop any manufacturing activity (it can't generate enough agricultural surplus). Hence population growth continues for both agglomerations with zero utility until agglomeration 1 hits its *critical value*  $\bar{N} = 1.667$ . Since it is now possible to produce manufactured goods in the first agglomeration, utility becomes strictly positive there and the factor flows from the other agglomerations will push the population size of agglomeration 1 up to its optimal value  $N^* = 2.0408$ . This corresponds to the first *disequilibrating adjustment phase with factor movements*. Once the optimal size of agglomeration 1 is reached, adjustment through *good flows* will take place. This works as follows.

Since agglomeration 1 has reached its optimal size, its agricultural surplus is equal to 0. The only possibility to raise this surplus is to import it from the other agglomerations. Suppose that there are  $k$  agglomerations of type 2, as depicted in

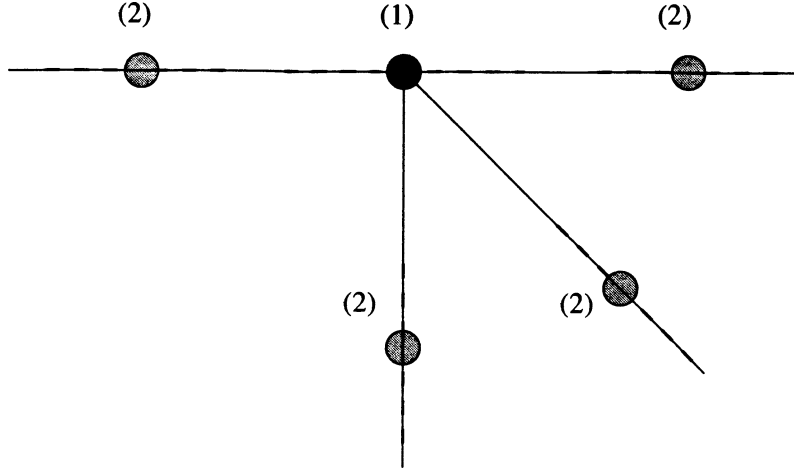
FIGURE 7 : AN EXAMPLE OF SPATIAL CONFIGURATION WITH  $k = 4$ 

Figure [7]. It is clear that the minimal distance between the agglomeration 1 and the other agglomerations must be greater than the extent of the agricultural areas (which must of course not overlap). In this case it is easy to check that

$$d_{\min}(A_1, A_2) = r_1^*(N_1^*) + r_2^*(N_1^*) = 2.3282 .$$

Hence

$$d(A_1, A_2) \geq 2.3282$$

must hold. Let us suppose that the agricultural surplus imports for agglomeration 1 are given by

$$\Delta S_A = \sum_{i=1}^k S_{A_i} e^{-\tau d(A_1, A_i)} = k S_{A_2} e^{-\tau d(A_1, A_2)} \quad (6.1)$$

since there are  $k$  identical agglomerations of type 2 located at the same distance  $d$  (see Figure [7]). In order for the “new”  $M$ -workers in agglomeration 1 to achieve the same utility level as the “old” workers, they must consume  $x_1^*$  units of the imported  $A$ -good. Hence

$$\Delta n^* = \frac{\Delta S_A}{x_1^*} \frac{1 - \rho}{f} \quad (6.2)$$

additional  $M$ -firms can be supported. This leads to a new total quantity of  $M$ -goods produced, which is given by

$$\bar{Q}^* = (n^* + \Delta n^*)^{\frac{1}{\rho}} \frac{f \rho}{m(1 - \rho)} \quad (6.3).$$

Since  $\rho < 1$ , increasing returns to scale are at work and the increase in the quantity produced will be stronger the smaller  $\rho$  is. Since the “new” workers must achieve the same utility as the “old” ones, they will consume a quantity

$$q_c = \frac{\Delta S_A}{x_1^*} x_2^*$$

of this  $M$ -good production. Hence the net increase in the local production of  $M$ -goods in agglomeration 1 is given by

$$\Delta q^* = \bar{Q}^* - Q^* - q_c \quad (6.4).$$

So far this is a straightforward reasoning. The two fundamental questions to be answered now are : is  $\Delta q^*$  non-negative (in order for this adjustment to be rational) ? is  $\Delta q^*$  sufficiently large in order for the agglomerations of type 2 to be compensated ? As we will see, those questions are difficult to investigate, even if several interesting results can be highlighted.

Let us start by examining the imports in agricultural surplus, given by equation (6.1). Since this expression can be seen as an attraction-accessibility measure, it is immediate that  $\Delta S_A$  is increasing with the *economic potential*  $S_{A_2}$  (as with the number  $k$  of rural agglomerations of type 2 in the region) and decreasing with both transport costs  $\tau$  and distance  $d$  (which should be seen as a measure of *accessibility*). **The higher the region's potential and the better the accessibility, the higher the agricultural surplus that can be collected in order to exploit economies of scale.** This is a very straightforward and important result with very strong historical pertinence as illustrated by the following remark.

**Remark 6.2** The emergence of the real urban phenomenon requires not only agricultural surplus but *exchangeable agricultural surplus*. Several historical examples show that agricultural surplus is a necessary but not a sufficient condition for urban agglomerations to grow and develop.  $M$ -goods must have a sufficiently high value and transport technologies must attain a certain level of development. The case of Jericho, as explained in BAIROCH P., [1] highlights this situation : “*Cette cité avait certainement une activité artisanale, attestée par des restes de poteries et autres articles manufacturés; mais, bien sûr, l'importance de celle-ci reste impossible à déterminer. En outre, et ceci est très important, dès -7000 environ, on est en présence d'une agriculture avancée avec irrigation et domestication des animaux. Mais village fortifié ou véritable ville ? La question reste posée; et le déclin ultérieur réduit le rôle de Jericho en tant que facteur d'urbanisation. Cette “défaillance” de Jericho peut s'expliquer par le caractère spécifique de la région immédiate de cette cité qui est, en quelque sorte, une grande oasis fertilisée par des sources locales, oasis à l'intérieur d'une région à caractère semi-désertique*”. In this case the local development had taken place and given rise to the emergence of a locally supplied  $M$ -good industry. Nevertheless, an isolated geographical position in a semi-desertic region, synonymous with bad accessibility and low economic potential prevented the emergence of a real urban agglomeration, collecting agricultural surplus and distributing manufactured goods.

By replacing all values by their values at  $N_1^*$  we can show that

$$\frac{\partial \Delta q^*}{\partial \Delta S_A} = \frac{1}{c} \left[ \frac{1-\rho}{\rho f} \left( n^* + \frac{k S_{A_2} e^{-\tau d(A_1, A_2)}}{c} \frac{1-\rho}{f} \right)^{\frac{1-\rho}{\rho}} - \frac{(1-\mu)\rho}{m} n^{*\frac{1-\rho}{\rho}} \right]$$

which shows that the quantity produced increases with the *A-good* imports if  $\rho$  is not too close to 1 (in which case scale economies are so small that the transport cost diseconomies decrease total net output available). This highlights the fact that **a certain degree of specialization and diversification (which goes hand in hand with increasing strength in the economies of scale) must be achieved before urban structures can be developed.**

Let us look at some concrete numerical examples right now. We will show under what conditions adjustments can be realized through *good flows* (the situation leading to economic development) and under which conditions adjustments will be made through *factor flows* (which will make even worse the current situation). We will illustrate the main results obtained by the means of two different examples. In the first one, we fix the number of agglomerations of type 2 to  $k = 4$  and investigate for variable levels of transport costs  $\tau$  the maximal distance at which those agglomerations can be located in order for the compensation mechanism with goods to work. In the second example, we fix the distance  $d$  at 2.5 and investigate how many agglomerations of rural nature are needed in order for the compensation mechanism to work for different levels of transport cost.

Let us start with the case of very high transport costs and consider that  $\tau = 1$ . With  $k = 4$  no compensation by goods can be realized since the maximal distance for the agglomerations of type 2 from the agglomeration 1 would be inferior to  $d_{\min}$  (which is impossible since the agricultural areas overlap). Therefore adjustment will have to take place by factor movements, which will result in the concentration of all the *A-workers* in the agglomeration 1 which will hit its malthusian upper bound. This situation can be avoided by restrictive policies of a local government but no adjustment is possible in this case. As transport costs decrease to  $\tau = 0.5$ , adjustment is still impossible since  $d_{\max} = 2.2 < d_{\min}$ . As  $\tau$  hits the value 0.435, adjustment becomes feasible : the surplus imported from the agglomerations of type 2 is large enough in order to support an additional *M-worker* population in the agglomeration 1 which

- achieves the same highest utility than all other agents in the agglomeration 1
- produces enough additional *M-goods* so that all agents in agglomerations of type 2 achieve also this same highest utility level

Hence the workers in the agglomerations of type 2 achieve the same utility as in the agglomeration 1 and therefore no longer have any incentive to migrate to agglomeration 1. As transport costs continue to decrease to  $\tau = 0.25$ , there is even an additional *M-good* surplus of 6.5 left at distance 2.5, which will allow all agents of the economy to be better off jointly than when all agglomerations act individually in isolation. In this case the maximal feasible distance for the agglomerations of type 2 is of  $d_{\max} = 4.35$ .

Consider now the second scenario where  $k$  is variable and where the distance  $d$  is fixed at 2.5. In the high transport cost scenario where  $\tau = 1$ , the astronomic number  $k = 80$  is needed for adjustment. As  $\tau$  decreases to 0.5, we still need 6 agglomerations of type 2 at a distance of 2.5 in order for the compensation through goods to be realizable. The case where  $\tau = 0.435$  corresponds of course to 4 agglomerations of type 2, as seen in the first scenario. Finally, when  $\tau = 0.25$ , a single agglomeration of type 2 is needed in order for adjustment through goods to be feasible.

Let us summarize these two scenarios as follows : when economic potential is low (i.e. agricultural surplus is low) and when accessibility is bad (i.e. transport costs are high and/or distances are great), no adjustment through good flows is possible. In this case we have either a) unhealthy growth of the *leader* agglomeration or b) a constant disequilibrium where a local government restricts entry to the *leader* agglomeration. This scenario still applies to most of today's developing countries, as highlighted in the following remark.

**Remark 6.3** Developing countries with low revenues are traditionally stuck in a certain dilemma. On the one hand, they need to increase their revenue by developing an autonomous and performant agriculture, but their demand for *A-goods* is relatively inelastic with respect to increases in revenue. On the other hand, they need an agricultural surplus (and hence a developed agriculture) in order to realize the self-sufficient provision of their economic agglomerations, which are needed in order to produce the additional *M-goods* demanded in response to the increase in revenue. So what happens at a first stage, as response to increasing revenues, is an increase in the imports of western countries' agricultural surplus in order to provide the cities with food instead of an effort of developing the agricultural sector. This leads to a distortion of revenue distribution and unhealthy agglomeration mechanisms : cities generate revenues which outweigh increases in revenues in the country-side, thus attracting more population; this in turn does not increase agricultural output and enhances those countries' dependency on western countries' agricultural surplus. The "natural" compensation mechanisms, that is to say factor and goods movements, are cut-short by the imports and the cities are artificially kept alive and can grow beyond the "point of no return". Although we will not enter a normative debate on over-concentration, we believe that unhealthy agglomeration mechanisms are at work in developing countries.

When economic potential is high (i.e. agricultural surplus is high due to increases in productivity) and/or when accessibility is good (i.e. transport costs are low and/or distances are small), adjustment through good flows is possible. In this case we observe the emergence of real *urban central places* that a) can't provide themselves with the foodstuffs they need and b) collect agricultural surplus and provide their region with manufactured goods.

## 7 Conclusions

In this paper we have developed a two stage model of the formation of agglomerations in a pre-industrial setting. At a first and local stage, agglomerations grow exogenously and try to develop sufficient agricultural surplus in order to support a local full-time manufacturing activity. At a second and global stage, agglomerations interact with each other and try to *realize the potential gains from specialization and division of labor (rural-urban)*. As we have shown, the formation of urban agglomerations can only result from the second stage of the model; an urban center emerges, producing and distributing manufactured goods while collecting the other agglomerations' agricultural surplus. This strongly resembles the hierarchical models of central-place theory, where a higher-order city provides lower-order agglomerations with manufacturing goods and services it produces. The main result of our model is to highlight the importance of *real strict indivisibilities*, namely the fact that a critical population size is needed before any manufacturing activity can take place. The introduction of this barrier to the production of manufactured goods gives rise to interesting dynamics and

interpretations, as we have shown in the two last sections of our work. This model is a first attempt to demonstrate how *non-differentiable* and *discontinuous* phenomena play important roles in (spatial) economic theory. This aspect has clearly been neglected until now, which is some kind of curious since we definitively do not live in a smooth world. We believe that many economic phenomena require the explicit introduction of *non-differentiabilities* and *discontinuities* in order to capture structural modifications of basic economic mechanisms. Since those two aspects are harder to analyze from a technical and logical point of view, more work is called for here.

Possible extensions of our model include the analysis of coordination costs and different forms of spill-over effects between the agricultural and the non-agricultural sector. A major part of our future work will be concerned with the analysis of the diffusion of the production of the non-agricultural good throughout the city system. How does the technological level influence the production structure of the system? Does increasing agricultural productivity strengthen or weaken polarization of the leader-follower structure? Can we endogenously model the diffusion of the non-agricultural production?

Despite the results obtained, some major limitations of our model need to be overcome in the future. First, this model is aimed at explaining urban development in a pre-industrial setting. It is clear that any application to more modern periods is not desirable, especially since the role of agriculture has diminished a lot these last 200 years. The absence of an urban land-market could also be criticized, although this hypothesis is often used in order to simplify complex issues in city-system models. The assumption of collective production, although pertinent in certain circumstances of early urbanization as we have mentioned, needs to be relaxed in the future. The major problem in this case is the need for another logical controlling device concerning the sizes of the agglomerations. We do not believe that the fiction of pure self-organization is very useful in order to understand the evolution of urban structures. Finally, some will argue that our model relies too heavily on numerical examples. While this is true to some extent, we remind the reader that our approach deals with *very strong* non-convexities, which render analytical investigation extremely difficult. Despite the relatively simple analytical expressions of most of the important functions, our model includes *non-differentiabilities*, *discontinuities*, *increasing returns to scale*, *fixed city locations* and a *constrained consumer problem*. Obtaining clear-cut analytical equilibrium results in such a spatial-sequential setting seems to be, if not impossible, at least very complicated.



## Appendix A : Market clearing proof

In this appendix we will show that equation (3.11) holds for all values of  $N$ . Hence the markets for  $M$ -goods clear for all values of  $N$ .

PROOF. We have to check that

$$Q^* = (N + n^*l^*)x_2^*(N) \quad (7.1)$$

for the cases of high and low revenue. Let us check the case where  $w(N) \geq c/\mu$  (we are above the *inf M-breakpoint*). We know by equation (3.10) that

$$Q^* = n^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)}$$

which can be written as

$$Q^* = (n^*)^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \left( \frac{(1-\mu)w(N)}{(1-\mu)w(N)} \right)$$

for the equilibrium value of  $n$ . Since  $w(N) \geq c/\mu$  we know by equation (3.2) that  $x_1^* = \mu w(N)$  so that

$$Q^* = (n^*)^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \left( \frac{(1-\mu)w(N)}{w(N) - x_1^*} \right)$$

and since  $x_2^* = (1-\mu)w(N)P^{-1}$  we have

$$Q^* = (n^*)^{\frac{1}{\rho}} \frac{f\rho}{m(1-\rho)} \left( \frac{Px_2^*}{w(N) - x_1^*} \right)$$

which simplifies to

$$Q^* = n^* \frac{f}{(1-\rho)} w(N) \left( \frac{x_2^*}{w(N) - x_1^*} \right)$$

when replacing  $P$  by its equilibrium value  $P^*$ . Since  $f/(1-\rho) = l^*$  and  $w(N) - x_1^* = SA/N$  we have

$$Q^* = L^*w(N)x_2^* \frac{N}{SA}$$

by using  $n^*l^* = L^*$ . Finally, by using the definition of  $SA$  we get

$$Q^* = L^* \left( \frac{SA}{N} + x_1^* \right) x_2^* \frac{N}{SA} = \left( L^* + L^*x_1^* \frac{N}{SA} \right) x_2^* .$$

Since  $L^*x_1^* = SA$  must hold if all the agricultural surplus is used we obtain finally the market clearing condition

$$Q = (L^* + N)x_2^* .$$

A strictly similar development can be applied to the case where  $w(N) < c/\mu$  in order to show that the market for  $M$ -goods clears for all values of  $N$  (one just has to take the appropriate expressions for  $x_1^*$  and  $x_2^*$ ).  $\square$

---

## References

---

- [ 1 ] BAIROCH P., De Jericho à Mexico – Villes et économie dans l’histoire, COLLECTION ARCADES, *Gallimard, Paris*, 1985
- [ 2 ] BECKER R., HENDERSON V., Intra-industry specialization and urban development (in “The Spatial Economy”), HURIOT J.-M., THISSE J.-F., EDS., *Cambridge University Press, Cambridge*, 2000
- [ 3 ] BECKMANN M., PUU T., Spatial Structures – Advances in Spatial and Network Economics, *Springer Verlag, Berlin*, 1989
- [ 4 ] BECKMANN M., Economic growth in a central place system in “Urban agglomeration and economic growth, GIERSCH H., EDS., *Springer Verlag, Berlin*, 1995
- [ 5 ] DUDLEY L., The Word and the Sword – How techniques of information and violence have shaped the world, STUDIES IN SOCIAL DISCONTINUITY SERIES, *Blackwell Publishers, Cambridge*, 1991
- [ 6 ] DURANTON G., Labor specialization, transport costs and city size , REGIONAL SCIENCE & URBAN ECONOMICS, *Blackwell Publishers, Cambridge, Massachusetts*, 1998
- [ 7 ] FUJITA M., KRUGMAN P., VENABLES A., The Spatial Economy – Cities, regions and international trade, *MIT Press, Massachusetts*, 1999
- [ 8 ] FUJITA M., THISSE J.-F., Economics of Agglomeration (forthcoming), *Cambridge University Press, Cambridge*, 2001
- [ 9 ] FUJITA M., THISSE J.-F., Agglomeration and market interaction , CORE WORKING PAPER, *Core, Louvain-la-Neuve, Belgium*, 2000
- [10] HENDERSON V., Urban development – Theory, fact and illusion, OXFORD UNIVERSITY PRESS, *Oxford, England*, 1988
- [11] HIRRIART-URRUTY J.-B., LEMARÉCHAL C., Convex analysis and minimization algorithms – Volume I and II, FUNDAMENTALS OF PURE AND APPLIED MATHEMATICS NO 305 – 306, *Springer Verlag, Berlin*, 1996
- [12] HOHENBERG P., LEES L., La formation de l’Europe urbaine 1000-1500 , COLLECTION HISTOIRES, *Press Universitaires de France, Paris*, 1992
- [13] HURIOT J.-M., Von Thünen , COLLECTION ECONOMIE, *Economica, Paris*, 1994

- [14] KALDOR N., The irrelevance of equilibrium economics , *ECONOMIC JOURNAL*, vol 82, pp. 1237–1255, London, 1972
- [15] KRUGMAN P., Geography and trade , *GASTON EYSKENS LECTURE SERIES*, *Leuven University Press & MIT Press, Cambridge & London*, 1991
- [16] LATOUCHE R., Les origines de l'économie occidentale , *SÉRIE L'ÉVOLUTION DE L'HUMANITÉ*, *Albin Michel, Paris*, 1970
- [17] MOKYR J., The lever of riches – Technological creativity and economic progress, *Oxford University Press, New York*, 1990
- [18] MOKYR J., Urbanization, Technological Progress and Economic History in “Urban Agglomerations and Economic Growth”, *GIERSCH H., ED., Springer Verlag, Berlin*, 1995
- [19] NORTH D., The Rise of the Western World (in “Political competition, innovation and growth”), *BERNHOLZ P., STREIT M., VAUBEL R., EDS., Springer Verlag, Berlin*, 1998
- [20] NORTH D., THOMAS R., The rise of the Western World – A new economic history, *Cambridge University Press, Cambridge*, 1973
- [21] NORTH D., Institutions, institutional change and economic performance , *THE POLITICAL ECONOMY OF INSTITUTIONS AND DECISIONS*, *Cambridge University Press, Cambridge*, 1990
- [22] OTTAVIANO G., TABUCHI T., THISSE J.-F., Agglomeration and trade revisited (Forthcoming in *International Economic Review*), 2001
- [23] OTTAVIANO G., THISSE J.-F., On Economic Geography in Economic Theory : increasing returns and pecuniary externalities (Forthcoming in *Journal of Economic Geography*), 2001
- [24] PAGE S., On the emergence of cities , *JOURNAL OF URBAN ECONOMICS* 45, 184 – 208, *Academic Press, New York*, 1999
- [25] SOUTHALL A., The City in Time and Space , *Cambridge University Press, Cambridge, UK*, 1998
- [26] TABUCHI T., Urban Agglomeration and Dispersion : A Synthesis of Alonso and Krugman , *JOURNAL OF URBAN ECONOMICS* 44, 333 – 351, *Academic Press, New York*, 1998
- [27] WEIBULL J., An axiomatic approach to the measure of accessibility , *REGIONAL SCIENCE AND URBAN ECONOMICS*, *Elsevier Science B.V., North-Holland*, 1976
- [28] ZANGWILL W., Nonlinear programming – a unified approach, *INTERNATIONAL SERIES IN MANAGEMENT*, *Prentice Hall, Englewood Cliffs, NJ*, 1969