J. Barkley Rosser

Complex Evolutionary Dynamics in Urban-Regional and Ecologic-Economic Systems

From Catastrophe to Chaos and Beyond

Springer

Complex Evolutionary Dynamics in Urban-Regional and Ecologic-Economic Systems

J. Barkley Rosser

Complex Evolutionary Dynamics in Urban-Regional and Ecologic-Economic Systems

From Catastrophe to Chaos and Beyond

Springer

J. Barkley Rosser
James Madison University
Department of Economics
South Main St. 800
Harrisonburg, VA 22807, USA
rosserjb@jmu.edu

ISBN 978-1-4419-8827-0 e-ISBN 978-1-4419-8828-7
DOI 10.1007/978-1-4419-8828-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011928060

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This is, therefore, the intensest rendezvous.
It is in that thought that we collect ourselves,
Out of all the indifferences, into one thing:

Within a single thing, a single shawl
Wrapped tight round us, since we are poor, a warmth,
A light, a power, the miraculous influence.

Here now, we forget each other and ourselves.
We feel the obscurity of an order, a whole...
Wallace Stevens, 1950, Final Soliloquy of the Interior Paramour

Within the past decade we have seen on the one hand the majority of the world's population come to reside in urban areas for the first time in world history, while on the other we have seen a dramatic rise in average global temperature widely thought to be mostly due to the economic activities humans carry out in those urban areas. Nonlinear complex dynamics are profoundly involved in both of these related developments. This book will open with considering the historical forces behind this rise of cities, including some critiques of certain recent ideas of "new economic geography," and will conclude (except for a mathematical appendix) with a consideration of the science and economics of global warming and how to deal with it in the face of the uncertainties arising from these nonlinear complex dynamics, which will reflect the unique perspective of this author having been involved with climatological research for more than 35 years. In between there will be discussions of broader regional economic dynamics, the foundations of evolutionary theory, and the dynamics of ecologic-economic systems. A deep theme is that in all of these areas, nonlinear complex dynamics can result in discontinuities, such as sudden changes of urban growth patterns or the sudden crashes of biological populations harvested by human beings. These discontinuities can appear with little warning and can serve as the basis for the theories of unexpected events such as "black swans" as proposed by Nassim Taleb (2010).

This book has been many years in the writing, having been contracted for originally over a decade ago by Kluwer, which was bought out some years later by Springer, who is publishing it now. It was originally supposed to be the second volume of the second edition of my 1991 book, *From Catastrophe to Chaos: A General*

Theory of Economic Discontinuities, published by Kluwer. That book was rejected by 13 publishers before Zack Rolnick at Kluwer took it up, after which it did quite well for such a monograph, leading to the contract for a second edition. Originally written as a critique of standard economic theory, after it came out it came to be viewed by many as a kind of reference work for nonlinear dynamics in economics. It was with this in mind that it was decided to break up the second edition into three volumes, given that I was thinking in such terms. It was in this vein that the first volume of the second edition was written and published in 2000 by Kluwer, expanding on the material in the first eight chapters of the first edition. I wrote the Preface for that book on September 9, 1999, exactly 11 years ago from my writing this Preface on September 9, 2010.

As it is, this book does fulfill in many ways the original intention of being that second volume, expanding on Chapters 9–14 of the 1991 book, which in fact are reproduced essentially directly in this book as, respectively, Chapters 1, 3, 4, 6, 8, and 10. In turn, the final chapter, Appendix A, which serves as a Mathematical Appendix here, is the second chapter from the first volume of the second edition from 2000. The remaining chapters are brand new, with each generally expanding on the chapter that precedes it.

Despite this, for various reasons this book does not carry the originally planned title. This is really a result of how long it has taken me to complete it, having gotten distracted by other projects and responsibilities during most of the past decade. One reason is that there has been such an explosion of literature in the areas of interest of this book that making this a reference book for all this has simply become undoable, at least as originally envisioned. Thus, the topics of urban-regional and evolutionary-ecological economic systems have been approached by focusing more on specific matters of particular interest in connection with the deeper themes of the original book, without attempting to be fully comprehensive.

There is also the matter that most economists now understand much better and are more open to the nonlinear complex dynamics mathematics that was the focus of the 1991 and 2000 books. Thus, there is less need to emphasize in this book an exposition of the mathematical techniques involved. This reinforces the idea of shifting more to focus on the issues of the topics themselves, even as nonlinear and complex dynamics continue to be an ongoing theme, including the original focus on discontinuities in economic systems. In any case, the relegation of the purely mathematical chapter to an appendix at the end symbolizes this shift, which fits with the change of title for this book.

I wish to thank the following for giving either advice, insights, or materials or general support for this book: Ehsan Ahmed, Craig Allen, Tim Allen, Åke Andersson, Dann Arce, Brian Berry, Gian-Italo Bischi, Buz Brock, Dan Bromley, the late Reid Bryson, Ken Button, Ping Chen, Carl Chiarella, Paul Christensen, David Colander, Bob Costanza, James F. Crow, Herman Daly, Herbert Dawid, Dick Day, Dee Dechert, Christophe Deissenberg, Domenico Delli Gatti, Dimitrios Dendrinos, Peter Dorman, Debbie Dove, Bill Duddleston, Steve Durlauf, Gustav Feichtinger, Duncan Foley, Carl Folke, Jamie Galbraith, Mauro Gallegati, Laura Gardini, Herb Gintis, John Gowdy, Steve Guastello, Roger Guesnerie, Weihong

Huang, Don Hester, Geoff Hodgson, Buzz. Holling, Cars Hommes, John Horgan, Bob Horn, Jean-Charles Hourcade, Yannis Ioannides, Börje Johansson, Jim Kahn, Shashi Kant, Steve Keen, Ali Khan, Alan Kirman, Michael Kopel, Roger Koppl, Ingrid Kubin, Honggang Li, Akio Matsumoto, Pat Michaels, Phil Mirowski, Peter Nijkamp, Dick Norgaard, Charles Perrings, Jason Potts, Tönu Puu, Otto Rössler, Kenshi Sakai, Larry Samuelson, Bill Sandholm, Tom Schelling, Frank Schneider, Willi Semmler, Ajit Sinha, Gene Smolensky, Michael Sonis, Roger Stough, Nassim Taleb, Vela Velupillai, Alessandro Vercelli, Nick Vriend, Florian Wagener, Dave Warsh, Wolfgang Weidlich, Roy Weintraub, Martie Weitzman, Ulrich Witt, Bill Wood, and Wei-Bin Zhang. Of course, none are responsible for any errors or misinterpretations that are contained in it.

I wish especially to thank my publisher, Jon Gurstelle, as well as his assistant, Talia Winch, and others at Springer involved in the production of this book. It was Jon who finally pushed me to complete this book after my long gathering of materials and writing papers without putting it together to make the effort. I went through several publishers at both Kluwer and Springer after Zack Rolnick stepped aside and before Jon took over who did not push me, although I cannot complain of their friendliness or sympathy. However, it was Jon who forced me to finally choose whether to do this or not, and I chose to do it.

Finally, I wish to thank my wife and colleague, Marina Rosser, source of much deep and wise advice. She also exhibited the tolerance that a writer of a book must have from a spouse for the distractions and disruptions of normal life that the writer of a book necessarily generates for those around and close. She has been my strong support throughout the long period of this endeavor and its many ups and downs. I am deeply grateful to her for all this and more.

I dedicate this book to my three lovely and loving daughters: Meagan, Caitlin, and Sasha. More than any other book I have written, this one is concerned with the future, their future, and that of their children and their childrens' children, on and on, hopefully.

Harrisonburg, VA J. Barkley Rosser, Jr.
9/9/10

Contents

Chapter 1
Discontinuous Evolution of Urban Historical Forms

It was in permanent containers that neolithic invention outshone all earlier cultures: so well that we are still using many of their methods, materials, and forms. The modern city itself, for all its steel and glass, is still essentially an earth-bound Stone Age structure. . . . With storage came continuity as well as a surplus to draw on in lean seasons. The safe setting aside of unconsumed seeds for next year's sowing was the first step toward capital accumulation.
Lewis Mumford (1961, The City in History, p. 16)

As far as the present record stands, grain cultivation, the plow, the potter's wheel, the sailboat, the draw loom, copper metallurgy, abstract mathematics, exact astronomical observation, the calendar, writing and other modes of intelligible discourse in permanent form, all came into existence at roughly the same time, around 3000 B.C. give or take a few centuries. The most ancient urban remains now known, except Jericho, date from this period. This constituted a singular technological expansion of human power whose only parallel is the change that has taken place in our own time.
Lewis Mumford (ibid., p. 33)

1.1 Introduction

Perhaps more than any other human institution, the city reflects the tension between continuity and discontinuity in our social existence. It is simultaneously the vessel of continuity containing the deep structures of the past (symbolized by the longevity of many urban buildings, roads, and other infrastructure) as well as the staging ground for most of the revolutionary upheavals of human history. This conflict is manifested in the above pair of quotations from Mumford (1961).

The scope and variety of possible discontinuities in urban systems can be seen from the following quote from J.C. Amson (1975, p. 178):

> Specific illustrations of the structural singularities typical of urban systems spring readily to mind: the onset of preurban clustering in a primitive dispersed society; the tenfold enlargement of an already large city within one generation; the emergence 'almost overnight' of a

J.B. Rosser, *Complex Evolutionary Dynamics in Urban-Regional and Ecologic-Economic Systems*, DOI 10.1007/978-1-4419-8828-7_1,

suburban fringe to a provincial town; the relatively instantaneous depopulation of its inner core after centuries of lively inhabitation; the reversals of income patterns, social patterns, or ethnic zoning patterns in a metropolis; the counterreversal of that zoning half a century later; the exchange of dominance roles between rival cities; the spontaneous formation of a conurbation from its constituent towns; and many more.

Amson (1974, 1975) initiated the formal use of discontinuous techniques of analysis in urban economics with catastrophe theory by analyzing urban density as a function of rent and "opulence" (an index of the attractiveness of a city) within a cusp catastrophe context. Casti and Swain (1975) modeled central place orders and urban property prices as Zeeman's (1974) cusp catastrophes and as butterfly catastrophes. Mees (1975) modeled the revival of cities in medieval Europe as a butterfly catastrophe. Wilson (1976)[1] examined transportation mode choice as a fold catastrophe and Poston and Wilson (1977) did so for retail center size. Isard (1977) modeled population related to agglomerative and deglomerative tendencies as a cusp catastrophe, which prefigured more rigorous models by Casetti (1980), Dendrinos (1980a), Papageorgiou (1980), and Papageorgiou and Smith (1983) explaining sudden urban growth.[2] Smith (1977) used the fold catastrophe to model the decline in the number of banks post-1920s that occurred due to the spread of automobiles.[3] Wagstaff (1978) tried to explain Greek settlement patterns between the second and seventeenth centuries as a result of changes in agricultural land quality and "external threat" using a cusp catastrophe model.[4] Dendrinos (1978) modeled intraurban manufacturing and residential activities as a five-dimensional hyperbolic umbilic catastrophe, and in 1979, he modeled the formation of slums in cities with the six-dimensional parabolic ("mushroom") catastrophe. Puu (1979, 1981a, b, c) used the five-dimensional hyperbolic and elliptic umbilic catastrophes to model changes in regional trading patterns. Nijkamp and Reggiani (1988) have shown that a generalized optimal control model of nonlinear, dynamic spatial interaction can generate a catastrophe theoretic interpretation.

In contrast to the largely deterministic approach of catastrophe theory, the synergetics approach emphasizes self-organization of systems through nonequilibrium phase transitions arising from stochastic fluctuations near critical bifurcation points. Foreshadowed by Fuller (1975, 1979), this approach was developed by Nicolis and Prigogine (1977) and Haken (1977, 1983) and first suggested as applicable to urban and regional systems by Isard and Liossatos (1977). The latter explicitly link this approach to the more deterministic catastrophe theory approach through the common emphasis on systemic structural changes arising from transitions through bifurcation points.

The major application of these ideas in urban and regional modelling has been due to members of the "Brussels School," particularly Allen and Sanglier (1978, 1979, 1981), Sanglier and Allen (1989), and their associates under the direct influence of Ilya Prigogine, winner of the 1977 Nobel Prize in chemistry, who was himself located in Brussels. Also influenced by Forrester (1969), their efforts have generally focused on the construction of large-scale simulation models of urban and regional systems, incorporating stochastic processes, nonlinear feedback effects, and threshold levels for specific economic activities. Model simulations tend to exhibit dynamic self-organization with some cities growing and others declining as they pass upper or lower economic thresholds, respectively. Intraurban variations of this approach[5] have been used to model the development of specific cities in France (Pumain, Saint-Julien, and Sanders, 1987).

Yet another approach of a deterministic nature has been that of mathematical ecology, focusing on applying Lotka–Volterra equation systems of predator–prey cycles to model urban and regional cycles. Such an approach was initiated by Dendrinos (1980b) and Dendrinos and Mulally (1981) and summarized by Dendrinos with Mulally (1985). Sonis (1981, 1983) used it to analyze technological diffusion in regional spaces. Dendrinos and Mulally (1983) empirically tested such a model for many US cities and found it useful as did Dendrinos (1984) for US regional systems. Haag and Weidlich (1983) developed such a model for interregional migration which also contains elements of the synergetics approach. The onset of instability or structural change in such dynamic systems can be studied using bifurcation theory and catastrophe theory.

Furthermore the work of May (1976) and May and Oster (1976) has shown that Lotka–Volterra systems can generate chaotic dynamics given certain parameter values in the equations. This directly inspired analysis of chaotic dynamics in urban and regional systems. The first such application was an intraurban one of retail and residential systems attributed to Beaumont, Clarke, and Wilson (1981a). But most have focused on multiple-region models, either involving migration or population more generally (Rogerson, 1985; Day, Dasgupta, Datta, and Nugent, 1987; Dendrinos, 1982, 1985; Dendrinos and Sonis, 1987, 1989, 1990) or generalized interregional business cycles (White, 1985; Puu, 1987, 1989, 1990). However, White (1985) combines the original Beaumont, Clarke, and Wilson (1981a) intraurban retail model with a Brussels School synergetics approach to argue that chaotic fluctuations near a bifurcation point can engender the self-organization process. Thus, it would seem that the practitioners in the urban and regional fields have generally been more willing to see the links between different methods of discontinuous analysis rather than focusing on what distinguishes them from each other.

It may well be that urban and regional economics is a subdiscipline more open to the use and application of discontinuous techniques than others in economics. One reason may be the obvious existence of spatial discontinuities in land use, land values, social classes, and other such categories. Another may be the multidisciplinary links to geography, urban and regional planning, environmental science, political science, sociology, and other areas. This has led to the multidisciplinary fields of "urbanology" and "regional science," which while frequently using neoclassical methods can be argued to possess no central theory and thus to be more open to alternative methodologies. This lack of a clearly defined central theory has led on occasion to analyses entirely lacking any rigorous foundation and notable for their extreme ad hocracy and general uselessness. Notorious examples of this have been some of the more flagrantly off-the-wall and out-of-the-blue catastrophe theory applications.

Although it might seem that a defense of openness toward new and different approaches in the advance of knowledge should be unnecessary, nevertheless I shall do so by quoting the godfather of nonequilibrium phase transitions (which can also occur in the structure of knowledge itself), Ilya Prigogine and his coauthor Isabelle Stengers from their 1984 book, *Order Out of Chaos*, p. 206:

> We believe that models inspired by the concept of 'order through fluctuations' will help us with these questions and even permit us in some circumstances to give a more precise formulation to the complex interplay between the individual and collective aspects of behavior. From the physicist's point of view, this involves a distinction between states of the system in which all individual initiative is doomed to insignificance on the one hand, and on the other, bifurcation regions in which an individual, an idea, or a new behavior can upset the global state. Even in those regions, amplification obviously does not occur with just any individual, idea, or behavior, but only those that are 'dangerous'—that is, those that can exploit to their advantage the nonlinear relations guaranteeing the stability of the preceding regime. Thus we are led to conclude that the same nonlinearities may produce an order out of the chaos of elementary processes and still, under different circumstances, be responsible for the destruction of this same order, eventually producing a new coherence beyond another bifurcation.

1.2 Agglomeration and the Formation and Sudden Growth of Cities

1.2.1 The Debate

According to Braudel (1967, p. 373), "The town-country confrontation is the first and longest class struggle history has known." It is "the oldest and most revolutionary division of labor." (ibid.) The depth of this division, this contradiction, has inspired an array of thinkers and actors who wished to overcome it somehow: from the utopian socialist, Fourier (1840), the "scientific socialists," Marx and Engels (1848), and the anarchist, Kropotkin (1899), through the seminal "new town" advocate of garden cities and greenbelts, Ebenezer Howard (1902), and his modern disciples such as the Goodmans (1960), James Rouse, the developer of Columbia, Maryland, and Robert E. Simon, the developer of Reston, Virginia. We shall not seek how to achieve this goal, but rather to explain why this division exists.

Much debate exists over why cities exist. Are they merely central places of regional trading areas where markets occur, such as the German tradition of von Thünen (1826), Christaller (1933), and Lösch (1940) believes? Or do they arise, especially large cities, from agglomerative external economies such as argued by Adam Smith (1776, Part III),[6] Marshall (1890, Chap. 8), Hotelling (1929), and Chinitz (1961)? And why do we observe in history sudden changes in city sizes during particular periods such as the original "urban implosion," (Mumford, 1961), 5,000 years ago, or the sudden growth of cities noted in Europe by Pirenne (1925) after 1000 A.D.? We shall begin by discussing the case for agglomeration.

1.2.2 *Instability and Agglomeration*

1.2.2.1 A General View

Let us contemplate the clustering out of an initially uniform distribution of population as resulting from an instability in the balance of agglomerative and deglomerative forces. We view our units or agents as households rather than as individuals since the sexual and reproductive drives provide a fundamental source of agglomeration at this lowest level. But for a primitive agricultural or hunting society without trade, the self-sufficient household (possibly an extended family group or even clan) will be dominated by a deglomerative dynamic in its behavior, relative to other households. It will seek to maximize the amount of land available to itself for its farming or hunting activities. Thus on a homogeneous plane with identical households, there will tend to be an even distribution of those households as congestion costs (deglomerative force) predominate.

The origins of agglomeration beyond the household unit were probably not originally economic. Mumford (1961) argues that the first fixed gathering points were ancestral burial grounds of tribes or clans, often in caves which became the sites of religious ceremonies and artistic endeavors (i.e., Lascaux). "The city of the dead predates the city of the living" (Mumford, 1961, p. 7). But such agglomerations tended to be temporary rather than permanent, except for the corpses themselves (the "city of the dead").

Economic agglomeration came about from the discovery of the advantages of some collective labor in agriculture, most likely in the building and maintenance of irrigation and drainage systems. Such activities become the basis for small agricultural villages. But congestion costs remained high, especially due to public health problems associated with human excretion which are not totally relievable by recycling as fertilizer of "nightsoil."[7]

The likely critical element leading to a shift in the balance was probably the greater development of hydraulic systems in all their variety, as has been argued by Wittfogel (1957): not only the extensive building of dams, irrigation canals, and drainage ditches which allowed the expansion of population, but also the construction of water supply systems for direct human use, and, most important of all, for reducing congestion costs, *sewers* for removing human waste products. An old cliche goes, "sewers are the foundation of civilization." As Mumford notes (1961, p. 18), all these systems are ultimately containers, and so, "a city is a container of containers." Once this crucial barrier was passed, technological changes could occur more rapidly, which tended to increase the strength of the agglomerative advantage and the further expansion of cities.

1.2.2.2 A Local Instability Model

However, let us return to the initially even distribution of identical households on the homogeneous plane and consider a general model of the effects described above. In so doing, we shall initially follow the model of Papageorgiou and Smith (1983) in which the switch from deglomeration (or congestion costs) outweighing agglomeration effects (positive, locational, external economies) to just the opposite occurs at a bifurcation point where the dynamic system becomes unstable (Amson's "structural singularity").

They consider a homogeneous square plane, transformed into a torus, with M cells and N identical actors. Let n_1 = population in cell 1 and the externality generated by those in 1 to someone in cell j to be given by

$$E1_j = \phi_{1j} n_1, \tag{1.1}$$

where ϕ_{1j} is a distance-response function whose sign indicates whether a cell-to-cell relation is agglomerative (positive) or deglomerative (negative). Given the homogeneity of the plane, the signed "volume" of the spatial interaction field,

$$Z = \sum_t \phi_{tj}, \tag{1.2}$$

will be constant for all cells j. However, the total externality experienced by someone in j depends on the distribution of population and is given by

$$E_j = \sum_t \ _{tj}. \tag{1.3}$$

Agents are assumed to migrate stochastically, obeying a rational expectations utility maximization condition where utility of an agent at 1 is given by

$$u_1 = f[q_1, E_1(n)], \tag{1.4}$$

with $\partial f/\partial q_1 > 0$ and $\partial f/\partial E > 0$, and q_1 is land per capita consumed at 1. Assuming no transportation costs and letting u and n, respectively, represent vectors of the distributions over cells of utility and population, the probability of moving to cell j is given by

$$p_j = h_j[u,\ n] \ \text{ with } \sum_j p_j = 1 \text{ and } p_j \geq 0. \tag{1.5}$$

Letting $\mathrm{d}n_1/\mathrm{d}t$ be the expected net population change (assumed to be due purely to migration) in 1 at time t, a steady state will be given by $\mathrm{d}n_1/\mathrm{d}t = 0$ for all 1 and a spatial equilibrium by $u_1 = u$ for all 1. The latter occurs when

$$\mathrm{d}n_1/\mathrm{d}t = n_1 \left[N \Big/ \left(\sum_k n_k u_k \right) \right] (u_1 - \bar{u}). \tag{1.6}$$

It readily follows from this that a uniform distribution of population (u, n) will be such a steady-state equilibrium.

Stability of this equilibrium can be examined by considering perturbations given by $n = 0_1$ such that

$$\begin{bmatrix} \mathrm{d}n'_1/\mathrm{d}t \\ \vdots \\ \vdots \\ \mathrm{d}n'_M/\mathrm{d}t \end{bmatrix} = \begin{bmatrix} c_0 c_1 c_2 \cdots c_1 \\ c_1 c_0 c_1 \cdots c_2 \\ c_2 c_1 c_0 \cdots c_3 \\ \vdots \qquad\quad \vdots \\ c_1 c_2 c_3 \cdots c_0 \end{bmatrix} \begin{bmatrix} n'_1 \\ \vdots \\ \vdots \\ n'_M \end{bmatrix} \tag{1.7}$$

and

$$\sum_1 n'_1 = 0. \tag{1.8}$$

The elements of the C matrix are given by

$$c_0 = \alpha + \beta\phi_0 - (\alpha + \beta Z)/M, \tag{1.9}$$

$$c_k = \beta\phi_{M-k} - (\alpha + \beta Z)/M, \tag{1.10}$$

for all 1 and $k \geq 1$, where

$$\alpha = (n/u)\,(\partial f/\partial q_1)\,(\mathrm{d}q/\mathrm{d}n_1)|_n, \tag{1.11}$$

$$\beta = (n/u)\,(\partial f/\partial E_1)|_n. \tag{1.12}$$

The equilibrium will be stable if the real parts of all the eigenvalues of C, $\lambda_j < 0$, form the well-known Lyapunov theorem. Papageorgiou and Smith show that if $M \to \infty$ and $\phi_k > 0$ for only a finite number of cells (the spatial externality extends only over a limited area), then $\lambda_0 = 0$ and $\lambda_j = \alpha + \beta Z$ for all 1 and for $j \geq 1$. Thus, the condition for stability of the uniform distribution of population is $\alpha + \beta Z < 0$. This is equivalent to

$$(\partial f/\partial q_1)(\mathrm{d}q_1/\mathrm{d}n_1)|_n + (\partial f/\partial E_1)(\mathrm{d}E_1/\mathrm{d}n_1)|_n = \mathrm{d}u/\mathrm{d}n_1|_n < 0, \tag{1.13}$$

which admits a helpfully intuitive interpretation.

The first term is the marginal effect of increased congestion from greater population on utility and is always negative because $\partial f/\partial q_1 > 0$ and $\mathrm{d}q_1/\mathrm{d}n_1 < 0$ and thus is the congestion cost. The second term is the marginal effect on utility of the spatial externality whose sign depends on the sign of Z because $\partial f/\partial E_1 > 0$. Thus if the spatial externality is negative ($Z < 0$), the uniform distribution will be stable. However, for the distribution to become unstable, the spatial externality must become *sufficiently positive* to outweigh the negative congestion effect. The bifurcation point beyond which clustering will begin occurs where the two effects are evenly balanced. Papageorgiou and Smith see the positive agglomeration effect increasing as population rises and technology improves.

1.2.2.3 A Catastrophe Theory Interpretation

Although Papageorgiou and Smith did not do so, it is straightforward to generate a catastrophe theory interpretation of this model. The state variable could be the maximum rate of population growth in any cell. Control variables could be Z, the agglomeration variable, and the marginal cost (negative marginal utility) of congestion. The combination of these two variables for which the maximum eigenvalues of C equals zero traces out a bifurcation set at which discontinuities in the rate of maximum cell population growth would occur.

It can also be argued that this model provides a more solid foundation for the sort of model developed by Walter Isard (1977). His is of a cusp catastrophe surface for a single center with population as the state variable and parameters representing the agglomerative externality factor and the level of technology, respectively, as the control variables. The surface represents a social welfare maximum and the control variables are claimed to operate through their impact on the marginal productivity of labor, although the surface is not derived more explicitly from an underlying theoretical model. In Isard's formulation, the cusp catastrophe manifold comes from the canonical Thom (1972) equation:

$$W = -1/4x^4 + 1/2\alpha x^2 + \beta x + C, \tag{1.14}$$

where W is social welfare, x is population, α is the agglomerative parameter, β is the technological parameter, and C is a constant. Setting $\mathrm{d}W/\mathrm{d}x = 0$ for utility maximization gives the manifold

$$-x^3 + \alpha x + \beta = 0. \tag{1.15}$$

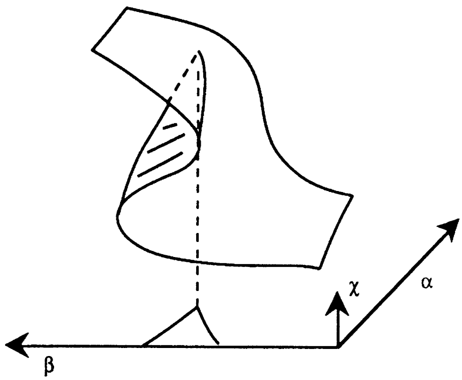
Isard interprets the first term as representing deglomerative diseconomies of scale which conflict with the agglomerative second term. Removing the third term would give something not all that different from what Papageorgiou and Smith derived more rigorously in a more disaggregated form. And the way the third term enters is not inconsistent with the verbal story Papageorgiou and Smith tell about the impact of technological change within their model. For a certain range of β, an increase in α (Z for Papageorgiou and Smith) will eventually lead to a discontinuous leap in population. This Isard cusp catastrophe manifold is depicted in Fig. 1.1.

It should be noted that a major difference between these approaches is that Papageorgiou and Smith emphasize migration between areas whereas Isard emphasizes production at a single site.

1.2.2.4 Some Further Variations

The Papageorgiou and Smith model has inspired several other efforts emphasizing the emergence of instability and the appearance of multiple equilibria. Anas (1988) has examined a variety of complex dynamics in a very similar model but shows that the presence of taste heterogeneity with respect to location can perform the same function as congestion costs do in the Papageorgiou and Smith model.

Fig. 1.1 Agglomerative instability as a cusp catastrophe



Weidlich and Haag (1987) present a three-region model with nonlinear migration equations driven by a single agglomeration parameter. They analyze their model in terms of equations of motion of the system, given by

$$\mathrm{d}n_1(t)/\mathrm{d}t = \mathrm{e}^{kn_1(t)} \sum_{i=1}^{L} kn_i{}^{(t)} \mathrm{e}^{-kn_i(t)} - kn_1(t)\mathrm{e}^{-kn_1(t)} \sum_{i=1}^{L} \mathrm{e}^{kn_i(t)}, \tag{1.16}$$

with the constant

$$\sum_{l=1}^{L} kn_1(t) = kN, \tag{1.17}$$

where $n_1(t)$ is the population in region 1 at time t, N is total population (assumed to be constant), k is the agglomeration parameter, and L is the number of regions (assumed to equal 3 in this case). They depict this dynamical system graphically with flux lines in a space of agglomeration-weighted, regional populations. They derive three separate cases for low, middle, and high values of k, the agglomeration parameter. These are depicted in Figs. 1.2, 1.3, and 1.4, in which each vertex represents a situation where all the population is in a single region and each side where one region has zero population.

In Fig. 1.2, $k < k_c{}^1$; the agglomeration parameter lies below the lower of the two bifurcation values of k. In this situation from any initial point the system moves to the center, representing our uniform distribution of population case, which happens to be a globally stable equilibrium.

In Fig. 1.3, $k_{c1} < k < k_c$; the agglomeration parameter lies between its two bifurcation values. In this situation the uniform distribution (center) is still locally stable but has lost its global stability. If the initial situation is sufficiently nonuniform, the system will converge on an equilibrium with agglomeration occurring (a city forms) in the nearest region to the initial point. In fact there are four basins of attraction

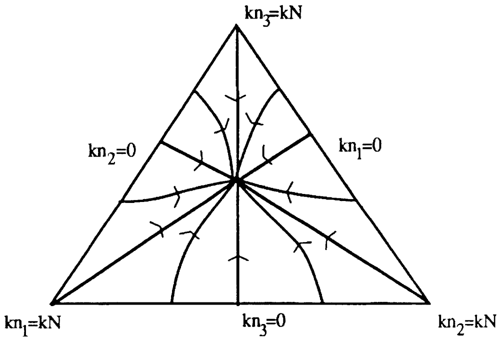


Fig. 1.2 Weidlich–Haag migration with weak agglomeration

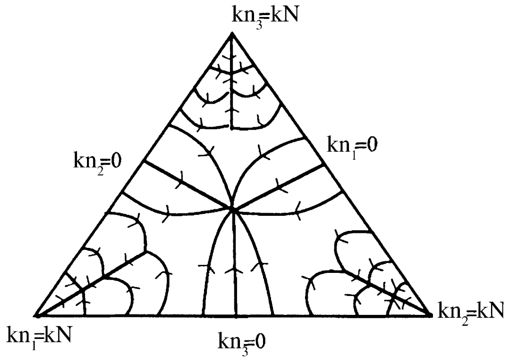


Fig. 1.3 Weidlich–Haag migration with intermediate agglomeration

around four, locally stable, stationary states, separated by lines intersecting at three unstable saddle points.

In Fig. 1.4, $k > k_c$; the agglomeration parameter now exceeds the upper bifurcation point. The uniform distribution has become an unstable node. In addition there are still three stable stationary states, each representing a "city" existing in a different region, and three unstable saddle points.

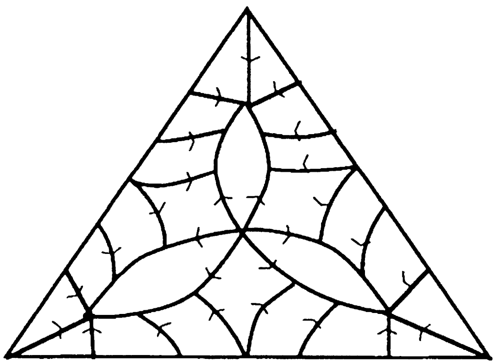


Fig. 1.4 Weidlich–Haag migration with strong agglomeration

We can directly compare this result with that of Papageorgiou and Smith. The upper bifurcation point in Weidlich and Haag corresponds to the condition given by Papageorgiou and Smith for the appearance of instability of the uniform distribution, and hence the condition for the original formation of cities. The lower bifurcation point may correspond with the value of Z, the locational externality in Papageorgiou and Smith, becoming positive. In their model, the uniform distribution remains locally stable, even when Z is positive, as long as marginal congestion costs remain greater. However, the Weidlich and Haag result suggests that if Z is high enough in the Papageorgiou and Smith model, one might find agglomerative behavior if the initial situation is sufficiently nonuniform (some sort of a city already exists), even though the uniform distribution is still locally stable.

Leonardi and Casti (1986) have examined a variation that includes moving costs in the form of "spatial discount factors." Assuming that $P_i =$ relative population in zone i, $P_i \geq 0$, $\Sigma i P_i = 1$, $Q_i =$ dwelling space in zone i, $f_{ij} =$ spatial discount factor associated with a move from zone i to zone j, $0 \leq f_{ij} \leq 1$, $\alpha_1, \alpha_2 =$ elasticity parameters, and $K =$ a nonnegative constant, the interzone migration equation is

$$P_i = Q_i\left(\sum_j f_{ij}P_j\right)^{\alpha} \Bigg/ \sum_k Q_k\left(\sum_j f_{kj}P_j\right)^{\alpha}, \tag{1.18}$$

where $\alpha = \alpha_1/\alpha_2$. This equation closely resembles equations studied by Beckmann (1976), Harris and Wilson (1978), and Andersson and Ferraro (1982). Although they do not make it at all clear, it would seem that α, the elasticity ratio, might be interpreted as an agglomerative parameter, and the f_{ij} "locational friction" parameters might also reflect congestion costs or other deglomerative effects. In any case they examine a dynamic generalization of this model in which the matrix Q is given by

$$q_{ij} = f_{ij}P_j \Bigg/ \sum_m f_{im}P_m. \tag{1.19}$$

The condition for a unique and stable equilibrium then becomes

$$\alpha < 1/\sup \mathrm{Re}\lambda_Q, \tag{1.20}$$

where the λ_Qs are the eigenvalues of Q. For α greater than this critical bifurcation value there will be multiple equilibria, stable ones alternating with unstable ones in the usual manner. They further argue that the smaller the locational friction parameters, the wider the zone of α for which there exists a unique and stable equilibrium. They conclude that this contradicts Papageorgiou and Smith, but this seems unclear due to basic differences between their respective models.

1.2.2.5 The Role of Production Made Explicit

An approach that combines elements of the Isard approach with those of a model of Casetti (1980) was developed by Papageorgiou (1980, 1981). Nonlinear productive returns to city size are modeled through their impact on the marginal product of labor. Migration depends on a comparison of equilibrium urban utility (u) with equilibrium rural utility (v) where the latter has linearity of that relation. If N is urban population, X is the urban good, and the ms are positive constants derived from various input costs (Papageorgiou, 1980, pp. 1039–1040), then the change in equilibrium urban utility with urban population is given by

$$dU/dN = -m_1 + m_2[d/dN(\partial X/\partial N)]. \tag{1.21}$$

The term $d/dN(\partial X/\partial N)$ describes how the marginal product of labor used in the urban good changes with city size. It is assumed that it increases initially at an increasing rate and later at a decreasing rate. For a sufficiently large m_1, this gives relations as shown in Fig. 1.5.

It is then posited that over time technological change will raise this u more rapidly than it raises the equivalent v for agriculture. In a world of constant population, this can be seen in Fig. 1.6.

In this system, urban population leaps discontinuously from N_3 to N_4. There is a hysteresis effect here in that if v were to rise more rapidly (or u were to decline), there would be a discontinuous drop from some urban population considerably below N_4 down to some level considerably below N_3. For u_1 and u_2, there are three equilibrium states with the outer two being stable and the middle one unstable.[8]

Despite its slightly different formulation, this model reflects the basic characteristics we have already seen. The nonlinear utility–population relation embodies the conflicting forces of agglomeration and congestion with the former "winning" at lower population levels and the latter winning at higher ones.[9] Changes in this "net

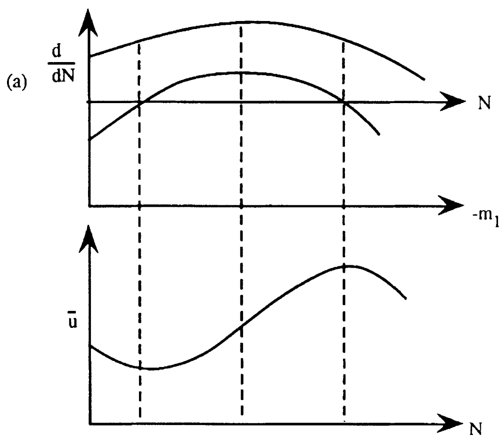


Fig. 1.5 Marginal and total urban utility as a function of population

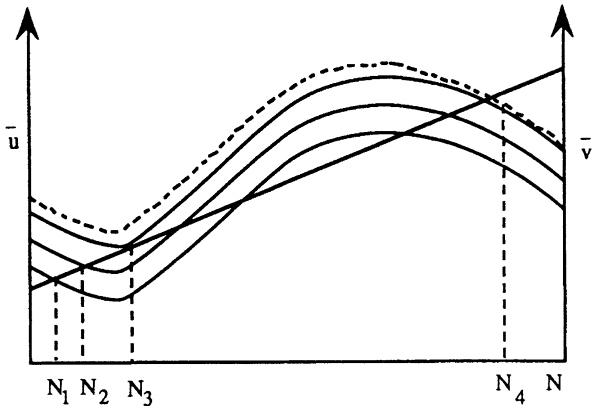


Fig. 1.6 Urban–Rural population distributions equilibrium

agglomerative tendency" are associated with changes in stability and number of equilibria, with discontinuous changes in equilibrium urban populations occurring as the tendency crosses certain critical levels. Thus, sudden changes in urbanization arise from interactions between agglomeration and technology in nonlinear systems.

1.3 Long-Distance Trade and Instability

1.3.1 Another View: Open Versus Closed Cities

An alternative to agglomeration as the driving force of urban formation and sudden growth is the expansion of long-distance trade. The German tradition view of cities as central places of trading areas naturally fits in with the idea that as the trading area expands, so does the city serving as its central place, irrespective of the presence or absence of agglomerative economies per se. In particular, the development of large cities is associated with the expansion of *long-distance trade*. It is Pirenne (1925) who is most frequently cited in this regard, due to his arguments linking the reemergence of long-distance trade and the relatively sudden expansion of medieval cities in Europe after 1000 A.D. But the argument has deeper roots than that.

Let us return to the question of the original formation of cities out of a uniform distribution, rather than the later question of the expansion of medieval European cities. Braudel (1973) distinguishes between "open" cities and "closed" cities. This distinction does not depend on the presence or absence of a wall, although a closed city is more likely to have one than an open city. An open city is one that is intimately connected economically to its immediately surrounding hinterland, archetypally, the self-sufficient neolithic village, recycling its night soil. The closed city is separated from its immediate hinterland and depends largely upon long-distance trade to sustain itself, archetypally, Renaissance Venice (also Singapore and Hong Kong). For Braudel, it is the existence of long-distance trade that allows for the separation of town and country and the "closure" of cities from their immediate hinterlands.

V. Gordon Childe (1942) argued that it was the ability to rely on long-distance trade that allowed ancient cities in Egypt and Mesopotamia to *dominate* their hinterlands in an exploitative manner, from the palace, the temple, and the granary, hidden within the walled citadel. Thus, it was not the economic externality of agglomeration, but the surplus accumulated from long-distance trade that allowed for the initial formation and growth of cities.[10] Braudel clearly sympathizes with this view in his remark, "[T]he town-country confrontation is the longest, running class struggle history has known."

Braudel refined this view and linked it explicitly to the German tradition through his vision of the hierarchy of central places. At the bottom are (open) agricultural villages, then the small town with its subordinate villages, then the provincial city with its ring of small towns, and then the national-imperial center, dominating the provinces and trading with the world at large.[11] Braudel (1984) eventually expanded this to a view of the entire world economy as a "set of sets," a series of such hierarchies, dominated by an ultimate central core (the "leading world city") with an ultimate periphery in the poorest and most isolated areas, a view further developed by Wallerstein (1979). Oscillations between core and periphery at this scale became for him the source of the longest of economic cycles, "la longue durée" (or long-term secular trend), lasting hundreds of years, an idea further discussed by Goldstein (1988). Without denying the association between large cities and long-distance trade, the advocates of the agglomerative perspective argue that the growth of long-distance trade followed the initial development of urban areas based on local agglomeration. Thus Mumford (1961) criticizes Pirenne explicitly for his theory of long-distance trade as the ultimate source of the growth of medieval European cities. Instead Mumford saw peace and protection of *local* markets afforded by the Church as the basis of medieval European, urban development.

> But note: the regular market, held once or sometimes twice a week under the protection of Bishop or Abbott, was an instrument of local life, not of international trade. So it should be no surprise that as early as 833, when long-distance trade was mostly in abeyance, Lewis the Pious in Germany gave a monastery permission to erect a mint for a market already in existence. The revival of trade in the eleventh century, then, was not the critical event that laid the foundations of the new medieval type of city: as I have shown, many new urban foundations antedate that fact, and more evidence could be added. Commercial zeal was rather a symptom of a more inclusive revival that was taking place in Western civilization; and that was partly a mark of the new sense of security that the walled town itself had helped to bring into existence. (Mumford, 1961, p. 254)

Of course an obvious response by Pirenne and his allies might be to agree that agglomeration laid a necessary foundation but that it was still the appearance of long-distance trade that was the stimulus and trigger of the actual urban growth.

1.3.2 The Mees Version of Pirenne's Hypothesis

A major effort to model the Pirenne hypothesis is due to Mees (1975). We begin with open, small towns in a localized landscape, after the Moorish and Viking invasions had broken apart the long-distance trading patterns of Western Europe. For a given region, population, P, is initially assumed to be constant and divided between farming, p_f, and towns, p_g. U_f and U_g are country and town utility levels, respectively, and t is the long-distance transportation cost. Above a certain level of t, there will be no long-distance trade, but as t declines below that level, such trade will appear and expand. Demographic dynamics depend on utility-maximizing migration, given by

$$\mathrm{d}p_f/\mathrm{d}t = p_f p_g \left(U_f - U_g\right) = -\mathrm{d}p_g/\mathrm{d}t. \tag{1.22}$$

Mees shows that when t is high and there is no trade, these dynamics can be depicted as in Fig. 1.7, where the only stable equilibrium is E_m, representing a mixture of town and country. He argues that as t declines, this curve shifts and may end up either as Fig. 1.8 or as Fig. 1.9, depending on the overall population density, the average productivity of the region, and the town–country productivity difference.

In Fig. 1.8, a region with high population density and a high town–country productivity difference has E_g as its only stable equilibrium and has completely specialized in urban production for long-distance trade.

In Fig. 1.9, just the opposite occurs because E_f is the only stable equilibrium and the region specializes in farming for long-distance trade. If the average level of

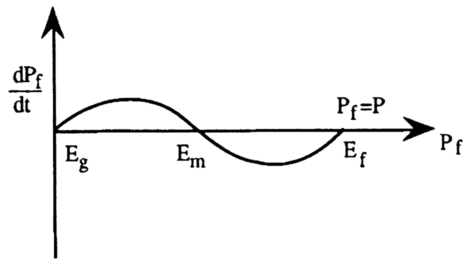


Fig. 1.7 No long-distance trade urban–rural equilibrium

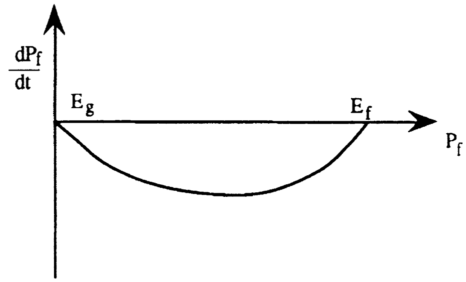


Fig. 1.8 Long-distance trading urban center

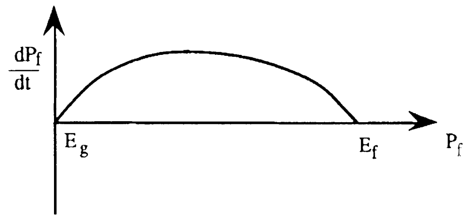


Fig. 1.9 Long-distance trading rural area

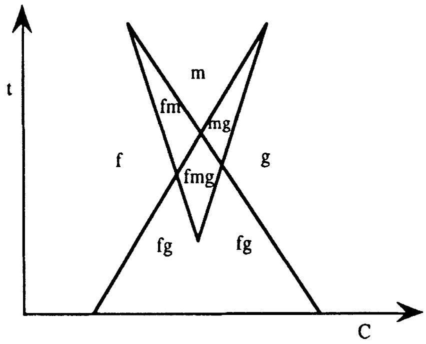


Fig. 1.10 Butterfly bifurcation set for long-distance trade model

productivity is high enough, and the other variables are at intermediate levels, then any of the three kinds of equilibria may be viable.

Mees argues that these results can be depicted in a Thom-type (1972) butterfly catastrophe whose four control variables form a four-dimensional bifurcation set. A projection of this set into the two-dimensional t and C space (where C is population density) is depicted in Fig. 1.10, with the possible equilibria, f, m, g, depicted for each zone of the space. The average productivity, q, produces the butterfly effect in which the triple equilibria occur. The sectoral productivity difference, Δq, is the bias factor that determines the tilt of the butterfly zone one way or the other.

Thus when t is high, the equilibrium will be m. But as t declines, it could become f or g, depending on C and the other factors, or even just remain mixed in the butterfly portion. Exogenous population change, by changing C, could trigger such shifts. In any case, it is the sudden shift of many local areas from all being mixed (small open towns in localized farming areas) to either purely urban (g) or purely rural (f); the decline in t represents the implicit dynamics of Pirenne's hypothesis.

1.3.3 Comparative Advantage and City Size

That there exist thresholds in the trading relations with the rest of the world which can be associated with sudden shifts in city size may reflect changes in the city's comparative advantage situation. Dendrinos (1980a) developed a model of such discontinuous city size changes arising from continuous changes in comparative advantage.

Given a production function and a position in the larger economy, the maximum utility level associated with different amounts of labor (and hence city sizes) is given by

$$V^* = v(L^{\mathrm{d}}), \tag{1.23}$$

where the inverse of v can be interpreted as the demand for labor, L^{d}. This V^* may be nonlinear, due to economies and diseconomies of urban scale (possibly reflecting variations in agglomerative effects), and even characterized by multiple local maxima.

Labor supply depends on factor rewards (v^{s}) and the level of comparative advantage of the city denoted by s. Thus L^{s} is labor supply

$$L^{\mathrm{s}} = S^{-1}(s, V^{\mathrm{s}}), \tag{1.24}$$

with S^{-1} smooth, twice differentiable, and possessing a smooth inverse. For a given s and L, there will be associated V^* and V^{s}. If $V^* > V^{\mathrm{s}}$, there will be immigration and vice versa. When $V^* = V^{\mathrm{s}}$, there will be an equilibrium. This s could depend on transportation costs of trade, thus linking this model to that of Mees.

If V^* has more than one local maximum, there can be multiple equilibria. In such cases, smooth changes in s, the level of comparative advantage, can lead to discontinuous changes in L, the equilibrium city labor force. This is depicted in Fig. 1.11.

The s curves represent labor supply curves for different levels of s. If labor supply were to increase from s_5 toward s_1, the city labor force would smoothly increase from L_5 to $L(b)_2$, beyond which it would discontinuously leap to slightly more than $L(a)_2$, and then on smoothly to L_1. In reverse, a decline in labor supply from s_1 to s_5 would see the labor force smoothly decline to $L(a)_4$, after which it would discontinuously decline to just below $L(b)_4$, beyond which it would smoothly decline to L_5. Clearly there is a hysteresis effect here. Also, of the three equilibria associated with s_3, the middle one is unstable. This model has a family resemblance to that of Papageorgiou (1980), as argued by Dendrinos and Rosser (1990).

1.3.4 Logistical Networks and Long-Distance Trade

Åke Andersson (1986) has proposed a major extension of the Pirenne–Mees hypothesis to the much broader context of what he calls "logistical networks." These are all

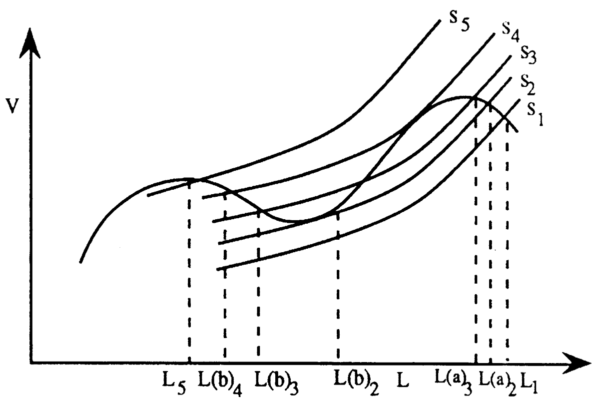


Fig. 1.11 Discontinuous urban comparative advantage and labor supply shifts

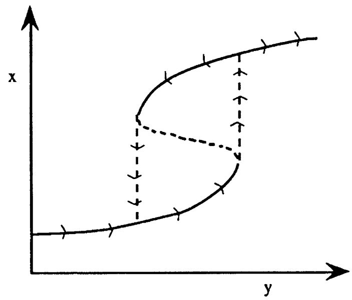


Fig. 1.12 Logistical revolutions

the factors that can facilitate interurban or interregional economic interactions, not just trade, but also financial transaction mechanisms and communication systems for information transfers. Andersson declares (1986, p. 1), "The great structural changes of production, location, trade, culture, and institutions are triggered by slow but steady changes in the logistical networks." In particular he argues that four such "logistical revolutions" have morphogenesized urban and regional structures over the last 1,000 years.

The first of these was that of Pirenne, a lowering of long-distance transportation costs and the rapid growth of cities such as Florence and Brussels from the eleventh century. The second logistical revolution arose from the development of state-backed, reliable banking and its associated transformation of trade relations, beginning in the sixteenth century, especially benefiting Amsterdam and London. The third involved a greater international division of labor based on the production technologies of the Industrial Revolution, beginning in the eighteenth century. Manchester especially manifested this. Finally in the late twentieth century we have the communications–information revolution with its headquarters in San Francisco and Tokyo. Each of these revolutions has been associated with a complete restructuring of the world system of city hierarchies.

To summarize this, he presents a general model of changes in commodity production, x, knowledge infrastructure, z, and network infrastructure, y, as functions of each other, depreciation rates, and various other constants and control variables. He argues that discontinuous changes in production, x, can arise from continuous changes in network infrastructure, y. This is depicted in Fig. 1.12 as a Thom-type (1972) fold catastrophe with hysteresis effects. He argues that this represents the general form of a logistical revolution, with the actual path of commodity production following a logistic during a discontinuity.

1.4 A Possible Synthesis: The Role of Technological Change

1.4.1 Agglomeration, Logistical Networks, and Technology

The preceding discussion has not resolved the debate between the advocates of agglomeration and the advocates of long-distance trade as to the principal driving force of urbanization or de-urbanization. Both are important and are affected at different times and in different ways by technological change. Thus the development of economies of scale in sewage treatment plants encourages agglomeration, while improved septic tank technologies encourage deglomeration. Economies of scale in manufacturing might encourage urban specialization and expansion along with increased long-distance trade (the third logistical revolution), whereas the spread of computers and fax machines might encourage de-urbanization and localization (the fourth logistical revolution).

So we seek a synthesis, taking into account both effects and the impact of technological change upon the urbanization process. We consider a synthesis of several of the models discussed earlier with a model by Nijkamp (1983), although some of the earlier models already possess the potential for such a synthetic interpretation. Thus in the Isard (1977) model, if the technology parameter is assumed to represent the impact on long-distance trade of technological change, then it fits the bill. Also in the Mees (1975) model, if the urban–rural productivity difference (the bias factor) is viewed as reflecting urban agglomerative economies, then that model might fit the bill.

We shall consider our model in per capita terms and assume that total, but not urban, population is constant. Also we shall assume that the urban sector is measured relative to the rural sector, thus allowing us to ignore explicit modelling of the latter. Let $U =$ utility, $X =$ the export base (assumed to be related to total urban output by a fixed multiplier, $b > 1$), $L =$ the logistical network determining ease of long-distance trade and other interurban interactions, $I =$ infrastructure (assumed to drive the net agglomerative effect), $T =$ the technology or information level, $K =$ capital stock (indexed by sector), $C =$ consumption, and δ_i equals the ith sector's depreciation rate (assumed to be zero for sector T).

$$U = U(C, K, I), \tag{1.25}$$

$$X = X(I, L, T, K_x), \tag{1.26}$$

$$I = I(T, K_1), \tag{1.27}$$

$$L = L(T, K_{\mathrm{L}}^{-}), \tag{1.28}$$

$$T = T(K_{\mathrm{T}}), \tag{1.29}$$

$$K = K_x + K_1 + K_{\mathrm{L}} + K_{\mathrm{T}}, \tag{1.30}$$

$$bX = \mathrm{d}K/\mathrm{d}t + C + \Sigma\delta_i K_i. \tag{1.31}$$

Following Nijkamp (1983), the production relations are considered to be nonlinear in all arguments, with zones of increasing marginal productivity alternating with zones of decreasing marginal productivity beyond critical inflection points and also including time asymmetries that generate hysteresis effects. A presentation of

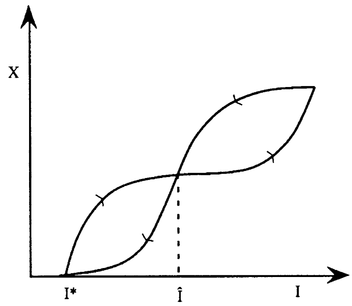


Fig. 1.13 Asymmetric nonlinear urban dynamics

such a partial effect can be seen in Fig. 1.13, with the relationship between I and X, drawn from Nijkamp. Changes in T can change the position of I, the critical inflection point, as well as I^*, the basic threshold level.

The dynamics will be driven by the solution to

$$\max U = \int_{t=0}^{\infty} U(C, K, I)\mathrm{e}^{rt}\,\mathrm{d}t, \tag{1.32}$$

where r is the time discount rate. Nijkamp presents a simpler version of this model because he has output purely as a function of technology (R & D), infrastructure, and capital, without any other interactions between sectors except for the intersectoral allocation of capital. His solution is that the shadow prices (marginal products) of capital in each sector should be equal. That holds here also, but the expression of these marginal products is more complicated because of the greater number of interactive effects between sectors.

Even in the simpler Nijkamp model, the presence of threshold effects, varying degrees of economies of urban scale, and time asymmetries lead to the possibility of "bang-bang" solutions to the optimal control problem, with discontinuous changes and patterns of growth followed by decline as technology and agglomeration interact, at times cooperatively and at times in conflict with each other. With our more complex model, these effects can also lead to chaotic dynamics as well.

1.4.2 Rome Was Not Built in a Day

And neither did it fall in a day.

Let us apply this synthetic view to one of the most spectacular of all urban expansions and declines, the rise and fall of Rome and the cities of its empire. This directly relates back to our discussion of the Pirenne hypothesis in which he argued that the political collapse of Rome itself in the fifth century did not cause the collapse of the Roman system of long-distance trade or the disappearance of Roman provincial cities. De-urbanization did not occur until the Moorish and Viking invasions finally broke down this trading system in the eighth century. However, we shall argue that