AGGLOMERATION ECONOMIES, TRADE, AND A SYSTEM OF CITIES: A GENERAL EQUILIBRIUM APPROACH

by

YOUNG GAK KWON

M.R.P., Syracuse University (1978)

Submitted to the Department of Urban Studies and Planning in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology September 1987

© Young Gak Kwon, 1987

Signature of Author

Department of Urban Studies and Planning
September 1987

Certified by

Alan Strout
Thesis Co-supervisor

Certified by

William Wheaton
Thesis Co-supervisor

Accepted by

Langley Keyes
Chairman, Department Committee
ACKNOWLEDGEMENTS

I would like to express my appreciation to the members of my committee, professors Karen Polenske, Alan Strout, and Bill Wheaton for their time, insights, and comments during the preparation of this dissertation. Alan Strout has kept the faith in me and provided encouragement in most difficult times. Bill Wheaton’s substantive inputs during our thoughtful discussions have been indispensable for this dissertation. During my long stay at MIT, Karen Polenske has always been generous in offering valuable suggestions and guidance. I owe many intellectual debts to them.
AGGLOMERATION ECONOMIES, TRADE, AND A SYSTEM OF CITIES: A GENERAL EQUILIBRIUM APPROACH

by

YOUNG GAK KWON

Submitted to the Department of Urban Studies and Planning in Partial Fulfillments of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

A general equilibrium approach that integrates the conventional supply- and demand-oriented approaches to the determination of city size and output is proposed. Among the prominent features of the approach are agglomeration economies and diseconomies, industrial structure of the city, and the interdependence between cities via migration and trade. The approach is then utilized to relate economic efficiency to alternative city size distributions in the nation.

The optimum city size distribution in which potential welfare levels of all cities are maximized is attained when all cities become autarkic and of size $N_0$, the single-city optimum size. However, under limited factor mobility, a hierarchical city size distribution is more likely to emerge in equilibrium. This latter city size distribution, though suboptimal, represents mutual benefits for both cities resulting from gains from trade.

Thesis Co-supervisor:
Alan Strout, Ph.D.,
Senior Lecturer in Urban and Regional Planning.

Thesis Co-supervisor:
William Wheaton, Ph.D.,
Associate Professor in Urban and Regional Economics.
CONTENTS

1. INTRODUCTION ________________________________________________ (5)

2. MODELS OF CITY SIZE AND OUTPUT DETERMINATION -------------- (11)
   2-1. Introduction ________________________________________________ (11)
   2-2. Demand- and Supply-oriented Models ____________________________ (11)
   2-3. Localization and Urbanization Economies ________________________ (14)
   2-4. Synthesis __________________________________________________ (22)

3. A SINGLE-CITY MODEL _________________________________________ (26)
   3-1. Introduction ________________________________________________ (26)
   3-2. Consumer Equilibrium ________________________________________ (29)
   3-3. Producer Equilibrium ________________________________________ (37)
   3-4. Market Equilibrium __________________________________________ (43)
   3-5. Migration and City Size Distribution ____________________________ (58)

4. A TRADE MODEL WITH SCALE ECONOMIES _________________________ (66)
   4-1. Introduction ________________________________________________ (66)
   4-2. Production under "Mild" Scale Economies ________________________ (67)
   4-3. A Balanced Free-Trade Model ________________________________ (75)
   4-4. Migration and Balanced Free-Trade Model ______________________ (98)

5. POLICY IMPLICATIONS _________________________________________ (115)

6. CONCLUSION __________________________________________________ (122)

7. BIBLIOGRAPHY ________________________________________________ (124)
LIST OF FIGURES

Fig. 2-1: Hypothetical Economies of Scale with Urban Size -- (20)
Fig. 3-1: Total Labor Supply with City Size ---------------------- (36)
Fig. 3-2: Supply and Demand Curves (i=1) ---------------------- (48)
Fig. 3-3: Equilibrium Price with City Size (i=1) -------------- (52)
Fig. 3-4: Utility Level with City Size ------------------------- (54)
Fig. 3-5: Net Urbanization Economies by Industry (i=1,2) ---- (56)
Fig. 3-6: Equilibrium City Sizes (N≤2N₀) --------------------- (61)
Fig. 3-7: Equilibrium City Sizes (N>2N₀) --------------------- (63)
Fig. 4-1-1: Price Ratio vs. q ------------------------------- (78)
Fig. 4-1-2: Factor Price Ratio vs. q ------------------------ (82)
Fig. 4-2: Excess Demand and Supply vs. q ------------------- (86)
Fig. 4-3-1: Numerical Solution of (4.11), p vs. q --------- (89)
Fig. 4-3-2: Numerical Solution of (4.11), q_A vs. q_B ------ (90)
Fig. 4-4-1: Numerical Solution of (4.11) and (4.20),
q_A vs. E_2 --------------------------------------------- (92)
Fig. 4-4-2: Numerical Solution of (4.11) and (4.20),
p vs. E_2 ----------------------------------------------- (93)
Fig. 4-5: Changes in Utility Levels, Before and After Trade (95)
Fig. 4-6-1: Numerical Solution of (4.33), z vs. q -------- (103)
Fig. 4-6-2: Numerical Solution of (4.33), z vs. p -------- (104)
Fig. 4-7: Numerical Solution of (4.34) --------------------- (106)
Fig. 4-8: Equilibrium Utility Levels vs. City Size
Distribution ----------------------------------------------- (107)
Fig. 4-9: Utility Level with City Size (Single City Model) (110)
Fig. 4-10: Population Share of City A vs. N (at Q) ------ (111)
Fig. 4-11: Equilibrium Utility Levels vs. National
Population ----------------------------------------------- (112)
LIST OF TABLES

Table 4-1: Solutions to Endogenous Variables
(Trade-Only Model) ---------------------------------- (97)

Table 4-2: Solutions to Endogenous Variables
(Migration-and-Trade Model) ----------------------- (114)
1. INTRODUCTION

With the continuing increase of worldwide urban populations, the importance of cities in regional and national economic development has become paramount. Thus, many countries now adopt urban policies in order to affect the spatial allocation of populations among cities as a part of their national economic development planning (Rodwin[1970], Thompson[1972], Reiner and Parr[1980], Richardson[1981], Renaud[1981]). According to recent UN reports, eighty-three percent of 114 developing countries indicated that the spatial distribution of their urban population was unacceptable (UN[1978, 1980]). The reports also noted that the reason for such unacceptability was attributed to a heavy concentration of population in a few large metropolitan centers caused by migrations from small cities and rural areas. These countries seem to believe that unchecked market forces give rise to excessive growth of their largest primate cities, causing unwanted results such as economic inefficiency, regional inequality, political unrest, and the familiar "big city headaches" such as congestion, pollution, and overburdened infrastructure and housing.

The emerging consensus among these countries, according to the reports, is that the uneven spatial distribution of population is the main obstacle to overall socioeconomic development. To the countries, the existence of their largest cities is reminiscent of economic inefficiency and inequality. In
practice, therefore, most urban policies have mainly been limited to alleviating the situation in their biggest cities, either to direct more public investments into them or to divert people away from them. Although many countries, such as Brazil, Egypt, and Mexico, have emphasized regional growth and development in small and medium size (or "secondary") cities, such efforts seem to reflect their desire to keep migrants out of large cities, such as São Paulo, Cairo, and Mexico City. Rarely has there been a serious commitment to develop the biggest city as well as smaller cities within the context of the national system of cities. What is lacking is the view that the unit of national urban policies must be the national economy rather than the city economy.

Aside from the equality consideration, if spatial urban policies to alter city sizes are needed for national economic development, one has to question the validity of the current policies on economic efficiency grounds: can the size containment of the largest city enhance the economic welfare of its own citizens and others in the rest of the country? At least two shortcomings of the current policies can readily be pointed out. First, a preoccupation only with the biggest city and its size merely indicates a greater concern for local development than for national development. By excluding the rest of the country, the current policies may be blamed for their artificially narrow areal scope. They also seem to lack an explicit consideration of the contribution of the biggest city to the national economy. Second, their focus on the relationship between one city size and
the economic efficiency and welfare within even that city is misleading. Unless the city exists in isolation, its size and efficiency are determined in interaction with the rest of the country. Moreover, a city without being related to the national size distribution of cities may never identify its size measure. In absolute terms, for example, Jakarta may be the biggest city in its country, but not so by Mexican standards. Of course, no implication regarding the economic efficiency and welfare in Jakarta, let alone its contribution to the national economy, can be drawn from its population size alone. The size and efficiency of a city are a relative concept that can be meaningful only within the framework of the national system of cities. It is clear that, unless urban policies are directed toward the national system of cities, they may not be useful in enhancing national economic welfare.

However inadequate the current urban policies may be, they nevertheless seem to reflect the current state of knowledge in the literature. Although stylized facts such as the rank-size rule and paradigms such as "concentrated decentralization" (Rodwin[1961], Alonso[1968]) exist, it is surprising that no study has analytically dealt with the basic issue: what spatial allocation of populations among cities is commensurate with national economic development? Ever since Isard raised the question almost three decades ago, it still remains unanswered (Isard[1956], p. 183).
In this dissertation, an attempt will be made to address the issue directly by devising an analytic spatial economic framework in which the economic efficiency and welfare of a city depend not only on its own size and internal characteristics, but also on those of other cities within the nation. It is hoped that this framework can serve as an analytic basis of national urban policies in enhancing the overall level of welfare. To this end, we will develop a spatial general equilibrium urban model in which cities are viewed as centers of production, consumption, and trade. Essential elements of the model will include sizes of cities, industrial composition and production technologies, and economic interactions of cities via migration and trade. With the model, it will then be possible to answer the following related questions. (1) Does an existing size distribution of cities leave room for welfare improvement, and justify a planned intervention? (2) If (1) is so, what would be an alternative size distribution? (3) What would be an available and effective means of planned intervention in a typical mixed economy?

Following the long tradition in the international trade literature, the model will emphasize comparative statics and will be of a 2 by 2 by 2 (factors, industries, and cities) type, with possible extensions for n-cities where appropriate. In addition to comparative advantage which is the basis of the standard trade model, we will incorporate agglomeration economies into our model. Thus, it represents an expanded interurban version of the standard international trade model. However, it is well known
that such differences as absence of tariffs and exchange rate concerns, greater factor mobility, similarity in tastes and production technology make the interurban version a unique and more plausible trade model than the international version under the usual set of assumptions.

In chapter 2, we will review two representative yet polar types of urban models regarding the city's output and size determination. Since both types are incorporated into our model, we will discuss their merits and shortcomings along with the rationale of our model specification. The single-city model to be developed in chapter 3 will be adapted from discussions made in the previous chapter. It will be the basic tool used throughout this dissertation. We will show how the city's internal production and consumption characteristics, including its size, determine its welfare level. In chapter 4, we will deal with two cities that are mutually interdependent in the economy by means of balanced trade and exchange of factors. Under different assumptions, equilibrium sizes and welfare levels of both cities will be determined. We will then examine changes in the equilibrium welfare levels as a result of changes in the size distribution of cities. In chapter 5, we will cover, by utilizing lessons from the previous chapters, some suggestive issues in national urban policies. Due to the scant nature of economic data at the city level in most countries, however, we will rely on numerical analyses rather than on empirical ones. In chapter 6, we will present the conclusions. After a brief summary of the
dissertation, we will delineate limitations of our model and possible directions for future research in urban modelling and policy making.
2. MODELS OF CITY SIZE AND OUTPUT DETERMINATION

2-1. Introduction

Before we build a sound urban model, it is necessary to investigate existing urban models that explain the determination of city size and output. Since the standard assumption is to equate city size with city laborers, most models rely on the mechanics of the urban labor market for explanations. Moreover, since the number of laborers is highly dependent on net migration into a city, urban models put a particular emphasis on the relationship between employment and migration. Two contrasting approaches, each of which emphasizes either the demand or supply side of the labor market, will be reviewed. We will then argue that incorporation of both views would be necessary to build a spatial general-equilibrium urban model. In addition, refinement of some key concepts for our model specification will be noted.

2-2. Demand- and Supply-oriented Models

Muth[1971], Engle[1974], Schaefer[1977], and Miron[1979] provide a useful classification of urban models on city size and output: the demand- and supply-oriented approach. The former type, which can also be labeled as the external approach, inherits the Keynesian heritage in urban modelling in that the main emphasis is on exogenous demands from other areas for the output of the city. Factor supplies (labor and capital) are pre-
sumed to be perfectly elastic so that factor in-migrations to the city are limited only by the city's factor demands at given real factor prices. It is further assumed that the demand for products is price inelastic and that no matter how much is demanded externally, it will be supplied by the city. Spatially, this implies that the size and output of the city are wholly determined by its output market areas combined with varying demand intensities over space. Migration becomes the consequence of the city's growth rather than the cause of it. City sizes will be in continuous equilibrium, because migration will always clear the interurban labor market. The familiar export-base model and the various urban hierarchy models, such as Lösch[1954], Beckmann[1958], and Beckmann and McPherson[1970], belong to this type.

Although there is a consideration in this approach of mutual interdependence among cities via exports and factor exchange, and, to a varying degree, of their locations and associated transport costs, it usually does not take account of the city's internal production and consumption characteristics. In Beckmann's model, for example, per capita consumption of a good is held to one unit regardless of its price. With no explicit consideration of comparative costs in production, in particular, the selection of export sectors becomes rather arbitrary. The level of exports is often exogenously determined, but there is a complete lack of explanation as to how the level of imports is determined. This approach essentially envisions a hierarchy of
cities based on trade imbalance. Goods are exported only down the hierarchy from large to smaller cities, but there are no corresponding reverse flows. In spite of such criticisms of the demand-oriented approach, it is to its credit that, unlike the supply-oriented approach, establishment of a city size distribution, albeit unsatisfactory, is feasible within the model.

The supply-oriented, or internal approach, on the other hand, is often concerned with one city in isolation, and represents the tradition in the international trade literature. The main focus of this approach is on the city's production technology, endowment of factors, and other internal characteristics. It is often assumed that factor demands are perfectly price elastic, but factor supplies are inelastic due to non-wage considerations. Further, since the city is presumed to face a perfectly elastic export demand at a given output price (or terms of trade), it is suggested that no matter how much the city produces, it will be purchased elsewhere. This is the familiar "small country approximation" in the international trade literature. Consequently, the output and size of the city are limited only by the internal cost consideration and available factor supplies. Migration now becomes the cause of the city's growth, rather than the response of it. Thus, the city's internal characteristics (including non-wage aspects), which presumably are a determinant for potential migrants, become of paramount importance for its growth.
In this approach, an exogenous shock, for example, to shift the demand for labor will induce migration into the city, and this will in turn result in an even greater equilibrium city size and output. This "growth-feeds-by-itself" hypothesis has been suggested by many authors including Borts and Stein[1964], Thompson[1968], and Muth[1971]. Although this approach emphasizes the city's various internal characteristics in production and consumption, its neglect of other cities as both output market and factor supply areas make it highly unrealistic as a basis of a spatial urban model. In general, due to its lack of the explicit linkage among cities, this approach cannot establish a distribution of city sizes endogenously. However, since this approach has been influential in the literature and pervasive in the minds of many governments, as it incorporates the concept of agglomeration economies, we will now look into it further.

2-3. Localization and Urbanization Economies

Following the usual Heckscher-Ohlin practice of constant returns to scale production technology, proponents of the pure supply-oriented approach try to find the raison d'être of the city in its resource endowment (natural resources included) that gives rise to its comparative advantage. However, the approach is generally extended to contend that when a city is assumed to be built on homogeneous land, the main economic justification for
the city is the presence of agglomeration economies in production and/or consumption. After Ohlin[1933], Hoover[1937] and Isard[1956], it is widely accepted that agglomeration economies can be realized at the firm (scale economies), the industry (localization economies), and the city level (urbanization economies) (for example, see Isard[1956]). In the urban context, however, the last two types of agglomeration economies, which are external to the firm, get the most attention. This is because in a city multiple firms of widely different sizes seem to coexist competitively within the same industry. Under the existence of external economies, therefore, firms are usually modelled to operate under constant returns to scale and to behave competitively (Kemp[1955,1964], Melvin[1969], Chipman[1970]).

Localization economies, which represent immobile external economies to the firms of each industry, are internal to the industry of the city. They are the result of the enlargement of one industry, which facilitates greater intra-industry specialization and division of labor in that industry. The enlargement of the industry also means enhanced intra-industry facilities for the firms in terms of collective research, marketing, communication, and innovation. Urbanization economies, which are external to both the firm and the industry but internal to the city, are the result of greater spatial concentration of all industries and their increased level of economic activities within the city. These economies reflect all tangible and intangible locational advantages for any firm of any industry in a large and
diverse city economy. Such advantages result from a large city-wide labor market, a well-developed infrastructure, and other public facilities, which are often lacking in a smaller, less diverse place. Thus, the sizes of industries, their composition, and inter-industry externalities are all implied to be the components of urbanization economies of the city. The usual practice of using the city size alone amounts only to a proxy measure of urbanization economies, however.

Mills[1967] was among the first formally to present an analytic supply-oriented model that incorporates urbanization economies in the production function of a single composite urban good. In his monocentric city where firms are located in the city center surrounded by residential areas, a unique equilibrium size is reached by a sharp increase of commuting costs and congestion (urbanization diseconomies) after certain city sizes are attained. The equilibrium city size represents an optimum size at which net urbanization economies (urbanization economies less diseconomies) are maximized. The equilibrium is essentially an autarkic one, and the optimum city size is determined entirely by the internal conditions of the city, regardless of the size distribution of cities in the rest of the country. Because the city is of the aforementioned "one-city-in-isolation" type, and its industrial composition is singular (i.e., city with one aggregate industry), however, the equilibrium city size should mean either of two paradoxical results: there is one city in the country; or all cities are identical. Of course, the naivety of one
universal optimum city size is well known.

An indirect way to sidestep the above dilemma, while staying within the Mills' framework, would be to specify a range of "efficient" city sizes in which net urbanization economies are positive (Alonso[1971], Richardson[1972]). However, efficiency here refers to neither an optimum nor an equilibrium, and the preoccupation here is to find city sizes for "self-sustaining growth". A more direct solution to avoid the above paradox of one efficient city size, however, is Henderson's recognition that different-sized monocentric cities perform different functions, and they may operate at different, but efficient, equilibrium sizes (Henderson[1974]). Each city, producing only one distinct export good, is engaged in free trade with other cities (but not with cities of its own type, the type being determined by the industry) at exogenously determined terms of trade. Thus, multiples of different-sized cities can coexist in equilibrium; however, cities of the same type must be identical in size.

It should be noted that Henderson's model is established, rather a priori, on complete specialization of cities at all times in which no distinction can be drawn between the city and the (export) industry (or between urbanization economies and localization economies). The sizes of both are identical when measured by laborers. According to this model, a city, however big it may be, must always be engaged in trade, and unlike the standard trade model, factor movements can never substitute for
trade. Henderson notes that complete specialization occurs when there are no production benefits or positive externalities from locating two different industries in the same place (i.e., the city center).

If they are located together, because workers in both industries are living and commuting in the same city, this raises the spatial area of the city and average commuting costs for a given degree of scale economy exploitation in any one industry. ... Thus, specialized cities will be larger and more efficient than potential non-specialized cities; and they will be able to pay higher factor returns (Henderson [1980], p.17).

However, the statement is premature, and cannot explain even the initial identification of the export sector. Thus, its complete specialization is totally unclear. The complete specialization is rather the result of Henderson's reliance on the so-called small city approximation in which the city produces (and consumes) on such a small scale that the effect of its production (and consumption) on the national and international markets can be ignored. With no consideration given to other cities as potential market areas, therefore, the complete specialization is presumed to be feasible. Of course, the identification and the degree of specialization of the export sector must be determined by production and trade equilibrium conditions, not
a priori. Although complete specialization is always a possibility, the actual degree of specialization is to be limited by the extent of the market. In light of the fact that the contemporary urban concern stems mainly from the alleged gigantism of the largest cities and their pervasive effects on national economies, the small city approximation seems particularly inappropriate.

Aside from inter-industry production externalities, the possibility of multiple industries in the city is more a rule than an exception. Later we will show that even under external economies, the production possibility curve of the city is generally concave. Thus, there exists a rare possibility of complete specialization by one industry. This is because, to the extent that the marginal products of factors are finite and diminishing even under scale economies, greater specialization by the city in a particular good will eventually increase the marginal (social) opportunity cost of that good. Although the specific degree of specialization by the city cannot be inferred in the absence of the production functions, even a casual observation will hold that incomplete specialization is more likely in a large market such as the city. Moreover, as the "home market effect" of the trade literature suggests, it can be argued that the city's size of the internal market determines the range of goods produced, or industrial diversity. Indeed, industrial diversity or "breadth", to quote Thompson[1968], must be a fundamental characteristic of the city.
It is thus not surprizing to note that the single-industry models of Mills and Henderson generate rather unrealistic results. Quite contrary to Henderson's prediction, for example, Dixit[1973], in an exercise of such a single-industry model, could not reach an equilibrium city size much in excess of one million, however large the assumed urbanization economies might be. Moreover, the existence of a single industry implies that the output will not respond to a change in price (or the terms of trade). Such a rigidity can usually be avoided once multiple industries are allowed for in the city.

For example, Isard[1956] presents a hypothetical illustration of multiple industries which are subject to urbanization economies and diseconomies as in Fig. 2-1. It shows four curves for the corresponding representative industries of the city in which net economies increase first, attain maxima, and finally decrease as city size increases. The top "total economies" curve is meant to represent an overall net economies figure for the city, hopefully after appropriate weighting of the industry curves. The total economies curve is akin to the single industry curve implied by the models of Mills and Henderson, but with two different possibilities: outputs can now become price-elastic; and industrial composition can play a role in determination of the equilibrium city size. Thus, the rigidity problem can now be nicely worked out.
Fig. 2-1: Hypothetical Economies of Scale with Urban Size.

Implicit in depiction of Fig. 2-1 are the usual assumptions of the supply-oriented approach, however. In particular, the curves drawn reflect either the autarkic situation of the hypothetical city, or no consideration of other cities in determination of the national system of cities. It is clear the curves will change readily once the city becomes interdependent with other cities via trade, for example.

2-4. Synthesis

The previous section has shown that the two polar types of urban models attribute city size and output determination mainly to one of two sources: either external demand for the city's output or internal characteristics of the city including agglomeration economies. In view of reality, however, it is abundantly clear that a truly general spatial urban model calls for an integration of both. For the size of the city, its output, and welfare level are all dependent on those of other cities within the nation.\(^1\) In particular, there is a need to incorporate into our model some refinement of the production conditions typified by the supply-oriented approach, as well as the consideration of demand for the city's output of either internal or external origin indicated by the demand-oriented approach.

\(^1\) International trade, an important factor in urban modelling, is beyond the scope of this study.
The single-industry, single-city models of Mills and Henderson nevertheless represent an improvement over the pure supply-oriented models in that they incorporate both urbanization economies and diseconomies albeit in less than satisfactory ways. However, their failure to allow for industrial diversity within the city seems to generate inconsistencies as noted above. It should also be noted that unless the industrial composition of the city is known, no efficiency implication can be drawn by the city size alone. This suggests that at least two industries must be introduced to allow for a meaningful industrial composition and the consequent agglomeration economies realized within the city.

Once the existence of, say, two industries is accepted, there remains the problem of specifying urbanization and localization economies within the model. Despite deliberations made by Hoover and Isard, however, the relationship between the two agglomeration economies that are external to the firm does not seem well established. As Isard aptly notes, "only a fine line of distinction" separates them. Although their conceptual differences have already been noted, their separate specifications within the model seem to be an intractable problem. Further work in this area is badly needed. Consequently, we will handle this problem by the following simplified, but reasonable, set of assumptions.

First, we envision the existence of a single homogeneous
labor market from which labor is allocated between the two industries. The size increase of the labor market along with its greater spatial concentration would enhance its efficiency and allow for a greater potential realization of localization economies for both industries. On the production side, this is made possible by the increased intra-industry specialization resulting from the greater division of labor within each industry. On the consumption side, the potential can be realized, because the bigger labor market means increased internal demands for the outputs of both industries.

Second, instead of introducing urbanization economies separately, we assume that they are the result of localization economies. Thus, urbanization economies are presumed to consist of localization economies only. Although this assumption excludes the possibility of inter-industry production externalities, it is argued that this nevertheless generates a conceptually more sound measure of urbanization economies than most measures based on the city size alone. Like Isard's industry curves of Fig. 2-1, this assumption will generate localization economies curves of the two industries. Unlike Isard's model in which the lack of a weighting function precludes measurement of urbanization economies at the city level, however, our model will generate a measure of them which is a weighted transformation of the two localization economies curves. The measure will be in utility terms, and it will reflect not only the city size but also the underlying industrial composition of the city in equilibrium. A typical
phenomenon of urbanization economies, as will be shown later, will be rising real wages and/or falling prices of goods. In the following chapter, we will discuss in detail a single-city model that incorporates the above points.
3. A SINGLE-CITY MODEL

3-1. Introduction

The purpose of this chapter is to formulate under specific assumptions a spatial general equilibrium model of a single city, which is engaged in both production and consumption under autarky. Under the assumptions that the two industries of the city experience localization economies and that the city experiences consequent urbanization economies and diseconomies, we will show how a unique equilibrium is obtained under a perfectly competitive setting. In the end, real wages, capital rentals, outputs and their prices, and the level of welfare are all determined endogenously, given the resource endowments and the prevailing internal production and consumption characteristics of the city. The fact that the equilibrium is attained totally by the internal characteristics puts this single-city model in the tradition of the supply-oriented approach. Later in this chapter, however, we will allow two cities of the nation to be engaged in factor exchange but not in trade. This will present us with a city size distribution which largely reflects the extension of the supply-oriented approach. A complete balanced-trade model will be the subject of the next chapter.

To build our single-city model, we will envision a hypothetical city based on the following characteristics and assumptions. More specific assumptions will follow in later sections
which will cover the equilibrium among consumers, producers, and markets.

a. By city, in general, it is meant that: geographically a homogeneous monocentric area the center of which is occupied by the point Central Business District (CBD) for firms, surrounded by residential areas; and economically the place where its people live and work.¹ In the single-city model in particular, the absence of interurban trade requires that the output market area be limited to that within the city.

b. The city "produces" one nontraded good, housing or land, for residential use only, and it produces two goods for own consumption and possible interurban trade. Of course, in the single-city case, trade is denied.

c. The city is spatial in that housing is differentiated by its distance from the city work center (CBD), and its rents are determined accordingly. The prices of the two goods are assumed to be invariant within the city, and they are the same as the FOB prices. In trade, we also assume that they are not subject to transport costs for deliveries from one city center to another.

¹ This highly simplified definition is based on the additional assumption that the city has its unequivocal members as both residents and workers. In reality, the situation is closer to what Alonso[1971] calls a "fuzzy set" in which membership is not unequivocal but a matter of degree.
d. Except for their actual housing locations, city residents are assumed to be identical in all other respects including location preferences: for example, skills, utility functions, vocational preferences, and capital ownership. Consequently, no distinction exists between workers and capitalists, and the usual distributional considerations between them are avoided. The city land is collectively owned by them, and each has an equal share in that collective ownership (à la land bank). Thus, total housing rents are to be equally divided among residents.

e. The two industries employ two homogeneous factors, labor and capital, which are both fully employed and fully mobile between the industries within the city. We also assume that firm sizes within an industry are identical.

f. All firms in the same industry share an identical production function, but those between industries are different.

The above list of assumptions is fairly traditional in that it depicts the requirements of resource ownership among citizens, and of perfect competition in the city in both output and factor markets. The special nature of housing, which is produced solely with land, deserves further clarification, however. Being a perfectly localized nontradable good, the prices of housing and the associated commuting costs are uniquely determined within the
city. Locational equilibrium for identical residents requires that the combined costs for housing and commuting be the same regardless of location. By specifying per capita consumption of housing as unity, we may view the combined costs roughly as a fixed cost of living unique to the city. Rather than treating housing and commuting as the objects of utility and disutility respectively as in Alonso[1964], we can thus simplify the combined costs as an "admission cost" for any resident to live in the city. Because the increase of city size would push up the cost, it can be the measure of urbanization diseconomies in our model.

3-2. Consumer Equilibrium

Suppose there are $N$ residents in the city, each with the following individual utility function, $u$, which incorporates the fixed housing consumption set to unity and the fixed demand for leisure:

$$u = x_1^{\theta_1} x_2^{\theta_2}, \quad (3.1)$$

$$\theta_1, \theta_2 > 0, \quad \theta_1 + \theta_2 = 1,$$

where $x_i =$ individual consumption of good $i$, $(i=1,2)$. We normalize to unity the fixed amount of total available time after leisure. The representative resident then will allocate it between work and commuting as in the following:

$$l(t) + gt^2 = 1, \quad 0 \leq t \leq m, \quad (3.2)$$
where $l(t) =$ labor supply at location $t$,

\[ g = \text{a technologically determined commuting parameter}, \]

\[ g t^2 = \text{time spent for commuting at location } t, \]

\[ m = \text{distance of the city boundary from the center}. \]

Consequently, although the demand for leisure is fixed, the supply of labor is variable with respect to location. For example, residents in the center, who spend no time in commuting, spend as much time for work as those in the boundary do for work and commuting, i.e., $l(0) = l(m) + gm^2$. Thus, time preference for either commuting or work is the same. The commuting time function, $gt^2$, is based on a plausible premise that the marginal commuting time is increasing with respect to distance. This is especially plausible when congestion is a realistic possibility in the modern city.\(^2\) In terms of its cost, we assume that at location $t$, commuting costs the resident the foregone wage of $wgt^2$ where $w$ is the wage rate. While this assumption explicitly considers the time aspect of commuting, it neglects the direct out-of-pocket cost and other psychic costs observed in reality. However, the assumption considerably simplifies our model in that

\(^2\) Although our choice here of a quadratic function, $gt^2$, is arbitrary, it is one way of introducing congestion into the city at the individual level. Under no congestion, the speed of a trip must be constant regardless of location. With congestion, however, it must decrease. For example, with a more general commuting time function, $gt^\beta$, the speed for a one-way trip is defined as $dt/d(gt^\beta) = (1/\beta g)(t^{1-\beta})$. In order for it to decrease, $\beta$ must be greater than 1. Thus any number greater than 1 can be used for $\beta$. The use of 2 for $\beta$, however, considerably simplifies our derivation of the labor supply function (3.17) to follow.
it does not require the introduction of a separate transportation sector into the city. Thus, the economic consequence of commuting, or of urbanization diseconomies, is simply a waste of some of the total available labor for the city as a whole.

The utility function is to be maximized subject to the following budget constraint:

$$p_1x_1 + p_2x_2 + p_h(t) - w[t] - I = 0,$$  \hspace{1cm} (3.3)

where $p_i$ = price of good $i$ \hspace{0.5cm} (i=1,2),

$p_h(t) = \text{price of housing service (rent) at location } t$,

$I = \text{nonwage Income}$.

Aside from location, residents will set the marginal utilities of the two goods proportionally to their respective prices, yielding the following individual demand functions:

$$p_ix_i = \theta_i[w[t] + I - p_h(t)] \hspace{0.5cm} (i=1,2).$$  \hspace{1cm} (3-4)

Substituting (3.2) and (3.4) into (3.1), we get the indirect utility function,

$$v = \left(\frac{\theta_1}{p_1}\right)^{\theta_1}\left(\frac{\theta_2}{p_2}\right)^{\theta_2}[w(1-gt^2) + I - p_h(t)].$$  \hspace{1cm} (3.5)

Spatial equilibrium in the housing market requires that $\partial v/\partial t = 0$. Differentiating (3.5), it turns out that

$$\partial p_h(t)/\partial t = -2wgt.$$  \hspace{1cm} (3.6)

Thus, at any location, increased foregone wages due to increased
commuting distance must just be offset by reduced rents, thereby leaving everyone indifferent as to location. Integrating the above, with the added assumption that \( p_h(m) = 0 \), or at the city edge the rent is zero,\(^3\) to determine the constant of integration, we can derive the following rent gradient, or the price of housing service at location \( t \):

\[
p_h(t) = wg(m^2 - t^2). \tag{3.7}
\]

With per capita housing consumption fixed at unity, the city boundary will be determined by the following equation:

\[
\pi m^2 = N, \tag{3.8}
\]

where \( N \) = population of the city. Given (3.8), we can rewrite the rent gradient that clears the housing market:

\[
p_h(t) = wg(N/\pi - t^2). \tag{3.9}
\]

We are now ready to calculate per capita nonwage income, \( I \), as a last step to determine per capita income. First, there is the rent income originating from the equal housing share provision. In our monocentric city, total housing rents are defined

\[^3\text{Since at the city edge urban rent is zero, this assumes that the opportunity cost of land in nonurban use is zero too. However, the outcomes of our model do not change even if we adopt a positive value for the opportunity cost. This is because the increase by that much of rent payments is exactly matched by the increase in rent incomes, leaving any resident with constant after-rent incomes.}\]
as

\[ R = 2\pi \int_0^m p_h(t) dt. \quad (3.10) \]

Using (3.8) and (3.9), \( R \) and the per capita amount, \( R/N \), become

\[ R = \frac{w g N^2}{2\pi}, \quad \frac{R}{N} = \frac{w g N}{2\pi}. \quad (3.11) \]

The remaining nonwage income comes from capital rentals. Under the equal capital ownership, per capita capital rental income is

\[ r \left( \frac{K}{N} \right), \quad (3.12) \]

where \( r = \) capital rental, \( K = \) capital stock in the city.

It is to be emphasized that with variable labor supply, the term \( K/N \) denotes the amount of per capita capital, not the usual capital-labor ratio of the city. Combining (3.11) and (3.12), per person nonwage income is

\[ I = \frac{w g N}{2\pi} + \frac{rK}{N}. \quad (3.13) \]

Substituting (3.2), (3.9), and (3.13) into (3.4), we can rewrite the individual demand functions in order to determine equilibrium conditions for the residents as

\[ p_i x_i = w_\theta \left[ 1 - \frac{g N}{2\pi} + \frac{rK}{w N} \right] \quad (i=1,2). \quad (3.14) \]

Of course, the market demand functions are obtained by aggregating the above across the \( N \) individuals:
\[ p_i x_i = w \theta_i N \left( 1 - \frac{gN}{2x} + \frac{rK}{wN} \right) \quad (i=1,2), \]  

(3.15)

where \( x_i \) = total (market) demand for good \( i \). The bracketed term of either (3.14) or (3.15) indicates per capita income after the housing rent adjustment deflated by the wage rate. We now have obvious consumer equilibrium conditions: the proportion of the resident's income allocated to good \( i \) is equal to \( \theta_i \).

The same substitution into (3.5) leads to an updated indirect utility function:

\[ v = \theta_1^{\theta_1} \theta_2^{\theta_2} \left( \frac{w}{p_1} \right)^{\theta_1} \left( \frac{w}{p_2} \right)^{\theta_2} \left[ 1 - \frac{gN}{2x} + \frac{rK}{wN} \right]. \]  

(3.16)

Housing location, \( t \), is no longer an argument in (3.16), and individuals, regardless of location, derive the same level of welfare. With \( K/N \) fixed, or when capital and city population are changing in the same proportion, utility becomes a linearly decreasing function of \( N \). This is what we have expected with the introduction of urbanization diseconomies alone. In the next section where the production sector of the city is concerned, urbanization economies will be introduced, and a more complete picture will subsequently emerge.

Although it has almost become self-evident by now, we now proceed to calculate the city-wide labor supply, \( L \), for later use. In the monocentric city, it is defined to be:
\[ L = 2\pi \int_{0}^{m} t \xi(t) dt. \]  
(3.17)

Substituting (3.2) and (3.8) into (3.17),

\[ L = N - \frac{gN^2}{2\pi}. \]  
(3.18)

Out of the total available time after leisure, \( N \), or the city size, \( gN^2/2\pi \) is spent on commuting, leaving the city with the above for \( L \). The ratio of labor to population, \( L/N \), decreases as city size increases. The labor supply function (3.18) is a quadratic function the relevant portion of which is limited, of course, to the rising part of it (see Fig. 3-1). For example, point A in Fig. 3-1 is the hypothetical extreme city size at which residents at the city edge would have to allocate all of their time to commuting and none to work. More practical limits to \( N \) are in order.

Since population density in housing is modelled to be uniform throughout the city, and since the parameter \( g \) is assumed to be invariant across city sizes, the upper bound of \( N \) is entirely determined by the situation of the person at the city edge. That is, one must be able to trade off the lower housing rent at that location against the increased commuting time and the consequent reduction in wage. Two reasonable behavioral assumptions are in order. We assume that on a daily basis each person sets aside 12 hours for leisure (including sleeping and other necessary daily routines) and another for work and commut-
Fig. 3-1: Total Labor Supply with City Size.

$g/2r = 0.25 \times 10^7$, $N_A = 2 \times 10^7$, $N_B = 1 \times 10^7$. 
ing the latter of which is normalized to unity for the labor unit. We further assume that the round-trip commuting time for the resident at the edge is likely to be limited to no more than 6 hours. From (3.2) and (3.8), this means

\[ 0 \leq N \leq \frac{\pi}{2 \gamma} \]  \hspace{1cm} (3.19)

and from (3.18),

\[ 0 \leq L \leq \frac{3\pi}{8 \gamma} \]  \hspace{1cm} (3.20)

which require that the curve of Fig. 3-1 be limited to the portion OB. Thus, at the largest practical city size, \( N_B \), for example, the city would lose 1/4 of the total available time to commuting. According to our model, if cities are endowed with equal capital per capita, \( K/N \), large cities will have higher capital-labor ratios, \( K/L \), than smaller ones. This is the direct consequence of urbanization diseconomies and the other assumptions made for this model.

3-3. Producer Equilibrium

In the city, there are two industries, each producing a distinct homogeneous good by using two homogeneous factors, labor and capital. The production function of the jth firm in the ith industry is specified to be
\[ Y_{ij} = \lambda_i L_{ij}^{\alpha_i} K_{ij}^{1-\alpha_i}, \quad (i=1,2) \quad (j=1,2,\ldots,k_i), \quad (3.21) \]

\[ 0<\alpha_i<1, \quad \lambda_i>0, \]

where: \( Y_{ij} = \) output of firm \( j \), industry \( i \),

\( L_{ij} = \) labor employed by firm \( j \), industry \( i \),

\( K_{ij} = \) capital employed by firm \( j \), industry \( i \),

\( k_i = \) number of firms in industry \( i \).

Given that all firms within each industry are identical in size and technology, the firm production functions can be aggregated into the industry production function,

\[ Y_i = \lambda_i L_i^{\alpha_i} K_i^{1-\alpha_i}, \quad (i=1,2), \quad (3.22) \]

where: \( Y_i = \) output of industry \( i \),

\( L_i, K_i = \) labor, capital of industry \( i \).

Into the industry production function, localization economies, or what Chipman[1970] calls "parametric external economies of scale", are introduced by \( \lambda_i \). According to this specification, the term \( \lambda_i \) is viewed by the firm as a parameter, or a Hicks neutral shift factor in making its business decisions. Because the remaining elements of (3.21) constitute a homogeneous-of-degree-one production function, the collective behavior of the firms assures that the industries have constant-returns-to-scale production functions. This guarantees the exhaustion of total revenues by factor payments. Thus the existence of localization economies is consistent with perfect com-
petition.

Although the term $\lambda_i$ is external to the firm, it is internal to the industry, and is actually related to the $i$th industry's output as in the following:

$$
\lambda_i = \frac{Y_i^{\varepsilon_i}}{\varepsilon_i}, \quad 0<\varepsilon_i<1, \quad (i=1, 2).
$$

(3.23)

This specification conforms to our previous observation: localization economies result from the advancement of industry output which, under spatial proximity of firms within the city, facilitates division of labor and intra-industry specialization. When (3.23) is substituted into (3.22), we obtain

$$
Y_i^{1-\varepsilon_i} = L_i^{\alpha_i} K_i^{1-\alpha_i}, \quad \text{or} \quad Y_i = \left( L_i^{\alpha_i} K_i^{1-\alpha_i}\right)^{\rho_i},
$$

(3.24)

$$
\rho_i = \frac{1}{1-\varepsilon_i}, \quad \rho_i>1, \quad (i=1, 2).
$$

The industry production function of (3.24) is homogeneous of degree $\rho_i$. With $\rho_i>1$ due to localization economies, it will actually exhibit increasing returns to scale. According to Chipman, this is called the "objective" production function, whereas that of (3.22) is called the "subjective" production function. The essential difference is that while the former is based on actual production properties, the latter is based on

---

4 For a similar specification of external economies of scale, see Kemp[1955,1964], Melvin[1969], and Chipman[1970], among others.
entrepreneurial behavior. Not only is the term \( \lambda_i \) external to the firms, but the relation (3.23) is assumed to be unknown to them. As we shall see later, the determination of competitive equilibrium in the presence of localization economies requires the use of both.

According to the first-order conditions for profit maximization, the reward to each factor is the value of its marginal product to the firm, not to the industry. For such an entrepreneurial decision, therefore, we compute the marginal private, or "subjective" products of labor and capital from (3.21), or (3.22) in the aggregate, holding \( \lambda_i \) constant:

\[
\frac{\partial Y_i}{\partial L_i} = \lambda_i \alpha_i \left[ \frac{K_i}{L_i} \right]^{1-\alpha_i}, \quad \frac{\partial Y_i}{\partial K_i} = \lambda_i (1-\alpha_i) \left[ \frac{K_i}{L_i} \right]^{-\alpha_i}. \tag{3.25}
\]

Substituting (3.22) into (3.25) for the elimination of \( \lambda_i \), we obtain

\[
\frac{\partial Y_i}{\partial L_i} = \alpha_i L_i^{\rho_i-1} \left[ \frac{K_i}{L_i} \right]^\rho_i (1-\alpha_i) = \alpha_i \left[ \frac{Y_i}{L_i} \right],
\]

\[
\frac{\partial Y_i}{\partial K_i} = (1-\alpha_i) L_i^{\rho_i-1} \left[ \frac{K_i}{L_i} \right]^{\rho_i (1-\alpha_i)-1} = (1-\alpha_i) \left[ \frac{Y_i}{K_i} \right]. \tag{3.26}
\]

It is to be noted that the above marginal private products of factors become less than the marginal social or "objective" products below, which can be computed directly from (3.24), by a factor of \( \rho_i \):
\[
\frac{\delta Y_i}{\delta L_i} = \rho_i \alpha_i \left(\frac{Y_i}{L_i}\right), \quad \frac{\delta Y_i}{\delta K_i} = \rho_i (1-\alpha_i) \left(\frac{Y_i}{K_i}\right).
\] (3.27)

Naturally, we assume that all marginal products are positive, finite, diminishing, and smaller than the average products.\(^5\)

From (3.24), (3.26), and (3.27), this requires additional constraints on \(\rho_i\) as follows:

\[\rho_i \alpha_i < 1, \quad \rho_i (1-\alpha_i) < 1.\] (3.28)

Under perfect competition, a uniform wage rate, \(w\), and capital rental, \(r\), must prevail between the industries. From (3.26), therefore, we arrive at the following producer equilibrium conditions in which the factor payments exactly meet the outputs:

\[L_i = \alpha_i \left(\frac{P_i}{w}\right) Y_i, \quad K_i = (1-\alpha_i) \left(\frac{P_i}{r}\right) Y_i.\] (3.29)

The above producer equilibrium conditions can now be used to reveal the actual production properties via the marginal private cost curve, or the industry supply curve. Combining the objective production function (3.24) and (3.29), we obtain

---

\(^5\) This implies that the objective production function is subject not just to increasing returns to scale, but to decreasing returns to factor proportions as well.
The supply curve has constant elasticity $\rho_i/(1-\rho_i)$. Because the degree of homogeneity $\rho_i>1$, the elasticity becomes negative, and the supply curve is negative-sloping. Therefore, the effect of localization economies is that at a given factor price ratio $w/r$, the supply price of $Y_i$ relative to the wage rate, $p_i/w$, is decreasing with respect to the output.

The above profit-maximizing decisions, however, also represent a socially efficient production schedule. This is because the private marginal rates of technical substitution calculated from (3.26) are identical to the social ones from (3.27):

$$\frac{\delta Y_i}{\delta L_i} = \alpha_i \frac{K_i}{1-\alpha_i L_i}.$$

Despite the presence of localization economies, production is efficient, and is operated along the city's production possibility curve.

Although production is efficient in the sense of the identical marginal rates of factor substitution for the two industries, the marginal rate of product substitution in production (or marginal rate of transformation) is generally not equal to the price ratio. Totally differentiating (3.24) and using (3.29), we derive the following marginal rate of transformation:

$$\frac{p_i}{w} = \alpha_i(1-\alpha_i)^{1-\alpha_i} \frac{1}{w} \left[ \begin{array}{c} 1-\alpha_i \end{array} \right]^{-1} Y_i^{-\epsilon_i}, \quad \epsilon_i = \frac{\rho_i-1}{\rho_i}.$$

(3.30)
This possible analytic complication, however, does not seem worth pursuing, and we limit ourselves to the kind of localization economies which ensure equality between the marginal rate of transformation and the price ratio. This is possible by adding the following constraint:

\[
\rho = \rho_i \quad (i=1,2).
\]  

Under this assumption, of course, the ratio of marginal social costs (i.e., the slope of the production possibility curve) is equated to the ratio of marginal private costs (i.e., the price ratio which is also identical to the marginal rate of substitution in consumption). The producer equilibrium conditions of (3.29) under (3.33) now result in both production efficiency and product mix efficiency.

3-4. Market Equilibrium

Urbanization diseconomies introduced by the consumption sector, and localization economies introduced by the production sector can now be combined to show the effect of the net

---

6 This is mainly because our overriding concern is to investigate economic efficiency of different-sized cities and of alternative city size distributions rather than of different industries.
urbanization economies on the city economy in equilibrium. In closing the model, we will first determine the factor market equilibrium under the full employment assumption, including the city's industrial composition in terms of factors, i.e., the allocation of factors between the two industries. This will give the complete solution to the output market equilibrium. In the end, factors, products, and their prices can be solved for in terms of the city's endowments of capital and population. The prices may then be substituted into the indirect utility function which will reflect the net effect of urbanization economies and diseconomies in terms of the exogenous variables $K$ and $N$.

Our single-city model can be summarized into three sets of equations. From (3.15), the consumer equilibrium conditions,

$$p_i X_i = w \theta_i N \left( 1 - \frac{qN}{2x} + \frac{rK}{wN} \right) \quad (i=1,2), \tag{3.15}$$

from (3.29), the producer equilibrium conditions,

$$p_i Y_i = \frac{wL_i}{\alpha_i} = \frac{rK_i}{1-\alpha_i} \quad (i=1,2), \tag{3.29}$$

and the production possibility set, which consists of: from (3.24) and (3.33), the objective production functions; and from (3.18), the city's resource endowments.\(^7\)

\(^7\) For space saving in notation, we omit subscript $i$ in the $\Sigma$ sign hereafter.
\[ Y_i = \left( L_i^\alpha K_i^{1-\alpha} \right)^\rho, \quad \text{for } i = 1, 2. \quad (3.34) \]

\[ \sum L_i = NL \left( 1 - \frac{gN}{2\pi} \right), \quad \sum K_i = K, \]

The model is now closed by the condition that consumption of each good equals its local production, or,

\[ X_i = Y_i \quad \text{for } i = 1, 2. \quad (3.35) \]

Therefore, by equating (3.15) to (3.29), we get

\[ L_i = \sigma_i \theta_i N \left( 1 - \frac{gN}{2\pi} + \frac{R K}{w N} \right), \quad (3.36) \]

\[ K_i = (1 - \sigma_i) \theta_i \frac{w}{r} N \left( 1 - \frac{gN}{2\pi} + \frac{R K}{w N} \right), \quad \text{for } i = 1, 2. \]

This equilibrium allocation of factors, however, must also clear the factor markets. For this, we choose the labor market. The total equilibrium labor must be the same as the city-wide labor supply. Thus, by using (3.18), (3.34) and (3.36), we obtain

\[ L = L_1 + L_2 = \sum \sigma_i \theta_i N \left( 1 - \frac{gN}{2\pi} + \frac{R K}{w N} \right) = N \left( 1 - \frac{gN}{2\pi} \right) \quad \text{for } i = 1, 2. \quad (3.37) \]

This will also determine the equilibrium factor price ratio, or the ratio of capital rental to wage rate,

\[ \frac{r}{w} = \frac{\sum (1 - \sigma_i) \theta_i N \left( 1 - \frac{gN}{2\pi} \right) (K)^{-1}}{\sum \sigma_i \theta_i \left( 1 - \frac{gN}{2\pi} \right) (N)^{-1}} \quad \text{for } i = 1, 2. \quad (3.38) \]

which is now solely a function of exogenous variables and
parameters. By Walras' Law, the equilibrium factor price ratio guarantees that capital is also fully employed. By substituting (3.38) into (3.36), we determine \( L_i \) and \( K_i \) as follows:

\[
L_i = \frac{\alpha_i \theta_i N}{\sum \alpha_i \theta_i N \left( 1 - \frac{gN}{2} \right)} \quad (i=1,2),
\]

\[
K_i = \frac{(1-\alpha_i) \theta_i}{\sum (1-\alpha_i) \theta_i K} \quad (i=1,2).
\]

The above leads to this observation: in autarky equilibrium, the allocation of a factor to industry \( i \) is proportional to the weighted average of the factor's elasticity in that industry's subjective production function, the weights being the consumer's budget shares on all goods.

The output can now be obtained by substituting (3.39) into the objective production functions (3.34):

\[
Y_i = X_i = \left[ \left( \frac{\alpha_i}{\sum \alpha_i \theta_i} \right)^{1-\alpha_i} \left( \frac{1-\alpha_i}{\sum (1-\alpha_i) \theta_i} \right)^{1-\alpha_i} \left( \frac{K}{N} \right)^{1-\alpha_i} \theta_i N \left( 1 - \frac{gN}{2} \right)^{\alpha_i} \right]^\rho.
\]

\[(i=1,2) \quad (3.40)\]

The prices are now derived by substituting (3.38) and (3.40) into the demand functions (3.15):

\[
\frac{P_i}{w} = c_1 \left( \theta_i N \right)^{1-\rho} \left( 1 - \frac{gN}{2} \right)^{1-\rho \alpha_i} \quad (i=1,2) \quad (3.41)
\]

where 
\[
c_1 = \alpha_i^{-\rho \alpha_i} \left( \sum \alpha_i \theta_i \right)^{\rho \alpha_i - 1} \left[ \frac{\sum (1-\alpha_i) \theta_i}{1-\alpha_i} \right]^{\rho (1-\alpha_i)} \left( \frac{K}{N} \right)^{\rho (\alpha_i - 1)}.\]

Due to the homogeneous nature of our model, only the ratio \( p_i/w \),
that is the real price of good $i$ in terms of labor, is determinate.

We now show that the equilibrium is both unique and stable by means of the supply and demand curves. Combining (3.29), (3.34), and (3.38), we obtain the supply curves as in (3.30)

$$\frac{p_i}{w} = \alpha_i \frac{\sum \alpha_i \theta_i K - \frac{\theta_i N}{\sum (1-\alpha_i) \theta_i N}}{(1-\alpha_i) \left[ 1 - \frac{gN}{2\pi} \right] ^{1-\alpha_i} y_i (1-\rho)/\rho}$$

\begin{equation}
(i=1,2),
\end{equation}

and combining (3.15) and (3.38), we obtain the demand curves

$$\frac{p_i}{w} = \frac{\theta_i N \left[ 1 - \frac{gN}{2\pi} \right]^{1-\alpha_i} y_i}{\sum \alpha_i \theta_i X_i}$$

\begin{equation}
(i=1,2).
\end{equation}

Both curves and the solution are depicted in Fig. 3-2 for industry 1. The supply curve has constant elasticity $\rho/(1-\rho)$, and the demand curve has constant elasticity -1. Since $\rho/(1-\rho) < -1$ with $\rho > 1$, the supply curve is always flatter than the demand curve, and they meet exactly once. Therefore, the equilibrium is always unique and stable.\(^8\)

---

\(^8\) The Marshallian stability condition is met in this typical case under external economies. Since firms will raise (lower) their output when the excess demand is positive (negative) at a given price, the equilibrium will always be restored. Under the Walrasian stability condition in which the firms will adjust the price according to the excess demand, however, the equilibrium is unstable. Under our assumption that the firms do not recognize the function of (3.23), the Marshallian condition seems more appropriate.
Fig. 3-2: Supply and Demand Curves (i=1).

\[ \alpha_1 = 0.7 \quad \theta_1 = 0.5 \quad \rho = 1.04 \quad N = 1 \times 10^6 \quad K = 2 \times 10^6 \]

\[ \alpha_2 = 0.3 \quad \theta_2 = 0.5 \quad g/2\pi = 0.25 \times 10^{-7} \]

In equilibrium, \( X_1 = Y_1 = 1.122 \times 10^6 \), \( p_1/w = 0.869 \).
In order to check rough orders of magnitude for the equilibrium, some tentative values are assigned to the parameters which are presented in Fig. 3-2. First, we arbitrarily assume that the upper limit of N mentioned in (3.19) to be \(10^7\), which in turn determines that \(g/2\pi=0.25\times10^{-7}\). For a city of size \(N=1\times10^6\), for example, this means that 2.5 percent of the total time is used for commuting, and \(L=0.975\times10^6\). The critical value for our model, however, is to come from the degree of homogeneity \(\rho\). Although not directly applicable to our model, an empirical study by Shefer[1973] in estimation of localization economies reports median values of \(\rho\) to be roughly between 1.03 and 1.05. This study is a rare attempt to measure directly the extent of localization economies in industry production functions by applying the US SMSA data of the two-digit SIC manufacturing industries to a variant of the CES production function.\(^9\) Other US studies based on urbanization economies (thus less relevant to our study), however, report slightly different values (see Sveikauskas[1975], Segal[1976], Moomaw[1981]). We initially choose \(\rho=1.04\).

The values of labor elasticities in the subjective production function \(\alpha_1=0.7, \alpha_2=0.3\) were deliberately chosen to allow the two industries to have markedly different production

\(^9\) The abbreviations are as follows. 
SMSA: Standard Metropolitan Statistical Area. 
SIC: Standard Industrial Classification. 
CES: Constant Elasticity of Substitution.
technologies, which along with "mild" scale economies, generally tend to retain the familiar concavity in the production possibility curve despite scale economies (see Kemp[1964], Melvin[1969]). With neutral preferences between the two goods \( \theta_1=\theta_2=0.5 \), and \( K/N=2 \), we get for the first industry, for example, \( L_1=6.825\times10^5 \), \( K_1=6\times10^5 \), \( p_1/w=0.869 \), and \( X_1=Y_1=1.122\times10^6 \). The output is remarkably large, because under constant returns to scale, i.e. \( \rho=1 \), it would be only \( 0.657\times10^6 \). Despite our allowance for multiple industries in the city, the existence of only two industries in a big city seems to cause considerable specialization by each industry.

Equation (3.41) shows that the equilibrium price (of good i in terms of labor) is a function of per capita capital in the city \( K/N \), city size \( N \), and a set of production, consumption, and transportation parameters. When \( K/N \) increases at given \( N \), the price curve shifts down. This is a normal factor proportion effect which is further accentuated here with localization economies. Now if we isolate the effect of \( K/N \), the effect of city size on the price can be identified. Taking the derivative of (3.41) with respect to \( N \), while holding \( K/N \) fixed, we arrive at the following elasticity of price with respect to city size

\[
\frac{d}{dN} \left( \frac{p_i}{w} \right) \frac{N}{p_i} = -\frac{\rho(1-\alpha_i)\frac{gN}{2\pi}}{1-\frac{gN}{2\pi}} < 0. \tag{3.44}
\]

As expected from the elasticities of the supply and demand.
curves, the equilibrium price becomes a monotonically decreasing function of $N$ within a range which is much broader than, for example, the one in (3.19). With equi-proportional increases in $K$ and $N$, the city will experience unequivocal decreases in prices. Higher degrees of homogeneity $\rho$ and greater capital elasticities in the subjective production function $(1-\alpha_i)$ will increase the above elasticity in absolute terms, and accentuate the relationship. This relationship is shown in Fig. 3-3 for the first industry ($i=1$) with parameters $\alpha_i=0.7$ and $\rho=1.04$.

We are now ready to determine the equilibrium utility level in the city. Substituting the factor price ratio (3.38) and the prices of goods (3.41) back into the indirect utility function of (3.16), we get

$$v = c_2c_3N^{\rho-1} \left(1-\frac{gN}{2\pi}\right)^{\rho} \alpha_i \theta_i,$$

(3.45)

where:

$$c_2 = \left(\prod_{i}^{\alpha_i(1-\alpha_i)^{1-\alpha_i} \theta_i}\right)^{\rho} \theta_i,$$

$$c_3 = \left[\left(\prod_{i}^{\alpha_i \theta_i}\right)^{-\Sigma \alpha_i \theta_i (\Sigma (1-\alpha_i) \theta_i)^{-\Sigma (1-\alpha_i) \theta_i (K \Sigma (1-\alpha_i) \theta_i)}\right]^{\rho}.$$

Unlike the price case of (3.41) and Fig. 3-3 in which urbanization economies are largely reflected, the effect of $N$ on the utility level is not always beneficial. Here, urbanization diseconomies are also reflected in the form of reduced labor with

---

10 For space saving in notation, we omit subscript $i$ in the $\prod$ sign hereafter.
Fig. 3-3: Equilibrium Price with City Size ($i=1$).

\[
\begin{align*}
\alpha_1 &= 0.7 & \theta_1 &= 0.5 & \rho &= 1.04 \\
\alpha_2 &= 0.3 & \theta_2 &= 0.5 & K/N &= 2 & g/2\pi &= 0.25 \times 10^{-7}
\end{align*}
\]
increased city size, which not only decreases the total resource base of the city, but also increases the capital-labor ratio K/L. Since all production functions are assumed to be subject to decreasing returns to factor proportions, increase in N eventually leads to diminished industry output. Thus, the equilibrium utility level of (3.45) also represents the equilibrium net urbanization economies for the city. This relationship is presented in Fig. 3-4. With increase in N, the utility level initially goes up rather rapidly, attains a maximum, and eventually goes down in a less rapid, asymmetrical manner.

It is to be noted that each point on the curve of Fig. 3-4 indicates both an equilibrium and an optimum, not just in the private sense but, on the account of \( \rho = \rho_i \) (3.33), in the social sense as well. The curve shows, for the city in autarky, the highest level of welfare attainable under the production possibility set, which is determined by the given city size N and other internal characteristics of the city. On the other hand, the "optimum city size", which corresponds to \( N_o \) of Fig. 3-4, relates to the global maximum of the curve. This particular size \( N_o \) and the resulting production possibility set assure that the city output and welfare are maximized in real terms on a per capita basis. It is found by \( dv/dN = 0 \) from (3.45), while holding \( K/N \) fixed, and turns out to be

\[
N_o = \frac{2\pi}{g} \frac{\rho - 1}{(\rho - 1) + \rho \alpha_i \theta_i}. \tag{3.46}
\]

For the illustrative parametric values employed before (see Fig.
Fig. 3-4: Utility Level with City Size.

\[ \alpha_1 = 0.7 \quad \theta_1 = 0.5 \quad \rho = 1.04 \]

\[ \alpha_2 = 0.3 \quad \theta_2 = 0.5 \quad K/N = 2 \quad g/2\pi = 0.25 \times 10^{-7} \]

\[ N_0 = 2.86 \times 10^6 \quad v_0 = 1.324 \]
3-4), \( N_0 = 2.86 \times 10^6 \), to arrive at \( v_o = 1.324 \). This size \( N_0 \) is below the practical upper limit of \( N \) of (3.19), assumed to be \( 10^7 \), and therefore remains as a potential attainable target.

Finally, it remains to show that the equilibrium welfare level can actually be attributed to the underlying industrial composition of the city. Monotonically transforming the direct individual utility function (3.1) into the logarithms, and entering the output (3.40) into it, while recognizing \( x_i = X_i / N \), we get

\[
\log u = \sum \theta_i \log x_i = \sum \theta_i \left[ c_4 + \log \left( N^{a - 1} \left(1 - \frac{gN}{2a} \right)^{\rho \alpha_i} \right) \right],
\]

where \( c_4 = \rho \log \left[ \left( \frac{\alpha_i}{\sum \alpha_i \theta_i} \right)^{\alpha_i} \left( \frac{1 - \alpha_i}{\sum (1 - \alpha_i) \theta_i} \right) \left( \frac{K}{N} \right)^{1 - \alpha_i} \theta_i \right] \).

Equation (3.47) is presented in Fig. 3-5 in which the top "TOTAL" curve exactly corresponds to the one in Fig. 3-4, albeit in the logarithmic scale. The lower two industry curves represent utilities, or net urbanization economies attributable to the respective industries, i.e., \( \theta_i (\log x_i) \).

Thus, vertical summation of them yields the total net urbanization economies curve for the city as a whole. Of course, the industry curves attain maxima at different city sizes \( N_i \), which can be found by setting \( d(\log x_i) / dN = 0 \), while holding \( K/N \) constant:
Fig. 3-5: Net Urbanization Economies by Industry (i=1, 2).

\[ \alpha_1 = 0.7 \quad \theta_1 = 0.5 \quad \rho = 1.04 \]
\[ \alpha_2 = 0.3 \quad \theta_2 = 0.5 \quad K/N = 2 \quad g/2\pi = 0.25 \times 10^{-7} \]
\[ N_0 = 2.86 \times 10^6 \quad N_1 = 2.08 \times 10^6 \quad N_2 = 4.50 \times 10^6 \]
\[ N_i = \frac{2\pi}{g} \frac{\rho^{-1}}{\rho^{-1} - \rho a_i} \]  \hspace{1cm} (3.48)

For the parameters used in Fig. 3-5, \( N_1 = 2.08 \times 10^6 \), and \( N_2 = 4.5 \times 10^6 \), setting respectively the upper and lower bounds to \( N_0 = 2.86 \times 10^6 \). Because industries are differentiated only by \( a_i \), a labor-intensive industry will have a smaller \( N_i \) than a capital-intensive industry.\(^\text{11}\)

We have now completed modelling of our single-city that explicitly acknowledges the existence of multiple industries, and therefore builds upon the single-industry, single-city models of the supply-oriented approach. In particular, the industrial composition, or the output mix of the city, which has been determined by the weights from the utility function, gives rise to the level of welfare or of net urbanization economies in the city.

We have thus proposed that urbanization economies and diseconomies be measured in utility terms. This is in the tradition of general equilibrium. Unlike other single-city models, therefore, our model indicates that in equilibrium, the criteria of efficiency are met, i.e., efficiency in consumption, production, and output mix.

\(^{11}\) Although we have maintained the existence of only two industries mainly for convenience, it is clear that the model can operate under \( n \) industries. In that case, there will be a continuum of \( N_i \) corresponding to \( a_i \) (i=1,2,...,n), and \( N_0 \) will reflect all \( N_i \).
It is to be recalled that Isard's illustration of the relationship between net urbanization economies and city size as in Fig. 2-1, is a noble concept owing to the emphasis on industrial composition. However, it is incomplete as an analytic tool, due to its lack of weights with which the industry curves could be combined into the total net economies curve for the city. Moreover, the industry curves largely reflect the production situations only, and are measured in (nominal) dollar terms. Unless (nominal) incomes are identical across different city sizes, the curves do not reflect any welfare implications. It is clear that our use of the utility function in the measurement can solve these problems quite handily.

3-5. Migration and City Size Distribution

Like the models of the supply-oriented approach in general, our single-city model was concerned with one city in isolation, and not with its relationship with other cities. The equilibrium of the city was achieved under the exogenous resource endowments including population \( N \), and nowhere was there any consideration of a city size distribution in the national economy. In this section, however, we establish a limited degree of inter-dependence between cities and derive a resulting city size distribution by allowing free factor movements but not trade between cities. In the Heckscher-Ohlin, constant-returns-to-scale world, it is well known that perfect factor mobility and free trade are substitutes as equilibrating forces. However, they are not in
the world of urbanization economies and diseconomies in which not only factor proportions but also the scale of factors are the objects of the equilibrium. A complete balanced-trade model, which is defined later, will be the subject of the next chapter.

In addition to the previous assumptions, we assume that all cities share identical production, utility, and commuting cost functions. We further suppose that people migrate costlessly from city to city to maximize utility, that each migrant in the nation owns capital equally and takes with him his share of it, and that capital rentals as well as wages are spent in the city where people work and live. Notice that the movement of people does not change the ratio of capital to population in either the source city or the host city, so that resources (capital and population) are always distributed equally between cities on a per capita basis. The utilization of resources, however, will generally differ between cities of different sizes.

Now suppose initially that two cities A and B are engaged in the exchange of factors by following the rules listed above. With people (along with their capital) moving in response to utility differences, an equilibrium is reached when a common utility level is achieved in the two cities. Equation (3.45) of our single-city model is directly applicable, and the equilibrium conditions are

---

12 The constant term \( c_3 \) contains an element with \( K/N \). Since \( K_A/N_A = K_B/N_B = K/N \), however, \( c_3 \) is the same in both cities.
\[ v_A = c_2 c_3 N_A^{\rho - 1} \left( 1 - \frac{Q N_A}{2\pi} \right)^\rho \xi \alpha \beta, \quad v_B = c_2 c_3 N_B^{\rho - 1} \left( 1 - \frac{Q N_B}{2\pi} \right)^\rho \xi \alpha \beta, \]
\[ v_A = v_B, \quad N_A + N_B = \bar{N}, \quad (3.49) \]

where: \( v_A, v_B = \) utility levels in cities A and B respectively,
\( c_2, c_3 = \) constant terms of parameters, see (3.45),
\( N_A, N_B = \) populations of cities A and B respectively,
\( \bar{N} = \) population in the nation.

The shapes of the utility curves and the solution for (3.49) generally depend on the national population parameter \( \bar{N} \) as it relates to the optimum city size \( N_0 \) of the single-city model. When \( \bar{N} \) is sufficiently small so that \( \bar{N} < 2N_0 \), a unique equilibrium is reached at point P in Fig. 3-6, in which the horizontal axis is fixed of length \( \bar{N} \), and measures \( N_A \) from left to right and \( N_B \) the other way around. Due to the symmetry between the curves \( v_A \) and \( v_B \), which are adapted from Fig. 3-4, the equilibrium city sizes \( N_A \) and \( N_B \) are identical but not greater than \( N_0 \) (\( N_A = N_B \leq N_0 \)). Because P is on the non-falling part of both \( v_A \) and \( v_B \) curves (i.e., net urbanization economies are not decreasing in both cities), however, we should suspect that this equilibrium is generally unstable, because slight migration from city B to A due to an exogenous shock will increase \( v_A \) as well as decrease \( v_B \).\(^{13}\)

The gap between \( v_A \) and \( v_B \) then becomes cumulative as the migra-

\(^{13}\) Due to the symmetry between the \( v_A \) and \( v_B \) curves, we will consider henceforth only the case in which \( N_A > N_B \).
Fig. 3-6: Equilibrium City Sizes ($\bar{N} \leq 2N_0$).

\[ a_1 = 0.7 \quad \theta_1 = 0.5 \quad \rho = 1.04 \]
\[ a_2 = 0.3 \quad \theta_2 = 0.5 \quad \bar{K}/\bar{N} = 2 \quad g/2\pi = 0.25 \times 10^{-7} \]
\[ N_0 = 2.86 \times 10^6 \quad \bar{N} = 5 \times 10^6 \quad N_A = N_B = 2.5 \times 10^6 \]
tion progresses. The migration will continue beyond Q at which \( v_A \) attains its maximum and \( N_A \) becomes \( N_0 \). It is also clear that even when \( N_A = N_B = N_0 \), the equilibrium is unstable due to the same disequilibrating migratory flow. Of course, when \( \bar{N} \) is even smaller, for example \( \bar{N} < N_0 \), there is no justification for the existence of the two cities, and the migration will proceed until city B is totally vacated so that \( N_A = \bar{N} < N_0 \). However, this is an extreme case which can hardly be imaginable in reality. To sum up, it is therefore clear that a unique and stable equilibrium in the city size distribution when \( \bar{N} < 2N_0 \) is not possible. A numerical example is given in fig. 3-6 for the unstable equilibrium at P.

When \( \bar{N} \) is larger, so that \( \bar{N} > 2N_0 \), multiple equilibria appear to be reached at three different sets of city sizes such as points P, Q, and R in Fig. 3-7. At Q which is on the rising part of \( v_B \) but also on the falling part of \( v_A \), the equilibrium city sizes are different for the two cities, the magnitude of the difference being determined by the difference between \( \bar{N} \) and \( 2N_0 \). For the parametric values used in Fig. 3-7, it represents a highly skewed city size distribution at \( N_A = 9.8 \times 10^6 \), and \( N_B = 0.2 \times 10^6 \). However, the equilibrium at Q is most unstable, because any movement toward another equilibrium point P unambiguously improves the utility levels for both cities. Migration from city A to B, once started, will proceed until point P is reached at which \( N_A = N_B > N_0 \). The equilibrium at P which is on the falling part of both \( v_A \) and \( v_B \), however, must be a stable
Fig. 3-7: Equilibrium City Sizes ($\bar{N} > 2N_0$).

\[ \sigma_1 = 0.7 \quad \theta_1 = 0.5 \quad \rho = 1.04 \quad \frac{g}{2x} = 0.25 \times 10^{-7} \]

\[ \sigma_2 = 0.3 \quad \theta_2 = 0.5 \quad \bar{K}/\bar{N} = 2 \quad \bar{N} = 10 \times 10^6 \]

At P, $N_A = N_B = 5 \times 10^6$; at Q, $N_A = 9.8 \times 10^6$, $N_B = 0.2 \times 10^6$
A slight deviation from P toward Q will be met by counter-migration from city A to B, thereby restoring the equilibrium. When \( \bar{N} > 2N_0 \), therefore, a stable equilibrium is reached at P where city sizes are identical but larger than \( N_0 \) (\( N_A = N_B > N_0 \)). Since the common utility level in the equilibrium falls as \( \bar{N} \) increases, however, there must be a practical limit in this equilibrium. It would be unreasonable to operate under the confines of two cities in the nation when \( \bar{N} \) is much greater than \( 2N_0 \).

If we extend the above further to a more realistic and general situation in which \( \bar{N} \) is much larger than \( 2N_0 \) and the existence of more than two cities are considered, however, we suspect that the stable equilibrium city size will be \( N_0 \) for all cities in the nation. As Henderson[1977] argues, this is because with \( \bar{N} \) being sufficiently large, the divisibility problem due to lumpiness can be avoided in replication of cities with size \( N_0 \). For example as in Fig. 3-7, suppose there are many cities with sizes slightly larger and smaller than \( N_0 \). With many cities in existence, the sum of any "excess" populations beyond \( N_0 \) from the larger cities could be added to the smaller cities, thereby helping to achieve \( N_0 \) among all cities. Thus, all cities eventually converge \( N_0 \) in size.

It is argued that this rather "trivial" distribution of city sizes in equilibrium is an inherent phenomenon in the models of the supply-oriented approach. The lack of opportunity to trade goods forces the city to produce locally what it consumes under
the equilibrium conditions that reflect the internal characteristics only. With no consideration given to other cities in both production and consumption, all cities tend to be "standardized", and are required to converge on the ideal city size $N_0$ for equilibrium. This is in a sharp contrast with the reality in general, and the trade model in particular, which is to be developed in the next chapter where cities of continuously different sizes can coexist and benefit from each other with trade albeit in a suboptimal manner.
4. A TRADE MODEL WITH SCALE ECONOMIES

4-1. Introduction

The purpose of this chapter is to allow trade between the two cities of our single-city type, and to review the resulting city size distribution in the nation. As a general equilibrium approach, this is an attempt to integrate both the supply- and demand-oriented approaches to the determination of city sizes and output in general and to the equilibrium in the inter-city labor markets in particular. Under the assumption that the nation consists of the two cities, the inter-city terms of trade \((p_2/p_1)\) becomes endogenous, and is to be determined in equilibrium. As such, we will retain all the assumptions of the previous chapter but the no-trade provision, and add more specific assumptions in due course. While housing consumption, commuting, and the labor supply will still depend on the city's internal characteristics even with trade, the production and consumption of traded goods and the level of welfare will also depend on those of the trading partner. In this way, the interdependence between the two cities becomes more direct and realistic than that of the simple migration model of the last chapter.

This will be carried out in two steps. First, we will develop a balanced free-trade model which meets such requirements as free trade, zero total value of excess demands in all markets within the nation (or Walras' Law), and no inter-city investment
of capital. In this step, migration is not allowed, and therefore city sizes are fixed. Under urbanization economies and diseconomies, although commodity prices will be the same in equilibrium, factor prices, incomes, and utility levels will in general differ between the two cities. The next step then will be to allow factor movements to arrive at a common level of utility between the two cities. Given the total endowments of population (and capital) in the nation, this will allow us to derive a system of cities with generally different sizes in equilibrium. Despite scale economies, the model will operate much in a similar manner as the standard trade model of factor proportions and comparative advantage. This is the result of our assumption to be made in which "mild" scale economies for both industries give rise to incomplete specialization for the city.

4-2. Production under "Mild" Scale Economies

When the two traded goods are produced under increasing returns to scale within the city, it is important initially to assess the degrees of scale economies from the society's point of view in each city. If they are sufficiently large and "severe", the marginal social (as well as private) opportunity cost that

---

1 Following the tradition in the international trade literature, the term "balanced" here means that no inter-city transfer (investment, reparations, and gifts) of capital is taking place. Under our assumption that capital and its rentals stay where its owners work and live, the requirement is always met.

2 It is to be recalled that under our assumption of (3.33), (i.e., \( \rho = \rho_i \), \( i=1,2 \)), the ratio of marginal social costs is the
measures the slope of the production possibility curve of the
city decreases, and the production possibility curve has a
convex, or "bowed-in" shape. This generally happens when the
difference in factor intensities of the two production functions
is relatively small compared to the degrees of scale economies.
Under the circumstances, specialization in any good pays both so-
cially and privately. Scale then becomes the dominant basis for
trade, and the usual stable equilibrium results in complete spe-
cialization by each city in one yet indeterminate good. This
situation, however, seems hardly plausible in the case of exter-
nal economies, because it would be naive to assume that the
entrepreneurs behave competitively despite such high degrees of
scale economies. This is especially so in the urban context, if
one accepts industrial diversity as a fundamental characteristic
of cities, particularly the largest ones.

On the other hand, "mild" degrees of scale economies combin-
ed with a large difference in the factor intensities yield the
familiar concave or "bowed-out" production possibility curve for

---

3 An indeterminate pattern of trade and complete specialization
are the usual outcomes when trade is caused by economies of
scale. For example, Melvin[1969] arrives at an equilibrium in
which two identically endowed trading partners end up with
having markedly different utility levels by complete specializa-
tion. Henderson's single-industry model, on the other hand,
tacitly assumes "severe" degrees of scale economies, and starts
from complete specialization. He, for example, uses a high
degree of homogeneity of 1.25 (akin to \( \rho \) in our model) in a
numerical example (see Henderson[1977], p.680).
the most part except for the extreme points close enough to both axes. Now comparative advantage arising from different factor proportions between cities of different sizes becomes the basis for trade. Because the marginal social opportunity cost is rising in this case, the equilibrium pattern of trade, under neutral demands for the two goods, will generally show a more modest specialization in the good for which the city has comparative advantage in production and export. In short, despite scale economies, incomplete specialization is more likely under moderate degrees of scale economies. In light of the empirical evidence pointing to the relatively low degrees of homogeneity (on the order of \( \rho = 1.03-1.05 \), and certainly much less than 1.1, for example, see section 3-3), and of the observed urban industrial diversity in reality, we assume that the city is subject to the concave production possibility curve at least for its relevant part in trade.

Because this assumption is critical to our model, we now probe it further. Specifically, the assumption of the concave production possibility curve requires that \( \frac{d^2 Y_1}{dY_2^2} = \frac{d(-p)}{dY_2} < 0 \).

---

4 Differences in tastes between cities can of course be another basis for trade. Because our focus here is on production under scale economies, we make explicit this assumption of demand neutrality between cities to separate the influence of tastes and to determine the pattern (or direction) of trade. This is also the usual practice in the Heckscher-Ohlin model in determining the pattern of trade. See Melvin[1969], Layard and Walters([1978], pp. 113-119). In our model, because cities are assumed to have identical log-linear utility functions, demands are apparently neutral, and they spend the same proportions of income on each good at any prices.
according to (3.32) and (3.33), where \( p \) is the price of the second good in terms of the first. For the representative city engaged in trade, using the marginal productivity conditions and full employment conditions, we first express all endogenous variables on the production side in terms of a single variable \( q \), defined below, and exogenous variables.

From the producer equilibrium conditions of (3.29), the factor price ratio is equal to the marginal rates of substitution in production, which, in turn, depend on the factor ratios alone, regardless of the scale of output, in both industries:

\[
\frac{r}{w} = \frac{1-a_i}{a_i} \left( \frac{K_i}{L_i} \right)^{-1}.
\]

(4.1)

Along with the full employment conditions of factors from (3.34), and letting

\[
q = \frac{1}{2 \pi} \left( 1 - \frac{gN}{2 \pi} \right) \left( \frac{K}{N} \right)^{-1},
\]

or,

\[
q = \frac{1 - \frac{gN}{2 \pi} + \frac{rK}{WN}}{1 - \frac{gN}{2 \pi}},
\]

(4.2)

where: \( 1 < \frac{1}{\alpha_1} < q < \frac{1}{\alpha_2} \), \( 0 < \alpha_2 < \alpha_1 < 1 \),

we derive the following:

\[\text{------}\]

\[\text{------}\]

5 Because of this reason and the resulting analytic convenience, we have elected to use either \( r/w \) or \( q \) as the key endogenous variable.
\[ L_1 = \frac{\alpha_1(1-qa_2)}{\alpha_1-a_2} N \left(1-\frac{gN}{2\pi}\right), \quad L_2 = \frac{\alpha_2(qa_1-1)}{\alpha_1-a_2} N \left(1-\frac{gN}{2\pi}\right), \]  
\[ \frac{K_i}{L_i} = \frac{1-a_i}{\alpha_i} \frac{K}{N} (q-1)^{-1} \left(1-\frac{gN}{2\pi}\right)^{-1}. \]  

Finally substituting (4.3) into (3.29) and (3.34), the remaining endogenous variables are expressed in \( q \):  

\[ Y_1 = \left[ a_1^\alpha_1 (1-\alpha_1)^{1-\alpha_1} \left(\frac{K}{N}\right)^{1-\alpha_1} \frac{N}{\alpha_1-a_2} \left(1-\frac{gN}{2\pi}\right)^{\alpha_1(1-qa_2)} (q-1)^{\alpha_1^{-1}} \right]^{\rho}, \]  
\[ Y_2 = \left[ a_2^\alpha_2 (1-\alpha_2)^{1-\alpha_2} \left(\frac{K}{N}\right)^{1-\alpha_2} \frac{N}{\alpha_1-a_2} \left(1-\frac{gN}{2\pi}\right)^{\alpha_2(qa_1-1)} (q-1)^{\alpha_2^{-1}} \right]^{\rho}, \]  
\[ p = c_1 \left[ \frac{K}{N} \right]^{-\rho (\alpha_1-\alpha_2)} \left[ \left(1-\frac{gN}{2\pi}\right)(q-1) \right]^{\rho (\alpha_1-\alpha_2)} \left[ \frac{1-qa_2}{q(a_1-1)} \right]^{\rho-1}, \]  

where: \[ c_1 = \left[ \frac{\alpha_1^\alpha_1 (1-\alpha_1)^{1-\alpha_1}}{\alpha_2^\alpha_2 (1-\alpha_2)^{1-\alpha_2}} \right]^{\rho}. \]

Because \( K/N \) is assumed to be fixed throughout, for the city of a given size \( N \), \( (1-gN/2\pi) \) is constant, and \( q \) becomes a measure of per capita total income deflated by the wage rate. Similarly, \( (q-1) \) measures the ratio of factor rewards \( r/w \). As (4.2) indicates, when the city is completely specialized in the labor-intensive first good (or capital-intensive second good), \( q \) becomes \( 1/\alpha_1 \) (or \( 1/\alpha_2 \)). When it is incompletely specialized, \( q \) takes an intermediate value between the two. In the autarky

---

6 The price ratio \( p \) can more readily be obtained from (4.8), soon to follow, as the ratio of the marginal product of labor (or capital) in the first industry to that in the second industry.
equilibrium derived before in (3.38), \( q = 1 / \sum \alpha_i \theta_i \), \( \sum \theta_i = 1.7 \)

Differentiating the endogenous variables with respect to \( q \), we derive the following elasticities:

\[
\frac{dY_1}{dq} \frac{q}{Y_1} = -\rho q \left( \frac{\alpha_2}{1-q \alpha_2} + \frac{1-\alpha_1}{q-1} \right) < 0,
\]

(4.6)

\[
\frac{dY_2}{dq} \frac{q}{Y_2} = \rho q \left( \frac{\alpha_1}{q \alpha_1 - 1} - \frac{1-\alpha_2}{q-1} \right) = \rho q \left( \frac{1-\alpha_1 + \alpha_2 (q \alpha_1 - 1)}{(q \alpha_1 - 1)(q-1)} \right) > 0,
\]

\[
\frac{dp}{dq} \frac{q}{p} = q (\alpha_1 - \alpha_2) \left( \frac{\rho}{q-1} - \frac{\rho-1}{(q \alpha_1 - 1)(1-q \alpha_2)} \right) > 0.
\]

Since \( d(-p)/dY_2 = -(dp/dq)(dq/dY_2) \), and \( dY_2/dq > 0 \) from (4.6), if \( dp/dq > 0 \), which from (4.7) is very likely when \( \rho - 1 \) is sufficiently small and close to zero, then \( d(-p)/dY_2 < 0 \) and the production possibility curve becomes concave. It is to be noted that under constant returns to scale, i.e., \( \rho - 1 = 0 \), the price ratio of (4.5) depends solely on the factor price ratio (the terms with the exponents \( \rho (\alpha_1 - \alpha_2) \)) that determines the factor proportions

\[7\] For the purpose of comparative statics, however, \( q \) is a more convenient measure than \( (q - 1) \) in deciding the direction of change in \( r/w \) as well as the degree of specialization by the city between the two goods.

\[8\] The elasticity of the price ratio with respect to the factor price ratio \( r/w \) must lie between zero and one, because from (4.2) and (4.7),

\[
0 < \frac{dp}{d(r/w)} \frac{r/w}{p} = \rho (\alpha_1 - \alpha_2) - \frac{(\rho - 1)(q - 1)}{(q \alpha_1 - 1)(1 - q \alpha_2)} < \rho (\alpha_1 - \alpha_2) < 1.
\]

Thus, changes in the price ratio will have a magnifying effect on the factor price ratio.
via (4.3). Under increasing returns to scale, however, it also depends on the relative degree of specialization in production, or the allocation of factors between the two goods (measured by the term with the exponent \((\rho-1)\)). Thus, under the assumption of small \((\rho-1)\), the factor proportions effect will tend to dominate over the "specialization effect".

In order to illustrate the above further, and for later use, there is a need to examine the responses of the marginal products of factors with respect to the change in \(q\). First, from the marginal productivity conditions (3.26), and (4.3), we express them in terms of \(q\),

\[
\frac{r}{p_1} = c_2 N^{\rho-1} \left(1 - \frac{gN}{2}\right) \rho^{\alpha_1} (q-1)^{1-\rho} (1-q^{\alpha_2})^{\rho-1},
\]

\[
\frac{r}{p_2} = c_3 N^{\rho-1} \left(1 - \frac{gN}{2}\right) \rho^{\alpha_2} (q-1)^{1-\rho} (1-q^{\alpha_2})^{\rho-1},
\]

\[
\frac{w}{p_1} = c_2 \left(\frac{K}{N}\right) N^{\rho-1} \left(1 - \frac{gN}{2}\right) \rho^{\alpha_1-1} (q-1)^{\rho} (1-q^{\alpha_2})^{\rho-1},
\]

\[
\frac{w}{p_2} = c_3 \left(\frac{K}{N}\right) N^{\rho-1} \left(1 - \frac{gN}{2}\right) \rho^{\alpha_2-1} (q-1)^{\rho} (1-q^{\alpha_2})^{\rho-1},
\]

---

9 We are dealing here with the marginal private products of factors. However, the marginal social products, which are greater than the corresponding private products by a factor of \(\rho\) according to (3.27) and (3.33), can also be used. Of course, the elasticities with respect to \(q\) are the same in both measures.
where: 

\[
c_2 = \left[ a_1^{\alpha_1} (1-\alpha_1)^{1-\alpha_1} \right]^{\rho} (\alpha_1 - \alpha_2) \left\{ 1 - \rho \left( \frac{K}{N} \right) \right\}^{\rho(1-\alpha_1)-1},
\]

\[
c_3 = \left[ a_2^{\alpha_2} (1-\alpha_2)^{1-\alpha_2} \right]^{\rho} (\alpha_1 - \alpha_2) \left\{ 1 - \rho \left( \frac{K}{N} \right) \right\}^{\rho(1-\alpha_2)-1}.
\]

We now consider an extension of the Stolper-Samuelson theorem on the terms of trade and the marginal products of factors under constant returns to scale. It is still applicable to our model. Suppose now that the city, which is currently incompletely specialized, is larger in size than its trading partner, and that, owing to its higher K/L ratio, exports the capital-intensive second good. Suppose also that the terms of trade, \( p \), is slightly improved. If, despite scale economies, the production possibility curve is concave, the production of the second good, \( Y_2 \), will rise, and that of the first good, \( Y_1 \), will fall in the new equilibrium, thereby increasing the \( r/w \) ratio. Then, the marginal product of capital which is employed relatively intensively by the city must rise, and that of labor must fall in both industries as in the following:

\[
\frac{d(r/p_1)}{dq} = q \left[ \frac{1-\rho(1-\alpha_1)}{q-1} - \frac{(\rho-1)\alpha_2}{1-q\alpha_2} \right] > 0,
\]

\[
\frac{d(r/p_2)}{dq} = q \left[ \frac{1-\rho(1-\alpha_2)}{q-1} + \frac{(\rho-1)\alpha_1}{q\alpha_1-1} \right] > 0,
\]

\[
\frac{d(w/p_1)}{dq} = -q \left[ \frac{\rho(1-\alpha_1)}{q-1} + \frac{(\rho-1)\alpha_2}{1-q\alpha_2} \right] < 0,
\]

\[
\frac{d(w/p_2)}{dq} = -q \left[ \frac{\rho(1-\alpha_2)}{q-1} - \frac{(\rho-1)\alpha_1}{q\alpha_1-1} \right] < 0.
\]

This is because from the definition of \( p \),
\[
\frac{dp}{dq} \frac{q}{p} = \frac{d(r/p_1)}{dq} \frac{q}{r/p_1} - \frac{d(r/p_2)}{dq} \frac{q}{r/p_2} \\
= \frac{d(w/p_1)}{dq} \frac{q}{w/p_1} - \frac{d(w/p_2)}{dq} \frac{q}{w/p_2}.
\]

(4.10)

and as long as \(dp/dq > 0\) as assumed, the square-bracketed terms of (4.9) must be positive.\(^{10}\) Again, it is apparent that this depends on \((p-1)\) being sufficiently small. Under the assumption of the concave production possibility curve, the endogenous variables are largely subject to the "factor proportions effect", and, therefore, our model behaves much in line with the traditional factor proportions model under constant returns to scale. More specific aspects of our model, including incomplete specialization, will be discussed in the following section.

4-3. A Balanced Free-Trade Model

In this section, we consider early-defined balanced free-trade between the two cities A and B of fixed sizes that comprise the nation. The cities share the identical industry production functions (4.4), and are subject to their own concave production possibility curves for the relevant ranges in trade as delineated before. They also share the identical demand functions (3.15) in which the expenditure for each good is a fixed proportion of the

\(^{10}\) The elasticities of the marginal products with respect to a change in the factor price ratio \(r/w\), on the other hand, must lie between plus one and minus one. Refer to (4.2) and (4.9).
income. We assume that city A is larger than B, i.e., \( N_A > N_B \).

One of the general equilibrium conditions is that the terms of trade be equal. Rearranging the terms of \((4.5)\), and assigning \( q_A \) and \( q_B \) for \( q \) to cities A and B respectively, we obtain

\[
\begin{align*}
\frac{p_A}{c_A} &= c_4 \left[ \left(1 - \frac{q_{NA}}{2\pi} \right) (q_A - 1) \right]^{\rho (\alpha_1 - \alpha_2) \frac{1 - q_A \alpha_2}{q_A \alpha_2 - 1}} \rho - 1, \\
\frac{p_B}{c_B} &= c_4 \left[ \left(1 - \frac{q_{NB}}{2\pi} \right) (q_B - 1) \right]^{\rho (\alpha_1 - \alpha_2) \frac{1 - q_B \alpha_2}{q_B \alpha_2 - 1}} \rho - 1, \\
\frac{p_A}{p_B} &= \left( \frac{q_{A}}{q_{B}} \right)^{1 - \frac{q_{NA}}{2\pi}} \frac{K_A}{N_A} \left(1 - \frac{q_{A}}{2\pi} \right) \frac{K_B}{N_B} \left(1 - \frac{q_{B}}{2\pi} \right). \tag{4.11}
\end{align*}
\]

where: \( p_A, p_B = \) price ratios of good 1 to good 2 in A and B,

\[
c_4 = \frac{c_2}{c_3} = \left[ \frac{\alpha_1 (1 - \alpha_1) - \alpha_2 (1 - \alpha_2)}{\alpha_1 (1 - \alpha_1) - \alpha_2 (1 - \alpha_2)} \rho \frac{\bar{K}}{N} - \rho (\alpha_1 - \alpha_2) \right],
\]

\[
\frac{r_A}{w_A} = (q_A - 1) \left(1 - \frac{q_{NA}}{2\pi} \right) \frac{K_A}{N_A}^{-1}, \quad \frac{r_B}{w_B} = (q_B - 1) \left(1 - \frac{q_{NB}}{2\pi} \right) \frac{K_B}{N_B}^{-1}. \tag{4.12}
\]

\( r_A, r_B, w_A, w_B = \) capital rentals, wage rates of A and B.

Equation \((4.11)\) shows that the price ratios are monotonically increasing functions of \( q \) under our assumption of the concave production possibility curves, i.e., \( \frac{dp}{dq} > 0 \) from \((4.7)\).

It is immediately clear that if city sizes are identical, \( N_A = N_B \), then the autarky price ratios are the same, i.e., \( p_A = p_B \) at \( q_A = q_B = 1/\Sigma \alpha_i \theta_i \), and therefore there is no basis for trade. The identical cities will remain in autarky. Only when city sizes are different, the autarky price ratios become different, and they will set the limits for the equilibrium terms of trade. As \( N_A \) is assumed to be greater than \( N_B \), \( p_A \) is lower than \( p_B \) in
autarky (q_A=q_B=1/\sum a_i \theta_i), and they become the lower and upper limits respectively. In equilibrium, therefore, q_A must be greater and q_B must be less than 1/\sum a_i \theta_i. This, along with (4.2), gives the ranges of q_A and q_B:

\[ 1 < \frac{1}{a_1} \leq q_B \leq \frac{1}{\sum a_i \theta_i} \leq q_A \leq \frac{1}{a_2}. \quad (4.13) \]

From the response of the input of (4.3), it is clear that changes in q_A and q_B directly affect the allocation of factors between the industries toward the direction that reflects the comparative advantage of each city. Moreover, according to the response of the output Y_i already noted in (4.6), the deviations of q_A and q_B away from their identical autarky value in (4.13) represent the relative degrees of specialization in equilibrium by cities A and B respectively: the relatively capital-rich city A produces more of the capital-intensive second good (and less of first good), and the labor-rich city B produces more of the labor-intensive first good (and less of second good) than under autarky.

Because the equilibrium terms of trade must lie within the autarky price ratios of the two cities, incomplete specialization by both cities is practically assured. This is because one city's degree of specialization, in the sense of the shift from autarky to the equilibrium production in trade, is to be limited by that of the other. In short, specialization is limited by the extent of the market. An illustration is given by Fig. 4-1-1 in which p_A and p_B are related to q respectively for the renewed parametric values of \( N_A=4\times10^6 \), \( N_B=0.4\times10^6 \), \( \rho=1.03 \), \( a_i=0.8 \),
Fig. 4-1-1: Price Ratio vs. q.

\[ \alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{q}{2\pi} = 0.25 \times 10^{-7} \]
\(a_1=0.2, \theta_1=\theta_2=0.5, g/2\pi=0.25\times 10^{-7}, \bar{K}/\bar{N}=2\). The \(p_A\) curve is located lower than the \(p_B\) curve because \(N_A>N_B\). \(S\) and \(T\) being the autarky equilibrium points for cities \(A\) and \(B\) respectively \((q_A=q_B=2.0)\), any value between them is a possible candidate for the equilibrium terms of trade.\(^{12}\) Thus, maximum \(q_A\) is attained when city \(A\) moves to \(U\) \((q_A=2.11, q_B=2.0)\), and minimum \(q_B\) is attained when city \(B\) moves to \(V\) \((q_A=2.0, q_B=1.90)\) at the parametric values of \(N_A\) and \(N_B\). To test the sensitivity of the model, we now vary the initial parametric values.

In the extreme case in which trade occurs between practically the largest and smallest cities \((N_A=10^7, N_B=10^4,\) for example), maximum \(q_A\) is 2.367 and minimum \(q_B\) is 1.731. Because the price ratio is a decreasing function of city size, the greatest deviation of either \(q_A\) or \(q_B\) from its autarky value is realized when the difference in city sizes is greatest. Even in this case, incomplete specialization by both cities is guaranteed, because

\(^{11}\) The values of parameters \(a_i\) and \(\rho\) are slightly changed here partly to reflect the assumed concavity of the production possibility curves. As we shall see later, however, this change is not very critical in the behavior of the model.

\(^{12}\) Any line parallel to the horizontal axis and cutting the line segment \(ST\) will also meet the \(p_A\) and \(p_B\) curves exactly once and may determine the equilibrium. The equalization of the price ratios, however, is only a necessary condition, and the equilibrium also requires the balance of trade.
\[ 2.367 = q_A < \frac{1}{\alpha_2} = 5.0, \quad 1.25 = \frac{1}{\alpha_1} < q_B = 1.731. \quad (4.14) \]

It is to be noted that at the previous parametric values of \( \rho=1.04, \alpha_1=0.7, \) and \( \alpha_2=0.3 \) which reflect more "severe" scale economies for both cities, the extreme city sizes yield maximum \( q_A=2.462 \) and minimum \( q_B=1.684 \). Although even greater specialization is now potentially possible, both cities will still retain concavity in the production possibility curves over the possible values of \( q_A \) and \( q_B \), \((q_A=2.0-2.462, q_B=1.684-2.0)\), because on the account of (4.7) the elasticities of the price ratios are still greater than zero:

\[ 0 < 0.551 < \frac{dp_A}{dq_A \ p_A} < 0.632 < \frac{dp_B}{dq_B \ p_B} < 0.719 < 1. \quad (4.15) \]

The elasticities are also less than 1, indicating that \( p_A \) and \( p_B \) are strictly concave functions of \( q_A \) and \( q_B \) respectively. Finally, (4.15) shows that the \( p_A \) curve must be less elastic than the \( p_B \) curve with respect to \( q_A \) and \( q_B \) in equilibrium.

As in (4.14), both cities will still be incompletely specialized in equilibrium due to the fact that the maximum \( q_A \) is less than \( 1/\alpha_2 \) (\( q_A=2.462<3.333=1/\alpha_2 \)), and the minimum \( q_B \) is greater than \( 1/\alpha_1 \) (\( q_B=1.684>1.429=1/\alpha_1 \)). It is thus clear that our assumption of concave production possibility curves and incomplete specialization is sustained under a wide range of city sizes, and that the behavior of our model is stable at varying parameter values.
Fig. 4-1-2, on the other hand, illustrates a linear functional relationship between the factor price ratio and $q$, i.e., $q_A$ or $q_B$, of (4.12) under the parametric values therein. Due to its larger city size and thus greater urbanization diseconomies, city A's factor price ratio, $r_A/w_A$, in autarky at $S$ is lower than that of city B at $T$. In trade equilibrium, of course, $r_A/w_A$ will rise up and $r_B/w_B$ will fall down along their respective lines as the arrow marks indicate.

However, will the factor price ratios be equalized in equilibrium? It is well known that according to the Heckscher-Ohlin model under constant returns to scale, trade equilibrium ensures the equalization of both factor prices and factor price ratios (more accurately both absolute and relative factor rewards$^{13}$) between the trading partners. Under increasing returns to scale, however, this will generally not happen.$^{14}$ Indeed, (4.11) and (4.12) show that $r_A/w_A$ will be higher than $r_B/w_B$ in equilibrium, reversing the autarky situation. Thus, the equilibrium will not be strictly Pareto optimal within the national economy, in the sense that the marginal rate of substitu-

---

$^{13}$ Although we use the terms factor "price" and factor "reward" interchangeably to denote wage rates and capital rentals, our use of the term "price" is simply traditional. Since the prices of capital or labor services are what we are after rather than the prices of capital or labor per se, the term "reward" is more appropriate. See Kemp([1964], pp. 48-49).

$^{14}$ Urbanization diseconomies, which merely determine the labor supplies of the cities, are not related to this phenomenon.
Fig. 4-1-2: Factor Price Ratio vs. q.

\[ \sigma_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{g}{2x} = 0.25 \times 10^{-7} \]
tion between factors, or the factor price ratio under perfect competition, is not the same between the cities. However, to the extent that \((p-1)\) is sufficiently small in (4.11), the specialization effect tends to be minimal, and there remains a strong possibility that the difference in the factor price ratios will be reduced in equilibrium. This will be probed later.

In addition to (4.11), equilibrium requires that trade be balanced. Because inter-city investment and transport costs are not considered, this requires that the national demand for any good must be equal to the national supply. Under (4.11) and according to Walras' law, we are free to choose either good for the balance of trade, and for this purpose we choose the second good. Already knowing the pattern of trade under neutral demands for the goods in both cities, we define the net excess supply of this good by city A, \(E_{2A}\), as the (nonnegative) difference between the local supply, \(Y_{2A}\), and the local demand, \(X_{2A}\). From (3.15), (3.29) and (4.3), this is calculated to be:

\[
E_{2A} = Y_{2A} - X_{2A} = \frac{w_A}{p_{2A}} \frac{q_A \bar{\theta}_i \theta_i - 1}{\theta_i - \bar{\theta}_i} \sum_{i=1}^{q_A} \left( 1 - \frac{\rho^A}{\rho^A} \right) \geq 0.
\]  

(4.16)

Substituting the marginal product of labor in the second industry, \(w_A/p_{2A}\), from (4.8) into the above, we express \(E_{2A}\) in terms of \(q_A\),\(^{15}\)

\(^{15}\)For completeness, the net excess demand for the first good by city A, \(E_{1A}\), is

\[
E_{1A} = X_{1A} - Y_{1A} = \frac{w_A}{p_{1A}} \frac{q_A \bar{\theta}_i \theta_i - 1}{\theta_i - \bar{\theta}_i} \sum_{i=1}^{q_A} \left( 1 - \frac{\rho^A}{\rho^A} \right) \geq 0.
\]
\[ E_{2A} = c_5 N_A^\rho \left( 1 - \frac{\rho N_A}{2\pi} \right) \rho \alpha_2 (q_A - 1) \rho (\alpha_2 - 1) \rho (q_A^\alpha_1 - 1) \rho -1 (q_A^\alpha_1 \theta_1 - 1), \]  
(4.17)

where: \( c_5 = \left[ \alpha_2^\alpha_2 (1 - \alpha_2)^{1 - \alpha_2} \right] \rho (\alpha_1 - \alpha_2)^-\rho \left( \frac{K_A}{N_A} \right) \rho (1 - \alpha_2) \).

Likewise, we define the net excess demand for the second good by city B, \( E_{2B} \), as the (nonnegative) difference between the local demand, \( X_{2B} \), and the local supply, \( Y_{2B} \), and express it in terms of \( q_B \),\(^{16}\)

\[ E_{2B} = X_{2B} - Y_{2B} = \frac{w_B}{P_{2B}} \frac{1 - q_B^\alpha_2 \theta_1}{\alpha_1 - \alpha_2} N_B \left( 1 - \frac{\rho N_B}{2\pi} \right) > 0, \]  
(4.18)

\[ E_{2B} = c_5 N_B^\rho \left( 1 - \frac{\rho N_B}{2\pi} \right) \rho \alpha_2 (q_B - 1) \rho (\alpha_2 - 1) \rho (q_B^\alpha_1 - 1) \rho -1 (1 - q_B^\alpha_1 \theta_1). \]  
(4.19)

Hence, the balance of trade:

\[ E_{1A} = c_6 N_A^\rho \left( 1 - \frac{\rho N_A}{2\pi} \right) \rho \alpha_1 (q_A - 1) \rho (\alpha_1 - 1) \rho (1 - q_A^\alpha_2) \rho -1 (q_A^\alpha_1 \theta_1 - 1), \]

where: \( c_6 = \left[ \alpha_1^\alpha_1 (1 - \alpha_1)^{1 - \alpha_1} \right] \rho (\alpha_1 - \alpha_2)^-\rho \left( \frac{K_A}{N_A} \right) \rho (1 - \alpha_2). \)

\(^{16}\) The net excess supply of the first good by city B, \( E_{1B} \), is

\[ E_{1B} = Y_{1B} - X_{1B} = \frac{w_B}{P_{1B}} \frac{1 - q_B^\alpha_2 \theta_1}{\alpha_1 - \alpha_2} N_B \left( 1 - \frac{\rho N_B}{2\pi} \right) > 0, \]

\[ E_{1B} = c_6 N_B^\rho \left( 1 - \frac{\rho N_B}{2\pi} \right) \rho \alpha_1 (q_B - 1) \rho (\alpha_1 - 1) \rho (1 - q_B^\alpha_2) \rho -1 (1 - q_B^\alpha_1 \theta_1) \rho (1 - q_B^\alpha_1 \theta_1). \]

Of course, the last terms of \( c_5 \) and \( c_6 \) here must contain the element \( K_B/N_B \) instead of \( K_A/N_A \), but it is to be reminded that \( K_A/N_A = K_B/N_B = \bar{K}/\bar{N} \).
Equations (4.17) and (4.19) are illustrated in Fig. 4-2 in which the arrow marks indicate the opposite responses away from the common autarky point \((q_A=q_B=0.0, E_{2A}=E_{2B}=0)\) by cities A and B along the \(E_{2A}\) and \(E_{2B}\) curves respectively for the parametric values employed therein. The reaffirmation by Fig. 4-2 of our presumed pattern of trade, however, is not limited to the particular parametric values employed. The pattern of trade is still maintained under a more general set of parametric values according to the following elasticities of \(E_{2A}\) and \(E_{2B}\) with respect to \(q_A\) and \(q_B\):

\[
\frac{dE_{2A}}{dq_A} = q_A \left[ \frac{\sum a_i \theta_i}{q_A \sum a_i \theta_i - 1} + \frac{(\rho - 1) a_i - \rho(1 - a_i)}{q_A a_i - 1} \right],
\]

(4.21)

which, under trade, i.e., \(q_A \sum a_i \theta_i > 1\), can be restated as

\[
\frac{dE_{2A}}{dq_A} > q_A \left[ \frac{1 - \sum a_i \theta_i}{(q_A \sum a_i \theta_i - 1)(q_A - 1)} + \frac{(\rho - 1) a_i}{q_A a_i - 1} \right] > 0;
\]

(4.22)

whereas \(q_B\) is to be lower than its autarky value, and the negative elasticity with respect to \(q_B\) is

\[
\frac{dE_{2B}}{dq_B} = q_B \left[ \frac{\sum a_i \theta_i}{1 - q_B \sum a_i \theta_i} + \frac{\rho(1 - a_i)}{q_B - 1} - \frac{(\rho - 1) a_i}{q_B a_i - 1} \right] > 0,
\]

(4.23)

because, on the account of (4.9), the sum of the last two terms of (4.23) must be positive.

Fig. 4-2 also suggests some clues to the possible degrees of
Fig. 4-2: Excess Demand and Supply vs. q.

\[ \sigma_1 = 0.8 \quad \sigma_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{q}{2\pi} = 0.25 \times 10^{-7} \]
specialization in production by both cities in equilibrium. Being monotonically increasing functions of parameters $N_A$ and $N_B$ within the relevant ranges of $q_A$ and $q_B$ respectively, the $E_{2A}$ and $E_{2B}$ curves will rotate up (or down) around the common autarky point when $N_A$ and $N_B$ increase (or decrease).\textsuperscript{17} Thus, as for the smaller city B, because in equilibrium $q_B$ will be located much farther away from the autarky point than $q_A$ will be in city A, the degree of specialization by city B in the first good will tend to be higher than that by city A in the second good. Even intuitively, this is a plausible result. To the extent that specialization is limited by the extent of the market, the realization of scale economies in trade by specialization will be greater for the smaller city B with its greater export market (for the first good) than city A. In the end, however, the smaller market size of city B will tend to limit the overall trade volume. This presents a contrast to (4.11) in which the greatest trade potential, in terms of the difference between the autarky price ratios, can be found between the largest and smallest cities.

Finally, the solution in equilibrium can be obtained by

\textsuperscript{17} Although increase in parameter $N_A$, with $K_A/N_A$ fixed, will eventually tilt down the $E_{2A}$ curve due to the dominance of urbanization diseconomies over scale economies, this will happen only in cities of sizes far greater than our largest city ($N_{\text{max}}=1 \times 10^7$). This is because $dE_{2A}/dN_A<0$ when $N_A>(2\pi/g)(1+\sigma)^{-1}$, which at the parametric values of Fig. 4-2, is $3.333 \times 10^7$. The same will apply to city B and the first good market as well.
solving the two sets of equations regarding the equalization of the price ratios (4.11), and the balance of trade consisting of (4.17), (4.19) and (4.20) simultaneously for the two unknowns \( q_A \) and \( q_B \). Due to the apparent difficulty in solving them algebraically, however, we will approach them by numerical iterative methods in two steps. First, we will derive numerically the functional relationship between \( q_A \) and \( q_B \) that satisfies both (4.11) and (4.13). Second, the above function then will be substituted into (4.17) and (4.19) so that \( E_{2A} \) and \( E_{2B} \) could be expressed as functions of only \( q_A \), and thus be amenable to finding the solution to (4.20). Since the two steps are essentially finding the roots of \( q_A \) and \( q_B \) numerically, in view of their magnitudes we will maintain a tolerance level of at least \( \pm 10^{-6} \).

As the first step, two graphs of Fig. 4-3 illustrate the numerical relationship between \( q_A \) and \( q_B \) that equalizes the product price ratios \( p_A \) and \( p_B \) in (4.11), and guarantees incomplete specialization by both cities. As noted before, Fig. 4-3-1 shows that \( q_A \) and \( q_B \) are monotonically increasing functions of the common price ratio while maintaining their ranges,

\[
\frac{1}{a_i} = 1.25 < 1.90 < q_B < \frac{1}{\sum a_i \theta_i} = 2.0 < q_A < 2.11 < 5 = \frac{1}{a_i},
\]

as the result of the difference in the sizes of cities A and B. The corresponding range of the common price ratio is

\[0.6105 \leq p_A = p_B \leq 0.6475,\]

indicating the limits given by the autarky price ratios of cities A and B respectively. Fig. 4-3-2 thus shows the opposite directions taken by \( q_A \) and \( q_B \) away from the
Fig. 4-3-1: Numerical Solution of (4.11), p vs. q.

\( a_1 = 0.8 \quad a_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \)

\( N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{g}{2w} = 0.25 \times 10^{-7} \)
Fig. 4-3-2: Numerical Solution of (4.11), $q_A$ vs. $q_B$.

\[ \alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{q}{2\pi} = 0.25 \times 10^{-7} \]
autarky point, i.e., \( \frac{dq_B}{dq_A} < 0 \) when \( p_A = p_B \). Again, any point on the curve is a possible equilibrium point. We now use the numeric data generated in this step as the input in the next step.

In the second step, the data represented by Fig. 4-3 are then substituted into (4.17) and (4.19) to generate the two graphs of Fig. 4-4. Of these, Fig. 4-4-1 confirms (4.22) and (4.23): whereas \( E_{2A} \) is a monotonically increasing function of \( q_A \), \( E_{2B} \) is a monotonically decreasing function of both \( q_A \) and \( q_B \). Moreover, under the concave production possibility curves in both cities, the price ratios and \( q_A \) must move in the same direction, as noted in (4.7). Thus, Fig. 4-4-2 is almost identical to Fig. 4-4-1. A unique equilibrium is therefore obtained at \( E \) that clears the markets in both cities for the second good, and on the account of Walras' Law, for the first good as well. By the virtue of the slopes of the \( E_{2A} \) and \( E_{2B} \) curves (\( \frac{dE_{2A}}{dp} > 0 \), \( \frac{dE_{2B}}{dp} < 0 \) and \( p = p_A = p_B \)), it is clear that the equilibrium is also stable (i.e., \( \frac{dE_{2B}}{dp} - \frac{dE_{2A}}{dp} < 0 \)). For the parametric values so far maintained, the equilibrium yields the following solutions for the endogenous variables: \( q_A = 2.0093, q_B = 1.9099, p_A = p_B = 0.6137, E_{2A} = E_{2B} = 5.024 \times 10^4 \). As was expected, the deviation of \( q_B \) from the autarky point is far greater than that of \( q_A \) (0.0901 vs. 0.0093). Thus, despite the significant shift away

\[18\] Notice that both the Marshallian and Walrasian stability conditions are met here. See Chapter 3, Footnote 8 for details.
Fig. 4-4-1: Numerical Solution of (4.11) and (4.20), $q_A$ vs. $E_2$.

$\alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03$

$N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad q_{2 \pi} = 0.25 \times 10^{-7}$

In equilibrium at $E$: $q_A = 2.0093 \quad q_B = 1.90991$

$p_A = p_B = 0.61372 \quad E_{2A} = E_{2B} = 5.024 \times 10^4$
Fig. 4-4-2: Numerical Solution of (4.11) and (4.20), \( p \) vs. \( E_2 \).

\[
\begin{align*}
\alpha_1 &= 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \\
N_A &= 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{\sigma}{2\pi} = 0.25 \times 10^{-7}
\end{align*}
\]

In equilibrium at \( E \):
\[
q_A = 2.0093 \quad q_B = 1.90991
\]

\[
\begin{align*}
p_A = p_B &= 0.61372 \quad E_{2A} = E_{2B} &= 5.024 \times 10^4
\end{align*}
\]
from the autarky situation in city B, the situation in city A has barely been changed, thereby limiting the overall trade volume.

We now turn to the gains from trade in each city by looking at the improvement in the level of utility. For the representative city, the indirect utility function (3.16) is denoted as a function of q by the use of (4.2) and (4.8),

\[ v = c_7 N^{\rho-1} \left( 1 - \frac{q N}{2\pi} \right)^{\rho} \sum \lambda_i \theta_i q(q-1) - \rho \Sigma (1-\alpha_i) \theta_i \left[ (1-q\alpha_2) \theta_1(q\alpha_1-1) \theta_2 \right]^{\rho-1}, \]

(4.24)

where: \( c_7 = \theta_1^{\rho_1} \theta_2^{\rho_2} \left[ \prod \alpha_i \theta_i (1-\alpha_i) \right] \rho (\alpha_1 - \alpha_2)^{1-\rho} \left( \frac{K}{N} \right) \rho \Sigma (1-\alpha_i) \theta_i. \)

Differentiating the above, we obtain

\[ \frac{dv}{dq} q = 1 - \rho (1-\Sigma \alpha_i \theta_i) \frac{q}{q-1} + (\rho-1) \left( -\frac{\theta_1 \alpha_2}{1-q\alpha_2} + \frac{\theta_2 \alpha_1}{q\alpha_1-1} \right), \]

(4.25)

which becomes negative, zero, or positive if q is less than, equal to, or greater than its autarky value \((1/\Sigma \alpha_i \theta_i)\) respectively. Thus, greater improvement in welfare is obtainable in both cities whenever q is located farther away from the autarky point, i.e.,

\[ \frac{dv_A}{dq_A} > 0, \quad \frac{dv_B}{dq_B} > 0, \quad \frac{1}{\alpha_1} < q_B < \frac{1}{\Sigma \alpha_i \theta_i} < q_A < \frac{1}{\alpha_2}. \]

(4.26)

The above relationship along with the utility levels of both cities in both autarky and equilibrium is illustrated in Fig. 4-5 in which the arrow marks indicate the opposite directions of movement by the cities from autarky to equilibrium. Because the
Fig. 4-5: Changes in Utility Levels, Before and After Trade.

\[ \alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ N_A = 4 \times 10^6 \quad N_B = 0.4 \times 10^6 \quad \frac{K_A}{N_A} = \frac{K_B}{N_B} = 2 \quad \frac{q}{2\pi} = 0.25 \times 10^{-7} \]

In equilibrium at A and B: \[ q_A = 2.0093 \quad q_B = 1.9099 \]

\[ v_A = 1.27553 \quad v_B = 1.2516 \]
equilibrium point B for city B is located considerably farther away from the autarky point than its counterpart A for city A, the improvement in welfare is accordingly greater for the smaller city. While both cities gain from trade, this suggests that not only is the smaller city more dependent on trade, the gains from trade is also greater for the smaller city. 19

Complete numerical solutions to all pertinent endogenous variables are listed in Table 4-1 for the parametric values maintained so far in this chapter. They are derived by substituting the results of Fig. 4-4 into the appropriate equations of sections 4-2 and 4-3. Notice in particular that the difference in the factor price ratios between the two cities, although considerable in autarky (0.45 vs 0.495), is kept to a minimum in equilibrium (0.45419 vs. 0.4504). The difference, albeit small, nevertheless suggests that the marginal rate of substitution between factors is not equated between the cities. Due to the externality involved in scale economies, therefore, the equilibrium is not strictly optimal in the national sense. It is, however, still optimal within the cities because not only is the marginal rate of substitution between factors is the same in the two industries, but the marginal rate of substitution in consumption and the marginal rate of transformation are the same within the cities. Thus, a slight discrepancy exists in local and na-

19 Notice that not only the utility levels of both cities but also the total national output in both goods is increased (see Table 4-1).
Table 4-1: Solutions to Endogenous Variables.

<table>
<thead>
<tr>
<th>Endogenous Variables</th>
<th>CITY A</th>
<th>Autarky</th>
<th>Equilibrium</th>
<th>CITY B</th>
<th>Autarky</th>
<th>Equilibrium</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>2.0</td>
<td>2.0093</td>
<td>2.0</td>
<td>1.9099</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L₁</td>
<td>2.88x10⁶</td>
<td>2.871x10⁶</td>
<td>3.168x10⁵</td>
<td>3.2631x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L₂</td>
<td>7.2x10⁵</td>
<td>7.2893x10⁵</td>
<td>7.92x10⁴</td>
<td>6.9686x10⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K₁</td>
<td>1.6x10⁶</td>
<td>1.5803x10⁶</td>
<td>1.6x10⁵</td>
<td>1.8112x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K₂</td>
<td>6.4x10⁶</td>
<td>6.4197x10⁶</td>
<td>6.4x10⁵</td>
<td>6.1888x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r/w</td>
<td>0.45</td>
<td>0.45419</td>
<td>0.495</td>
<td>0.4504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y₁</td>
<td>3.9865x10⁶</td>
<td>3.9662x10⁶</td>
<td>4.0243x10⁵</td>
<td>4.2304x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y₂</td>
<td>6.5298x10⁶</td>
<td>6.563x10⁶</td>
<td>6.2148x10⁵</td>
<td>5.888x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>0.6105</td>
<td>0.6137</td>
<td>0.6475</td>
<td>0.6137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₁</td>
<td>3.9865x10⁶</td>
<td>3.997x10⁶</td>
<td>4.0243x10⁵</td>
<td>3.922x10⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₂</td>
<td>6.5298x10⁶</td>
<td>6.5128x10⁶</td>
<td>6.2148x10⁵</td>
<td>6.3904x10⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td>0</td>
<td>3.08x10⁴</td>
<td>0</td>
<td>3.08x10⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>0</td>
<td>5.024x10⁴</td>
<td>0</td>
<td>5.024x10⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>1.27551</td>
<td>1.27553</td>
<td>1.2503</td>
<td>1.2516</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exogenous Variables:

- \( N_A = 4x10^6 \)
- \( K_A/N_A = 2 \)
- \( N_B = 0.4x10^6 \)
- \( K_B/N_B = 2 \)

Parameters:

- \( \alpha_1 = 0.8 \)
- \( \alpha_2 = 0.2 \)
- \( \theta_1 = \theta_2 = 0.5 \)
- \( \rho = 1.03 \)
- \( g/2\pi = 0.25x10^{-7} \)
tional optimality.

4-4. Migration and Balanced Free-Trade

The balanced free-trade model of the last section showed that while the utility levels of both cities were improved after trade, they were not equalized, thereby leaving a tendency for factors to move. In that model, city sizes were exogenous, and factor movements were not considered. In this section, however, we incorporate factor movements into the model, so that city sizes become endogenous, and are to be determined in equilibrium given the exogenous variable $\bar{N}$, the national population. We are particularly interested in finding out the equilibrium city size distribution and the welfare implications thereof.

Under urbanization economies and diseconomies, factor movements and trade are not perfect substitutes for each other, and consideration of both of them is essential. We have already noted that whereas factor movements alone yield identical city sizes in equilibrium that offer no basis for trade, trade alone does not guarantee the same utility levels between the cities of different sizes unless by pure chance. Because both trade and factor movements are simultaneously considered under the general equilibrium setting, this model represents a more direct and realistic rendition of the interaction between cities. Hence a more refined integration of both the supply- and demand-oriented approaches.
As for the behavioral assumption in factor movements, this new model will still operate under our somewhat restrictive assumption in which capital and its rentals are both invested and spent where the capital owner/laborer lives and works. The location decision thus becomes multi-dimensional. In his costless move from city to city to maximize utility, the capital owner/laborer has to consider not only the marginal product of labor, but also the marginal product of capital and the net housing costs (after the rebate from the land bank), all specific to each city. Despite its restrictiveness, however, it is argued that our assumption is more suitable to our purpose in our model than the usual practice in the trade literature in which workers have no claims for capital including land, and utility essentially refers only to the real wage rate or the marginal product of labor. Because capital always moves with the laborer under our assumption, we will use the term "migration" to denote the combined movements of factors.

City sizes being endogenous now, we define at the outset a new variable $z$ as the population share of the larger city $A$ out of the national population $\bar{N}$ in equilibrium. With two cities in the nation, we obtain that

$$z = \frac{N_A}{\bar{N}}, \quad 1-z = \frac{N_B}{\bar{N}},$$

$$0 < 1-z < \frac{1}{2} < z < 1.$$  \hspace{1cm} (4.27)

Following the last section, we can determine the equilibrium of
our migration and trade model by solving simultaneously three equations regarding: the equalization of the product price ratios, $p_A=p_B$; the balance of trade, $E_{2A}=E_{2B}$; and the equalization of the utility levels, $v_A=v_B$. The last equation, however, needs further discussion.

As for city $A$, the indirect utility function (3.16) can be rewritten by the use of (4.2) and (4.27) as

$$v_A = \theta_1^\theta_1 \theta_2^\theta_2 p_A \left[ \frac{w_A}{p_{2A}} \right] q_A \left( \frac{1-q_N z}{2\pi} \right),$$

which, by (4.16), can again be specified as

$$v_A = \theta_1^\theta_1 \theta_2^\theta_2 (\alpha_1-\alpha_2)(\bar{N}) \left( \frac{q_A}{2} \right) E_{2A} \left( z \frac{q_A}{q_A \Sigma \alpha_i \theta_i} - 1 \right).$$

Likewise, for city $B$, we obtain

$$v_B = \theta_1^\theta_1 \theta_2^\theta_2 (\alpha_1-\alpha_2)(\bar{N}) \left( \frac{q_B}{2} \right) E_{2B} \left( \frac{q_B}{1-z} \frac{1}{1-q_B \Sigma \alpha_i \theta_i} \right).$$

From (4.29) and (4.30), provided that the price ratios are equalized ($p_A=p_B$) and the trade is balanced ($E_{2A}=E_{2B}$), the utility levels are equalized only if

$$\frac{z}{q_A} + \frac{1-z}{q_B} = \Sigma \alpha_i \theta_i,$$

or,

$$q_B = \frac{q_A (1-z)}{q_A \Sigma \alpha_i \theta_i - z},$$

which further limits the deviations of $q_A$ and $q_B$ from their autarky value, $1/\Sigma \alpha_i \theta_i$, but is nevertheless in line with their required directions of change in trade, i.e.,
\[
\frac{\delta q_A}{\delta q_B} \frac{q_B}{q_A} = - \frac{q_A - (1/\sum \alpha_i \theta_i)}{(1/\sum \alpha_i \theta_i) - q_B} < 0, \text{ at constant } z. \quad (4.32)
\]

Therefore, one of the necessary conditions for the equalization of the utility levels requires that the weighted average of the reciprocals of \( q_A \) and \( q_B \) be \( \sum \alpha_i \theta_i \), the weights being the population shares of cities A and B respectively. Of course, the equilibrium achieved in the last chapter via migration only \( (q_A = q_B = 1/\sum \alpha_i \theta_i \text{ and } z = 1/2) \) satisfies this as a special case.

By utilizing (4.31), we can thus reduce one variable \( q_B \) and express the relevant equations in terms of the two unknowns \( q_A \) and \( z \). Substituting (4.31) into (4.11), we obtain

\[
\begin{align*}
\rho_A &= c_4 \left[ \frac{1-\frac{q_N}{2\pi}}{(q_A-1)} \right] \rho (\alpha_1 - \alpha_2) \left( \frac{1-q_A \alpha_1}{q_A \alpha_1 - 1} \right)^{\rho-1}, \\
\rho_B &= c_4 \left[ \frac{1-\frac{q_N(1-z)}{2\pi}}{q_A(1-z) - q_A \sum \alpha_i \theta_i + z} \right] \rho (\alpha_1 - \alpha_2) f_1, \\
f_1 &= \frac{(q_A \theta_1 (\alpha_1 - \alpha_2) - z(1-q_A \alpha_2))}{(q_A \theta_2 (\alpha_1 - \alpha_2) - z(q_A \alpha_1 - 1))} \rho^{-1}, \\
\rho_A &= \rho_B. \quad (4.33)
\end{align*}
\]

Likewise from (4.17-20), we get

\[
\begin{align*}
E_{2A} &= c_5 (q_N) \rho \left[ \frac{1-\frac{q_N}{2\pi}}{(q_A-1)} \right] \rho \alpha_2 (q_A-1) \rho (\alpha_2 - 1) (q_A \alpha_2 - 1) \rho^{-1} (q_A \sum \alpha_i \theta_i - 1), \\
E_{2B} &= c_5 [q_N(1-z)] \rho \left[ \frac{1-\frac{q_N(1-z)}{2\pi}}{q_A(1-z) - q_A \sum \alpha_i \theta_i + z} \right] \rho \alpha_2 \left( \frac{q_A(1-\sum \alpha_i \theta_i) - z(q_A-1)}{q_A \sum \alpha_i \theta_i - z} \right) \rho (\alpha_2 - 1) f_2, \\
f_2 &= \frac{(q_A \theta_2 (\alpha_1 - \alpha_2) - z(q_A \alpha_2 - 1))}{(q_A \sum \alpha_i \theta_i - z)} \rho^{-1} \left( \frac{z(q_A \sum \alpha_i \theta_i - 1)}{q_A \sum \alpha_i \theta_i - z} \right),
\end{align*}
\]
Finally from (4.29-30) and (4.34), we derive for later use

\[ V_A = \frac{c_7(Nz)^{\rho-1}}{1 - gNz(1-z)} q_A (a_i - 1) \theta_i \left( \frac{q_A (1 - q_A a_2) (a_i - 1) \theta_i}{q_A^2 a_i \theta_i - z^2} \right)^{\rho}, \]

\[ V_B = \frac{c_7(N(1-z))^{\rho-1}}{1 - gN(1-z)} q_A (1 - q_A a_2) \theta_i \left( \frac{q_A (a_i - 1) \theta_i}{q_A^2 a_i \theta_i - z^2} \right)^{\rho}, \]

\[ f_3 = \frac{1 - z}{\sum a_i \theta_i - a_2 z} \left[ \frac{q_A (1 - q_A a_2) (a_i - 1) \theta_i}{q_A^2 a_i \theta_i - z^2} \right]^{\rho-1}, \]

\[ V_A = V_B. \]  

The solution in equilibrium can now be obtained by solving simultaneously (4.33) and (4.34) for \( q_A \) and \( z \). As before, we apply the two-step numerical methods to them. First, we derive numerically from (4.33) the relationship between \( q_A \) and \( z \) in which the price ratios are equalized. It is illustrated in Fig. 4-6-1 where \( P \) indicates the "trivial" solution point with no trade (\( q_A = q_B = 1/\sum a_i \theta_i \)) and identical sizes between the two cities (\( z = 1/2 \)). As the population share of city A increases rightward from \( P \) thus making trade feasible, \( q_A \) also increases initially, then attains a peak, and finally decreases. On the other hand, \( q_B \) calculated from (4.31) decreases monotonically. Trade volume being largely determined by the smaller city under increasing returns to scale, this is an expected result. As city B gets smaller and smaller, the extent of specialization in city A
Fig. 4-6-1: Numerical Solution of (4.33), $z$ vs. $q$.

$a_1 = 0.8 \quad a_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03$

$\bar{N} = 4.8 \times 10^6 \quad \bar{K}/\bar{N} = 2 \quad \frac{a}{2\pi} = 0.25 \times 10^{-7}$
Fig. 4-6-2: Numerical Solution of (4.33), \( z \) vs. \( p \).

\[ a_1 = 0.8 \quad a_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ \bar{N} = 4.8 \times 10^6 \quad \frac{\bar{K}}{\bar{N}} = 2 \quad \frac{a}{2\pi} = 0.25 \times 10^{-7} \]
measured by \((q_A - 1/\sum a_i \theta_i)\) eventually decreases, but that in city B \((1/\sum a_i \theta_i - q_B)\) increases due to its increasing export market size. Fig. 4-6-2 also shows that the equalized price ratio decreases with respect to increases in \(z\). Because the equilibrium price ratio is determined largely by the autarky price ratio of the larger city, and because its autarky price ratio decreases as city A gets larger, this is a plausible result.

We now apply the above numeric data represented by Fig. 4-6 to the balance of trade equation (4.34), and generate Fig. 4-7 which shows the relationship between the equalized price ratio and the net national excess supply of the second good \((E_{2A} - E_{2B})\). Because the data already reflect both (4.31) and (4.33), it is to be noted that our application of the data for the determination of the balance of trade will also reflect the equalization of utility levels on the account of (4.29) and (4.30). In short, Fig. 4-8, which shows the relationship between the utility levels of the two cities vs. \(z\), by the application of the same data to (4.35), is another equivalent graphic manifestation of Fig. 4-7. In both figures, we have two equilibrium points: whereas \(P\) refers to the autarky equilibrium with identical city sizes \((z=0.5)\), \(Q\) refers to the trade equilibrium with different city sizes \((z=0.75793)\).

Are they stable? Because the "trivial" equilibrium at \(P\) involves no trade and is determined entirely by migration, its stability is determined by the shapes of the utility curves of
Fig. 4-7: Numerical Solution of (4.34), $E_{2A} - E_{2B}$ vs. $p$.

$\alpha_1 = 0.8, \alpha_2 = 0.2, \theta_1 = \theta_2 = 0.5, \rho = 1.03$

$\bar{N} = 4.8 \times 10^6, \bar{K}/\bar{N} = 2, \frac{g}{2\pi} = 0.25 \times 10^{-7}$

At $P$, $E_{2A} = E_{2B} = 0, p_A = p_B = 0.62713$

At $Q$, $E_{2A} = E_{2B} = 8.5373 \times 10^4, p_A = p_B = 0.62037$
Fig. 4-8: Equilibrium Utility Levels vs. City Size Distribution.

\[ \alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ \bar{N} = 4.8 \times 10^6 \quad \bar{k}/\bar{N} = 2 \quad \frac{g}{2\pi} = 0.25 \times 10^{-7} \]

At P, \( z = 0.5 \) \quad \( v_A = v_B = 1.28449 \)

At Q, \( z = 0.75793 \) \quad \( v_A = v_B = 1.27843 \)
the two cities as they are related to their respective city sizes.\textsuperscript{20} As was noted before in the last chapter, in a two-city framework the equilibrium at P is stable only if the equilibrium city sizes are greater than the optimum city size of the single-city model $N_0$ which at the current parametric values is $2.201 \times 10^6$. At P in Fig. 4-8, for example, $N_A=N_B=0.5\tilde{N}=2.4 \times 10^6 > N_0$, and it is thus stable. More on this will follow later. The equilibrium at Q, on the other hand, concerns both trade and migration, and its stability test requires the examination of both Fig. 4-7 and Fig. 4-8. According to Fig. 4-7, $d(E_2A-E_2B)/dp$ is negative in the neighborhood of Q, so it is stable in terms of trade under the Marshallian criterion which we have adopted in place of the Walrasian one for our models under external economies of scale. However, it is unstable in terms of migration, because, according to Fig. 4-8, any migratory disturbance from Q will be cumulative in either direction, and set a stage for further disequilibrating migration away from the equilibrium.\textsuperscript{21} It is to be noted that whereas a more even city size distribution than the one at Q is welfare-improving for both cities, a more skewed distribution toward a greater primacy of the larger city is welfare-deteriorating for both cities.

\begin{flushleft}
\textsuperscript{20} Refer to Fig. 3-7, Chapter 3. The P here corresponds exactly to the P of Fig. 3-7, albeit at different parametric values.
\end{flushleft}

\begin{flushleft}
\textsuperscript{21} This equilibrium is akin to the one at Q of Fig. 3-7, but, of course, is different from it due to the existence of trade here.
\end{flushleft}
While the equilibrium at P is meaningful and stable only when the national population is larger than twice the optimum city size of the single-city model (\(N \geq 2N_o\)) within the two-city framework, the equilibrium at Q cannot be obtained otherwise. This is because from (4.34) or (4.35), as long as trade occurs, the balance of trade or equalization of the utility levels cannot be achieved unless \(N_A > N_o\) and \(N_B < N_o\). Due to the complexities involved in both equations, however, it would be instructive to look at Fig. 3-4 instead. The single-city relationship between the utility level vs. city size, which is analogous to the one implied by (4.35), is reproduced here for our current parametric values as Fig. 4-9. Although (4.35) would additionally reflect the gains from trade associated with specific city size distributions especially for smaller cities, its curvature with respect to the individual city size is essentially the same as that of Fig. 4-9. According to Fig. 4-9, it is clear that \(N_A\) and \(N_B\) must be on the opposite sides of \(N_o\) and \(\bar{N} = N_A + N_B > 2N_o\) for the existence of the equilibrium. 22

Moreover as \(\bar{N}\) increases, accommodation of the increased national population within the two cities requires that \(z\) must increase as well, thus leading to a tendency toward a greater primacy. This relationship from (4.35) is illustrated in Fig. 4-

---

22 Due to the asymmetry of the curve, at any common level of utility the difference between the equilibrium city size and \(N_o\) will be greater for city A than city B. Hence, in equilibrium, \(N_A + N_B > 2N_o\). For example, refer to Table 4-2.
Fig. 4-9: Utility Level with City Size (Single-City Model).

\[ \alpha_1 = 0.8 \quad \theta_1 = 0.5 \quad \rho = 1.03 \]
\[ \alpha_2 = 0.2 \quad \theta_2 = 0.5 \quad K/N = 2 \quad g/2x = 0.25 \times 10^{-7} \]
\[ N_0 = 2.201 \times 10^6 \quad v_0 = 1.28449 \]
Fig. 4-10: Population Share of City A vs. $\bar{N}$ (at Q).

\[ \begin{align*}
\alpha_1 &= 0.8 & \alpha_2 &= 0.2 & \theta_1 &= \theta_2 &= 0.5 & \rho &= 1.03 \\
\bar{K}/\bar{N} &= 2 & \frac{q}{2\pi} &= 0.25 \times 10^{-7}
\end{align*} \]
Fig. 4-11: Equilibrium Utility Levels vs. National Population

\[ \alpha_1 = 0.8 \quad \alpha_2 = 0.2 \quad \theta_1 = \theta_2 = 0.5 \quad \rho = 1.03 \]

\[ \bar{K}/\bar{N} = 2 \quad \frac{\sigma}{2\pi} = 0.25 \times 10^{-7} \]
10, and the consequent deterioration in the utility levels for both equilibria at P and Q is given in Fig. 4-11. According to Fig. 4-11, because the utility levels at P are always higher than those at Q at any relevant $\bar{N}$, it is tempting to denote the point O (where $N_A=N_B=N_0$) as the "optimum national population". Complete solutions to all relevant endogenous variables at $\bar{N}=4.8\times10^6$ and other parametric values employed in this chapter are given in Table 4-2 for the equilibria at P and Q and for both cities.
Table 4-2: Solutions to Endogenous Variables.

<table>
<thead>
<tr>
<th>Endogenous Variables</th>
<th>CITY A</th>
<th></th>
<th>CITY B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Autarky Equilibrium</td>
<td>Trade Equilibrium</td>
<td>Autarky Equilibrium</td>
<td>Trade Equilibrium</td>
</tr>
<tr>
<td></td>
<td>at P</td>
<td>at Q</td>
<td>at P</td>
<td>at Q</td>
</tr>
<tr>
<td>q</td>
<td>2.0</td>
<td>2.01750</td>
<td>2.0</td>
<td>1.94712</td>
</tr>
<tr>
<td>$L_1$</td>
<td>1.8048x10^6</td>
<td>2.6303x10^6</td>
<td>1.8048x10^6</td>
<td>9.1846x10^5</td>
</tr>
<tr>
<td>$L_2$</td>
<td>4.512x10^5</td>
<td>6.7687x10^5</td>
<td>4.512x10^5</td>
<td>2.0973x10^5</td>
</tr>
<tr>
<td>$K_1$</td>
<td>9.6x10^5</td>
<td>1.4219x10^6</td>
<td>9.6x10^5</td>
<td>4.9937x10^5</td>
</tr>
<tr>
<td>$K_2$</td>
<td>3.840x10^6</td>
<td>5.8543x10^6</td>
<td>3.840x10^6</td>
<td>1.8245x10^6</td>
</tr>
<tr>
<td>$r/w$</td>
<td>0.47</td>
<td>0.4525</td>
<td>0.47</td>
<td>0.4598</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>2.4414x10^6</td>
<td>3.6106x10^6</td>
<td>2.4414x10^6</td>
<td>1.223x10^6</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>3.893x10^6</td>
<td>5.9908x10^6</td>
<td>3.893x10^6</td>
<td>1.8007x10^6</td>
</tr>
<tr>
<td>$p$</td>
<td>0.6271</td>
<td>0.6204</td>
<td>0.6271</td>
<td>0.6204</td>
</tr>
<tr>
<td>$X_1$</td>
<td>2.4414x10^6</td>
<td>3.6635x10^6</td>
<td>2.4414x10^6</td>
<td>1.1701x10^5</td>
</tr>
<tr>
<td>$X_2$</td>
<td>3.893x10^6</td>
<td>5.9054x10^6</td>
<td>3.893x10^6</td>
<td>1.8861x10^6</td>
</tr>
<tr>
<td>$E_1$</td>
<td>0</td>
<td>5.29x10^4</td>
<td>0</td>
<td>5.29x10^4</td>
</tr>
<tr>
<td>$E_2$</td>
<td>0</td>
<td>8.54x10^4</td>
<td>0</td>
<td>8.54x10^4</td>
</tr>
<tr>
<td>$v$</td>
<td>1.28449</td>
<td>1.27843</td>
<td>1.28449</td>
<td>1.27843</td>
</tr>
<tr>
<td>$z$</td>
<td>0.5</td>
<td>0.75793</td>
<td>0.5</td>
<td>0.24207</td>
</tr>
<tr>
<td>$N$</td>
<td>2.4x10^6</td>
<td>3.6381x10^6</td>
<td>2.4x10^6</td>
<td>1.1619x10^6</td>
</tr>
</tbody>
</table>

Exogenous Variables

$\bar{N} = 4.8x10^6$  \hspace{1cm} $\bar{K}/\bar{N} = 2$

Parameters

$\alpha_1 = 0.8$  \hspace{1cm} $\alpha_2 = 0.2$  \hspace{1cm} $\theta_1 = \theta_2 = 0.5$

$\rho = 1.03$  \hspace{1cm} $g/2\pi = 0.25x10^{-7}$
5. POLICY IMPLICATIONS

On the basis of our findings from the last two chapters in general and the migration and trade model in particular, we can derive some broad policy implications regarding economic efficiency with respect to city size distributions in the nation. It is hoped that these points will be useful in devising contemporary national urban development policies that are concerned with affecting the spatial allocation of populations among cities. It is however to be noted at the outset that economic equality which generally is no less important in such national policies is not elaborated here in view of our models in which homogeneous individuals own an equal amount of capital as well as an equal access to housing. Because cities are also assumed to share human and capital resources equally on a per capita basis as well as identical production technologies, the consideration of economic equality becomes irrelevant in equilibrium when the equilibrium involves perfect mobility of factors between cities. Also not elaborated in this connection are equally important socio-political considerations such as national unity and political integration.1

(a). The optimum city size distribution in which potential level of welfare for all in the nation is maximized is char-

1 For specific goals of national urban development policies, see the literature cited in Chapter 1.
acterized by the identical size $N_0$, the single-city optimum size, among all cities. A higher (lower) degree of homogeneity associated with localization economies $r$ and/or a lower (higher) weighted average of the labor elasticities of the industry subjective production functions $\Sigma a_i h_i$ will increase (decrease) $N_0$. It is also a stable equilibrium with many cities in existence in the economy. In the two-city framework, however, stability in equilibrium requires the common city size be larger than $N_0$, thereby causing a slight suboptimum compared to the theoretical optimum above. This optimum is achieved totally by migration of factors and absence of trade. Cities are identical in all aspects and self-sufficient. This suggests that the common view that optimum city systems would probably be hierarchical in size such as the rank-size distribution (for example, Richardson[1981]) may be mistaken.2

(b). For small and/or underdeveloped countries with a sufficiently low level of urbanization so that the total urban population is less than $N_0$, the optimum as well as equi-

---

2 Because scale economies differ among industries, Richardson seems to argue, net urbanization economies (or what he calls "agglomeration economies") are maximized at different city sizes. He, therefore, implicitly assumes that the number and composition of industries are necessarily different among cities. Unless industrial activities or factors are geographically immobile, however, there is no reason why this must happen under our general equilibrium framework. Henderson's similar argument, on the other hand, is based on complete specialization by cities in a single but different industry. The pitfalls of his model were indicated in Chapter 2.
librium city size distribution is characterized by the existence of a single city in the nation due to cumulative migration. Because the level of welfare of the city is still on the rising part of the curve with respect to city size, the primary concern for such countries must be greater urbanization and industrialization before any consideration is taken into city size distribution.

If factor mobility is sufficiently low in such countries so that existing city sizes can be regarded as fixed, a trade equilibrium is feasible that offers gains from trade over autarky albeit at different welfare levels between cities. However, trade between small cities under scale economies is generally characterized by low trade volume and offers little improvement. Thus promoting greater factor mobility under the circumstances, thereby helping to achieve the single equilibrium city size, seems a lot more sensible approach than, for example, investment in inter-urban transportation systems that would presumably facilitate trade rather than factor mobility. Similarly, insitu development strategies aimed at the existing small cities and towns such as the agropolitan development approach (Friedmann and Weaver[1979]) would appear ineffective.3

3 This is because spatial segregation of populations among cities and towns rather than integration - be it "independent development" or "protection from external exploitation" - is the main motive of such strategies.
(c). In the above situation (b), therefore, the existence of a single large city or its dominance over smaller cities in equilibrium must not be regarded as being of economic inefficiency. Although the city size distribution may appear highly primate, the dominant city is still small in absolute terms (i.e., less than No), and the equilibrium city size distribution represents close to the optimum. This suggests that relative primacy measures of city size distributions such as the ratio-based primacy index, or its more elaborate variants such as the one by El-Shakhs[1972], or the statistically-determined Pareto coefficient (see Rosen and Resnick[1980]) all have inherent shortcomings in terms of associating economic efficiency with city size distributions. Clearly there is a need to complement them with some measures of absolute city sizes.4

(d). The more general equilibrium city size distribution with different sizes is obtained by our migration and trade model, and it is characterized by the larger city being greater than and the smaller city less than No within the two-city framework. It is required that the combined sizes be no less than 2No. This situation is thus likely to emerge in countries which are larger in size and/or have a

---

4 For this matter, refer to Wheaton and Shishido[1981] and Kwon[1982] among others.
higher level of urbanization than those in (b). In equilibrium, trade is balanced and welfare levels are equalized between the two cities. This is in sharp contrast with the common central-place-based hierarchical models, notably Beckmann and McPherson [1970] which is characterized by trade imbalance and a functional presupposition between the higher-order city size and its lower-order market areas.

It is to be noted that in order to arrive at this equilibrium, trade is a precondition; otherwise, migration alone would result in a distribution of identical city sizes as in the above (a). It is in this sense that while the equilibrium is stable with respect to trade, it is unstable with respect to factor mobility. Although this equilibrium was shown in Fig. 4-8 to be inferior to the optimum city size distribution of (a), the almost universal occurrence of hierarchical city size distributions in reality leads us to suspect that factor mobility may not be as perfect as we have assumed. It may also be speculated that trade is so pervasive that migration occurs just enough to offset the post-trade welfare differentials between cities.

(e). The equilibrium city size distribution in (d) seems to closely reflect the current "primate" situation in many countries in that the very fact of hierarchical city sizes indicates a divergence from the optimum. Market outcomes under limited factor mobility clearly entail economic in-
efficiency, and a planned intervention that facilitates greater factor mobility and movement toward the optimum is consequently justified. Indeed, even a small difference in size between the larger and smaller cities, although not primate at all, must be indicative of economic inefficiency. Again, we can easily find the fallacy of relative primacy measures.

Within the confines of two cities in the nation, the equilibrium situation gets worse and moves toward even greater primacy as the national urban population increases. However, the supposition is arbitrary and may exaggerate reality especially in view of the fact that many countries probably have national urban populations far greater than $2N_0$. Within the framework of many cities and sufficiently large urban populations, our model would instead predict an equilibrium in which multiples of pairs of different-sized trading cities exist. It is not difficult to envision that this new equilibrium may probably be closer to the optimum than the one under the two-city framework.

(f). It is well observed that with enough urbanization and/or development, urban primacy eventually falls and the importance of a city size distribution diminishes. And our last observation in (e) seems to support this.\(^5\) In connec-

\(^5\) For seminal works on this subject, see Berry[1961], and Alonso[1968].
tion with (d) and (e), however, for countries with not enough overall urbanization and/or development, the problem of urban primacy seems real and may indeed incur economic costs. Countries like Egypt, Mexico, and Brazil in which one or two large cities appear well beyond and other cities well below $N_0$ may belong to this group. Although further migration into the large cities may eventually slow down as the countries approach "lower level equilibrium points" in city size distributions, the utility levels in both large and small cities will deteriorate.

According to our model, a higher level equilibrium point under trade can be achieved only when the small cities grow larger in size but remain less than $N_0$. With the total national urban population fixed, this suggests that the smaller cities must be fewer in number as well. This suggestion seems in line with the paradigm of "concentrated decentralization" advocated by Rodwin[1961], and Alonso[1968], among others. Finally, the current near monopoly by the big cities of localized noneconomic functions, such as central government services and quality higher education, undoubtedly seems to exacerbate the primacy problem. To the extent that such functions also determine the welfare levels of other cities, redistribution of them among all cities may seriously be considered.
6. CONCLUSION

This dissertation has proposed a general equilibrium approach to city size and output determination that integrates and builds upon the supply- and demand-oriented approaches currently available. As such, some prominent but often overlooked features of both approaches — agglomeration economies and dis-economies of the former and interdependence among cities via trade of the latter, for example, — have been retained in the new approach. However, the allowance for multiple industries, or industrial diversity, in the city and for trade balance is unique to our approach, and is deemed to be a theoretically more sound and empirically more accurate rendition of reality.

While the new approach may find many uses in urban modelling in general, our main objective with it has been to associate economic efficiency with various city size distributions in the nation. Given all the interdependence among cities, especially via migration and trade, it has been argued, the current pre-occupation with only the largest city and its size in national planning must be redirected toward the whole spectrum of cities of different sizes in the nation. This must be especially so when the efficacy of the national planning is seriously taken into account.

Based on the new approach, our single-city model has given rise to the optimum city size at which net urbanization economies
in autarky are maximized. With many cities in existence, it is also an equilibrium size for all cities when the equilibrium is achieved totally by migration among cities. The lack of such a uniform and optimum city size distribution in reality, however, has been attributed to the lack of complete factor mobility and/or the pervasive nature of inter-urban trade. Our migration and trade model has shown that with trade and migration occurring simultaneously, an equilibrium city size distribution with different sizes can be obtained. This hierarchical city size distribution, both suboptimal and more realistic, is to be differentiated from a similar result of some of the more restricted demand-oriented approaches based on the presumed hierarchical orders of goods as well as cities. Insightful policy implications follow the city size distribution and are delineated in the last chapter.

Our general equilibrium approach, however, undoubtedly suffers from important simplifications of reality. In particular, the specification of localization economies as the sole component of urbanization economies considerably simplifies both the scope and nature of urbanization economies. The assumption of full employment, coupled with the lack of consideration of rural areas, may significantly limit the use of our approach in many developing countries with high rates of rural-urban migration. It is hoped that further research with greater realism could extend our approach.
7. BIBLIOGRAPHY


El-Shakhs, Salah, "Development, Primacy, and Systems of Cities",


Ohlin, Bertil, Interregional and International Trade, Cambridge, MA: Harvard University Press, 1933


Thompson, Wilbur R., "The National System of Cities as an Object

