Large metropolises in the Third World: an explanation

Olga Alonso-Villar[*]

July 2000

Abstract

In this paper we have developed a model that sets out to explain the existence of megacities in developing countries, in the context of a core-periphery model à la Krugman. As in Krugman and Livas Elizondo (1996), this paper also suggests that agglomeration can be fostered by manufacturers mainly serving the domestic market. However, the analysis goes further by emphasizing that megacities are not only the result of protective trade policies, but also the consequence of the relative position of a country, in terms of industrialization, with respect to the rest of the world.

JEL classification: F12; R12

[*]Universidade de Vigo; Departamento de Economía Aplicada; Lagoas-Marcosende s/n; 36200 Vigo; Spain. Tel.: 34 986 812507; fax: 34 986 812401; e-mail: ovillar@uvigo.es. Financial support from the Spanish Ministry of Education through DGICYT grant PB98-0613-C02-01 is gratefully acknowledged.

1 Introduction

Concentration of population in cities appears as one of the most significant modern features. As observed by Bairoch (1988, p. 213): "For where the urban way of life had for thousands of years been the exception, it now became the rule." As a matter of fact, in 1995, 2.5 billion people lived in urban areas, which means that about 45 percent of the world's population was urban (Pugh, 1997).

Urban centers had been, however, an isolated phenomenon until the 19th century. Capitals of the old empires (such as Rome), centers of commerce and craftsmanship of the Middle Ages (such as Venice and Bruges), and capitals of the new Absolutist States of the 17th and 18th centuries (such as London and Paris) were some of the important cities of those eras. In fact, reinforcement of the power of the state and the growth of international trade led increasingly, between 1500 and 1700, to the overall urban population in Europe being mainly concentrated in a few cities, all of which were capitals, some being also ports (Bairoch, 1988). We could say that modern urbanization started in the United Kingdom after the Industrial Revolution, a revolution that affected not only the urbanization rate but also the role of cities. Whereas in traditional societies the functions of cities were mainly administrative, commercial, religious and craft-related, with industrialization the number of people working in industry increased notably. Employers, in need of labor force, attracted millions of workers from the countryside to the periphery of cities. This is how Birmingham, Leeds and Manchester expanded. This process rapidly extended across other countries: Germany (along the Ruhr); the North of France, and the East coast in the USA. In addition to the industrial cities, port-cities (Liverpool, Rotterdam, or New York), and those located in the core of communications networks (such as Chicago) grew notably.

In 1900 this process was still mainly European. In fact, 95 out of 140 cities with more than 200,000 inhabitants were located in Europe. Things have changed, however, during the 20th century, and concentration of population has become an increasing phenomenon, not only in Europe but all around the world. Starting in Latin America, followed more recently by Asia and Africa, during the past few decades, the urban population of developing countries has increased by 600% (Pugh, 1997). In this vein, by the end of the year 2000 there are expected to be as many as 35 megacities (cities with a population of over 5 million people) in the developing world, while only Shanghai was in this category in 1950.[1] Moreover, at the turn of the next century, seven of the ten largest cities in the world are expected to be in the less developed countries (LDC) (Sao Paulo, Bombay, Shanghai, Mexico City, Beijing, Lagos and Jakarta), while in 1950 there were only two (Shanghai and Calcutta) (Seitz, 1995). It seems, therefore, that there is a trend towards the concentration of urbanization into large agglomerations, especially in the developing countries.

Following Krugman's (1991) seminal paper, some recent articles have attempted to explain this trend by emphasizing the role of increasing returns to scale and imperfect competitition in explaining agglomeration in LDC. In particular, Krugman and Livas Elizondo (1996) explain the existence of large cities in developing countries, such as Mexico[2] D.C., by the strong backward and forward linkages that emerge from selling to the domestic market. Because of the existence of transport costs, …rms tend to choose production sites with good access to consumers (back linkages) and to other …rms (forward linkages), whether they produce goods for their workers or intermediate goods. So, concentration of economic activity is the result of a self-reinforcing process of agglomeration. Said authors suggest that this process can be fostered by the rise of import-substituting industrialization policies, arguing that the shift away from those policies may limit the growth of Third world megacities. Furthermore, Puga (1998) suggests that low transport costs, strong economies of scale and a large pool of agricultural workers available to migrate into cities could explain why major cities dominate in LDC.

This paper follows the same line of work in giving an explanation of the high degree of concentration that the poorer countries of the world are experiencing. Thus, as in the above papers, agglomeration emerges from the interaction between increasing returns to scale, transport costs and labor migration between locations. However, instead of emphasizing rural-urban migration, as in Puga (1998), it stresses the role that international trade can play on the agglomeration process of economic activity.[3] In this respect, the paper is most closely related to Krugman and Livas Elizondo (1996). That paper shows that a low degree of openness in an economy creates a tendency for the spatial concentra-

[1] Even though, Seitz (1995) also includes Buenos Aires in this group, it should be noted that Argentina was a relatively developed country some decades ago.

[2] The case of Mexico has inspired recent works both in terms of population concentration and of industrial location. See for example Dehghan and Vargas Uribe (1999), and Hanson (1996).

[3] See also Alonso Villar (1999), where international trade plays an important role in the urban pattern of a country, both in terms of formation and location of cities.

tion of manufacturing activities. Even though that gives us new insights into the way that trade policies and urban development are linked, this paper argues that it is the relative position of a country in itself, relative to others in terms of industrialization, which may be affecting agglomeration. In this vein, it develops a model that provides a formal framework in which to analyze the interplay between industrialization and agglomeration. As will be shown, the less developed a country is (i.e., the fewer manufactures it produces), the more likely it is that overall production/ population will be concentrated within a single city. Firms in such a country would tend to agglomerate, since any deviating firm would lose part of its national market, and this loss would not be offset by proximity to the foreign market, since it would have to compete with a large number of foreign firms. Krugman and Livas Elizondo (1996) suggest that when an economy turns outward, domestic demand is less important and, therefore, firms have little incentive to locate near the domestic demand. Even though this is true, there is an element which has not been considered: the competition effect. As a matter of fact, if a country is less developed, its firms are more dependent on the domestic market, since in the large foreign markets they would have to compete with many firms.

Unlike most recent economic geography papers that emphasize the role of an immobile demand as the force limiting agglomeration, the present model considers congestion costs, which include urban traffic congestion, pollution and high housing prices associated with large cities. This will allow us to discuss the differences that the consideration of different kinds of centrifugal forces and different immobile demands have on the results.

The paper is organized as follows. Section 2 describes the assumptions of the model. The analysis of the short run is presented in section 3, while the long-run equilibrium is characterized in section 4. Finally, section 5 concludes.

2 Assumptions of the model

The model consists of a long and narrow economy with three countries: A, B and C. Country B, which is the one on which we are focusing, being between countries A and C. One fundamental aspect that distinguishes the relationship between regions from that between countries is restriction on the movement of some factors. Usually, labor is not freely mobile between countries. For this reason, we consider that every country has a fixed population, and that it is impossible for a worker to change the country where she/he is working. Population in each country A and C is assumed to be concentrated in a single city. However, in country B there are two different locations and, in the long run, workers in this country can move between them without restriction. Without loss of generality, we assume that the distance between one city and the next is one. Hence, cities in country B are evenly distributed, so that each of them is one unit away from the nearest foreign city and two units away from the more distant one. World population is normalized to 1, and λ_B represents the proportion of world population in country B, where λ_1 denotes the proportion which is in city 1 and λ_2 that of city 2. Conversely, λ_A (respectively λ_C) represents the proportion of world population in country A (respectively C).

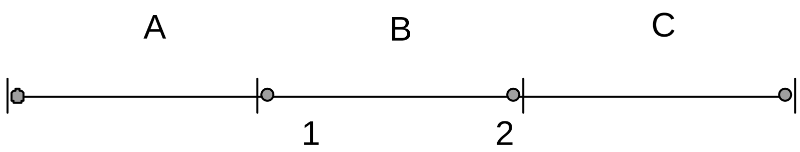

Figure 1. Possible locations in country B: 1 and 2.

In this context, we study which elements favor the agglomeration of economic activity in country B in only one city, and which limit this process. To do so, we first analyze the factors that intervene when country B does not trade with the others (autarky), and second, we note which new elements come into play when there is international trade.

2.1 Manufacturing

This sector produces a large number of differentiated varieties under increasing returns to scale, and firms are assumed to compete in a monopolistic regime of the Dixit and Stiglitz

(1977) type. So, each firm produces a different good and its production takes place in a single location. The number of firms in each city is endogenous, and denoted by n_j. We assume that all goods are produced with the same technology

$$L_{ij} = \alpha + \beta x_{ij}; \qquad (\alpha; \beta > 0);$$

where L_{ij} is the number of workers required to produce x_{ij} units of good i in location j: Labor is, therefore, the only production factor, and any variety i in city j requires the same fixed (α) and variable (βx_{ij}) quantities of labor. It follows that any firm producing good i in location j has a cost function

$$C(x_{ij}) = W_j(\alpha + \beta x_{ij});$$

where W_j is the local wage rate.

2.2 Individuals

Turning to the demand side, consumers in this economy are assumed to share a CES utility function

$$U = \left(\sum_i c_i^{\frac{\sigma - 1}{\sigma}} \right)^{\frac{\sigma}{\sigma - 1}};$$

where c_i is the consumption of good i and σ is the elasticity of substitution between any two goods, which is assumed to be greater than 1. This utility function implies love-for-variety à la Dixit-Stiglitz. Hence, individuals will buy not only the goods produced in their city, but also other cities' goods.

We assume that in transporting goods from other locations, consumers incur transport costs that take the usual iceberg form: a proportion of the good being shipped melts away. When a unit is delivered from city j to city k the amount that arrives is only $e^{-\tau D_{jk}}$, τ being the transport parameter, and D_{jk} being the distance between them. On the other hand, inside each city there are some additional costs, termed congestion costs, as a consequence of urban transport, pollution, or housing prices that mean that the higher the city size, the higher the value of these costs. Congestion costs are also assumed to take the iceberg form. So when a unit of a good is produced in, or arrives at, city k any consumer living in that city can only obtain a proportion $e^{-\gamma\lambda_k}$; γ being the congestion parameter.[4]

Finally, as in Krugman (1992), we assume that in the long run individuals in country B move towards locations with higher real wages, according to the law of motion

$$\frac{d\lambda_j}{dt} = \rho\lambda_j\lambda_k(\omega_j - \omega_k);$$

where ω_j is the real wage in city j, which implies that if city 1 offers a lower real wage than city 2, population in the former will decrease in favor of the latter.

3 Short-run equilibrium

In this section, we assume that there is no labor (firm) mobility between locations. Given an initial distribution of population between locations, we calculate prices, amounts of goods, number of firms and wages in each city of the world.

Drawing on Starrett's (1978) spatial impossibility theorem, Fujita (1993) indicates that there are only two basic types of models which can explain the endogenous formation of cities: non-price interaction models and non-competitive models. The model discussed here is included in the latter group.

Scale economies (due to the existence of fixed costs) in production imply that every good is produced in only one location, so that different cities have different goods. To determine the profit-maximizing behavior of firms, it is important to stress the fact that there are two types of demand: the demand of individuals living in the city where the

[4] We could treat intra-urban congestion in a more explicit way, such as land consumption and/or traffic congestion in cities. We could consider, for instance, cities as long and narrow. Workers, in need of land to live on, locate along a line at the central point of which production takes place. The commuting distance of the worker living on the outskirts of the city is offset by paying no land rents. Conversely, the worker living at the center does not incur commuting costs, but has to pay a land rent equal to the commuting costs of the former. Hence, the distance from the outskirts of the town to the center gives us information about both commuting costs and land rents (see Krugman and Livas Elizondo, 1996). If each worker consumes a unit of land, distance and population are equivalent. Therefore, we could use the above congestion costs to mean both commuting costs and land rents. However, such an extension would not substantially change the main conclusions of the paper. Therefore, we take the simplest form of urban congestion.

good is produced (domestic demand) and the demand from other cities (export demand). The important point to note is that both demands have the same price elasticity, σ, so that transportation and congestion costs (which mean that consumers in different cities pay different prices for the same good) do not alter the behavior of firms. Then it can be shown that the f.o.b. price charged by the firm that produces good i in city j is:

$$p_{ij} = W_j \beta \frac{\sigma}{\sigma - 1}. \qquad (1)$$

We can see that this price (which is a constant mark-up over marginal cost) only depends on the wage rate, W_j, offered in city j. Therefore, all goods produced in the same city have the same price. Monopolistic competition implies that firms enter the market until profits are zero. All this implies that

$$x_{ij} = \alpha\left(\frac{\sigma - 1}{\beta}\right) \quad \text{for every good } i \text{ and city } j. \qquad (2)$$

Since every firm produces the same quantity and has the same technology, the number of firms in city j, n_j, will be proportional to its population: $n_j = n\lambda_j$, n being the number of goods in the whole economy (this value can be obtained by dividing the number of workers in the economy by the number needed in each firm). Hence, if the population in a city is twice the other, it also doubles the number of firms (and varieties) of the former.

In order to obtain the wage rate in city j, W_j, we normalize the units of goods such that $p_{ij} = W_j$; which means that β should equal $\frac{(\sigma-1)}{\sigma}$. Let us suppose that we have a numeraire good at $j = A$. Therefore, $W_A = 1$.

We can prove that[5]

$$W_j = \left[\sum_k Y_k \left(e^{-(\tau D_{jk} + \gamma\lambda_k)} T_k\right)^{\sigma - 1}\right]^{\frac{1}{\sigma}}; \qquad (3)$$

where

$$T_j = \left[\sum_k \lambda_k \left(W_k e^{\tau D_{jk} + \gamma\lambda_j}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \qquad (4)$$

[5] See Appendix A.

is a price index[6] at j, and

$$Y_j = \lambda_j W_j; \tag{5}$$

is the disposable income of city j.

Therefore, for a given allocation of labor between cities we can now calculate the wage rate in each location.

4 Long-run equilibrium

We are now interested in knowing what happens in our economy if workers in country B can move across its national boundaries. The force that moves workers from one place to another is the real wage, defined as the ratio between the wage rate and the price index, namely $\omega_j = W_j T_j^{-1}$.

Let us define a long-run equilibrium as any distribution of the population between the two locations such that either both cities offer the same real wage, or there is concentration in the city that offers the highest real wage. Using the dynamic process described above, we know that workers tend to move to the city with the highest real wage and move away from the other. Taking this into account, equilibrium will be stable if, when population in city 1, λ_1; slightly increases then $\omega_1 < \omega_2$; which implies that workers leave city 1, and when it slightly falls then $\omega_1 > \omega_2$; which implies that workers move to city 1.

As is well known, the system of equations that define the wage rate cannot be solved analytically because of its strong non-linearity. However, basic understanding of the behavior of this model can be obtained by numerical simulations. To understand the main forces at work we distinguish two cases: autarky and free trade.

[6] See Dixit-Stiglitz (1977). Notice that W_k is the f.o.b. price charged by a firm located in city k. On the other hand, because of transport and congestion costs, a proportion of the good disappears before it reaches the consumer. Hence, the c.i.f. price paid by an individual of city j for 1 unit of the good produced in city k is $W_k e^{\tau D_{jk} + \gamma \lambda_j}$. Therefore, T_j can be interpreted as an average price facing an individual of city j. Krugman (1991) calls it the true price index.

4.1 Autarky

In this section we assume that country B does not trade with the others. This means that we have an economy with only two locations, between which individuals can move.

In what follows we show the basic understanding of the model with numerical examples, where parameters are σ (elasticity of substitution); τ (transport) and γ (congestion). All the figures are obtained, following Krugman (1992), with $\sigma = 4$. In terms of transport costs, three different values are considered: $\tau = 0.26$ (low), $\tau = 0.6$ (intermediate) and $\tau = 1.5$ (high). Finally, congestion costs take values between $\gamma = 0$ (no congestion costs) and $\gamma = 1$ (high). Figure 2 has been obtained keeping $\tau = 0.26$ and calculating the real wage differential for different values of congestion. Also, in figure 3 the congestion parameter is kept constant, $\gamma = 1$; and we allow for changes in the transport parameter.

We plot the real wage differential ($\omega_1 - \omega_2$) against the labor force in city 1.[7] Any point where the wage differential is zero is an equilibrium. This equilibrium is stable if the curve is downward-sloping and is unstable if it is upward-sloping. There may also be corner equilibria: concentration in city 1 (or respectively 2) when $\omega_1 - \omega_2 > 0$ (or respectively $\omega_1 - \omega_2 < 0$).

[7] For simplicity, in these figures we have written the population in each city of country B relative to its national population, which implies that population in city 2 is equal to 1 minus population in city 1, namely $\lambda_2 = 1 - \lambda_1$.

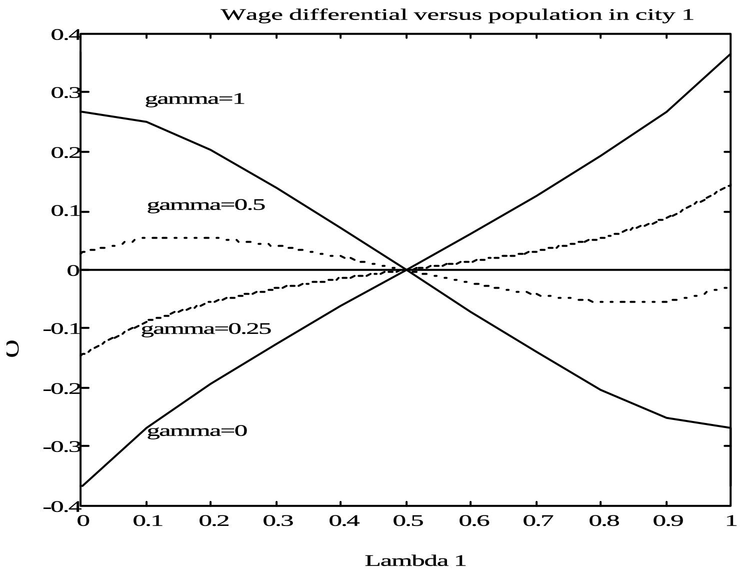

Figure 2. Only congestion parameter varies

We note that for high congestion costs ($\gamma = 0.5$; and $\gamma = 1$) concentration does not emerge as a long-run equilibrium, since if all individuals are in city 1, $\lambda_1 = 1$, the real wage in city 2 is higher than in city 1 (see figure 2). Therefore, concentration in city 1 is an equilibrium when the wage rate differential curve ends above zero. The analysis suggests that the lower the congestion costs, the more likely agglomeration of economic activity.[8] On the other hand, because of the symmetry between cities 1 and 2, we can also see that if a spatial distribution of production/population is in equilibrium, the symmetric will also be. In other words, if concentration in city 1 is an equilibrium, concentration in city 2 is also an equilibrium, and reciprocally.

Furthermore, for high congestion costs ($\gamma = 0.5$; and $\gamma = 1$) an even distribution of population between the two cities emerges a stable spatial configuration, whereas if congestion costs are low ($\gamma = 0$; and $\gamma = 0.25$) it does not. It follows, then, that the higher the congestion costs, the more likely is an even distribution of population.

Next we study what happens if transport cost varies.

[8] When the congestion parameter decreases, the curve of wage differentials moves counter-clockwise.

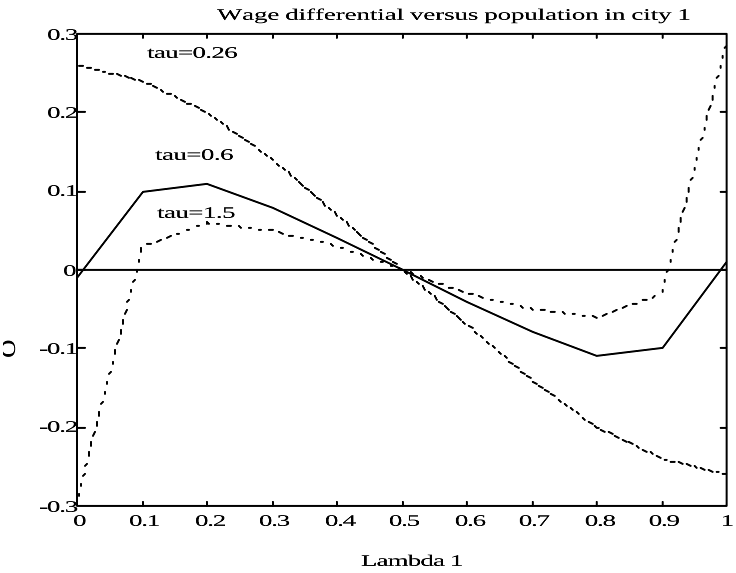

Figure 3. Only transport parameter varies

For a given value of the congestion parameter, the higher the transport parameter, the more likely it is to find agglomeration in one city (see figure 3).[9] It also follows that by decreasing the transport costs, an even distribution of production is more likely. When $\tau = 1.5$; we find that both concentration in either city and an even distribution between them are stable equilibria (there are also two unstable equilibria between them). As usual in models with increasing returns, we find multiplicity of equilibria, which means that there is scope for government intervention, as discussed in Alonso Villar (2000).

From all this, it follows that transport costs are a centripetal force that fosters the agglomeration of economic activity, while congestion costs are a centrifugal force favoring dispersion (for a formal analysis of this see Alonso Villar (2000)). In Ades and Glaeser (1995) we can find empirical evidence of this feature: urban concentration falls with improvements in transportation networks.

4.2 Free trade

In this section we assume, then, that the three countries trade between themselves and that countries A and C have the same population, namely $\lambda_A = \lambda_C$: Our model is based

[9] When the transport parameter decreases, the curve of wage differentials rotates clockwise.

on Krugman (1980), where gains from trade occur because the world economy produces a greater diversity of goods than would either country alone, thus offering each individual a wider range of choice. Following Krugman (1991, 1992), we would expect that the existence of an immobile demand would limit the agglomeration of country B's activity to a single location. In our case, the two foreign markets are symmetrically located with respect to country B and, since there is no international labor mobility, they represent an immobile demand for firms in country B.

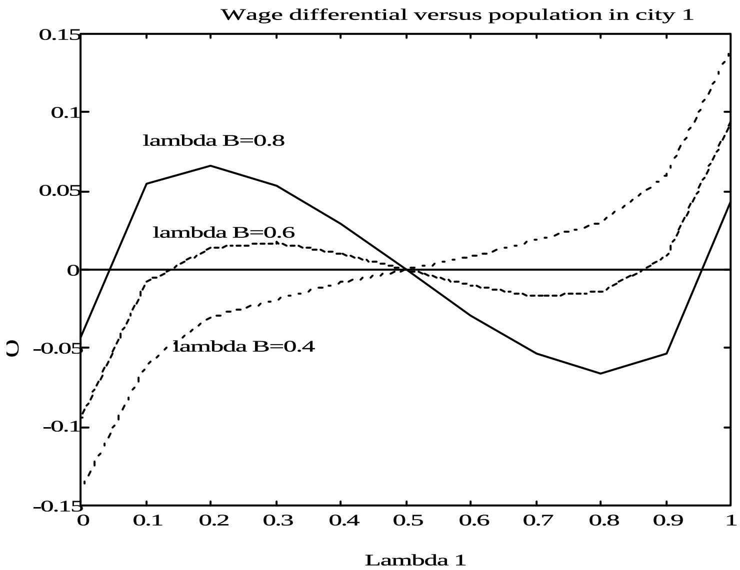

Figure 4. Only the relative level of country B's industrialization varies

In figure 4 can see that dispersion between the two cities is a stable equilibrium when population in country B is high enough. Otherwise, all economic activity in country B would concentrate in a single location. As opposed to Krugman (1991,1992), the analysis suggests that the existence of an immobile demand, in this case represented by the two foreign markets, seems to favor the agglomeration in country B, instead of halting it.

Let us now characterize the spatial pattern where two cities emerge in country B as a stable equilibrium analytically. We can prove that the symmetric equilibrium is stable if and only if $\frac{dT_1}{d\lambda}T_1^{-1} - \frac{dW_1}{d\lambda}W_1^{-1} > 0$, where λ denotes the share of country B's population living in city 1.[10] In the appendix we derive

[10] See Appendix B.

$$x_3 - x_2 = \lambda_B \left(\gamma + \frac{z_2 \frac{2\sigma - 1}{1-\sigma}}{\sigma + z_2 \lambda_B \frac{\sigma - 1}{2}} \right) - \frac{z_1 W_1^{-1} \lambda_A (\sigma - 1)}{\sigma + z_2 \lambda_B \frac{\sigma-1}{2}} x_1; \qquad (6)$$

where $x_1 = \frac{dT_A}{d\lambda} T_A^{-1}$; $x_2 = \frac{dT_1}{d\lambda} T_1^{-1}$ and $x_3 = \frac{dW_1}{d\lambda} W_1^{-1}$. Also, z1 and z2 , which depend on parameters directly, and through the expressions for the price indices and the wage rate in location 1, are defined in the appendix. To study when the symmetric equilibrium is stable we have to see when $x_3 - x_2 > 0$.

When transport costs are very low, z_1 and z_2 tend to zero, so that expression (6) takes the form

$$x_3 - x_2 = \lambda_B \gamma. \qquad (7)$$

This implies that the higher the congestion parameter (γ), and the higher the population/production in country B (λ_B), the more likely it is that the symmetric equilibrium will be stable. Besides, the stability of the symmetric equilibrium is warranted whenever $\gamma > 0$.

When transport costs are very high, z_1 tends to zero and z_2 tends to $\left(\frac{\lambda_B}{2}\right)^{-1}$. Therefore

$$x_3 - x_2 = \lambda_B \left[\gamma + \frac{2}{\lambda_B (1 - \sigma)} \right]. \qquad (8)$$

In other words, when transport costs are very high two cities can coexist as a stable configuration if and only if the population in country B and congestion costs are high enough; i.e., if and only if, $\gamma \lambda_B > \frac{2}{(\sigma - 1)}$.

Let us now analyze what happens with expression (6) when country B is either very large (λ_B tends to 1) or when it is very small (λ_B tends to 0). In the former case,

$$x_3 - x_2 > 0 \Leftrightarrow \gamma > \frac{2(2\sigma - 1)}{\sigma - 1} \frac{1 - e^{\tau d_{12}(1-\sigma)}}{2\sigma - 1 + e^{\tau d_{12}(1-\sigma)}}. \qquad (9)$$

It follows then that the higher the transport cost, the higher the congestion cost has to be if we want to guarrantee stability. In other words, if country B is large and congestion costs are high enough, a two-city spatial pattern can emerge as a stable configuration.

In the latter case,

$$x_3 - x_2 \to 0. \tag{10}$$

Hence, when population/production in country B is very small it is unlikely that two cities will be found as a stable spatial pattern.

To summarize, we can conclude that transport costs work against dispersion while congestion costs and country B's population favor dispersion. Therefore, as opposed to Krugman (1991, 1992), the existence of an immobile demand, represented by the two foreign markets, seems to favor agglomeration in country B, instead of halting it. The analysis suggests that the less developed country B is (i.e., the fewer manufactures it produces), the less likely it is that production/population will spread between the two cities. In other words, the more populated foreign countries are, the more likely it is that concentration in country B will be found, which means that the existence of this immobile demand does not limit agglomeration. On the contrary, Krugman (1991,1992) points out that the greater the number of farmers (who represent an immobile demand sited in two symmetric locations analogous to our foreign markets), the more dispersion emerges since firms can obtain benefits by moving closer to this dispersed and immobile market where competition is lower.

In this paper, however, the immobile demand represented by the two foreign markets, consists of workers who also produce manufactures, with which firms in country B will have to compete. From Section 3, we already know that the greater the size of the foreign markets, the more varieties there are produced. So, the more difficult it is for firms in country B to find benefits by moving away from concentration. If the two foreign markets are very large, production in country B tends to agglomerate in a single city, since any deviating firm would have to compete with a large number of foreign firms and would lose part of its national market. This suggests that the effect of an immobile demand on the concentration of production is not always the same. The fact that the potential market does or does not produce other varieties with which to compete appears to be a crucial factor.

5 Conclusions

World's population is becoming increasingly urban: in 1994 about 75% of the population in the more developed countries, and 35% in the less developed countries lived in cities, while in 1950 only 15% of the population in the poorer countries was urban (Seitz, 1995).

Why are individuals concentrated in cities? In the last few years some papers have tried to explain this fact through formal microeconomic models where cities emerge from interaction between individuals. Some of them are non-price interaction models in which agglomeration is generated by technological externalities (see Henderson (1974) and Rauch (1991) among others). Others are noncompetitive models, especially monopolistic competition models (see, for example, Krugman (1991, 1992) and Fujita (1993)). In these models, agglomeration emerges from three sources: economies of scale at ...rm level, transport costs, and the mobility of the industrial labor force. Increasing returns to scale means that the production of each good will take place in a single location. On the other hand, the existence of transport costs implies that the best locations for a ...rm will be those with easy access to markets, and the best locations for workers, those with easy access to goods. Thus, concentration is the result of a self-reinforcing process of agglomeration.

But if it is obvious that concentration is becoming a quite common pattern of the spatial distribution of population, it is not less well-known that urban growth is dramatically intense in the less developed countries, where a large number of huge cities have started to appear: Sao Paulo is expected to surpass 23 million people in 2000, while Bombay, Shanghai or Mexico will have populations of over 16 million. There are clearly more than a few problems in analyzing the consequences of these huge concentrations of people: high poverty rates, crime, pollution and congestion.

In this paper, we have developed a model that sets out to explain the existence of megacities in developing countries, in the context of a core-periphery model à la Krugman. It has been shown that increasing returns to scale and transport costs are factors that favor agglomeration, while congestion costs prevent it. The centripetal forces are the same as in Krugman (1991), but congestion costs substitute for farmers as the centrifugal forces. The two centrifugal forces, congestion costs and the immobile demand represented by farmers, have different effects on concentration and it should be emphasized that the effects of other parameters, such as that of transportation, can differ depending on the kind of centrifugal force one considers. By considering immobile farmers, concentration is more likely when transport costs are low, because in that case firms do not increase their benefits by moving closer to the dispersed farmers. Conversely, by considering congestion costs, when transportation costs between locations decrease concentration is more difficult, since more citizens will want to move to a smaller city where congestion is lower, without paying much for transport costs in delivering goods.

As in Krugman and Livas Elizondo (1996), this paper also suggests that agglomeration can be fostered by manufacturers mainly serving the domestic market. However, the analysis goes further by emphasizing that megacities are not only the result of protective trade policies, but the consequence of the relative position of a country, in terms of industrialization, with respect to the rest of the world. It is very difficult for a developing country to break these huge agglomerations if its relative position does not improve. Since the goods produced by the less developed country have to compete with the products of the rest of the world, it would not be profitable for its firms to choose locations more distant from its national market and closer to international markets. Therefore, contrary to Krugman (1991, 1992), we find that when a country has a low level of industrialization, an immobile demand represented by foreign markets leads to concentration instead of dispersion. This allows us to emphasize that different assumptions about an immobile demand which represents a potential market may drive the economy to different spatial patterns. The fact that this demand consists of individuals who do, or do not, produce goods with which to compete seems to be crucial to the results.

Appendix A

In order to obtain the system of equations that define wages, price indices and income in any city, we should begin by solving the consumer's problem

$$\max \left(\sum_i c_i^{k\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{11}$$
$$\text{s.t.} \quad \sum_{ij} p_{ij}^k c_i^k = m;$$

where c_i^k is the consumption of good i by an individual of city k, p_{ij}^k is the c.i.f. price paid by this individual for a unit of good i produced in j, and m is the income of this individual. From the first order condition, we have

$$c_i^k = \frac{p_{2j'}^{k\ \sigma}}{p_{ij}^{k\ \sigma}} c_2^k. \tag{12}$$

Denoting total consumption of variety i in city k by C_i^k, $C_i^k = \lambda_k c_i^k$, it follows that

$$p_{ij}^k C_i^k = \frac{p_{2j'}^{k\ \sigma}}{p_{ij}^{k\ \sigma-1}} C_2^k. \tag{13}$$

Income of city k, given by equation (5), is used to pay for goods consumed in this city, i.e.,

$$Y_k = \sum_{ij} p_{ij}^k C_i^k. \tag{14}$$

Using expression (13), we can write $Y_k = p_{2j'}^k C_2^k \left[\sum_{ij} \left(\frac{p_{2j'}^k}{p_{ij}^k} \right)^{\sigma-1} \right]$. Rearranging this, we have

$$p_2^k C_2^k = \frac{Y_k p_2^{k\ 1-\sigma}}{\sum_j n_j p_j^{k\ 1-\sigma}}; \tag{15}$$

where n_j is the number of varieties produced in location j.[11]

Consider now the total sales in city k of all goods produced in city 2, namely,

$$S_{2k} = n_2 p_2^k C_2^k. \tag{16}$$

[11] Note that from equation (1) we already know that any variety in city j has the same f.o.b. price. Therefore, we can write prices in terms of locations, instead of doing it in terms of both varieties and locations. Hence, we can drop subscript i. We are identifying good 2 with any good produced in city 2.

By using the fact that the c.i.f. price paid by any individual of city k for a unit of any good delivered from city j is $p_j e^{(\tau D_{jk}+\gamma_k \lambda_k)}$ and taking into account that $p_j = W_j$, we have that revenues in city 2, from selling all the goods manufactured there, are

$$\sum_k S_{2k} = \lambda_2 \sum_k Y_k \left(W_2 e^{\tau D_{2k}+\gamma\lambda_k} T_k^{-1}\right)^{1-\sigma}; \tag{17}$$

where T_k is the price index at k given by expression (4).

Since labor is the only factor of production, one way to write the market clearing condition for workers at location 2 is that economy-wide expenditure on the workers' products must equal their income, which means that

$$\sum_k S_{2k} = W_2 \lambda_2. \tag{18}$$

From (17) and (18) it follows that the wage rate in city 2 is

$$W_2 = \left[\sum_k Y_k \left(e^{-(\tau D_{2k}+\gamma\lambda_k)} T_k\right)^{\sigma-1}\right]^{\frac{1}{\sigma}}. \tag{19}$$

This proof could be repeated for a generic city j.

Appendix B

Now we want to characterize the situation in which two cities emerge in country B as a stable equilibrium. In order to do so we have to differentiate totally the equilibrium – given by equations (3)-(5) – with respect to λ at the symmetric equilibrium ($\lambda = 0.5$), and then analyze when $\frac{d(\omega_1 - \omega_2)}{d\lambda} < 0$. By λ we mean the proportion of country B's population in city 1, namely, $\lambda = \frac{1}{\lambda_B}$, and by $1-\lambda$ that of city 2.

Since we are differentiating around the symmetric equilibrium, we have some simplifications. On one hand, any change in a variable (wage, price index) in country C is matched by an equal but opposite sign change in the respective variable in country A. For example, $\frac{dW_C}{d\lambda} = -\frac{dW_A}{d\lambda}$ (= 0 since the numeraire is at location A and, therefore, $W_A = 1$). The same happens with respect to locations 1 and 2 in country B. So, for instance, $\frac{dW_1}{d\lambda} = -\frac{dW_2}{d\lambda}$; and $\frac{dT_1}{d\lambda} = -\frac{dT_2}{d\lambda}$. On the other hand, by symmetry, $T_1 = T_2$ and $T_A = T_C$. Analogously for other variables.

It follows then that

$$\frac{d(\omega_1 - \omega_2)}{d\lambda} = 2T_1^{-1}\left(\frac{dW_1}{d\lambda} - W_1 T_1^{-1}\frac{dT_1}{d\lambda}\right): \tag{20}$$

Therefore, the symmetric equilibrium will be stable if and only if

$$\frac{dW_1}{d\lambda}W_1^{-1} < \frac{dT_1}{d\lambda}T_1^{-1}: \tag{21}$$

We want to find under what conditions the above expression holds. In order to do so, we first calculate all the endogenous variables at $\lambda = 0{:}5$.

$$T_A = e^{\gamma\lambda_A}\left[\lambda_A\left(1 + e^{\tau d_{AC}(1-\sigma)}\right) + \frac{\lambda_B}{2}W_1^{1-\sigma}e^{\tau d_{1A}(1-\sigma)}\left(1 + e^{\tau d_{1A}(1-\sigma)}\right)\right]^{\frac{1}{1-\sigma}} = e^{\gamma\lambda_A}\tilde{T}_A; \tag{22}$$

$$T_1 = e^{\gamma\frac{\lambda_B}{2}}\left[\lambda_A e^{\tau d_{1A}(1-\sigma)}\left(1 + e^{\tau d_{1A}(1-\sigma)}\right) + \frac{\lambda_B}{2}W_1^{1-\sigma}\left(1 + e^{\tau d_{12}(1-\sigma)}\right)\right]^{\frac{1}{1-\sigma}} = e^{\gamma\frac{\lambda_B}{2}}\tilde{T}_1; \tag{23}$$

$$W_1 = \left[\lambda_A e^{\tau d_{1A}(1-\sigma)}\left(1 + e^{\tau d_{1A}(1-\sigma)}\right)\tilde{T}_A^{\sigma-1} + \frac{\lambda_B}{2}W_1\tilde{T}_1^{\sigma-1}\left(1 + e^{\tau d_{12}(1-\sigma)}\right)\right]^{\frac{1}{\sigma}}: \tag{24}$$

Secondly, by differentiating the price indices and the wage rate around the symmetric equilibrium we can find the effects of a small change in the allocation of labor between locations 1 and 2. These derivatives are as follows

$$\frac{dT_A}{d\lambda}T_A^{-1} = z_1\frac{\lambda_B}{1-\sigma}\left(1 + \frac{1-\sigma}{2}\frac{dW_1}{d\lambda}W_1^{-1}\right); \tag{25}$$

$$\frac{dT_1}{d\lambda}T_1^{-1} = \lambda_B\left[\gamma + z_2\frac{1}{1-\sigma}\left(1 + \frac{1-\sigma}{2}\frac{dW_1}{d\lambda}W_1^{-1}\right)\right]; \tag{26}$$

$$\sigma \frac{dW_1}{d\lambda} W_1^{-1} = z_1 W_1^{-1} \lambda_A (\sigma - 1) \frac{dT_A}{d\lambda} T_A^{-1} + z_2 \lambda_B + z_2 \frac{\lambda_B}{2} \frac{dW_1}{d\lambda} W_1^{-1} + \tag{27}$$
$$+ z_2 \frac{\lambda_B}{2} (\sigma - 1) \left(\frac{dT_1}{d\lambda} T_1^{-1} - \gamma \lambda_B \right);$$

where z_1 and z_2 were defined to simplify the above expressions (see the similarity with the corresponding expressions given in Fujita, Krugman, and Venables (1999), chapter 18)

$$z_1 = \left(\frac{T_A}{W_1} \right)^{\sigma - 1} \left(1 - e^{-\tau d_{1A}(1-\sigma)} \right) e^{\tau d_{1A}(1-\sigma)}; \tag{28}$$

$$z_2 = \left(\frac{T_1}{W_1} \right)^{\sigma - 1} \left(1 - e^{-\tau d_{12}(1-\sigma)} \right): \tag{29}$$

The system of equations (25)-(27) is linear in variables $\frac{dT_A}{d\lambda} T_A^{-1}; \frac{dW_1}{d\lambda} W_1^{-1}$, and $\frac{dT_1}{d\lambda} T_1^{-1}$: Let us call them $x_1; x_2$ and x_3, respectively. We can now write the system with this new notation and after some arithmetic we have that

$$x_3 - x_2 = \lambda_B \left(\gamma + \frac{z_2 \frac{2\sigma - 1}{1 - \sigma}}{\sigma + z_2 \lambda_B \frac{\sigma - 1}{2}} \right) - \frac{z_1 W_1^{-1} \lambda_A (\sigma - 1)}{\sigma + z_2 \lambda_B \frac{\sigma - 1}{2}} x_1 : \tag{30}$$

Analyzing when the symmetric equilibrium is stable is then equivalent to seeing when $x_3 - x_2 > 0$:

References

[1] ADES, A. and GLAESER, E. (1995) Trade and circuses: explaining urban giants, The Quarterly Journal of Economics, 110 (1), pp. 195-227

[2] ALONSO VILLAR, O. (2000) Metropolitan areas and public infrastructure, Investigaciones Económicas, forthcoming.

[3] ALONSO VILLAR, O. (1999) Spatial distribution of production and international trade: a note, Regional Science and Urban Economics, 29, pp. 371-380.

[4] BAIROCH, P. (1988) Cities and Economic Development. From the Dawn of History to the Present, The University of Chicago Press.

[5] DEHGHAN, F. and VARGAS URIBE, G. (1999) Analysing Mexican population concentration: A model with empirical evidence, Urban Studies, 36 (8), pp. 1269-1281.

[6] DIXIT, A.K. and STIGLITZ, J.E. (1977) Monopolistic competition and optimum product diversity, American Economic Review, 67(3), pp. 297-308.

[7] FUJITA, M. (1993) Monopolistic competition and urban systems, European Economic Review, 37, pp. 305-315.

[8] FUJITA, M.; KRUGMAN, P., and VENABLES, A. (1999) The spatial economy. Cities, regions and international trade, The MIT Press.

[9] HANSON, G.H. (1996) Economic integration, intraindustry trade, and frontier regions, European Economic Review, 40, pp.941-949.

[10] HENDERSON, J.V. (1974) The sizes and types of cities, American Economic Review, 64, pp. 640-656.

[11] KRUGMAN, P. (1980) Scale economies, product di¤erentiation, and pattern of trade, American Economic Review; 70, pp. 950-959.

[12] KRUGMAN, P. (1991) Increasing returns and economic geography, Journal of Political Economy, 99, 483-499.

[13] KRUGMAN, P. (1992) A dynamic spatial model, NBER Working Paper No 4219.

[14] KRUGMAN, P. and LIVAS ELIZONDO, R. (1996) Trade policy and the third world metropolis, Journal of Development Economics, 49, pp. 137-150.

[15] PUGA, D. (1998) Urbanization patterns: European versus less developed countries, Journal of Regional Science, 38(2), pp. 231-252.

[16] PUGH, C. (1997) Poverty and progress? Reflections on housing and urban policies in developing countries, 1976-96, Urban Studies 34(10), pp. 1547-1595.

[17] RAUCH, J.E. (1991) Comparative advantage, geographic advantage and the volume of trade, The Economic Journal, 101, pp. 1230-1244.

[18] SEITZ, J.L. (1995) Global Issues: An Introduction, Blackwell Publishers.

[19] STARRETT, D. (1978) Market allocations of location choice in a model with free mobility, Journal of Economic Theory, 17, pp. 21-37.