Trade Restrictions, Migration, and Economic Geography

DANIEL HELLER*

1. INTRODUCTION

When Thomas Sankara took office as President of Burkina Faso in 1983 he decreed that all the luxury limousines owned by the government be sold immediately, and outlawed emigration from the countryside into the country’s capital of Ouagadougou. With the first measure, he tried to signal his disapproval of government waste. By restricting the mobility of Burkina Faso’s citizens he had hoped to avert a development many other Third World countries had previously experienced, namely population «overgrowth» in their largest cities. Between 1950 and 1990, cities in developing countries grew at an average rate of more than 4% p.a. – compared to about 2% in the industrialized world. This population explosion in poor countries can be traced back statistically to two reasons. The first is a high overall population growth rate. The second, and more important, reason is substantial rural-urban migration of the type that Sankara tried to stop by decree (BRUECKNER 1990, RICHARDSON and TOWNROW 1986).

There is a large body of literature that tries to explain the observed increasing spatial disparity. One branch of literature focuses on the effects of economic policy that have brought forth a distorted economic geography. According to this view, an «urban bias» exists due to policy interventions that favor agglomerations, for instance, domestic terms of trade that are skewed against agriculture, tariffs and exchange rate management that protect the urban industry, or heavy investments in the infrastructure of a single prime city (HAMER and LINN 1987). HENDERSON (1982) and ADES and GLAESER (1995) argue along similar lines, stating that in non-democratic states spatial policies have been used to assure maximal control over a country and its population.

Another important strain of the literature points out that cumulative processes or circular causations are important in understanding the shape of the economic landscape (MYRDAL 1957, HIRSCHMAN 1958, PRED 1966, and later FRIEDMANN 1973). The origin of a large city can lie either in a competitive advantage given by particularly good natural conditions (port or natural resource), or it can lie in a mere historical accident such as the birth place of a innovative entrepreneur. Due to an initially high marginal productivity of capital, this location experiences an inflow of capital, and investment increases. This,
in turn, leads to a rise in real wages. More people with varied skills are attracted from other regions. The circular causation continues since the relatively high skill level of workers is conducive to the formation of new companies. Economically speaking, there are agglomeration economies for both producers and consumers. This cumulative process stems entirely from market forces. Hence, the result of laissez-faire policies is a core-periphery pattern of the economic landscape. Regional disparities are reinforced until the diseconomies of agglomeration become too strong. MYRDAL'S (1957) analysis with respect to these diseconomies of scale is quite vague. However, he states that the core-periphery dichotomy becomes less acute with improvements of interregional transport and communication (p.34).

In the spatial model to be presented in this paper we try to incorporate the circular causation process mentioned in the last paragraph. Our basic framework is borrowed from Krugman-type models of economic geography (for instance, KRUGMAN 1991). In these models interregional transport costs play a crucial role in determining the shape of the economic landscape. Transport costs give rise to agglomeration economies. Consumers and producers benefit from agglomerating in a large city since this minimizes interregional transport costs. This circular relationship between producers and consumers induces a cumulative process that may eventually lead to a core-periphery geography. In the spirit of Myrdal, we will develop a North-South model in which each hemisphere consists of two regions, East and West. All four regions engage in trade with one another. Transaction costs in the form of transport costs or tariffs occur when goods are shipped from one region to another. The ultimate goal of the model is to analyze how changes in either transport costs or tariffs influence the geographic distribution of economic activity in North and South.

The remainder of the paper is structured as follows. In the next section, we present the formal outline of our model. In Section 3 the short and long-run equilibria of the model are derived. Section 4 contains some results we obtained from numerical simulations of the model and discusses their major implications. Concluding remarks are provided in Section 5.

2. THE MODEL

Let us assume that the world consists of two hemispheres or continents called North and South. Each continent is divided into two regions, East and West. The regions are of equal size $T$. Total population $L$ is normalized to one, with half of the population living in North and the other half in South. No migration between the continents is possible. There is, however, complete and costless mobility between East and West:

1. MYRDAL (1957) speaks in this context of «external dis-economies».
\[ L_{nw} + L_{ne} = 0.5 \]
\[ L_{sw} + L_{se} = 0.5 \]  
(1)

All individuals share the same utility function \( U \). They derive utility from consumption goods and from quality of life:

\[ U = C^\psi Q^{1-\psi} \]  
(2)

The term \( C \) is an aggregate of a large number of different consumption goods. The idea behind the aggregate \( C \) is that people value both quantity and diversity in consumption. Following DIXIT and STIGLITZ (1977), we assume that \( C \) takes the form:

\[ C = \left( \sum_{i} c_i^{\sigma-1} \right)^{\frac{\sigma}{\sigma-1}}. \]  
(3)

In equation (3), \( c_i \) denotes the quantity of good \( i \) consumed by an individual. Good \( i = 34 \), for example, could be Camembert cheese produced in the eastern region of North, and good \( i = 121 \) could be coffee beans produced in South West. If \( \sigma \) is larger than one \( C \) increases with the variety of goods consumed.

The second component of the utility function is quality of life \( Q \). Quality of life can be interpreted in many ways. It consists of the availability of recreational facilities, cleanliness of the air, commuting time, safety, and so forth. Empirical studies suggest that congestion and pollution – which contribute to a lower quality of life – usually increase with the population size in a location. Correlations also exists for crime rates (positive) and education (negative). Crime rates tend to increase whereas the quality of the schools tends to decrease with the size of a location (FUJITA 1989 and DICKEH and LLOYD 1990). Hence, it seems reasonable to assume \( Q \) to be the inverse of the population density in the location in which an individual lives.

Every individual maximizes her utility subject to her budget constraint, which is, in our case, equal to the wage rate \( w \). This maximization yields the individual demand function for any good \( j \) (see HELPMAN and KRUGMAN (1985), ch. 6, for details):

\[ c_j = \frac{P_j^\sigma w}{\sum_i P_i^j - \sigma} \]  
(4)

In this demand function, \( P_j \) denotes the price of good \( j \) in the region in which it is consumed. The price \( P_j \) is the c.i.f. price of \( j \), i.e. it includes transport costs or tariffs that
are incurred between the locations of production and consumption. The relationship between f.o.b. and c.i.f. prices will be discussed in more detail below.

Several points about equation (4) should be noted. Demand for any good depends linearly on the wage rate. Since we assume that quality of life has no market price, the entire wage is spent on consumption goods. Also, the demand for $j$ decreases with the number of differentiated goods. Furthermore, if the number of different goods is large, the demand elasticity is equal to $\sigma$.

As mentioned earlier, a central element in recent models of economic geography is that transport costs are an important factor in determining the economic landscape. With a constant return production structure and limited transport costs, there would be no economic incentive to agglomerate. Since the scale of production does not play a role in this case, everyone would produce all the goods herself in order to minimize transport costs. But with increasing returns of scale in production and no transport costs, there would be no agglomerations either. Production would occur wherever costs are lowest. A combination of increasing returns in production and limited transportation costs, however, provides an economic incentive to agglomerate. Transport costs are minimized if consumers and producers cluster in a single location. Consumers benefit because the less they have to pay for getting the goods to their homes, the higher their real wage. At the same time, producers can sell more because no purchasing power is «wasted» on transportation costs.

---

**Figure 1**

![Diagram](attachment:image.png)

---

2. The abbreviation c.i.f. stands for «cost, insurance, freight». A price net of transport costs is called an f.o.b. price which is short for «free on board».

3. If $Q$ were treated like a good that could be bought like, for instance, housing – a fraction $\phi$ of income would spent on $C$ and $(1 - \phi)$ on $Q$. 

In order to make the constant demand elasticity \( \sigma \) workable, transport costs, or tariffs, \( \tau \) take an «iceberg» form. For every unit shipped from one region to another only \( 1 / \tau \) units arrive \( (\tau > 1) \). No transport costs are added to goods that are produced and consumed in the same location. As shown in Figure 1, three different kinds of \( \tau \) exist in our model: transport costs for goods traded between East and West in North are denoted by \( \tau_n \). Analogously, \( \tau_s \) are the shipping costs for goods sent between the two regions of South. Finally, \( \tau_{ns} \) denotes the cost of sending goods from one continent to the other. The term \( \tau_{ns} \) includes of both transport costs as well as tariffs.

As in Rivera-Batiz (1988), production takes place under increasing returns to scale that are internal to a firm. The scale economies originate from indivisibility in production, which is captured by a fixed cost term \( a \) in the production function. With labor \( l_i \) as the only input, the production function for good \( i \) takes the form:

\[
l_i = a + bx_i ,
\]

where \( x_i \) is the output of \( i \). Due to the increasing returns each variety is produced by a single monopolist. Every monopolist maximizes her profits. The relevant price for the monopolist is not the c.i.f. price \( P \) paid by the consumers, but the f.o.b. price \( p \) (with \( P = \tau p \)). The profit-maximizing price is determined by the standard mark-up over marginal costs. Since the marginal costs are equal to \( bw \), the price set by a monopolist depends on the region in which it was produced. For a producer in North East, for instance, the optimal price is:

\[
\begin{align*}
\pi_{i,ne} &= \sigma - 1
\end{align*}
\]

Equation (7) states that the quantity produced of every good is independent of the region of production. Also, equations (5) and (7) show that labor input per company is \( l = a \sigma \), irrespective of the location. Hence, the number \( n \) of different goods manufactured in each region is proportional to the number of workers who live there. So the \( L_{lw} \) workers of North West produce \( n_{nw} = L_{nw} (a \sigma ) \) different goods.
The model presented here is similar to KRUGMAN and LIVAS-ELIZONDO (1992). In their model, KRUGMAN and LIVAS-ELIZONDO distinguish among three locations: two are domestic locations (comparable to the South in our model) and one is the rest of the world. The domestic market is only about one tenth of the size of the rest of the world. Transport costs are incurred in trade between the domestic locations. Imports from the rest of the world to their «South» are subject to transport costs and tariffs. No costs are, however, added to exports from the South. The centrifugal force, $Q$ in our case, is given in their model by the commute time to the central business district. People living in the larger location experience a real wage loss from longer commutes.

3. SHORT-TERM DYNAMICS AND LONG-RUN EQUILIBRIA

In order to solve the model, we can now set supply equal to demand. Supply $x$ of any consumption good is given by (7), and the demand function is determined by (4). For example, the demand of an individual living in South East for a good produced in North West is:

$$c_{se,nw} = \frac{(p_{nw} \tau_{ns})^{-\sigma}}{n_{nw}(p_{nw}\tau_{ns})^{1-\sigma} + n_{ne}(p_{ne}\tau_{ns})^{1-\sigma} + n_{sw}(p_{sw}\tau_{s})^{1-\sigma} + n_{se}p_{se}^{1-\sigma}} \cdot \frac{p_{se}(\sigma - 1)}{\sigma b}.$$  

(8)

Also, a firm's total sales are the sum of its sales in the four locations. For a firm located in South East, this means:

$$p_{se}x = p_{se} \left( L_{nw}c_{nw,se} \tau_{ns} + L_{me}c_{me,se} \tau_{ns} + L_{sw}c_{sw,se} \tau_{s} + L_{se}c_{se,se} \right).$$  

(9)

In equation (9) each of the four terms within the parentheses denotes the total quantity of a particular good shipped from the production site in South East to a particular region. Substituting in the expressions for $x$ and $n$, the terms from the production function, $a$ and $b$ disappear from equation (9). In a market equilibrium, there are four equations like (9), one per representative producer in each region. This system of four equations can be solved for the equilibrium prices $p_{nw}$, $p_{ne}$, $p_{sw}$ and $p_{se}$, given the population distribution, the transport costs and tariffs.

In order to be able to make utility comparisons among regions, we have to know the consumption aggregate $C$ and the quality of life $Q$ in each region. The consumption aggregates can be calculated via the equilibrium prices and equations (3) and (8). As discussed earlier, quality of life term $Q$ is assumed to be the inverse of the population

4. KRUGMAN (1991), for instance, calculates real wages in the different locations via price indices. The method of calculating the consumption aggregates is, of course, equivalent to the method used in constructing a price index.
density. For someone who lives in North West, $Q$ is equal to $T/L_{nw}$. Furthermore, we impose a migration pattern according to which the region in each hemisphere with the higher utility will attract some migrants from the other region (remember that $L_{nw} + L_{ne} = 0.5$ and $L_{sw} + L_{se} = 0.5$). The magnitude of this population influx depends on both the relative utility difference and the population size of the smaller region. After the first round of migration we apply the same procedure again for the new population distribution. We repeat this iterative process until a population distribution is found in which the workers' utility in West and East is (almost) identical. By convention, we also define West to be the more populous of the regions.

Table 1: Equilibria with $\tau_n = 1.1$, $\tau_{ns} = 1.3$

<table>
<thead>
<tr>
<th>$\tau_s$</th>
<th>North</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>West</td>
<td>50%</td>
</tr>
<tr>
<td>1.3</td>
<td>West</td>
<td>50%</td>
</tr>
<tr>
<td>1.4</td>
<td>West</td>
<td>50%</td>
</tr>
</tbody>
</table>

$\sigma = 4$, $\varphi = 0.9$

Tables 1 through 3 summarize the results of some numerical simulations of this model. Table 1 shows how different transport costs within the Southern continent influence the spatial equilibrium. The parameters held constant in this example were $\tau_n = 1.1$, $\tau_{ns} = 1.3$, $\sigma = 4$ and $\varphi = 0.9$. For $\tau_s = 1.2$, the only stable equilibrium is complete symmetry between the East and West. At any distribution other than symmetry in North and South, the less densely populated region is more attractive than the more crowded region. Low transport costs mean that the agglomeration economies are not sufficiently large to generate a core-periphery pattern. With a slightly worse infrastructure within the South ($\tau_s = 1.3$), the outcome changes. For $\tau_s = 1.3$ a spatial equilibrium is reached in the South with 72.5% of its population living in West, while no concentration occurs in the North. This result essentially stems from the existence of agglomeration economies or a <<home market effect>>. Since the number of different goods produced in a region depends linearly on its population size, a large region does not import as many different goods as a small one. For consumers, this means that the larger the home market the less money they have to spend on transport costs, i.e. the higher is their consumption. Producers are subject to this <<home market effect>> too. With output $x$ being equal for all firms, a large home market means that only small fraction of $x$ is exported. Since $\sigma > 1$ and $\tau > 1$, the f.o.b. price that a company can charge increases with the size of the home

---

5. This migration pattern assures convergence to a population distribution where the utility levels are the same in both regions.

6. Since no migration is allowed between hemispheres, there is no necessity for utility differences between North and South to disappear.
market. With increasing transport costs, the size of the home market gains importance. At \( \tau_s = 1.3 \) the agglomeration economies in the South are so strong that the West has to be more than 2.5 times as crowded and congested as the East for utility to be equalized. Similarly, a further increase of \( \tau_s \) to 1.4 induces an even more concentrated economy in the South. 90% of the Southern population will live in a single region. The spatial equilibrium in the North remains unaffected by these changes in \( \tau_s \), i.e. the agglomeration economies in the North are not strong enough to outweigh the pollution effect.

It is worth noting, that for \( \tau_s = 1.3 \) and 1.4, additional equilibria in the South exist with exactly 50% of the population living in each region. However, these equilibria turn out to be unstable. A small deviation from the symmetric equilibrium rises utility in the (slightly) larger region relative to the one in the smaller region. People migrate up to the point at which the large region becomes congested enough to equalize the utility levels again. Because of the Cobb-Douglas-type utility function no region will ever be completely abandoned, even if transport costs are extremely high, for the first (infinitesimally) small unit of quality of life provides infinite marginal utility.

Table 2: Equilibria with \( \tau_n = 1.1, \tau_s = 1.3 \)

<table>
<thead>
<tr>
<th>( \tau_{ns} )</th>
<th>North</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>West</td>
<td>East</td>
</tr>
<tr>
<td>1.2</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>1.3</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>1.4</td>
<td>50%</td>
<td>50%</td>
</tr>
</tbody>
</table>

\( \sigma = 4, \varphi = 0.9 \)

Table 2 shows the impact of varying tariffs or intercontinental transport costs \( \tau_{ns} \). Again, it is assumed that the infrastructure within the South is less advanced than within the North (\( \tau_n < \tau_s \)). For low tariffs (\( \tau_{ns} = 1.2 \)) we find a similar result as before: transport costs are too low to make the gains from agglomeration large enough to outweigh the disutility from pollution and congestion. The only stable population distribution is, therefore, the symmetric equilibrium. This changes if \( \tau_{ns} \) is increased to 1.3. Then, the spatial equilibrium in South lies at a point at which 72.5% of the South's population lives in West. In fact, we have already discussed this result for identical parameter values in Table 1. It does not come as a big surprise that for \( \tau_{ns} = 1.4 \) the equilibrium economic landscape in the South must be even more concentrated.
Finally in Table 3, the impact of different transport costs within the North is analyzed (with $\tau_s = \tau_{ns} = 1.3$). With $\tau_n = 1.2$, the centripetal force is not strong enough to prevent a symmetric equilibrium in the North. After an increase of $\tau_n$ to 1.25, the symmetric population distribution in North still prevails, while the South becomes somewhat less concentrated. A higher $\tau_n$ reduces the purchasing power of the North and producers in North will have to lower their f.o.b. prices. Apparently, consumers in South East benefit from this price reduction more than their compatriots in the West. If $\tau_n$ is equal to 1.28, agglomeration economies in the North are large enough to produce a 60:40 population distribution, while concentration in the South further decreases. Under the assumption that $\tau_n = 1.3$, all transport costs are identical and the equilibrium population distribution in North and South is expected to be the same. If $\tau_n$ is higher than $\tau_s$ ($\tau_n = 1.4$) 89% of North’s and 71.5% of South’s population is concentrated in one location. The results listed in Table 3 suggest that the spatial equilibrium in the South is indirectly influenced by $\tau_n$ in two ways. First, an increase in $\tau_n$ seems to reduce the tendency toward concentration in the South. But, second, as soon as a higher $\tau_n$ gives rise to an uneven economic landscape in the North, the centripetal forces in the South are enhanced.

The results reported in Tables 1 and 2 are consistent with those found by KRUGMAN and LIVAS-ELIZONDO (1992): lower transport costs within the developing country and lower tariffs between countries tend to lead to a less concentrated economic landscape in the South. Because our model explicitly contains an industrialized country we are able to point out a relationship that goes beyond KRUGMAN and LIVAS-ELIZONDO (Table 3). We show that the infrastructure (or $\tau_n$), as well as the magnitude of concentration in the North can have an impact on the spatial structure in the South. If the North is highly concentrated, it is likely that we find a core-periphery geography in the South too.

### 4. CONCLUSIONS

Recent models of economic geography have focused on the impact of transport costs on the spatial distribution of economic activity. In some models, such as KRUGMAN (1991)
and Krugman and VENABLES (1993), a reduction in transport costs increases the monocentric tendencies. The model presented here, along with Krugman and Livas-Elizondo (1992) suggests exactly the opposite, namely that in particular smaller locations would become more attractive after a reduction in transaction costs.

Our model suggests two distinct measures to counter «overgrowth» of Third World cities. The first measure is investment in the infrastructure of developing countries. Better roads and telecommunication facilities among the regions of a developing country enhance the attractiveness of the location outside the center for both producers and consumers. If it is difficult to transport goods to the large markets, even low wages and low land prices are not large enough an incentive for companies to move to a peripheral location. The second suggestion that comes out of the model is a reduction of tariffs between industrialized and developing countries. Since remote and small locations in the South depend heavily on exports, they are especially hurt by tariffs and they become less appealing locations. In this sense, our model supports the policy recommendation that Myrdal (1957) made several decades ago. As mentioned in the introduction, he argues that developing countries will remain more monocentric than industrialized countries as long as the infrastructure of the former was not improved.

REFERENCES

This paper investigates why Third World cities have been growing in the last several decades much more quickly than cities in industrialized countries. For this purpose, we develop a Krugman-type model of economic geography with two continents, North and South, each of which consisting of two regions, East and West. We study the impact different levels of transport costs and tariffs exert on the distribution of economic activities among the regions. We find that lower costs for transport between the regions in the South, for instance, induced by an improved infrastructure, as well as lower tariffs on intercontinental trade tend to lead to less concentrated economies.

ZUSAMMENFASSUNG

Cet article tente d’expliquer pourquoi les villes des pays du tiers monde ont enregistré une croissance de leur population sensiblement plus rapide que celles des pays industrialisés, au cours des dernières décennies. Nous développons un modèle de géographie économique semblable aux modèles de Krugman, constitué de deux continents, le nord et le sud, et de deux régions, l’est et l’ouest. Nous analysons l’impact qu’exercent différents niveaux de coûts de transport et de tarifs douaniers sur la distribution des activités économiques entre les régions. Il résulte par exemple que des coûts de transport réduits entre les régions du sud ainsi que des tarifs inférieurs entre les continents conduisent à des économies moins concentrées.