NHH
Norges
Handelshøyskole
Norwegian School of Economics
and Business Administration

Economic geography and development

by

Kjetil Bjorvatn

A dissertation submitted for the degree dr. econ

Acknowledgements

I would like to express my gratitude to my supervisor, Karl Rolf Pedersen, for excellent supervision. His comments and suggestions have significantly improved the quality of my work. I would also like to thank the other members of my dissertation committe, Karl Ove Moene and Rögnvaldur Hannesson, for their guidance and encouragement. My friends and colleagues at the Institute of Economics have made my years as a Ph.D. student a pleasant experience, both in an intellecutal and social sense. Thank you all.

Financial support from the Research Council of Norway, the Norwegian School of Economics and Business Administration, and the Foundation for Research in Economics and Business Administration is gratefully acknowledged.

Finally, my love and affection goes to my wife, Afsaneh, and our son, Daniel. They mean everything to me.

Bergen, June 1996

Kjetil Bjorvatn

Contents

Chapter 1

Introduction

1. Background

Why are some nations rich and others poor? To me, this is the fundamental question of economics. Indeed, it was the question which inspired Adam Smith (1776) to write «The wealth of nations», thus providing the foundation for contemporary economic thought. Economic geography, i.e., the study of the location of production in space, is almost as old as economic science itself. The field was initiated by von Thünen (1826) in his «Der isolierte Staat», published only fifty years after «The wealth of nations». For a modern exposition of von Thünen's model, see Samuelson (1983).

The fundamental question of economic geography is closely related to the one addressed by Adam Smith, namely: Why are some *regions* rich and others poor? The regional approach provides a natural starting point for the analysis of the nature and causes of the wealth of nations. As argued by Krugman (1991: 3), «... one of the best ways to understand how the international economy works is to start by looking at what happens *inside* nations.» This is because regional data are likely to be better and pose fewer problems of compatibility than national data. Furthermore, the importance of political institutions as determinant of divergence in wealth is probably reduced by using regional rather than national data.

On a theoretical level, regional and international economics are certainly closely related. The theory of international trade may in fact be

viewed as a special case of economic geography, one which abstracts from factor mobility and transportation costs. With greater economic integration across national borders, the distinction between regional and international economics becomes increasingly blurred. In the words of Fujita and Thisse (1995): «As market integration dissolves economic barriers between nations, national boundaries no longer provide the most natural unit of analysis.»

Despite its long history within economic thought, economic geography has remained in the intellectual periphery of economics until very recently. Krugman (1991: 4) explains the marginalization of location theory in the following way:

> The neglect of spatial issues in economics arises for the most part from one simple problem: how to think about market structure. Essentially, to say anything useful or interesting about the location of economic activity in space, it is necessary to get away from the constant-returns, perfect-competition approach that still dominates most economic analysis. As long as economists lacked the analytical tools to think rigorously about increasing returns and imperfect competition, the study of economic geography was condemned to lie outside the mainstream of the profession. Indeed, as standards of rigor in economics have risen over time, the study of location has been pushed further and further into the intellectual periphery.

The new developments in economic geography are closely related to recent progress made in trade and growth theory. Common denominators are increasing returns to scale and imperfect competition, with models based on the Dixit and Stiglitz (1977) monopolistic competition approach. For a survey of the classical and new literature on economic geography, see Fujita and Thisse (1995) and Knarvik (1995).

An important source of motivation for the increased interest in location theory has been the integration of national economies within trading blocks such as the European Union and NAFTA. Although developments in the predominantly industrialized world in this way motivate theory, Krugman (1995: 241-242) argues that new economic geography is better suited for analysing questions in the less developed world. First, problems of urbanization and regional inequalities appear to be greater in poor countries. Second, policy changes that are taking place, or are likely to take place, in Third World countries may be expected to have quite dramatic effects on their economic geography. For instance, according to Ades and Glaeser (1995), a reduction in the degree of government intervention and in the concentration of political power lead to a greater regional balance in resources and incomes. Third, the relevance of the new theories, focussing on transportation costs and increasing returns to scale, is likely to be greater in poor countries. One reason for this is the limited purchasing power and hence limited market size in these countries, a second reason being their larger share of tangible goods in production, goods which must be transported by road or rail.

While these justifications for the relevance of economic geography to the analysis of economic development may well be true, the methodology used by the overwhelming majority of recent contributions to the field, namely the Dixit and Stiglitz monopolistic competition model, appears less convincing in many Third World contexts. In two of the papers of this dissertation, an alternative formulation to the monopolistic competition model is presented, namely the dual technology model. For reasons described in the outline of the thesis, this model seems particularly suitable for analysing questions of location and industrialization in poor countries.

The central motivation for this thesis is the application of economic geography to development issues. The first article focuses on location of industrial activity. It analyses the effects of transportation

costs and technology on locational choice, and draws welfare implications. The second article addresses questions of city size. Cities in the model may be either too small or too large in efficiency terms, and countries with similar economic structure may have quite dramatic differences in degree of urbanization. The third article deals mainly with political factors in explaining urbanization and industrialization. It derives the optimal tax policy of a surplus maximizing government, and shows that while such policies may reduce inefficiencies in the rural area due to a reduction in surplus labor, political weakness may cause inefficiencies in the urban area through the creation of an oversized public sector. In the fourth article, a more detailed model of political economy is presented, focussing on aid efficiency in a rent-seeking society. Drawing on the rotten kid theorem by Becker (1974, 1976), the article demonstrates that aid efficiency is likely to be high when transfers are of a discretionary kind.

2. Outline

The first two essays of the thesis analyse traditional causes for locational choice, such as increasing returns to scale, transportation costs and labor immobility. The basic contribution of these articles is the application of the dual technology model introduced by Shleifer (1986) to questions of location. The dual technology model has been used by Murphy, Shleifer and Vishny (1989a, 1989b) in studies of Third World industrialization. An important assumption in the Dixit and Stiglitz model, which is the dominant analytical framework in recent contributions to economic geography, is the existence of a large number of firms and free entry and exit which drive profits to zero. While this may be a reasonable approximation to reality in rich countries, in poor countries the number of firms and variety of manufactured goods are typically relatively small, and market-entry is often very difficult due to imperfections in capital markets and bureaucratic barriers. Such imperfections on both the economic and political level give rise to a

limited number of large scale producers, the formal sector, with a potential for making profits and a competitive fringe of cottage producers, the informal sector. These realistic features of a Third World economy are captured by the dual technology model.

«Industrialization and regional inequality» focuses on the regional location of large scale technology in a two region model with inter-regional migration and trade. Labor moves to the region offering the higher expected wage, in accordance with Harris and Todaro (1970). Wage bargaining takes place between trade union and employer in the formal sector, while income from employment in the informal sector is determined competitively.

Two scenarios emerge in this model, one in which the large scale producers choose to locate in a single region (centralization, or equivalently the center-periphery scenario) and one in which the large scale producers choose to locate in separate regions (decentralization). In the decentralization outcome, expected income is equalized across regions through migration. The center-periphery scenario, on the other hand, is characterized by an expected wage in the center region exceeding labor income in the periphery, the latter inhabited by only immobile labor. This scenario not only creates regional income inequalities. Concentration of factories and labor in one region may also be Pareto-inferior to a situation with a more even distribution of labor. Considerations of both efficiency and equality may therefore be invoked in the support of government policies aiming at a regional balance of industrial base.

Gains from trade in this model arise from increasing returns to scale in production. By locating in a single region, the size of the market is effectively reduced, since demand in the periphery is satisfied by local, small scale supply. Locating in separate regions increases market size, but at the same time involves transportation costs since goods now have to be shipped across regional borders. Locational choice is thus based on a tradeoff between market size and

transportation costs. Transportation costs here act as a centripetal force, which means that an increase in transportation costs makes decentralization of large scale technology less likely.

This result is rather interesting because monopolistic competition models of similar structure generate the opposite result, namely that transportation costs are a centrifugal force, see for instance Krugman (1991) and Krugman and Venables (1993). The reason for this contrast in results relates to the fact that with dual technologies, the degree of competition is uniquely determined by the availability of small scale technology. Under monopolistic competition, on the other hand, high transportation costs reduce the degree of competition between regions, thereby making it more profitable for some firms to leave the center and serve local markets. When transportation costs are low, interregional competition is more fierce and there is less to gain by serving local markets. Instead, firms will have an incentive to reap the benefits of scale economies by locating in the center region where the market is relatively large.

The policy implications derived from the two modelling approaches differ accordingly. With monopolistic competition, an investment in infrastructure is not such a good idea if one is concerned with equality in income and industrial base between regions, while the dual technology approach would lend support to such a policy on both efficiency and equality grounds. The empirical study by Ades and Glaeser (1995: 213), claiming that «well developed transportation facilities lower the size of central cities», supports the conclusion from the dual technology approach. Casual observation points in the same direction: The degree of urban concentration in African countries is quite large and increasingly so although the quality of infrastructure is poor and in some cases appears to have been deteriorating during the last decades.

Whereas the above article focussed on transportation costs on goods crossing regional borders, the next paper in the dissertation, «City

size and economic development», deals with congestion costs within regions, such as urban commuting costs. There are two regions in the model; the city and the rural area. Manufacturing, for which dual technologies are available, takes places in the city and agriculture in the rural area. Market imperfections may arise due to monopolistic pricing, congestion externalities and coordination failure.

Multiple equilibria and the possibility of coordination failure is perhaps the most interesting case. Intuitively, the presense of multiple equilibria within the present context can be explained in the following way. Consider a small city. This city size involves a relatively efficient manufacturing sector due to low congestion costs. The large surplus generated from manufacturing creates a large demand for the agricultural good which makes it profitable for a large number of people to stay in the rural area supplying these goods. Hence, the small city and large rural region is an equilibrium configuration. But consider also the case of a large and therefore inefficient city. Here, congestion costs have crowded out a great deal of the potential surplus from large scale manufacturing, thus depressing demand for the agricultural good and making it profitable for only a limited number of people to stay in the rural area. The large and inefficient city may therefore also be an equilibrium situation.

In this way, two structurally identical countries may differ in a rather dramatic way in terms of economic efficiency and geography. The country coordinating on the inefficient equilibrium will be characterized by excessive city size with manufacturing taking place in the informal sector. The country coordinating on the efficient equilibrium, on the other hand, will be characterized by a profitable formal sector and a smaller degree of urban concentration. One rather paradoxical policy implication from this model is that industrialization may be encouraged by subsidizing agriculture. This policy would attract labor from the city to the rural region, thereby reducing congestion costs and laying the foundation for profitable large scale production.

The third article, «Leviathan in a dual economy», deals with migration and industrialization in a political economy context. Duality here refers to political and social institutions rather than to technology as in the first two articles. The rural area is characterized by income sharing, a common institution in many poor countries, see Fafchamps (1992) for a game-theoretical analysis of such solidarity networks and Platteau (1991) for a survey of the anthropological literature. Although income sharing takes various forms, its basic function is to limit the risk of starvation amongst the members of the community. With complete income sharing, peasants are remunerated according to average rather than marginal product, thereby creating overemployment in the rural area. This hypothesis is in accordance with the seminal contribution of Lewis (1954).

The urban area offers employment opportunities in manufacturing and bureaucracy. Manufacturing is competitive while bureaucracy is characterized by overemployment and an institutionally fixed wage markup. Migration between the regions equates the rural income with expected wages in the city. In this way, the model draws inspiration also from Harris and Todaro (1970). Shifting the focus of attention from the rural area to institutional rigidities in the urban area, their conclusion with respect to the efficiency of the regional allocation of labor is exactly opposite to that of Lewis. In the Harris-Todaro framework the formal sector wage markup coupled with migration according to expected wages create excessive rural-urban migration, as illustrated by the resulting pool of urban unemployed. Since my paper includes institutions which may create overemployment in both the rural and urban area, this article can be seen as presenting a synthesis between the Lewis and Harris-Todaro models.

The central contribution of the paper is however the introduction of a political duality. Based on the assumption that the rural population is less influential than the urban population, the

government for political reasons is limited to tax only the agricultural good. Furthermore, by exerting pressure on the government, the city dwellers are able to extract a share of government income. This kind of urban «income sharing» is organized through the government offering the urban population well paid positions in the bureaucracy, a common way to exercise political patronage in many Third World countries, see the World Bank (1979). Since income sharing is both a rural and an urban phenomenon, it is not possible to determine a priori whether the Leviathan's policies of urban bias increase or reduce the distortions in the economy.

From an empirical viewpoint, the causes of rural-urban migration analysed in the «Leviathan» paper are probably more significant than those of the first two papers in the dissertation. The following quote from Ades and Glaeser (1995: 224) captures much of the essence of my paper:

> Our political results are stronger than our results on trade. They display a robust causality running from dictatorship to urban centralization. Urban giants ultimately stem from the concentration of power in the hands of a small cadre of agents living in the capital. This power allows the leaders to extract wealth out of the hinterland and distribute it in the capital. Migrants come to the city because of the demand created by the concentration of wealth, the desire to influence the leadership, the transfers given by the leadership to quell local unrest, and the safety of the capital. This pattern was true in Rome, 50 B.C.E., and it is true in many countries today.

The last article in my dissertation, «Rent seeking and foreign aid», is only remotely related to economic geography. Rather, it develops further the issue of political economy. In the «Leviathan» paper, the interaction between lobbyists and government was treated rather

crudely, represented simply by an «income sharing» function. In «Rent seeking and foreign aid» a game theoretical approach is introduced in which interest groups compete with each other for the favor of local governments, responsible for the implementation of economic policy. Such influence activities are assumed to involve real resources, call it lobbying, and therefore constitute economic waste.

The paper analyses the effects of foreign aid on the political economic equilibrium in this rent-seeking society. It turns out that the degree of aid efficiency, i.e., the ability of the donors to reduce inefficiency and inequality in the recipient country, depends in a crucial way on the timing of aid relative to the lobbying decisions of the interest groups. When aid is allocated to the groups prior to their lobbying decision, i.e., when aid is given under commitment, it triggers an increase in rent seeking. Part of the aid is therefore dissipated. Furthermore, due to the endogeneity of economic policy, the aid will in effect end up being allocated according to the political influence of the recipient groups rather than according to the objectives of the donor. Aid efficiency in this scenario is therefore low.

This rather pessimistic result, noted by Pedersen (1995), should be contrasted with the second scenario, in which aid is given under discretion. In the discretion scenario, the flexibility of the donors in their implementation of aid projects is assumed to be greater than the flexibility of the pressure groups in their organising of a lobbying response. In practice, aid given under discretion involves short term projects which in some sense «surprise» the recipients. Drawing on the rotten kid theorem by Becker (1974, 1976), discretionary aid may in fact eliminate income inequalities and lobbying. The intuition is rather straightforward: when all groups qualify for aid, any asocial activity such as lobbying which increases personal benefit but at a cost for society at large, will be «punished» by a reduction in aid. As long as all groups qualify for aid, this punishment will be greater than the initial gain from lobbying and everybody will refrain from these acitivities. To

the degree that inequalities in income are caused by the influence efforts of groups with an unequal distribution of political skill, eliminating rent seeking also eliminates inequalities in income. These rather extreme results are of course the consequence of a highly stylized model. As always in economic modelling, the results should be interpreted as tendencies rather than accurate predictions of reality.

References

Ades, A. F. and Gleaser, E. L. (1995). «Trade and circuses: Explaining urban giants,» *Quarterly Journal of Economics* 110: 195-227

Becker, G.S. (1974). «A theory of social interactions,» *Journal of Political Economy* 82 6: 1063-1093

Becker, G.S. (1976). «Altruism, egoism, and genetic fitness: Economics and sociobiology,» *Journal of Economic Literature*: 817-826

Dixit, A.K. and Stiglitz, J.E. (1977). «Monopolistic competition and optimum product diversity,» *American Economic Review* 67: 297-308

Fafchamps, M. (1992). «Solidarity networks in preindustrial societies: Rational peasants with a moral economy,» *Economic Development and Cultural Change* 41 1: 147-174

Fujita, M. and J.-F. Thisse (1995). «Economic geography: old problems and new perspectives,» mimeo

Harris, J. and Todaro, M. (1970). «Migration, unemployment and development: a two-sector analysis,» *American Economic Review* 60 (March): 165-187

Knarvik, K.H.M. (1995). «Economic geography and trade: a survey of the literature,» Chapter 1 in the Ph.D. dissertation *On the location of economic and academic activity*, The Norwegian School of Economics and Business Administration

Krugman, P. (1991). *Geography and trade*, MIT Press

Krugman, P. (1995). «Urban concentration: the role of increasing returns and transport costs,» Proceedings of the World Bank annual conference on development economics, The World Bank: 241-277

Krugman, P. and Venables, A. (1993). «Integration, specialization, and adjustment,» NBER Working Paper No. 4559

Murphy, K.M., Shleifer, A. and Vishny, R.W. (1989a). «Industrialization and the big push,» *Journal of Political Economy* 97 5: 1003-102

Murphy, K.M., Shleifer, A. and Vishny, R.W. (1989b). «Income distribution, market size, and industrialization,» *Quarterly Journal of Economics* 104: 537-564

Pedersen, K.R. (1995). «Rent-seeking, political influence and inequality - a simple analytical example,» *Public Choice* 82: 281-305

Platteau, J-P. (1991). «Traditional systems of social security and hunger insurance: Past achievements and modern challanges,» Chapter 4 in Ahmad, E. et al. *Social security in developing countries,* Clarendon Press, Oxford

Samuelson, P.A. (1983). «Thünen at two hundred,» *Journal of Economic Literature* 21: 1468-1488

Shleifer, A. (1986). «Implementation cycles,» *Journal of Political Economy* 94: 1163-1190

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations,* The modern library, New York, 1994.

von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie,* Hamburg, Perthes. English translation: The isolated state, Oxford, Pergamon Press, 1966

World Bank (1979). *World development report, 1979,* The World Bank,Washington D.C.

Chapter 2

Industrialization and regional inequality[†]

Abstract

This paper studies the effects of changes in transportation costs on industrialization, inter-regional trade and migration. Using a dual technology model which seems particularly well suited for developing countries, the main result is that lower transportation costs improve the prospects for a decentralized equilibrium. This result is contrary to a standard conclusion from the recent literature on economic geography where both trade and labor mobility are involved. A threshold level of transportation costs is defined above which centralization will arise. In order to create regional equality in terms of income per capita and industrial base, the quality of infrastructure must be such that costs of transportation fall below this critical level. Wage bargaining and monopolistic pricing may create market failure and regional inequality, justifying government intervention in the market.

[†] I would like to thank Karl Rolf Pedersen, Karl Ove Moene, Røgnvaldur Hannesson, Anthony Venables, Svein Aage Aanes, and Hans Jarle Kind for valuable comments. Remaining errors are mine. Financial support from the Research Council of Norway is gratefully acknowledged.

Introduction

In recent years we have witnessed a renewed interest in economic geography, the location of production in space. So far, however, little attention has been offered on the situation in developing countries when applying this theory. A recent exception is Livas and Krugman (1992) who study the effect of trade policy on city size, inspired by the case of Mexico. In addition, Krugman and Venables (1995) analyse economic integration and its effect on the location of firms in a North-South context, and Puga (1994) focuses on differences in the pattern of urban concentration between more and less developed countries. The relative neglect of development issues is surprising given the fact that the most dramatic expressions of regional imbalance are found in the Third World, an indication of this being the increasing degree of urban concentration in poor countries. According to United Nations (1991), in 1990 seventy percent of the worlds' 20 cities comprising a population of more than eight million people were located in the Third World. The increasing degree of urban concentration in the Third World should be contrasted with the situation in Europe, where the trend is towards a more balanced urban system.

The overwhelming majority of recent contributions to economic geography theory use monopolistic competition à la Dixit and Stiglitz (1977) as analytical tool. An important assumption in that framework is the existence of a large number of firms and free entry and exit which drive profits to zero. While this may be a reasonable approximation to reality in rich countries, in poor countries the number of firms and variety of manufactured goods are typically relatively small, and market-entry is often very difficult due to imperfections in capital markets and bureaucratic barriers.

This paper proposes an alternative approach to analysing issues of economic geography which may be particularly relevant for developing countries. Its central feature is dual technologies which can be seen as capturing the distinction between formal and informal sector. This model was introduced by Shleifer (1986) and has been used

in for instance Murphy, Shleifer and Vishny (1989a) and (1989b). To my knowledge, the present paper represents the first application of the model to economic geography. Like the monopolistic competition approach, the dual technology model features increasing returns to scale. The fundamental difference between the two is that while the price setting in Dixit and Stiglitz is determined by the threat of entry by firms using similar technology for the production of substitutes to existing goods, in Murphy, Shleifer and Vishny the price is determined by the threat of entry by producers using a different technology, the small scale producers, offering goods identical to those of the monopolists.

Extending Murphy, Shleifer and Vishny's model, I introduce many of the features of Krugman (1991a), namely two regions and two goods, mobile and immobile labor, and transportation costs between regions. Although of similar structure, transportation costs in the present model have an opposite effect on locational choice to that of Krugman (1991a). In the dual technology setup, an increase in transportation costs crowds out profits generated by trade while leaving the degree of competition unchanged. Higher transportation costs in this way reduce the potential for market expansion through trade, and make concentration of labor more likely. At a threshold value for transportation costs the model outcome changes from one characterized by agglomeration of mobile labor and factories in one region, the socalled centralization scenario, to one characterized by a more balanced regional distribution of labor and factories, the decentralization scenario. Note that the terms «centralization» and «decentralization» in this context refer not to production as such, which will always take place in both regions due to the presence of immobile labor, but rather to the location of factories and hence the location of mobile labor. The present paper also differs from most recent contributions to economic geography in that it draws welfare implications. Government intervention in order to stimulate decentralization and correct for monopolistic pricing may be called for

in order to secure efficiency. Income inequality provides further justification for intervention.

The paper is organized as follows. Section 1 presents the model. Migration is motivated by regional differeces in expected wage, and sections 1.1 and 1.2 define expected wage in the two location scenarios, centralization and decentralization. Section 2.1 defines the threshold value of transportation costs at which the outcome of the model changes from one location scenario to the other. The market's choice of location is not necessarily the right one from society's viewpoint; section 2.2 draws welfare implications. A brief discussion of the results in relation to existing theoretical and empirical literature is presented in 2.3. Section 3 concludes.

1. Model

Let there be two regions A and B and two goods a and b. Each good can be produced by means of two technologies: Constant returns to scale, which can be thought of as informal sector production, and increasing returns to scale, which may represent formal sector production. Labor (L) is the only factor of production, each worker being endowed with one unit of time. Informal sector technology is assumed to be the same for both goods and given by:

$$X_{ij}^{CRS} = L_{ij}^{CRS} \tag{1}$$

where $i \in \{a,b\}$, $j \in \{A,B\}$, and superscript CRS indicating constant returns to scale. The second production technology is increasing returns to scale, indicated by superscript IRS:

$$X_{ij}^{IRS} = \alpha\left(L_{ij}^{IRS} - F\right) \tag{2}$$

where $\alpha > 1$ is the marginal product of labor available for production, that is, after a fixed cost F has been subtracted. Both of these technology

parameters are assumed to be identical for the two goods. This technology is associated with constant marginal costs and asymptotically decreasing average costs. It is thus a natural monopoly. In the following it is assumed that due to for instance government intervention in order to avoid costly duplication of fixed costs or due to capital market imperfections, large scale producers will have a monopoly position in the market. In other words, oligopoly does not arise. Such market imperfections and bureaucratic barriers to entry seem realistic in a poor country context. As will become evident in the following, however, the market power of the monopolists in this dual technology framework is very limited. Let preferences of a representative individual in region j be described by the Cobb-Douglas utility function:

$$U_j = \left(C_{aj}\right)^{\beta}\left(C_{bj}\right)^{1-\beta} \tag{3}$$

The pricing strategy for a monopolist is easy to determine given the utility- and production functions chosen. Since the Cobb-Douglas utility function generates a unit-elastic demand curve with marginal revenue equal to zero, an unconstrained monopolist would raise the price without limit in order to save costs. The monopolist's pricing policy is however constrained by the availability of small scale technology. The idea is that if a large scale producer should set the price of its output above the supply price in the competitive informal sector, this will attract a sufficient number of cottage producers to the production of that good, making such a pricing policy unprofitable for the large scale producer.[1] In other words, the threat of market entry by

[1] The number of cottage producers is endogenously determined in the model. Interpreting the model in a literary way, the threat of entry facing the large scale producers should therefore also be endogenous. In fact, with symmetries in both taste and technology labor is either absorbed in the formal sector or stuck on the countryside, and hence strictly speaking does not pose the same threat of entry to the large scale producers as cottage producers do. Although the symmetry case due to its mathematical tractability will receive considerable attention, a realistic interpretation of the model would allow for asymmetries and a number of goods and hence the existence of a competitive fringe of cottage producers which makes the «threat of entry»-assumption plausible.

small scale producers defines a price ceiling facing the monopolists. Charging a price below this price ceiling would of course not be profitable, since increased production only increases costs.[2]

Let the informal sector supply price be numéraire, its value set at unity. The limit-pricing strategy of the monopolistic firm is therefore to charge a price marginally less than unity and thereby capture the whole market for the good. In this model, then, regional production of any one good never takes place by means of both technologies. Note that the existence of small scale technology severly limits the degree of monopoly power. The monopolists in effect act as price takers, their profits being determined more or less as a residual. Approximate the monopolistic price to unity. Irrespective of the location of large scale production and the level of transportation costs, the threat of entry by cottage producers will make sure that consumer prices are always unity. This in turn implies that goods demand is a function of income alone:

$$C_{aj} = \beta Y_j \tag{4a}$$

and

$$C_{bj} = (1-\beta) Y_j \tag{4b}$$

where Y_j is income in region j, consisting of wage income and profits. Assume that those who earn profits do not take part in production, which again implies that labor earns only wage income. Let us now turn to the determination of these two categories of income, starting with wage formation.

The formal sector is assumed to be unionized, with wages determined through a bargaining process between union and factory owners. Following Nash (1950), let this bilateral bargaining be a cooperative game. The disagreement pay-off for the firm is assumed to

[2] Of course, any utility function with elasticity of substitution less than unity, or slightly above unity, will generate the same limit pricing rule as the Cobb-Douglas case, the condition being that marginal costs exceed marginal revenue.

be zero, since in the absence of an agreement with the union no formal sector production takes place. For the union, the disagreement pay-off is assumed to equal the workers' alternative wage, which is informal sector income. This implies that factory workers during a strike may engage in cottage production but do not receive any strike support. These assumptions are certainly debatable, but are not critical for the model. Alternative formulations of the Nash bargaining game may alter the division of factory-income between owners and employees, but will not affect the important assumption, namely that there exists a formal sector wage-markup which is positively correlated with profits per employee. Using the information above, the Nash bargaining solution described in Flanagan, Moene and Wallerstein (1993: 96) for firm i in region j can be expressed as:

$$w_{ij}^{IRS} - w^{CRS} = \gamma \frac{\Pi_{ij}}{L_{ij}^{IRS}} \tag{5}$$

where Π denotes profits and $\gamma > 0$ captures the bargaining strength of the union relative to that of the firm. Since the supply price of small scale producers has been set at unity, $w^{CRS} = 1$. Define $L_j^{CRS} \equiv L_{aj}^{CRS} + L_{bj}^{CRS}$ and $L_j^{IRS} \equiv L_{aj}^{IRS} + L_{bj}^{IRS}$, all of which are greater than or equal to zero. Resource constraints in the labor market can then be expressed as $L_j = L_j^{CRS} + L_j^{IRS}$ for each region and as $L = L_A + L_B$ for the economy as a whole. Using (5), regional income can be expressed as:

$$Y_j = (1+\gamma)\Pi_j + L_j \tag{6}$$

Note that only an industrialized region generates profits. Let $X_j^{IRS} \equiv X_{aj}^{IRS} + X_{bj}^{IRS}$. Profits in region j can then be expressed as:

$$\Pi_j = X_j^{IRS} - w_{aj}^{IRS} L_{aj}^{IRS} - w_{bj}^{IRS} L_{bj}^{IRS} - T_j \tag{7}$$

where w is wages and T_j denotes transportation costs arising from region j's exports. These are assumed to be of the «iceberg»-kind, a share τ of the traded goods «melting» away during transportation to the neighboring region. Since consumer prices are unity, transportation costs have to be borne entirely by the producers. Naturally, in the absence of trade, transportation costs are zero.

From (5) we know that $w_{ij}^{IRS} = \gamma \Pi_{ij} / L_{ij}^{IRS} + 1$, and from (2) that $L_{ij}^{IRS} = X_{ij}^{IRS} / \alpha + F$. Total fixed costs in region j are $F_j \in (0, 2F)$, depending on the number of factories that choose to establish in the region. Equation (7) can then be written:

$$\Pi_j = \frac{1}{1+\gamma}\left[X_j^{IRS} - \frac{X_j^{IRS}}{\alpha} - F_j - T_j\right] \tag{8}$$

where $(1+\gamma)^{-1}$ measures the share of value added which accrues to the capital owners, and the terms in the parenthesis are income, variable cost, fixed cost and transportation costs, respectively.

Labor is assumed to be homogenous in all respects except one, namely the degree of mobility. Differences in labor mobility may be due to preferences for living in a specific region, lack of skills required to operate effectively in both regions such as language, lack of capital to finance costs of migration and starting up in a new place, etc. To capture this in the simplest possible way, assume that there are only two kinds of workers, the immobile (L^{IM}) and the mobile. A description of migration wraps up the model. Mobile labor is assumed to move to whatever region offers the higher expected wage. Following Harris and Todaro (1970), let expected wage equal average wage. Expected wages in region j, Ew_j, can be expressed as:

$$Ew_j = \frac{\left(w_{aj}^{IRS} - w^{CRS}\right)L_{aj}^{IRS} + \left(w_{bj}^{IRS} - w^{CRS}\right)L_{bj}^{IRS}}{L_j} + w^{CRS} \tag{9}$$

Combining equations (5) and (9) gives us:

$$Ew_j = \gamma\pi_j + w^{CRS} \tag{10}$$

where $\pi_j \equiv \Pi_j / L_j$.[3] Note that π_j is the only variable in (10) affected by the location of labor and firms. Hence, the expected wage in the different location scenarios is uniquely determined by profits per capita. In this model with two goods and two regions there are only two location scenarios with large scale production, one involving large scale production in both regions, call it decentralization, and one involving large scale production in a single region, call it centralization. Decentralization is characterized by an interior migration equilibrium, $Ew_A = Ew_B$, which implies equalization of regional profits per capita:

$$Ew_A = Ew_B \Leftrightarrow \pi_A^d = \pi_B^d \tag{11}$$

Superscript d indicates the decentralized equilibrium. Per capita profits in the decentralization scenario are calculated in section 1.2. Centralization of large scale production, on the other hand, is characterized by a corner solution in which all mobile labor gathers in the center region since this is the only region offering prospects for well paid jobs in the formal sector. The periphery, only populated by immobile labor, is assumed to be too small a market to justify large

[3] The above formulation captures the important assumption in the model that individual labor income in a region is positively related to profits in that region. Alternative formulations of region specific income redistribution could work equally well. For instance, assume there is a proportional taxation on profits, t, with the tax revenue distributed equally between the households in the region. Such lump sum redistribution may represent investments in local public goods, such as city infrastructure and health services. The equivalent of (10) would then be: $w_j = w^{CRS} + t\pi_j$, and the analysis would carry through exactly as in the wage bargaining and Harris-Todary migration version of the story.

scale production, see footnote 4. Since labor differs only with respect to mobility, the immobile labor living in the center region will have the same chances of formal sector employment and hence the same expected wage as the mobile labor.

The model does not determine which region will become the center and which will become the periphery in the center-periperhy scenario. If greater profits can be generated by locating in one particular region, due to for instance region specific technological advantages or a larger immobile labor force, then this region might attract factory production and labor, thus establishing it as the center. Psychology may also play a role. If people believe a certain region will industrialize, then this may become a self fulfilling prophecy, see Krugman (1991b). For notational convenience we shall in the remainder of the paper assume that in the center-periphery scenario, region A will be the center and region B the periphery. Expected wages in the two regions can be expressed as:

$$\begin{aligned} Ew_A &= \gamma\pi_A^c + w^{CRS} \\ Ew_B &= w^{CRS} \end{aligned} \qquad (12)$$

where $\pi_A^c = \Pi_A^c/\left(L - L_B^{IM}\right)$, superscript c indicating the centralization scenario. Expected wage in the center region differs from the formal sector wage defined in (5) only in that profits in (12) are on a per regional capita basis, whereas in (5) on a per formal sector employee basis. Clearly, assuming that factory production in the center is profitable, $Ew_A > Ew_B$ in this scenario and all mobile labor will gather in the larger region. Per capita profits in the centralized scenario are calculated in section 1.1.

When calculating profits per capita in the centralized and decentralized scenarios, equal expenditure shares will be assumed, i.e., $\beta = 1/2$. This simplifying symmetry assumption allows us to focus with greater clarity on the main issue, namely the relation between

transportation costs and locational choice. Together with the technological symmetry assumptions made initially, equal budget shares implies that the two large scale producers are equally profitable. The symmetry assumptions bias the model in favor of the decentralized outcome. To see this, note that the symmetry assumptions imply complete specialization in the decentralization scenario, given profitable trade. With asymmetries in taste and/or technology, this is no longer generally true. Some labor will then be involved in small scale production, which depresses overall profits in the decentralization scenario. Hence, the symmetry assumptions imply that profits and expected wages are at a maximum under decentralization. Contrast this with the centralization scenario, where all labor in the center region is involved in large scale production, given that the market is large enough to support factory production of both goods. See the appendix for a discussion of migration equilibrium in the case of asymmetries.

1.1 Centralization

This section determines profits per capita in an industrialized region when the other region does not industrialize. In this center-periphery scenario, all mobile labor gathers in the center since this is the only region offering any profits.[4] In the center both goods are produced by means of large scale technology and in the periphery by small scale technology. The symmetry assumption regarding the production technology for the two goods implies that the marginal rate of transformation in the two regions is unity. Hence, there are no gains from trade between the two regions. Of course, if production technology for the two goods were asymmetric, there could be gains

---

[4] The assumption that the larger region (region A), and only that region, will industrialize in the center-periphery scenario, i.e., $\Pi_A^c > 0$ and $\Pi_B^c < 0$, is justified for $L - L_B^{IM} > \frac{2\alpha F}{\alpha - 1} > L_B^{IM}$. This can easily be seen from (13), noting that the population in the center region is $L - L_B^{IM}$ and in the periphery L_B^{IM}.

from trade based on comparative advantage à la Ricardo, given that transportation costs between the regions were not too high.

In the absence of trade between center and periphery we have $X_{ij} = C_{ij}$. From (4a) with equal expenditure shares we know that $C_{ij} = Y_j/2$. Using the fact that centralization creates a corner solution in which all mobile labor gathers in the center, i.e., $L_A = L - L_B^{IM}$, (6) can be written as $Y_A = (1+\gamma)\Pi_A + L - L_B^{IM}$. Plugging all this information into (8), and dividing by L_A, profits per capita in region A can be written:

$$\pi_A^c = \frac{1}{1+\gamma}\left[\frac{\alpha\left(L - L_B^{IM} - 2F\right)}{L - L_B^{IM}} - 1\right] \tag{13}$$

From the expression above it is evident that the smaller is the immobile population left behind in the periphery, i.e., region B, the larger are profits per capita in the center region. Labor immobility therefore constitutes a centrifugal force in this economy. Intuitively, the larger the share of immobile labor, the larger is the share of goods demand which is supplied by cottage producers in the periphery, and hence the less profits for the large scale firms located in the center region. The firms would like to eliminate the competition from the small scale producers, but doing so by means of exports from the center is not profitable due to high wage costs. It may however be profitable for a firm to move to the periphery and conquer the entire market for its output by employing the relatively cheaper labor located there. Evidently, the size of the local market is a decisive factor in determining the profitability of such a move.

1.2 Decentralization

This section determines profits per capita with factory production taking place in both regions. While labor market equilibrium in the previous section implied maximum migration to the center, in the decentralized equilibrium labor will be allocated between the regions so

that $\pi_A^d = \pi_B^d$. Due to the symmetry assumptions made, both regions will be equally large in the decentralized equilibrium, i.e., $L_A = L_B = L/2$ and their incomes will therefore be the same, i.e., $Y_A = Y_B$. Given profitable trade between the regions, there will be full specialization in production, and so $X_j^{CRS} = 0$ and $Y_j = X_j^{IRS}$. Profitable trade requires a cif price below unity, i.e., $\tau \leq (\alpha - 1)/\alpha$ where $(\alpha - 1)/\alpha$ is the markup on marginal cost charged by the monopolistic firm. The symmetry assumptions imply that each region will export half of its output and therefore $T_j = \tau Y_j/2$. Using (6) and the above information, profits per capita under decentralization in region A can be expressed as:

$$\pi_A^d = \frac{1}{1+\gamma}\left[\frac{2\alpha L - 4\alpha F}{L(2+\alpha\tau)} - 1\right], \quad \tau < \frac{\alpha - 1}{\alpha} \tag{14a}$$

and

$$\pi_A^d = \frac{1}{1+\gamma}\left[\frac{(\alpha - 1)L - 4\alpha F}{(\alpha + 1)L}\right], \quad \tau \geq \frac{\alpha - 1}{\alpha} \tag{14b}$$

The situation in region B is of course identical. Naturally, higher costs of transportation lead to lower profits per capita in the decentralized equilibrium with trade, given by (14a). Transportation costs are therefore a centripetal force. Intuitively, by locating in separate regions, the firms may capture the entire demand for their output, but at the same time such decentralization implies that access to each other's markets becomes costly, due to transportation costs. Clearly, the higher are the transportation costs, the more there is to gain by staying together.

2. Centralization versus decentralization

2.1 Threshold value of τ

In the following a critical level of transportation costs is identified at which the locational outcome of the model changes. Due to

indivisibilities and the formation of expectations, models of this kind typically generate several stable equilibria. However, the conclusion «anything can happen» is not particularly appealing. In order to analyse the effects of transportation costs on industry location and migration, we have to make some specific, and hopefully plausible, assumptions concerning the determination of equilibrium.

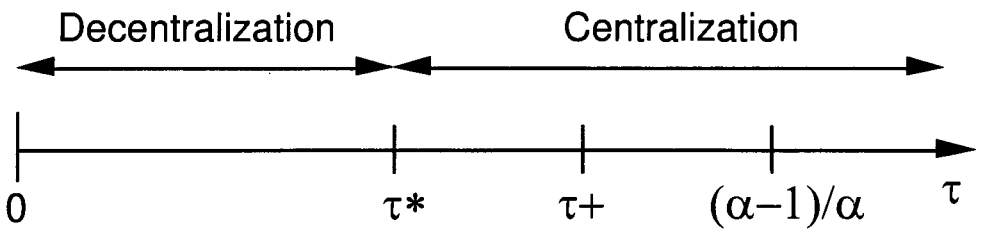
Assume first that the firms take regional wage levels as given, which seems plausible with the realistic assumption that there are a number of firms operating in the economy. Second, a firm will be able to move from one region to another, say from A to B, only if it can offer a competitive wage, i.e., a wage which makes expected wages in B at least as high as those of A. Third, assume that the location of labor is determined by the global migration optimum, i.e., the equilibrium associated with the highest expected wages.

Define τ^* as the threshold level of transportation costs at which profits per capita, and hence exptected wages, under centralization and decentralization are equal. Transportation costs above τ^* create centralization and below it decentralization. The existence of such a threshold can easily be verified by comparing (13) and (14) and noting that $\tau > (\alpha - 1)/\alpha \Rightarrow \pi_A^c > \pi_A^d$ and $\tau = 0 \Rightarrow \pi_A^c < \pi_A^d$. Hence, $\tau^* \in [0,(\alpha - 1)/\alpha]$, which rules out the case of decentralization without trade, i.e., (14b), as possible outcome. Equating (13) and (14a), the threshold level of transportation costs can be expressed as:

$$\tau^* = \frac{4FL_B^{IM}}{\alpha L\left(L - L_B^{IM} - 2F\right)} \tag{15}$$

For $\tau < \tau^*$, $\pi_A^d > \pi_A^c$ and decentralization will be the outcome, while for $\tau > \tau^*$, $\pi_A^d < \pi_A^c$ and only one region will industrialize. Note that if all workers were perfectly mobile, any level of transportation costs would lead to centralization, i.e., $L_B^{IM} = 0 \Rightarrow \tau^* = 0$.

Figure 1. Centralization versus decentralization



The figure above can be interpreted as dynamic story of development. Assume that the economy starts out in a situation with high transportation costs and centralization of large scale production. This may reflect the current state of affairs in many low income countries, particularly in Africa where urban concentration is high. With economic development and consequently reduced transportation costs, the agglomoration forces are weakened. Below a certain level of interregional trade costs, both factory owners and employees gain from a move to decentralized production. At this point, the increase in profits generated by such a move is sufficiently great to attract labor to the (former) periphery.

We have seen that immobile labor reduces the gains from centralization, and therefore constitutes a centrifugal force.[5] This is reflected by the fact that $(\partial\tau^*/\partial L_B^{IM}) > 0$. In other words, the larger is the number of immobile labor in the economy, the wider is the range of transportation costs for which decentralization will come about. The centripetal forces include transportation costs, technological progress and population growth. High transportation costs encourage the colocation of labor in order to reduce these costs. Technological progress in the form of an increase in α and/or a reduction in F reduces the economies of scale in large scale production and makes the centralization outcome more likely, i.e., $(\partial\tau^*/\partial\alpha) < 0$ and $(\partial\tau^*/\partial F) > 0$.

[5] An earlier version of the model included an additional centrifugal force, namely congestion costs. Such costs affected the threshold value of transportation costs but did not alter the basic insights from the model and were therefore left out.

Intuitively, a reduction in economies of scale makes market size less decisive, which implies that the benefits to the factories from establishing in separate regions and thus conquering the entire market for their goods, are reduced. This means that for a larger interval of transportation costs, centralization will be the outcome. An increase in total labor supply, for a given number of immobile workers, has a similar effect on the economy, i.e., $(\partial\tau^*/\partial L) < 0$. Reducing the relative importance of immobile labor in the economy also reduces the relative strength of the centrifugal force, and hence promotes centralization. Finally, as argued earlier, any asymmetries in taste or technology would also favor the centralized outcome.

The model also suggests that rather small differences in interregional transportation costs between two otherwise identical countries may create large differences in their economic geography. Accordingly, the country with transportation costs slightly below a critical value will be relatively decentralized and the country with transportation costs slightly above this value will be relatively centralized.

2.2 Welfare

The locational outcome of the market is not necessarily the optimal one from society's viewpoint. When ranking the two locational scenarios in Pareto efficiency terms, note first that due to our symmetry assumptions there are no distortions caused by monopolistic pricing. In the decentralized scenario there is complete specialization in production, which means that the marginal rate of transformation is unity and therefore equal to the marginal rate of substitution at market prices. In the centralization scenario, the marginal rate of transformation within each region is also unity, since in the center region only large scale production takes place and in the periphery only small scale production takes place. In this scenario, too, monopolistic pricing creates no inefficiencies. This implies that for a Pareto ranking of the two locational scenarios we can compare overall income in the

two scenarios at market prices. And since the formal sector wage bonus is determined uniquely by profits, profits is the only category of income varying according to locational choice.

Profits in the decentralized equilibrium can easily be found as $\Pi_A^d + \Pi_B^d = \pi_A^d L$ and in the centralized equilibrium $\Pi_A^c = \pi_A^c\left(L - L_B^{IM}\right)$. The critical level of transportation costs from an efficiency viewpoint, call it τ^+, can be found by plugging this information into (13) and (14a) and equating the two. The result is:

$$\tau^+ = \frac{2L_B^{IM}(\alpha-1)/\alpha}{\alpha\left[L - 2F - L_B^{IM}(\alpha-1)/\alpha\right]} \tag{16}$$

Overall profits are higher in the decentralized equilibrium for $\tau < \tau^+$ and higher in the centralized equilibrium for $\tau > \tau^+$. Thus, for transportation costs smaller than τ^+, economic efficiency calls for decentralization, and vice versa. It is straightforward to demonstrate that given $\pi_A^c > 0$, which we assume is true, then $\tau^+ > \tau^*$. Note that if all labor were mobile, $\tau^* = \tau^+$ and there would be no distortion in locational choice. Since the market outcome for $\tau > \tau^*$ is centralization, while $\tau < \tau^+$ on efficiency grounds calls for decentralization, the interval $\tau^* < \tau < \tau^+$ is characterized by market failure.

This distortion is due to formal sector wage bargaining. Consider a move from centralization to decentralization. This would increase the number of formal sector employees, since the immobile labor in the (former) periphery are now hired as factory workers. From (5) it is evident that, *ceteris paribus*, an increase in the number of formal sector employees reduces the formal sector wage. If profits increase only marginally from such a change in factory location, formal sector wages probably go down. Given our assumption that firms take the wage level as given and move from A to B only if they can offer a wage matching the one in A, the presence of wage bargaining may preclude

profit-maximizing locational decisions and therefore create socially inefficient outcomes. There may thus be a case for government intervention, encouraging the establishment of factory production in the periphery and supporting migration of labor to the smaller region. As we have seen, one policy which may achieve regional balance is investments in infrastructure. Below a certain critical level of transportation costs, τ^*, decentralization is the market outcome. In case of coordination failure, however, it may be important not only to create the necessary infrastructure but also to coordinate the actual transfer of resources to the periphery. On self-fulfilling prophecies, government coordination and industrialization, see Murphy, Shleifer and Vishny (1989a).

Transportation costs higher than τ^+ make centralization the efficient solution since it saves on transportation costs. But centralization involves an income gap between the regions, the immobile labor in the out-migrated region of course being poorer than the average worker in the centralized region. If lump sum taxation is not possible, this inequality is a cost which must be weighed against any efficiency gains from centralization. Income inequality, then, provides an additional argument for government involvement in the industrialization process in favor of regional balance.

The absence of distortions from monopolistic pricing is of course a special case. In the case of asymmetries in taste or technology, the decentralized equilibrium will not necessarily be characterized by complete specialization. Then a larger region will emerge, in which small scale producers supply the residual demand for the good imported from the smaller region, see the discussion in the appendix. Since in this case production by means of both technologies coexist within one region, the marginal rate of transformation (which equals the marginal cost in large scale production, i.e., $1/\alpha$) is strictly less than unity in that region. Not surprisingly, monopolistic pricing will in this case lead to the level of consumption of the monopolist's good being too low.

Recall that the strategy of the large scale producer is to supply the whole market at a price defined by the supply price of small scale producers. The way to implement a Pareto efficient allocation of resources is therefore to either increase demand through a consumption subsidy or to lower the price ceiling facing the monopolist by offering a subsidy to small scale producers. Note that a production subsidy to the large scale producer reduces its marginal costs but does not affect the quantity supplied. The consumption or small scale production subsidy that would realize a Pareto efficient allocation of resources is equal to $(\alpha-1)/\alpha$, i.e., the difference between the price of unity and marginal cost. With such a subsidy, a share $\alpha/(\alpha+1)$ of income will be spent on good a and a share $1/(\alpha+1)$ on good b, which should be contrasted with equal expenditure shares in the absence of subsidies.

2.3 Discussion

An important conclusion from recent models in economic geography is that centralization is negatively correlated with costs of transportation, see for instance Krugman (1991a) and Krugman and Venables (1993). In these models high transportation costs reduce the degree of competition between regions, thereby making it more profitable for some firms to leave the center and serve local markets. When transportation costs are low, interregional competition is more fierce and there is less to gain by serving local markets. Instead, firms will have an incentive to reap the benefits of scale economies by locating in the center region where the market is relatively large. One policy implication from these models, then, is that an investment in infrastructure is not such a good idea if one is concerned with equality in income and industrial base between regions.

Although the logic of the argument seems convincing, casual observation suggests that the above correlation may not necessarily be true. The degree of urban concentration in African countries is quite large and increasingly so although the quality of infrastructure is poor

and in some cases appears to have been deteriorating during the last decades. Furthermore, measuring the quality of infrastructure by density of roads and government expenditure on transportation and communications, the empirical study by Ades and Glaeser (1995: 213) confirms that «well developed transportation facilities lower the size of central cities.» Interpreting the regions in the present model as cities, this observation harmonizes well with the result in this dual technology model and questions the validity of the above mentioned monopolistic competition models. The observation also suggests that an investment in a country's infrastructure is likely to promote regional equality.

A small digression is perhaps in place here. According to Ades and Glaeser (1995: 197): «Urban theorists from von Thünen (1826) to Krugman (1991) have argued that when transportation is expensive activities will group together to save on travel costs. This theory predicts that urban concentration will be higher when transportation is more costly.» With respect to Krugman's article, which in the present paper is referred to as Krugman (1991a), this is clearly not true. For instance, Krugman (1991a: 496) states that: «high transportation costs militate against divergence», where by «divergence» is meant concentration of economic activity in a single region. As illustrated in figure 2, page 496, in Krugman's article, the relation between transportation costs and profitability of a firm establishing in the periphery relative to staying in the center forms a U-curve. This curve is everywhere below unity at transportation costs below a threshold level, which implies that concentration of economic activity will take place when transportation costs are small. When transportation costs are above the critical level, a regional balance in economic activity will emerge. Note however the border case with zero inter-regional transportation costs which makes location irrelevant.

There are exceptions to the standard result of transportation costs as a centrifugal force, and which therefore provide a theoretical underpinning for the empirical results of Ades and Glaeser. For

instance, in Krugman and Livas (1992) congestion costs rather than the presence of immobile farmers constitute the model's centrifugal force. That model produces a positive relation between transportation costs and centralization, exactly as in the dual technology approach. At low transportation costs, trade costs are lower than congestion costs, and hence it is profitable for firms to avoid large clusters of production. When transportation costs are high, trade costs exceed congestion costs and firms find it profitable to colocate, thus forming a core-periphery pattern.

Yet another exception is Krugman and Venables (1995). In this model, too, the relation between transportation costs and factory location forms a U-shaped curve. This curve would however intersect *twice* the «equal profitability line» in Krugman (1991a). In other words, low levels of transportation costs may produce decentralization rather than centralization. At intermediate levels of transportation costs, centralization in the North is likely, while high levels of transportation costs increase the prospects for industrial production in both North and South. The reason for this result is the absence of labor mobility in the model, creating a wage gap between center and periphery. With sufficiently low transportation costs, this wage gap becomes large enough to warrant industrialization also of the periphery, i.e., the South.

The dual technology model presented here includes neither congestion costs nor an entirely locked in labor force and is therefore similar in structure to standard models of new economic geography such as Krugman (1991a). Hence, even though we have seen that monopolistic competition models can be designed so as to generate all kinds of effects of transportation costs on locational choice, the most reasonable comparison of the present model is to models of similar structure. In that light, the correlation between transportation costs and locational choice reported here stands out as a contrast to the contributions mentioned above.

3. Conclusion

The present model predicts a positive relation between infrastructure quality and regional equality. More specifically, reductions in transportation costs between regions are likely to promote a regional balance both in terms of location of labor and industrial production and in terms of standard of living. If the government, or a donor, seeks a regional balance in its development efforts, and there may be reasons of both efficiency and equality for wishing so, then the model suggests that investing in infrastructure is a good idea. This policy implication is in accordance with the empirical results reported in Ades and Glaeser (1995), but is contrary to the predictions from recent economic geography models with a structure similar to the one in the present paper.

The model presented here also suggests that there is a discontinuity in the effect infrastructure quality has on locational choice. If we take centralization as point of departure, the model suggests that a «big push» in infrastructure investment is required in order to achieve decentralization and equality. Once transportation costs are reduced below a threshold value, decentralization will come about. Indeed, if we interpret the model literary, investments in infrastructure which fail to lower the costs of transportation below this point is of no value, since no trade takes place between core and periphery.

The market may fail to achieve an optimal allocation of resources both between regions and within regions. In particular, for some values of the model-parameters, the market outcome is centralization whereas efficiency calls for industrialization in both regions of the economy. Moreover, centralization involves income inequality, the immobile labor in the periphery being the losers. Thus, when a transfer of income to the periphery is not possible, or very costly, income equality provides an additional argument for policies seeking to stimulate a regional balance in industrial base.

Let me conclude with some suggestions for further research. First, the limit pricing strategy of the monopolist excludes the possibility of a given good being produced by means of both formal and informal sector within a region. Allowing for coexistence of formal and informal sector production within industries would probably strengthen the realism of the model. Second, the model could certainly be extended to allow for additional costs and benefits of agglomeration. One attempt in that direction has been made in Bjorvatn (1995), which introduces congestion costs within the present dual technology framework, but alternative extensions would certainly be of interest.

Appendix

In this appendix I first demonstrate that full specialization in production with asymmetric budget shares is not compatible with equality in expected wages and therefore is not a migration equilibrium. I then ask the question: Does an interior solution with asymmetric budget shares and therefore asymmetric distribution of labor across regions exist?

Full specialization in production under decentralization is conditioned on trade between the regions being profitable, i.e., $\tau \leq (\alpha - 1)/\alpha$. For concreteness, consider region B specializing in the production of good b. The case of region A specializing in a would be entirely similar. The more general formulation of (14a), allowing for asymmetries in budget shares, is:

$$\pi_B^d = \frac{1}{1+\gamma}\left[\frac{\alpha L_B - \alpha F}{L_B(1+\alpha\tau\beta)} - 1\right], \quad \tau < \frac{\alpha - 1}{\alpha} \tag{A1}$$

which by setting $\beta = 0.5$ and $L_B = 0.5L$ equals (14a). A higher budget share of good b, i.e., a lower β, increases profits per capita in the region specializing in the production of this good, i.e., region B, for two reasons: First, by reducing the share of income spent on imports the costs of transportation are also reduced. This effect is captured by the $\alpha\tau\beta$-term in the denominator of (A1). Second, with full specialization in production, population is distributed across regions according to the budget shares, or more precicely according to:

$$\frac{X_A^{IRS}}{X_B^{IRS}} = \frac{L_A - F}{L_B - F} = \frac{\beta}{1-\beta}$$

A lower β is therefore associated with a larger number of workers in region B, which reduces the average cost in production, captured by a

reduction in F/L_B in (A1). Together, these effects imply that profits per capita, and hence expected wages, in the larger region are greater than those of the smaller region, given full specialization in production. From this we can conclude that full specialization is not compatible with migration equilibrium in the case of asymmetric budget shares. This leads naturally to the second question, concerning the existence of equilibrium in the decentralized scenario.

Migration from the smaller region leads to a situation in which the larger region supplies both goods. As long as production in the smaller region is conducted by means of large scale technology, the residual demand of this good will be supplied in the larger region by small scale producers. Now let A be the larger region, due to $\beta > 0.5$. In the absence of full specialization, profits per capita in this region are then given by:

$$\pi_A^d = \frac{1}{1+\gamma}\left[\frac{(\alpha - 1 - \alpha\tau)\beta Y - \alpha F + \alpha\tau\beta L_A}{L_A\left(1+\alpha\tau(1-\beta)\right)}\right], \quad \tau < \frac{\alpha - 1}{\alpha} \qquad \text{(A2)}$$

where I have used the fact that $X_{aA}^{IRS} = \beta Y$ with Y denoting national income. Profits per capita in the smaller region, region B, are given by (A1). Does this migration reduce or increase the gap in expected wages between the two regions? This question is difficult to answer analytically. By using some simplifying assumptions, however, we can identify the forces determining whether or not a decentralized equilibrium exists.[6] First, consider the case in which the asymmetry in budget shares is not very large, which again implies that with full specialization the distribution of labor across regions is fairly even and hence profits per capita do not differ a great deal between regions. Assume also that fixed costs in production are small relative to total

[6] Krugman (1991a: 492-493) encounters the same problem of analytical solution to the question of stability, and chooses to illustrate the properties of the model numerically.

population, which means that the F/L_j-terms in (A1) and (A2) carry little weight. For simplicity, assume these terms are negligible. This means that emigration from the smaller region has only a negligible impact on profits per capita in that region, as evident from (A1). The effect of such migration on profits per capita in the larger region is however clearly negative. First, by moving from large scale production in the smaller region to small scale production in the larger region, national income is reduced, thereby reducing profits in the larger region. Second, migration to the larger region of course increases its population. All in all, profits per capita and hence expected wage must decline in the larger region. Migration thus creates a convergence of expected wages between the two regions when asymmetries are modest, and a stable interior solution is then likely to exist.

Consider now the case of a larger asymmetry in budget shares. Full specialization is now associated with a highly unequal distribution of labor and profits per capita across regions. The negative impact on the expected wage from the smaller out-migrating region now cannot be ignored. In addition, the «pull-factor» from the larger region is rather strong, due to the initially higher expected wage in this region. In this case, migration may not lead to a convergence of expected wages across regions, and a stable decentralized equilibrium may not exist. High transportation costs fortify the centripetal effect of asymmetry in budget shares, since the savings from colocation of labor in this case may be substantial. In other words, when transportation costs are high, the negative impact on national income of migration to the larger region is small, and hence profits per capita in this region are not greatly affected by an increase in its size. With asymmetry in budget shares, the higher are transportation costs, the less likely is the existence of an interior equilibrium.

References

Ades, A. F. and Gleaser, E. L. (1995). «Trade and circuses: Explaining urban giants,» *Quarterly Journal of Economics* 110: 195-227

Bjorvatn, K. (1995). «City size and development,» Discussion paper 4/96, The Norwegian School of Economics and Business Administration.

Dixit, A.K. and Stiglitz, J.E. (1977). «Monopolistic competition and optimum product diversity,» *American Economic Review* 67: 297-308

Flanagan, R. J., Moene, K. O., and Wallerstein, M., ed. (1993). *Trade union behaviour, pay bargaining, and economic performance.* Claredon Press, Oxford

Livas Elizondo, R. and Krugman, P. (1992). «Trade policy and the third world metropolis,» NBER-Working paper No. 4238

Harris, J. and Todaro, M. (1970). «Migration, unemployment and development: a two-sector analysis,» *American Economic Review* 60 (March): 165-187.

Krugman, P. (1980). «Scale economies, product differentiation, and international trade,» *American Economic Review* 70: 950-959

Krugman, P. (1991a). «Increasing returns and economic geography,» *Journal of Political Economy* 99 3: 483-499

Krugman, P. (1991b). «History versus expectations,» *Quarterly Journal of Economics* 106: 651-667

Krugman, P. and Livas, R. (1992). «Trade policy and the Third World metropolis,» NBER Working Paper No. 4238

Krugman, P. and Venables, A. (1993). «Integration, specialization, and adjustment,» NBER Working Paper No. 4559

Krugman, P. and Venables, A. (1995). «Globalization and the inequality of nations,» *Quarterly Journal of Economics* 110 4: 857-880

Murphy, K.M., Shleifer, A. and Vishny, R.W. (1989a). «Industrialization and the big push,» *Journal of Political Economy* 97 5: 1003-102

Murphy, K.M., Shleifer, A. and Vishny, R.W. (1989b). «Income distribution, market size, and industrialization,» *Quarterly Journal of Economics* 104: 537-564

Nash, J. F. Jr. (1950). «The bargaining problem,» *Econometrica* 18: 155-162

Puga. D. (1994). «The dynamics of urbanisation,» Center for Economic Performance Working Paper no. 652, London

Shleifer, A. (1986). «Implementation cycles,» *Journal of Political Economy* 94: 1163-1190

United Nations (1991). *World Urbanization Prospects 1990*, United Nations, New York

Chapter 3

City size and economic development†

Abstract

This paper studies city size in a developing country. The country consists of two regions, a city producing a manufactured good and a rural area producing an agricultural good. Manufacturing can take place in either formal or informal sector, the formal sector being characterized by increasing returns to scale production located in a central business district, while the informal sector uses constant returns to scale technology, production being decentralized. Agricultural production is constant returns to scale. Equilibrium city size is calculated and comparative static results discussed. Market failure may arise due to coordination failure, increasing returns to scale in production or congestion externalities. Government intervention may therefore be called upon in order to secure an optimal city size.

† I would like to thank Karl Rolf Pedersen, Karl Ove Moene, and Anthony Venables for valuable comments. Remaining errors are mine. Financial support from the Research Council of Norway is gratefully acknowledged.

Introduction

Third world cities tend to be relatively large. For instance, more than 35 percent of Argentina's population lives in the capital, Buenos Aires, a city of 12 million inhabitants. By the end of this decade, 17 of the world's urban giants with populations exceeding 10 million people will be located in the Third World, see United Nations (1991), and for an informal discussion, The Economist (1995). In an empirical paper, Ades and Glaeser (1995) find that the most important factors driving urban centralization are political ones. In particular, dictatorships tend to cause concentration in a single metropolis. The focus of this paper is not on political forces as determinant of city size, but rather on such issues as urban infrastructure quality, technology, formal versus informal sector, migration and population size. See however Bjorvatn (1995b) for a model dealing with dictatorship and urbanization.

Many observers feel that the large cities in developing countries are in some sense too large. In a recent paper, Livas and Krugman (1992) discuss the effect of trade policy on Third World metropolises, concluding that a move to free trade leads to a reduction in city size. With Mexico as a case in point, the authors evidently feel that such decentralization is a good thing. In their informal discussion, they describe the problems of Mexico city as «apparant at first sight and first breath» (p. 4). However, while providing the analytical tools for discussing optimal city size, the authors stop short of making any welfare analysis, probably due to the complexity of the model.

Optimal city size is a central issue in the traditional literature on economic geography, the standard reference being Henderson (1974). The main questions addressed in that paper are: When are new cities formed and when should they be formed? The main answer is that markets do not form new cities frequently enough. In other words, cities tend to be too large. Indeed, with continuous population growth, cities will be of the worst kind, producing no surplus at all. The problem is due to coordination failure. Firms have no incentive to leave the established city on their own, but everybody would benefit if

a group of firms relocated and together formed a new city. This inefficiency calls for government intervention, either directly by coordinating the relocation of firms, or indirectly by making relocation more attractive through taxing firms in existing cities and subsidizing those who choose to leave.

The most important explanation for geographical concentration is increasing returns to scale. In Henderson (1974) the economies of scale are captured by knowldege spillovers between firms. The agglomeration force in this model hinges on these externalities being city specific. If the spillover effects reached the entire economy, cities would never form since only costs would be incurred by colocation. Recent contributors to economic geography tend to favor the monopolistic competition approach à la Dixit and Stiglitz (1977) which involves pecuniary externalities, see for instance Abdel-Rahman (1988). This model has the obvious advantage of making the agglomeration effect more tangible, since increasing returns to scale appear on firm level rather than by reference to abstract knowledge spillovers between firms. The main results in the two approaches are however similar, as argued by Fujita (1989, chapter 8). In both cases there are increasing returns to scale on an aggregate level which must be weighed against diseconomies of city size. At a certain city size, the marginal costs and marginal benefits of further immigration to the city are equal. Moving beyond this point reduces city output and crowds out profits.

The agglomeration forces in the present paper are internal to the firm, and the paper therefore follows the more recent contributions to economic geography. However, rather than using the standard monopolistic competition model, a dual technology model known from Murphy, Shleifer and Vishny (1989) is applied.[1] Goods can be

[1]In Bjorvatn (1995a) the dual technology framework is used to study how changes in transportation costs between regions affect migration, trade and industrialization. That paper abstracts from commuting costs within regions and therefore does not address the issue of optimal city size.

produced either by increasing returns to scale or constant returns to scale technology. Dual technologies may be seen as capturing the dichotomy between formal and informal sector. This dichotomy is particularly significant in poor countries due to widespread market imperfections. The threat of entry by small scale producers defines a price ceiling for the large scale monopolist. Contrast this approach with monopolistic competition, where the monopolistic pricing rule is determined by reference to threat of entry by firms producing substitutes to the existing range of goods, a threat which drives profits to zero.

Diseconomies of city size are due to commuting costs, entering the analysis through two assumptions. First, formal sector production takes place at a certain point in the city, call it the central business district. Second, the distance to the central business district for the average worker, and hence the formal sector commuting cost, increases with city size. In the present model, these diseconomies of city size are entirely external to the workers. Wages are tied to productivity in small scale technology, which due to its decentralized nature is independent of city size. Hence, potential migrants to the city do not consider their impact on congestion costs when choosing location, and these costs are fully carried by the large scale producer.

The basic structure of the model does not differ from traditional or modern urban economic theory. There are increasing returns to scale in production and there are certain diseconomies of city size, and hence the main results from the literature carry through: First, optimal city size is determined as a tradeoff between the increasing returns to scale in production and the decreasing returns to scale due to commuting costs. Second, government intervention may be called upon in order to correct market failure. Third, continuous population growth crowds out any surplus created by the city. Although the qualitative results are familiar, the dual technology framework has to my knowledge not been applied in analysing questions of city size and is therefore of interest in its own right. Moreover, as argued above, this

approach may be particularly well suited for analysing urban economics in a development context. The paper is organized as follows. Section 1 presents the model. Section 2 considers issues of welfare and policy implications. In section 3 extensions to the model are briefly discussed. Section 4 concludes.

1. Model

The model has two regions, denoted by subscript $i = A,B$, and two goods, a manufactured good (X_A) and food (X_B). Labor (L) is the only factor of production, each laborer being endowed with one unit of time. Manufacturing takes place in region A, the city, and agriculture in region B, the rural area. Region specific production may be explained by reference to availability of land in the case of agriculture and a preference of workers to be close to their colleagues, due to for instance positive learning effects, in the case of manufacturing. Such effects are however not modelled explicitely here.

By considering only one possible city location, issues of urban concentration and city formation are not addressed. City formation is the key issue in Henderson (1974) and urban concentration is studied in Puga (1994) and, interpreting the model's regions as cities, in Bjorvatn (1995a). In the present paper we take urban concentration as a fait accompli, and study the interaction between a country's urban center and the rural area.

Agriculture

Since we wish to focus on the city, rural production technology is kept as simple as possible and assumed to be constant returns to scale. In reality one would perhaps expect decreasing returns to scale in agriculture due to scarce land resources. On the other hand, as suggested by Boserup (1965), population pressure may stimulate technological progress in agriculture which at least to some extent outweighs the decreasing marginal productivity of labor caused by scarcity of land. Although not modelled explicitely, such technological

improvements could explain constant returns to scale in agriculture. The primary motivation for this technology assumption is however analytical simplicity. Decreasing returns to scale would certainly affect relative prices and thereby the regional allocation of labor, but would not affect the qualitative results of the model, which are due to forces specific to the urban sector. Let therefore each farmer produce one unit of food:

$$X_B = L_B \tag{1}$$

Manufacturing technology and commuting time

The manufactured good can be produced by means of two technologies, constant returns to scale (superscript *CRS*) and increasing returns to scale (superscript *IRS*). Due to for instance government intervention in order to avoid costly duplication of fixed costs or capital market imperfections, large scale production takes place by a single firm, which thereby has a monpoly position in the market. In other words, oligopoly does not arise. As will become clear below, the market power of the monopolist is very limited due to the presence of the alternative technology.

Small scale manufacturing, or cottage production, can be associated with informal sector. Its technology is identical to that of farming, allowing each informal sector worker to produce one unit of the manufactured good:

$$X_A^{CRS} = L_A^{CRS} \tag{2}$$

Small scale manufacturing is a decentralized activity and commuting costs therefore do not apply to these producers. The availability of informal sector technology defines a wage floor and a price ceiling for the large scale manufacturer. Large scale factory production can be associated with the formal sector. Following a standard assumption in urban economic theory, factories are located at a particular point in the

city; the central business district. Commuting costs arise since factory workers are assumed to live at some distance from their place of work. Due to congestion effects, the commuting time is a positive function of the number of people in the city. In section 3 an alternative formulation is suggested, where commuting time is a function of factory workers alone. As in Livas and Krugman (1992), let the average time at the disposal of factory workers net of commuting be a negative linear function of the population in the city:

$$z_A = 1 - 0.5\gamma L_A \tag{3}$$

where $L_A = L_A^{IRS} + L_A^{CRS}$ is city size, $0.5\gamma L_A$ is the commuting time and γ can be interpreted as an inverse measure of region A infrastructure quality. The 0.5-term in the expression above is included for notational convenience, simplifying equation 6. A reduction in γ represents an improvement in city infrastructure since it reduces the time spent commuting for a given city size, and vice versa. In the absence of commuting costs, i.e., $\gamma = 0$, the time available to each worker for factory production is simply the time endowment of labor, which is equal to unity. The aggregate labor input in formal sector, net of commuting time, is

$$Z_A = z_A L_A^{IRS} = L_A^{IRS}(1 - 0.5\gamma L_A) \tag{4}$$

Commuting costs as described in (3) and (4) represent the disadvantages of city size in this model. This is not to say that time spent commuting is the only kind of disadvantage associated with a large city in real life. Pollution and increased marginal costs of basic services such as water supply and electricity are other examples which are perhaps even more important in this respect. However, the formulation above captures the essential part of the story, namely that an increase in city size crowds out the surplus created by the city. Furthermore, since wages are tied to income opportunities in the informal sector and therefore

insensitive to congestion effects in this model, congestion costs enter in the form of a negative externality, which seems to be plausible also for the other kinds of disadvantages of city size mentioned above, i.e., pollution and increased marginal costs of services (the latter due to for instance average rather than marginal cost pricing of these services).

As is evident from (4), there are constant marginal commuting costs in the number of formal sector employees. An increase in city size and/or a deterioration of its infrastructure would shift the commuting cost schedule upwards, thus crowding out profits, and vice versa. What implications do congestion costs of this kind have for formal sector employment? First of all, note that the formal sector firm cannot control city size by lowering its wage and is assumed unable to control migration directly. Second, there are increasing returns to scale in formal sector production, given by:

$$X_A^{IRS} = \alpha(Z_A - F_A), \quad \alpha > 1 \tag{5}$$

where α is the marginal product of labor available for production, that is, after both commuting costs and the fixed cost F_A have been subtracted. Constant marginal commuting costs in the number of formal sector employees, the inability of the firm to affect city size, and increasing returns to scale in production together imply that, as long as profits are positive, all migrants are absorbed in formal sector employment, i.e., $\Pi_A \geq 0 \Rightarrow L_A^{IRS} = L_A$, see A1 in the appendix for proof.[2] This also implies that the two technologies in manufacturing will

---

[2]The fact that congestion costs in the present model do not affect migration should be contrasted with that in Livas and Krugman (1992) where wages are a decreasing function of city size and the marginal worker's impact on average labor income is fully captured by increasing land rents. Since land rents only determine income distribution, no negative externalities arise from migration. Interpreting their model literary, then, cities are never too large in efficiency terms. This is quite surprising given the fact that the authors motivate their model by referring to the allegedly excessive size of Third World metropolises.

never be in use simultaneously. Inserting (4) into (5) and taking the first derivative, the contribution to manufacturing of an additional migrant to the city can be found as:

$$MPL_A^{IRS} = \alpha(1 - \gamma L_A) \tag{6}$$

Factory production thus forms a «Laffer-curve» which peaks at $L_A = 1/\gamma$ where $MPL_A^{IRS} = 0$. At this point only half of the time at the disposal of formal sector workers is allocated to work, the rest being spent commuting, i.e., $L_A = 1/\gamma \Rightarrow Z_A = 1/2\gamma$. An increase in region A labor beyond $1/\gamma$ will reduce labor input net of commuting costs and therefore reduce formal sector output in the city, i.e., $L_A > 1/\gamma \Leftrightarrow MPL_A^{IRS} < 0$.

Limit pricing

Let preferences of a representative individual be described by the Cobb-Douglas utility function

$$U = (C_A)^{\beta}(C_B)^{1-\beta} \tag{7}$$

The pricing strategy of the monopolist is easy to determine given the utility- and production functions chosen. Since the Cobb-Douglas utility function generates a unit-elastic demand curve with marginal revenue equal to zero, an unconstrained monopolist would raise the price without limit in order to save costs. The monopolist in this model is however constrained by the threat of entry by small scale producers, be they cottage producers in the city or peasants in the rural area. The informal sector supply price in this way defines a price ceiling for the monopolist.[3] Charging a price below this price ceiling would of

[3] Of course, any utility function with elasticity of substitution less than unity, or slightly above unity, will generate the same limit pricing rule as the Cobb-Douglas case, the condition being that marginal cost exceed marginal revenue.

course not be profitable, since increased production only raises the costs. Hence the limit-pricing strategy of the monopolistic firm is to exactly match the supply price of the cottage producers, i.e., $p_A^{IRS} = p_A^{CRS} = p_A$. Since in informal sector each worker produces one unit of the manufactured good, $w_A^{CRS} = p_A^{CRS}$ and therefore $p_A = w_A$.

Goods-demand and income

Let the price of food be numéraire, its value set at unity, in other words, $p_B = w_B = 1$. The demand functions can be written simply as:

$$C_A = \frac{\beta I}{p_A} \tag{8a}$$

and

$$C_B = (1-\beta)I \tag{8b}$$

Income consists of labor income and, whenever applicable, profits:

$$I = I_A + I_B = w_A L_A + \Pi_A + L_B \tag{9}$$

where profits can be expressed as

$$\Pi_A = p_A^{IRS}\alpha\left[L_A(1-0.5\gamma L_A) - F_A\right] - w_A^{IRS}L_A \tag{10}$$

Labor demand

The optimal output and hence employment decision of the formal sector producer is given by the limit pricing rule described above. Using the facts that $p_A = w_A$, $L_B = L - L_A$, and that in equilibrium $C_A = X_A$, and combining equations (5), (8a), (9) and (10), the inverse demand function for labor, given profitable factory production, can be expressed as:

$$w_A = \frac{\beta}{(1-\beta)}\frac{(L-L_A)}{\alpha\left[L_A(1-0.5\gamma L_A) - F_A\right]}, \qquad \Pi_A \geq 0 \tag{11a}$$

and, using (2) rather than (5), the inverse labor demand in manufacturing when factory production is not profitable becomes:

$$w_A = \frac{\beta}{(1-\beta)} \frac{(L - L_A)}{L_A}, \qquad \Pi_A \leq 0 \qquad (11b)$$

The slope of the inverse demand curve for factory labor is:

$$\frac{\partial w_A}{\partial L_A} = \frac{\beta}{\alpha(1-\beta)} \frac{\left[\gamma L_A(L - 0.5L_A) - (L - F_A)\right]}{\left[L_A(1 - 0.5\gamma L_A) - F_A\right]^2}, \qquad \Pi_A \geq 0 \qquad (12a)$$

Similarly, the slope of the wage-labor relation in manufacturing when only small scale technology is in use, is given by:

$$\frac{\partial w_A}{\partial L_A} = -\frac{\beta}{1-\beta} \frac{L}{L_A^2}, \qquad \Pi_A \leq 0 \qquad (12b)$$

The slope given by (12b) is clearly negative. In (12a), however, the expressions in the numerator parantheses are both positive, and the slope of the demand curve for factory labor depends on which of the expressions dominates. At low levels of urban employment the value of the first paranthesis is small and the numerator is therefore likely to be negative, and hence we have the conventional negatively sloped demand curve. At higher levels of urban employment, however, its value is large and the numerator is likely to be positive. Here there is a positive relation between city size and city wage. The intuition behind the upward sloping part of the factory's labor demand curve is that when the city becomes sufficiently large, an increase in city size reduces manufacturing output. This in turn puts an upward pressure on the price of the manufactured good and therefore an upward pressure on city wages.

Turning now to the factors which determine the level of the labor demand curve, it is evident from (12) that both an increase in the

budget share of manufacturing (β) and an increase in total labor supply (L) shift the demand curve to the right, irrespective of production technology. An improvement in the quality of city infrastructure (a reduction in γ), an increase in the productivity of formal sector labor (an increase in α), and a reduction in formal sector fixed costs (a reduction in F_A) all shift the factory labor demand curve given by (12a) to the left. Intuitively, countries which are technologically fairly advanced and/or have high quality city infrastructure have a very profitable manufacturing sector. Part of this income will be spent on the agricultural produce, given that this is a normal good. This again bids up wages in the rural area and leads to migration out of the city. A highly productive manufacturing sector is therefore a centrifugal force in the economy, reducing labor demand in manufacturing for any given wage.

The predicted negative correlation between city size and city infrastructure is supported by Ades and Glaeser (1995). They find that a one percent increase in the share of GDP spent on government transportation reduces main city size by 10 percent. Furthermore, their study suggests that higher GDP per capita increases main city size, but reduces main city growth. This evidence appears neither to confirm nor contradict the prediction in the present paper of a negative relation between city size and technological development. Note however that in the present model labor is the only factor of production and no technological progress takes place in the rural area. Hence, technological progress is of the labor saving kind and must necessarily reduce city population. A richer model including physical capital and the possibility of technological progress in the rural area would modify this result.[4]

---

[4]In Bjorvatn (1995a), technological progress increases the likelihood of labor and large scale producers locating in a single region. In that model, both goods can be produced by means of large scale technology, and this technology is not tied to any specific region. These features account for the difference in results between the two papers with respect to technological progress and locational choice.

Not surprisingly, Ades and Glaeser report a positive correlation between main city size and the country's population, which is predicted by this model too. Urban centralization, on the other hand, falls with population growth. This finding does not necessarily contradict the results from this model, since population growth up to the maximum profit city size is associated with a smaller urban population relative to total population. Population growth above this level, however, would increase urban concentration, contrary to the empirical evidence reported by Ades and Glaeser.

Equilibrium

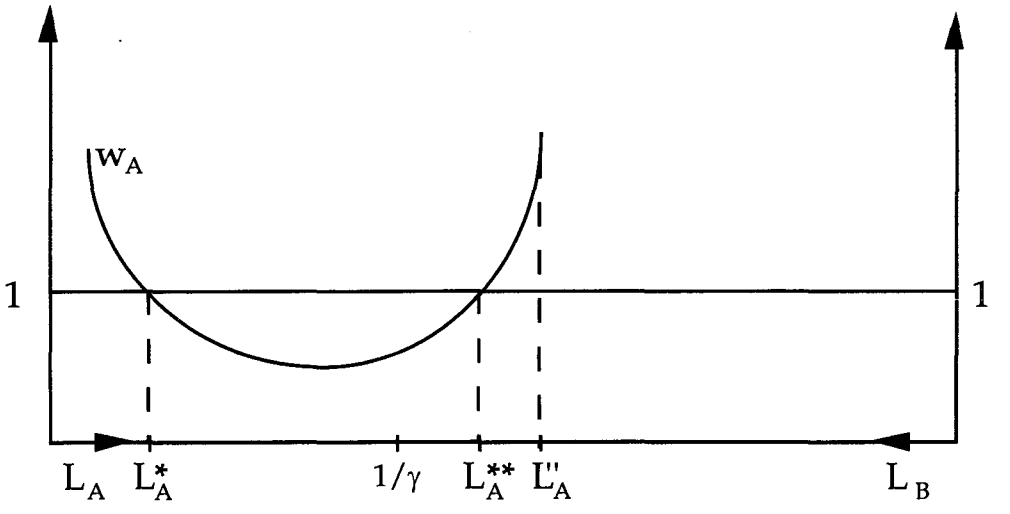
In equilibrium, wages in the two regions must be equal, i.e., $w_A = w_B = 1$, which again implies that $p_A = p_B = 1$. Inserting the equilibrium wage rate into (11a), i.e., assuming profitable large scale production, gives us two equilibria:

$$L_A^* = \frac{\alpha + \beta/(1-\beta) - \sqrt{\left[\alpha + \beta/(1-\beta)\right]^2 - 2\alpha\gamma\left[\alpha F_A + \beta/(1-\beta)L\right]}}{\alpha\gamma} \qquad (13a)$$

$$L_A^{**} = \frac{\alpha + \beta/(1-\beta) + \sqrt{\left[\alpha + \beta/(1-\beta)\right]^2 - 2\alpha\gamma\left[\alpha F_A + \beta/(1-\beta)L\right]}}{\alpha\gamma} \qquad (13b)$$

The urban labor demand curve and possible migration equilibria are illustrated in the figure below.

Figure 1. Equilibria



Of the two positive-profits-equilibria, only L_A^* is stable. This is evident by noting that $L_A^{**} > 1/\gamma$, which means that at this urbanization level the economy is on the downward sloping side of the «Laffer» productivity-curve, and hence the labor demand curve is upward sloping at this point.[5] This means that if a worker left the city to work in the rural region, food output would increase by one unit and manufacturing output by more than one unit. In other words, migration out of the inefficient city creates an excess demand for food, places an upward pressure on the income of peasants and thereby stimulates further migration out of the city.

Similarly, migration to the city from the rural region reduces manufacturing output by more than the reduction in food production, and therefore creates excess supply of food. This would depress the income of peasants and hence trigger further migration to the city. In this case the city aquires properties of «a black hole», attracting labor from the rural area with detrimental effects on output.

---

[5] This is also demonstrated in appendix A2, which shows that the marginal product of labor at L_A^{**} is strictly less than minus one.

The process of rural to urban migration would come to an end when factory profits have been entirely crowded out due to the congestion effects. Hence, the stable equilibrium associated with the large and inefficient city is one in which all manufacturing takes place in the informal sector, as indicated by L_A'' in figure 1, see (14b) for a derivation. In this case, labor will locate across regions in accordance with the budget shares in consumption, as is clearly evident by inserting the equilibrium wage of unity into (11b). For instance, if $\beta = 0.6$, sixty percent of the labor force will live in the city in the L_A''-equilibrium. In the more efficient L_A^*-equilibrium, of course, city population will be lower.

With existence of multiple equilibria, the economy may be be caught in a low-income trap with excessive city size. In this way, two structurally identical countries may look very different according to which equilibrium their economies coordinate on. The low-income country would be characterized by excessive city size with widespread informal sector manufacturing, while the high-income country would be characterized by profitable formal sector manufacturing and a smaller degree of urban concentration.

There is a role for the government in pushing the economy to the high-income, efficient city equilibrium. One rather paradoxical policy implication from this model is that industrialization may be encouraged by subsidizing agriculture. This policy would attract labor from the city to the rural region, thereby reducing congestion effects and making formal sector production possible. Yet another way to deal with the low-income trap is to invest in city infrastructure, thereby reducing congestion costs. If congestion costs become sufficiently low, this will improve the competitiveness of the large scale producer relative to the cottage producers, and improve the prospects of industrialization.

Finally, a note on the existence of multiple equilibria, which is contingent on $\Pi_A \geq 0$ at L_A^{**}. With negative profits at this level of city population, there will be a unique relation between the parameters of

the model and city size, L_A^* being the only surviving equilibrium. From (10) evaluated at equilibrium prices, the city size associated with the highest profits is easily found as $L_A = (\alpha - 1)/\alpha\gamma$, at which point $MPL_A^{IRS} = 1$. Similarly, zero profits occur at two levels of city size, a lower level (L_A') due to fixed costs in production, and a higher level (L_A'') due to congestion costs. The lower and upper bound of city size for profitable factory production, i.e., $\Pi_A \geq 0 \Leftrightarrow L_A \in [L_A', L_A'']$, are shown below. These zero profit limits can alternatively be expressed in terms of total population, by using the fact that $L_A = \beta L$ in a zero profit market equilibrium:

$$L_A' = \frac{\alpha - 1 - \Delta}{\alpha\gamma} \qquad \text{or} \qquad L' = \frac{L_A'}{\beta} \tag{14a}$$

and

$$L_A'' = \frac{\alpha - 1 + \Delta}{\alpha\gamma} \qquad \text{or} \qquad L'' = \frac{L_A''}{\beta} \tag{14b}$$

where $\Delta = \sqrt{(\alpha - 1)^2 - 2\alpha^2\gamma F_A}$. Existence of multiple equilibria is contingent on $L_A^{**} < L_A''$. Clearly, this is not always the case. If for instance $L_A'' \leq 1/\gamma$, then $L_A^{**} > L_A''$ (since as noted above $L_A^{**} > 1/\gamma$) and only the high income equilibrium survives. Comparing (13b) and (14b), it is evident that a larger population and/or stronger preferences for the manufacured good increase the likelihood for $L_A^{**} < L_A''$ and thereby make the existence of multiple equilibria more likely.

Population growth not only increases the chances of the economy ending up in a low income trap, *ceteris paribus* it will eventually lead to a situation where large scale production no longer is profitable. Population growth beyond the city size associated with maximum profits, $L_A = (\alpha - 1)/\alpha\gamma$, increases urban concentration, (L_A/L), and crowds out profits. In order to avoid deindustrialization, a continuous population growth must be accompanied by a continuous improvement in urban infrastructure and formal sector technology.

Technological improvements, in the sense of increasing α and reducing F_A, and improved urban infrastructure increases the range of population levels for which large scale production is profitable, i.e., increases $L \in [L', L'']$.

2. Welfare

In addition to the problem of coordination in a situation of multiple equilibria, there are two sources of market failure in this model. On the one hand there are increasing returns to scale in formal sector production, which may lead to a city which in equilibrium is too small in efficiency terms. On the other, there are congestion externalities which may lead to a city which is too large. Pareto efficiency requires the marginal rate of substitution to be equal to the marginal rate of transformation. With market prices equal to unity, the marginal rate of substitution is of course also unity. The marginal rate of transformation equals the marginal cost in manufacturing, i.e., $MRT_{A,B} = MC_A^{IRS}$. Since the marginal cost in large scale production is simply the inverse of the marginal product of factory labor, given by (6), the critical level of urban labor force at which $MRT_{A,B} = 1$, alternatively expressed in terms of total labor supply, can easily be found as:

$$L_A^P = \frac{\alpha - 1}{\alpha\gamma} \tag{15a}$$

$$L^P = \frac{\alpha - 1}{\alpha\gamma}\left[\frac{1 + \beta + \alpha(1-\beta)}{2\beta}\right] - \frac{(1-\beta)\alpha F_A}{\beta} \tag{15b}$$

where superscript P indicates Pareto-efficiency, and (11a) for equilibrium wages has been used in deriving (15b). Since both monopolistic pricing and congestion externalities are involved in this model, the market outcome is generally Pareto-inefficient. At levels of population and urbanization exactly equal to the critical values in (15),

however, monopolistic pricing and congestion externalities exactly offset each other and the market outcome is Pareto-efficient. Note also that L_A^P is the city size which generates the maximum profits. At levels of urban or total labor lower than those given by (15), the marginal rate of transformation is below unity, i.e., the city is too small in efficiency terms, and hence Pareto efficiency calls for a subsidization of the urban sector relative to the rural one, either by subsidizing manufacturing output or, equivalently, urban labor input. Taxing rural output or, equivalently, rural labor input would achieve the same results. These are policies of «urban bias».

At levels of urban or total labor higher than those given by (15), the marginal rate of transformation is above unity, i.e., the city is too large in efficiency terms, and hence the optimal policy involves taxing the urban sector relative to the rural sector. There may thus be a case for a policy of «rural bias». From (15) it is clear that urban bias is likely to be the optimal policy for a country with a more advanced technology and/or with a more developed urban infrastructure, while rural bias is likely to be the optimal policy in a country which is less developed in these respects.

Policies of urban or rural bias correct for distortions associated with large scale production, i.e., monopolistic pricing and congestion costs. But when is it optimal for society to change from one production technology to the other? Intuitively, for very small or very large population levels, it may be more cost efficient to leave manufacturing to small scale producers, thereby avoiding fixed costs and congestion costs. Clearly, as long as formal sector production at market prices generates a profit, which is given by the population interval defined in (14), this technology is superior to informal sector production which does not create any profit.

However, as noted above, equilibrium at market prices is generally not Pareto efficient. When ranking the two technologies in terms of Pareto efficiency, small scale production should be contrasted with large scale production corrected for market failure. The

population interval for which large scale production is Pareto superior to small scale production can be found by inserting the zero-profit uban population levels given by (14) into the optimality condition $MRS_{A,B} = MRT^{IRS}_{A,B}$. This defines a lower and upper limit of population size at which the two technologies generate the same utility. Within this interval profits are positive, and hence average costs are lower for the large scale technology relative to the small scale technology. Outside this interval, small scale technology is more cost efficient. The population interval for which large scale technology is Pareto superior is given by:

$$L \in \left[\frac{L'_A(1+\beta\Delta)}{\beta(1+\Delta)}, \frac{L''_A(1-\beta\Delta)}{\beta(1-\Delta)} \right] \qquad (16)$$

see appendix A3 for a derivation. The population interval defined by (16) is larger than that in (14), which reflects the fact that the government by means of optimal policies of urban bias in the low population case and rural bias in the large population case extends the population interval for which large scale production is profitable relative to the market solution. This also means that implementation of small scale technology outside this interval is achieved simply by the government refraining from intervening in the market.

3. Extensions

The model presented so far excludes two important features we would expect to find in a Third World city, namely coexistence of formal and informal sector and urban unemployment. The most convenient way to incorporate these phenomena into the model is to reformulate the congestion costs. So far commuting time has been determined by the size of the city, compare (3). Assume now that the average commuting time for formal sector employees is determined not by city size but rather by the number of such employees. To justify this assumption, imagine that labor is located at different distances from the central

business district, and that as production increases the factory has to employ workers living ever further away from this location. In other words, there are increasing marginal commuting costs in the number of formal sector employees. At the same time, these costs are unaffected by new arrivals to the city, who presumably settle down at the outskirts of the city. Equation (3) can then be reformulated as:

$$z_A = 1 - 0.5\gamma L_A^{IRS} \qquad (3')$$

Clearly, in this case the factory will never employ more labor than the profit maximizing level, $L_A = (\alpha - 1)/\alpha\gamma$. With city size in the interval $L_A' \leq L_A \leq (\alpha - 1)/\alpha\gamma$ only formal sector manufacturing production takes place, while urban population exceeding this level will be employed in cottage production, i.e., $L_A - (\alpha - 1)/\alpha\gamma = L_A^{CRS}$. This means that for sufficiently large cities, formal and informal sector will coexist.

The reformulation of commutig costs also allows us to include urban unemployment in the analysis. Assume that there is a wage premium in formal sector employment, $\overline{w}_A^{IRS} > 1$. In order to have a chance of obtaing a job in the factory, people have to live in the city. As in Harris-Todaro (1970), this chance is determined by the ratio between factory employees and total city labor. Furthermore, assume that participation in the competition for formal sector employment is a full time job, in other words, the alternative to formal sector employment to the participants is not cottage production but rather unemployment. Let L_A^U denote the number of urban umemployed. The expected wage in the city can then be expressed as

$$Ew_A = \frac{\overline{w}_A^{IRS} L_A^{IRS} + w_A^{CRS} L_A^{CRS} + w_A^U L_A^U}{L_A^{IRS} + L_A^{CRS} + L_A^U} \qquad (17)$$

where the denominator is the urban population. Since $w_A^U = 0$ and $w_A^{CRS} = w_B = 1$ and migration equilibrium involves $Ew_A = w_B$, urban unemployment equals:

$$L_A^U = L_A^{IRS}\left(\overline{w}_A^{IRS} - 1\right) \qquad (18)$$

which simply states that the number of urban unemployed equals the formal sector wage markup times the number of factory workers. These simple extensions to the model are not without sacrifice. By assuming congestion costs as in (3') rather than (3), congestion externalities are assumed away and hence there is no room for either multiple equilibria or market failure associated with these externalities.

5. Conclusion

This paper discusses the determinants of city size in a developing country, characterized by dual technology in the city and small scale agriculture in the rural area. Large scale production, or industrialization, requires a population which is not too small or too large. Population growth eventually crowds out profits and may lead to de-industrialization. Countries which are technically more advanced and/or have good quality city infrastructure will tend to be less urbanized and more likely to be industrialized. Market failure is likely to occur, due to coordination failure, increasing returns to scale in production or congestion externalities. The city may be either too small or too large in efficiency terms, the former case calling for a policy encouraging migration to the city and the latter calling for a policy encouraging migration to the rural area. The paper also identifies the population levels at which it is Pareto efficient to switch from one technology to the other. If the population is either too small or too large relative to these critical levels, nothing can be gained by government intervention in the market. Urban unemployment and coexistence of formal and informal sector are briefly discussed in an extension to the model.

Appendix

A1: Taking the first derivative of the profit function $\Pi_A = p_A^{IRS}\alpha\left[L_A^{IRS}(1-0.5\gamma L_A)-F_A\right]-w_A^{IRS}L_A^{IRS}$ with respect to factory workers (L_A^{IRS}), acknowledging that the limit pricing strategy of the factory implies $p_A^{IRS}=p_A^{CRS}=p_A=w_A$ and taking city population (L_A) as given, it turns out that

$$\frac{\partial\Pi_A}{\partial L_A^{IRS}}\geq 0 \Leftrightarrow L_A \leq \frac{2(\alpha-1)}{\alpha\gamma}\equiv L_A^0$$

Since $L_A^0 > L_A''$ defined in (14b), this shows that as long as profits are non-negative, the factory will absorb all workers in region A.

A2: Substituting (13b) into (6), the marginal product of labor at L_A^{**}, $MPL_A^{IRS}(L_A^{**})$, becomes

$$MPL_A^{IRS}(L_A^{**}) = -\left\{\beta/(1-\beta)+\sqrt{\left[\alpha+\beta/(1-\beta)\right]^2-2\alpha\gamma\left[\alpha F_A+\beta/(1-\beta)L\right]}\right\}$$

For equal budget shares, this expression is obviously less than minus unity. An increase in β would make it even more negative. However, it is easy to demonstrate that even when β is close to zero, the marginal product of labor at this city size must be less than minus unity. For $\beta=0$, the expression reduces to

$$MPL_A^{IRS}(L_A^{**}) = -\sqrt{\alpha^2-2\alpha^2\gamma F_A}$$

Since the expression in the square root of (11) must be positive, it is straightforward to demonstrate that the expression above must be less than minus unity. Hence, irrespective of budget shares, $MPL_A^{IRS}(L_A^{**})<-1$.

A3:

The marginal rate of transformation, or equivalently the marginal costs, associated with the two zero-profit levels of large scale production given by (14), can be found as

$$MRT_A^{IRS}(L'_A) = MC_A^{IRS}(L'_A) = \frac{1}{1+\Delta}$$
$$MRT_A^{IRS}(L''_A) = MC_A^{IRS}(L''_A) = \frac{1}{1-\Delta}$$

The marginal rate of substitution is

$$MRS_{A,B} = \frac{\beta C_A}{(1-\beta)C_B}$$

which for zero profits can be written in terms of labor allocation as

$$MRS_{A,B} = \frac{\beta}{(1-\beta)}\left[\frac{L}{L_A} - 1\right]$$

The population levels for which the two technologies generate the same utiliy can then be found by inserting the two zero profit urban population levels $L_A = L'_A, L''_A$ into the Pareto optimality condition $MRS_{A,B} = MRT_{A,B}^{IRS}$:

$$L = \frac{L'_A(1+\beta\Delta)}{\beta(1+\Delta)} \quad \text{for} \quad L_A = L'_A$$
$$L = \frac{L''_A(1-\beta\Delta)}{\beta(1-\Delta)} \quad \text{for} \quad L_A = L''_A$$

This defines the lower and upper limit of population size for which large scale manufacturing is the more efficient technology for society.

References

Abdel-Rahman, H.M. (1988). «Product differentiation, monopolistic competition and city size,» *Regional Science and Urban Economics* 18: 69-86

Ades, A. F. and Gleaser, E. L. (1995). «Trade and circuses: Explaining urban giants,» *Quarterly Journal of Economics* 110: 195-227

Bjorvatn, K. (1995a). «Industrialization and regional inequality,» Discussion paper 3/96, The Norwegian School of Economics and Business Administration.

Bjorvatn, K. (1995b). «Leviathan in a dual economy,» *Public Choice* 84: 137-151

Boserup, E. (1965). *The conditions of agricultural growth,* Aldine Publishing Company, Chicago

Dixit, A.K. and Stiglitz, J.E. (1977). «Monopolistic competition and optimum product diversity,» *American Economic Review* 67: 297-308

The Economist (July 29, 1995). «Turn up the lights: A survey of cities.»

Livas Elizondo, R. and Krugman, P. (1992). «Trade policy and the third world metropolis,» NBER Working paper No. 4238

Fujita, M. (1989). *Urban economic theory: Land use and city size,* Cambridge University Press

Harris, J. and Todaro, M. (1970). «Migration, unemployment and development: a two-sector analysis,» *American Economic Review* 60 (March): 165-187.

Henderson, J. V. (1974). «The sizes and types of cities,» *American Economic Review* 64: 640-657

Krugman, P. (1980). «Scale economies, product differentiation, and international trade,» *American Economic Review* 70: 950-959

Krugman, P. (1991a). «Increasing returns and economic geography,» *Journal of Political Economy* 99 3: 483-499

Krugman, P. (1991b). *Geography and trade,* MIT-Press

Krugman, P. and Livas, R. (1992). «Trade policy and the Third world metropolis,» NBER Working Paper No. 4238

Krugman, P. and Venables, A. (1993). «Integration, specialization, and adjustment,» *mimeo.*

Murphy, K.M., Shleifer, A. and Vishny, R.W. (1989). «Industrialization and the big push,» *Journal of Political Economy* 97 5: 1003-102

Puga. D. (1994). «The dynamics of urbanisation,» Center for Economic Performance Working Paper no. 652, London

United Nations (1991). *World urbanization prospects 1990,* New York: United Nations

Chapter 4

Leviathan in a dual economy[†]

Published in Public Choice 1995, 84: 137-151

Abstract

This paper studies the policies of a leviathan[1] state with an ambition to make money and to stay in power. The economic framework is a developing country characterized by dual markets. Government revenue is raised by taxing the agricultural sector. Part of this revenue is offered to the politically influential urban population in the form of low productive but well paid public sector employment. I calculate the optimal level of rural taxation from the point of view of the Leviathan and investigate its effects on migration, industrialization, public sector employment and urban unemployment in different political scenarios.

[†] I would like to thank Karl Rolf Pedersen and Anne O. Krueger for valuable comments to earlier versions of this paper.

[1]Leviathan is the sea monster that Jahve according to the Bible destroyed when creating earth. It is also the name of the whale which swallowed the prophet Jonah. The concept was introduced to the economics literature by Brennan and Buchanan (1980).

1. Introduction

The concept of dual markets has a central place in development economics. Economic development, according to a pathbreaking article by Arthur Lewis (1954), is achieved by moving resources from traditional forms of production and economic organization in rural areas to the capitalist economy of the urban areas where saving and investment take place. In accordance with the Lewis' plan for economic development, poor countries have introduced extensive taxation of rural areas. Schiff and Valdés (1992) estimate the average annual income flow from agriculture in the developing countries included in their study for the years 1960-1984 to 46% of agricultural GDP. However, the enormous outflow of resources from agriculture to central areas has failed to bring about successful industrialization and the expected big push. On the contrary, this transfer of resources has created a pool of low-productive but politically highly influential people in the cities, whose demand for economic privileges translates into policies which in many cases have been detrimental to economic development.

Many observers blame bad government policy for this lack of success, see for example The Economist (1989). This emphasis on the "negative state" suggests a public choice rather than a welfare economic approach to the problem of development. Inspired by Findlay (1991), this paper studies a self-interested, leviathan government which aims at maximizing government surplus.[2] This government may consist of a single dictator, or a group of people such as a military junta or a clan, whose members may hold senior positions in the military, the bureaucracy, etc. The Leviathan extracts an income from society through taxation, which can be interpreted as a rent related to the ruler's monopoly in use of force. For political reasons Findlay assumes that it is possible to tax the rural sector only.

---

[2]In development literature this surplus is usually referred to as "investible surplus", due to the importance of the government as capitalist investor in the early stages of development. In this paper it will be appropriate to interpret government surplus as the Leviathan's private income which will be spent on investment and consumption according to the ruler's own preferences.

Extending Findlay's analysis, I borrow from Richard Posner (1975) and suggest that government income may at least partly be subject to rent dissipation. The idea is that the urban sector is influential enough not only to resist taxation but also to capture part of the government income. I do not model political pressure explicitely, but simply postulate that there exists a positive relationship between the Leviathan's wealth, and possibly the degree of urbanization, and the pressure from influential groups on the government to share of its wealth. In this way political pressure is operationalized as a tax on government income. Although leaving aside very important questions concerning the formation of pressure groups, the lobbying technology and the government response to lobbying, this formulation seems reasonable as a reduced form expression of the political process. More wealth in the hands of the Leviathan increases the potential gains from lobbying, and is therefore likely to intensify the political pressure forcing the government to hand over part of the increase in wealth. In addition, a larger urban population means a larger group of people that can be rallied against the government, a fact which may increase the pressure on the government.

Transfers from the state treasury to the urban community can of course take many forms, such as gifts in cash or kind, subsidized food, and so on. Here we assume that the Leviathan purchases loyalty by offering well paid employment in the public sector, including bureaucracy, parastatals, police and military. The approach is inspired by Gelb, Knight and Sabot (1991), who stress the importance of public sector employment in developing countries as a political loyalty device. Since the creation of these positions are motivated by political consideration, it is reasonable to believe that their contribution to the national product is low.

The authors find support for their approach in the World Bank "World Development Report" (1979), page 64: "Overmanning at all levels is common since public undertakings are often viewed as employers of last resort; hiring decisions frequently result from the

exercise of political patronage while dismissal procedures are cumbersome and ineffectual.... These practices lead to frequent losses which are almost invariably financed from the national treasury or the banking system." The excessive size of and relatively high wage level in the public sector of LDCs, reported by Heller and Tait (1983), underscore the severity of this problem. In addition to the low productivity of the public sector, it is reasonable that the high wages in public sector attract rent seekers. People spend time lobbying for a position in the public sector, with those failing in their lobbying efforts suffering unemployment. Rent seeking triggered by above market wages in the public sector in this way adds to the economic waste.

An oversized public sector and extensive rent seeking are of course not problems limited to developing countries. Indeed, it is a phenomenon found in most societies, irrespective of political system or level of economic development. However, it might well be that the problem is particularly severe in developing countries due to weak property rights protection, failed economic policies which limit the potential for economic success through production for the market, and lack of political legitimacy which translates into weak states vulnerable to lobbying from influential groups. The paper is organized as follows: Section 2 describes the private sector and section 3 political pressure and unproductive activities. Section 4 derives the Leviathan's choice of rural taxation, and section 5 presents some comparative static results. Section 6 concludes.

2. The private sector

The economy under scrutiny is small and open with two constant returns to scale productive sectors, agriculture and manufacturing. Inputs in the agricultural production are labor and land and in manufacturing labor and capital, where land and capital are sector specific inputs. Labor is mobile between sectors while capital is mobile internationally. Both sectors compete for the same homogenous pool of labor, taking wages as given. The goods are traded on the

international market at fixed world prices. Domestic producer prices, however, may differ from world prices due to taxation of local agricultural supply. In many developing countries this kind of taxation is carried out by government marketing boards with monopolistic power in the market for rural produce and inputs.[3] An increase in taxation will reduce domestic supply of agricultural goods and lead to an increase in imports or a reduction in exports of such goods, leaving consumer prices unchanged.

The number of people involved in the leviathan government is assumed to be fixed, and will not be included explicitely in the analysis. Labor available for productive and non-productive activities, $\bar{L}$, is therefore equal to the total labor supply minus labor input in government. The resource constraint in the labor market can be expressed as

$$\bar{L} = L_m + L_a + L_p + L_u, \tag{1}$$

where L_m is labor input in manufacturing, L_a is labor input in agriculture, L_p is public sector employment, and, when applicable, L_u is unemployment, i.e. the unsuccessful rent seekers. For capital we have the following resource constraint:

$$K_m^d + K_m^f = K_m \tag{2}$$

where the capital used in manufacturing K_m comes from domestic (K_m^d) and foreign (K_m^f) sources, where K_m^f is positive when the country is a net importer and negative when it is a net exporter of capital.

i) Agriculture

[3]According to The Economist (1989) some boards in Africa are handling three quarters or more of their country's export earnings, which gives an indication of the dominant position of this institution in Africa. See also Schiff and Valdés (1992).

As in the traditional dual economy models initiated by Lewis, we shall assume that there is surplus labor in agriculture. Surplus labor is generated by income sharing, due to the extended family type of agriculture, an institutional setup commonly found in Africa. The peasants own the land they work on and share the land rent evenly between members of the group. Here labor is remunerated according to after tax average rather than marginal product. Let $\bar{p}$ denote the international price ratio between agricultural goods and manufacturing goods, with the latter chosen as numéraire with price equal to unity. Let t be the tax rate on the agricultural output. Land per farmer is given by q_a, and $g(q_a)$ is the average product. With income sharing, income per farmer (w_a) is given by

$$(\bar{p} - t)g(q_a) = w_a \qquad (3)$$

where the left-hand side of (3) is the after tax value of average product in agriculture, measured in units of the manufactured good.

ii) Manufacturing

With competitve markets in manufacturing, labor demand can be expressed as

$$f(k_m) - f'(k_m)k_m = w_m \qquad (4)$$

where the left-hand side of equation (4) is the marginal revenue product of labor in industry. Here, k_m is the capital-labor ratio in manufacturing, $f(k_m)$ is the average product of labor, $f'(k_m)$ is the marginal productivity of capital, and $f'(k_m)k_m$ is profit per manufacturing sector worker. The right hand side of (4) is the manufacturing wage rate. The slope of the labor demand curve in manufacturing is given by $f''(k_m)k_m^2/L_m$. When capital mobility is high, the second derivative in this expression will have a low negative

value, while limited degree of capital mobility implies that the second derivative has a higher negative value. Hence, the the more responsive are international capital movements to investment opportunities, the more elastic is the manufacturing labor demand curve, and vice versa.

Equilibrium

In equilibrium the peasant and the industrial worker earn the same income:

$$(\bar{p} - t)g(q_a) = f(k_m) - f'(k_m)k_m \qquad (5)$$

Note that with perfect capital mobility between countries the rate of return to capital is determined on the international market as $\bar{r}$, that is; $f'(k_m) = \bar{r}$. With competitive wage setting in manufacturing, w_m will then be determined uniquely by $\bar{r}$. Intuitively, this is because a constant marginal product of capital requires a constant capital-labor ratio which in this constant returns to scale economy uniquely determines the marginal product of labor and hence the manufacturing wage. To simplify the algebra, we shall assume perfect capital mobility when calculating the political economic equilibrium in section 4. The qualitative results, however, are not affected by this assumption. Gross domestic product and national income can be written as

$$GDP = w(\bar{L} - L_p - L_u) + \bar{p}g'(q_a)q_aL_a + f'(k_m)K_m$$

$$National\ income = w(\bar{L} - L_p - L_u) + \bar{p}g'(q_a)q_aL_a + f'(k_m)K_m^d + Aid$$

In both equations the first term is wage income of productive labor and the second is land rent in the rural sector. In the GDP expression the third term is remuneration to capital, which in the national income expression is remuneraton to capital in the hands of domestic owners.

Finally, national income includes foreign aid. Note that in the absense of government intervention, GDP and national income are at a suboptimal level. Income sharing in agriculture implies that the marginal product of labor in manufacturing exceeds that of agriculture. A reallocation of labor from agriculture to manufacturing would increase the productivity of the economy.

3. Political pressure and rent-seeking

The goal of the leviathan government studied here is to make money and to stay in power. Government income from domestic sources is generated through taxes and can be viewed as a rent stemming from the government's monopoly in use of force, as noted above. Rents attract rent seekers who place political pressure on the government by threatening it with demonstrations and the like. This pressure is operationalized as a tax on government income. The proceeds of this tax are distributed to the influential groups in the form of well paid employment in the public sector. We shall first consider the determinants of the size of public sector employment, and then deal with lobbying and urban unemployment.

3.1 Political pressure and the public sector

The alternative to manufacturing in the urban sector is public employment. For simplification we shall assume this activity is completely unproductive. The amount of money put into the public sector is determined by the "strength" of the state and the size of government income (G). My assumption is that the richer the Leviathan, the larger the potential for political pressure aiming at conquering part of this wealth. High government income therefore goes hand in hand with a high level of transfers to the public sector. Political pressure in this way works as an income tax on the government, with the "political pressure tax rate" denoted by θ. Total

transfers (T) from the Leviathan to the public sector can then be expressed as

$$T = \theta G \qquad \text{with} \qquad 0 \leq \theta \leq 1 \tag{6}$$

The size of the public sector depends on state autonomy, with autonomy defined as the government's ability to withstand pressure from urban lobbyists and hold on to the income generated from taxing agriculture. The lower the value of θ, the stronger is the state vis á vis society. State autonomy depends on a number of variables which may be both exogenous and endogenous relative to this model. Here I limit myself to studying two political scenarios. In the first scenario, θ depends only on exogenous factors, such as superpower support and/or personal charisma of the Leviathan. In the second scenario θ is endogenously determined, and described as a positive function of urbanization. The idea is that a larger urban population increases the number of people that can be mobilized for demonstrations and the like, thereby increasing the political pressure on the government. Public sector labor input L_p is determined by

$$L_p = \frac{T}{w_p} \tag{7}$$

Equation (7) shows possible wage-labor combinations for a given level of transfers. This relationship forms a rectangular hyperbola in the wage-labor space which we may call an iso-transfer curve. The public sector wage level is assumed to be exogenously determined, for example through negotiations with trade unions. With w_p fixed, the number of public servants is uniquely determined by the level of transfers. It can be argued that the fortune of the individual public sector servant is likely to be positively linked to that of the Leviathan, and not separate, as the fixed w_p would imply. Endogenizing w_p, explaining it as a positive function of G, would not change the

important features of this model. In fact, in the case of lobbying, to which we will turn next, since the expected wage remains the same, this endogenization would only affect the division of the urban labor force between public employment and unemployment. In the absense of lobbying, a similar endogenization of w_p would in addition to the above stated effect also strengthen the monopsony power of the Leviathan and thereby increase the optimal tax rate. This is because an increase in taxation which adds to the Leviathan's income now partly is absorbed by a public sector wage rise (and not solely by increasing demand for bureaucrats), which again takes away some of the pressure on private sector wages. Assuming perfect capital mobility, of course, the Leviathan has no market power and so the optimal tax rate would not be affected by endogenization of w_p even in the absense of lobbying. For simplicity, we shall stick to the assumption of a fixed public wage throughout the rest of the analysis.

3.2 Lobbying and urban unemployment

In the description above, we have said nothing about the competition for the lucrative public sector positions. One way of rationing these jobs is to make use of some lottery mechanism. In this case the competition does not involve economic waste. Still, the consequence of rent seeking is economic waste since the prizes of the lottery are low productive public sector employment. More likely, however, the competition for public sector employment in itself is resource consuming. Assume that application for public sector employment is a full time job, call it lobbying. By spending time on rent seeking the lobbyists exclude themselves from a position in the private sector, implying that those who fail in their lobbying efforts suffer unemployment. Following Harris and Todaro (1970), the simplest way of determining the extent of urban unemployment is by introducing a probability function which states that the expected urban wage equals the average urban wage. This is also the specification of rent seeking used in Krueger (1974). The expected wage in lobbying (w^e) is then

$$w^e = \frac{w_p L_p}{L_p + L_u} \tag{8}$$

where the denominator gives the number of rent seekers, successful and unsuccessful, the unsuccessful becoming unemployed. This means that the possible combinations of expected wage and number of rent seekers forms a rectangular hyperbola identical to the iso-transfer curve derived in the section above. In equilibrium, the expected wage must equal the market wage. Clearly, both the low productivity and the high wage features of the public sector contribute to the waste of resources, the latter by creating urban unemployment.

4. Rural taxation

The government wishes to maximize its surplus $\Pi = G - T$, where G consists of tax income (tX_a) and exogenously given foreign aid (Aid) and where T, as noted earlier, is the amount of money the Leviathan spends on the public sector. Since $T = \theta G$ government surplus can be written as

$$\Pi = (1-\theta)G = (1-\theta)\left(tX_a + Aid\right) \tag{9}$$

The decision variable for the government is the tax rate on rural output t. Taxing the rural sector causes migration to the "tax free" urban sector. Migration constitutes an **economic cost** to the government since it reduces the tax base. This limits the potential for the Leviathan's exploitation of the rural sector. In addition, when political pressure is a function of urbanization, rural to urban migration creates a **political cost** to the government. This further limits the scope for taxation of the rural sector.

4.1 Exogenous θ

I shall start out by analysing the case of exogenous political pressure, which implies a given level of the political pressure tax rate θ;

$$\theta = \overline{\theta} \tag{10}$$

Since $\overline{\theta}$ is fixed, surplus maximization becomes analogous to revenue maximization. Here, the political tax does not distort the Leviathan's decision on rural taxation. Note that migration to the cities in this scenario is no political burden to the government. However, there is what we have called an economic cost to the government of migration, since less people in the rural area means a smaller tax base. Maximizing (9) with respect to the tax level implies that at optimum, the elasticity of supply with respect to the rural tax should equal minus one.

$$\frac{\partial X_a}{\partial L_a}\frac{\partial L_a}{\partial t}\frac{t}{X_a} = -1 \tag{11}$$

With perfect capital mobility, the wage rate is uniquely determined by the international interest rate, and we are left with two unknowns, t and L_a, defined by two equations (11) and (3). It can easily bee seen that the political economic equilibrium implies

$$\overline{p}MP_a = p^* AP_a \left(= w_a\right) \tag{12}$$

Equation (12) says that the value of the marginal product of labor in agriculture, measured at world prices, should equal the marginal cost in agricultural production. In other words, the Leviathan will replicate

a profit maximizing strategy in the market for rural labor.[4] Profit maximization in agriculture means maximizing the land rent, which implies that the government will act as the country's landlord, the land rent functioning as tax income for the government. Since by definition agriculture with income sharing yields no land rent, the state "land owner" will reduce rural employment by taxing its output.

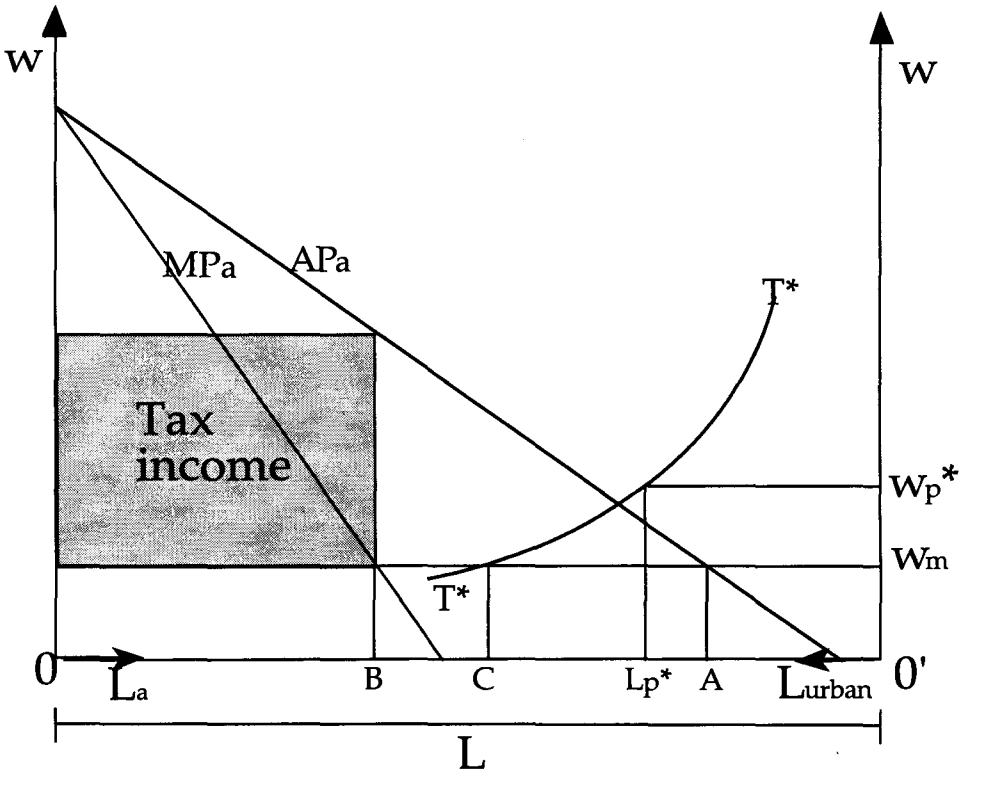
With perfect capital mobility the state will act as a price taker in the labor market and will therefore realize a Pareto efficient allocation of labor between rural and urban areas given that there are no distortions in the manufacturing sector. In this case, by taking the role of the profit maximizing land owner, the Leviathan has in fact corrected the distortion in the economy due to income sharing in agriculture. This potential welfare gain must be weighed against inefficiencies in the urban sector caused by the Leviathan's purchase of political support. The more autonomous the government, the less rent dissipation in the form of public employment and urban unemployment, and hence the greater the chance for an overall positive effect on the economy of the Leviathan's policies. A politically weak Leviathan, on the other hand, will be forced to create a large bureuacracy, thereby in effect transferring the inefficiency of surplus labor from the rural to the urban area.

How does this compare with Findlay's study? In his 1991 article, page 30, Findlay concludes that the Leviathan's taxation of the rural sector contributes to the industrialization of the economy but makes it worse off than before due to welfare losses from distortionary taxation. These results do not necessarily conform to the findings reported in the present work. First, industrialization **may** materialize, but it depends critically on the autonomy of the state vis à vis influential groups in

[4] Less than perfect capital mobility implies that the Leviathan achieves some degree of monopsonistic power in the labor market, which increases the revenue optimizing tax level. The more rigid are capital movements accross borders, the less elastic is the labor demand in manufacturing and hence the stronger is the monopsony power of the Leviathan.

society. It may well be that the policies of the Leviathan in sum have lead to a reduction in private manufacturing. This is due to political pressure and rent seeking absent from Findlay's analysis. Second, the economy **may** be worse off, but this again depends on how many resources are wasted on low productive activities in the urban areas and the degree of capital mobility. Indeed, the policies of a relatively autonomous Leviathan are likely to improve the overall economic efficiency, more or less by coincidence. The reason for the divergence in this second result is based on my assumption of income sharing and hence surplus labor in agriculture. With a capitalist rural sector as in Findlay's model, the policies of the Leviathan will unambiguously harm economic efficiency. Figure 1 summarizes the discussion so far.

Figure 1. The political economic equilibrium under perfect capital mobility



Total labor supply is given on the horisontal axis, with labor input in agriculture measured from left to right and labor input in urban

activities ($L_{urban} = L_m + L_p + L_u$) from right to left. Urban wages and income per farmer are measured on the right and left vertical axes, respectively. In the absense of rural taxation, labor input in agriculture is given by $0A$. Labor input in urban sector activities is therefore $0'A$. As shown in (12), the optimal strategy for the Leviathan is to act as the country's landlord, maximizing land rent in agriculture. This implies equating the value of the marginal product with the marginal cost of production in agriculture. Labor input in agriculture is now reduced to $0B$ with tax income marked by the shaded area. The urban labor force has increased to $0'B$. As argued above, part of the government revenues is used to finance the public sector. In the figure, this amount is given by the TT-curve, the iso-transfer curve, its level depending positively on government income and negatively on the degre of government autonomy. The public sector employs $0'L_p^*$. When the government jobs are rationed through lottery, the remainder of the urban workforce (L_p^*B) is employed in manufacturing. When lobbying is involved, however, unemployment equals L_p^*C and only CB is labor input in manufacturing. Clearly, the non-productive public sector together with lobbying for these positions have limited the potential for industrialization created by resource allocation from rural taxation.

4.2 Endogenous θ

In this section political pressure is endogenized by postulating that political pressure tax rate θ is a positive funciton of urbanization. The more people that can be organized for demonstrations and the like, the larger is the potential threat for the government. Assume that the political pressure tax rate is equal to the ratio of city dwellers to the total population, i.e.

$$\theta = \frac{\bar{L} - L_a}{\bar{L}} , \qquad (13)$$

For example, according to this formulation, an equal distribution of labor between the rural and urban sector would give a fifty percent political pressure tax rate. An increase in rural taxation causes migration to urban areas and thereby affects θ. This is the political cost of migration which adds to the economic cost studied above. The equivalent of (11) now becomes

$$\frac{\partial X_a}{\partial L_a}\frac{\partial L_a}{\partial t}\frac{t}{X_a}+\frac{\partial L_a}{\partial t}\frac{1}{L_a}(t+\frac{Aid}{X_a})=-1 \qquad (14)$$

where the second term on the left hand side captures the political cost of a marginal increase in t. Since the political pressure tax rate is a function of L_a, we still have a system of only two unknowns, t and L_a, defined by the equations (14) and (3). The political economic equilibrium in this case can be described as

$$\frac{\bar{p}(MP_a+AP_a)+Aid/L_a}{2}=p^*AP_a\ (=w_a) \qquad (15)$$

Since the average product is larger than the marginal product, the agricultural wage rate in (15) must be higher, and therefore taxation lower, than in (12) even in the absense of aid. Introduction of aid strengthens this result. Hence, not surprisingly, when migration carries a political cost the Leviathan chooses lower rural taxation compared to the situation when political pressure is exogenous. The welfare implications of this is ambiguous. Remember that the resource allocation between the rural and urban sector in (12) is optimal. Introduction of a political cost of urbanization reduces rural taxation and therefore create overemployment in the rural sector. This inefficiency must be weighed against the efficiency gains in the form of a smaller public sector and less rent seeking in the cities due to reduced government income.

5. Some comparative static results

5.1 Foreign aid

Independent of the formulation of the political pressure tax rate, aid given to the Leviathan reduces GDP by increasing the public sector and rent seeking. The Leviathan will also gain from an increase in aid since part of it ends up as government surplus. The reverse is of course true for a reduction in aid. In this model, the only kind of aid that makes sense from a development perspective is aid aimed at productivity enhancing projects in manufacturing, that is, technology transfers. This is based on the assumption that manufacturing cannot be taxed. Increased output in this activity does not raise government income and therefore does not increase the public sector or the potential for rent seeking. On the contrary, improved productivity in manufacturing reduces economic waste since a more competitive manufacturing sector attracts people from agriculture and reduces the tax income for the government. This means a lower level of transfers to the private sector, which means that the TT-curve shifts to the right. Fewer public sector jobs also reduces rent seeking by reducing the expected wage in lobbying.

Note that endogenization of the public pressure tax rate leads to a situation where the optimal level of rural taxation is negatively correlated with the level of aid. This is because the government by making it more profitable to stay on the countryside will put downward pressure on θ and keep a larger portion of the aid for itself. Indeed, if the level of aid is high enough, the government might even want to spend some of the aid to subsidize rural output. On the other hand, when the rent seeking tax is outside government control as in (12), there is no incentive to make rural taxation responsive to changes in aid, ceteris paribus.

5.2 Government autonomy

A reduction in foreign support and a fading personal charisma are factors which are likely to weaken government autonomy. In this model, such a development is captured by an increase in the political pressure tax, leaving less surplus for the Leviathan. This development seems to be a fair description of what has happened in Zaire under Mobutu[5]. Running the country as his own private business, President Mobutu has been able to amass enormous amounts of money, by some estimated to equal half of the country's external debt, currently at USD 10 billion.

With the weakening of the Mobutu regime, the political pressure tax rate for Zaire has increased, implying that most of income that reaches the state is handed over to the public sector, particularly military and police. Together with the loss in authority there has been a sharp decline in government revenues, due to a virtual halt in the inflow of foreign aid, a persistent decline in national product, and the breakdown of the country's infrastructure. Compensating for the shortfall in transfers from the state, public sector employees such as the military and police demand bribes and threaten with brutal force against civilians to make a living. The experience in Zaire and other African countries, such as the recent case of Rwanda, are dramatic expressions of the social disorder and economic inefficiencies that may result from a weakening of Leviathan. The situation in these countries are perhaps best described by the Hobbesian state of nature, where life is "solitary, poor, nasty, brutish and short."

6. Concluding comments

Despite the enormous transfer of resources from agriculture to urban areas in Third World countries, industrialization and economic success have not materialized. Many observers blame bad policies for this lack

[5]On the current state of affairs in Zaire, see for example the article "Zaire drifting into anarchy as authority disintegrates" by John Darnton in The New York Times of May 24, 1994, page 1.

of success. In many LDCs, particularly in Africa, policies appear to be carried out by a ruler or a political élite whose main goal is personal enrichment and power. This model endogenizes economic policy by studying a leviathan state operating in a traditional dual developing economy. We have derived the optimal policy of the Leviathan under different political assumptions and seen how these policies affect migration of labor, industrialization, public sector employment and urban unemployment.

Generally speaking, transfer seeking activities on part of the private sector tend to moderate the industrialization effect of rural taxation, since at least part of the increased urban labor force is absorbed in unproductive activities. Political pressure and rent-seeking, as materialized in a low productive public sector and urban unemployment, can explain the lack of success of the Lewis inspired plan for economic development. Going beyond the model, it may be argued that a large public sector in itself reduces the productivity of manufacturing. A large public sector often implies time consuming, and perhaps bribe consuming, procedures which may serve as serious impediments to business.

In addition, to the degree that the public sector absorbs the most able people in the economy, this reduces the skill level and hence the productivity in manufacturing and farming. Since the most talented people most likely also constitute the highest political risk, this problem of adverse selection of talent in productive activities seems a real threat to economic efficiency. If, as suggested by Murphy, Shleifer and Vishny (1991), there is a link between innovation and the quality of the most able people in manufacturing, this problem of adverse selection may also have detrimental effects on economic growth. Left for further research is an analysis of the policies of Leviathan within a dynamic economic setting. Another important improvement to the model would be to include a specification of the interaction between state and pressure groups. Grossman and Helpman (1993) have taken important steps in this direction, but more needs to be done.

References

Bates, R.H. (1988). «Governments and agricultural markets in Africa,» In R.H. Bates (Ed.), *Toward a political economy of development,* University of California Press.

Brennan, G. and J.M. Buchanan (1980). *The power to tax: Analytical foundations of a fiscal constitution.* Cambridge University Press.

Findlay, R. (1991). «The new political economy: Its explanatory power for LDCs,» In G.M. Meier (Ed.), *Politics and policy making in developing countries - perspectives on the new political economy,* San Fransisco, California: ICS Press.

Gelb, A., Knight, J. B. and Sabot R. H. (1991). «Public sector employment, rent seeking and economic growth,»*The Economic Journal* 101: 1186-1199

Grossman, G. and E. Helpman (1993). «Protection for sale,» CEPR Discussion paper no. 827, Centre for Economic Policy Research, London.

Harris, J. and M. Todaro (1970). «Migration, unemployment and development: a two-sector analysis,» *American Economic Review* 60 (March): 165-187.

Heller, P.S. and A. A. Tait (1983). «Government employment and pay: Some international comparisons,» IMF- occasional paper no. 24, International Monetary Fund, Washington D.C.

Krueger, A.O. (1974). «The political economy of the rent-seeking society,» *American Economic Review* 64(3): 291-303

Lewis, W.A. (1954). «Economic development with unlimited supplies of labour,» *Manchester School* 22: 139-91.

Murphy, K.M., A. Shleifer and R. Vishny (1991). «The allocation of talent: Implications for growth,» *Quarterly Journal of Economics* 2: 503-530.

Posner, R.A. (1975). «The social cost of monopoly and regulation,» *Journal of Political Economy* 83: 807-827.

Schiff, M. and A. Valdés (1992). *A Synthesis of the economics in developing countries.* Volume 4 in the World Bank comparative study: *The political economy of agricultural pricing policies.* The Johns Hopkins University Press.

The Economist (1989). *Poor man's burden,* A survey of the Third World, (September 23).

World Bank (1979). *World development report, 1979,* The World Bank, Washington D.C.

Chapter 5

Rent-seeking and foreign aid

Abstract

This paper analyses the impacts of foreign aid in a rent-seeking society in which income distribution is determined by political competition amongst influential groups. When given unconditioned on the actions of the recipients, aid will be ineffective, part of it being dissipated due to increased rent-seeking and the remainder allocated according to political influence rather than the objectives of the donor. When conditioned on the actions of the recipients, however, foreign aid may eliminate rent-seeking and inequality. This result is in accordance with Becker's rotten kid theorem. The central policy implication from the model is that aid offered to a rent-seeking society should be flexible, which in practice means emphasising small-scale and short-term projects. Furthermore, if the distribution of political skill between the groups is too disparate relative to the aid budget, political reform should preceed aid.

Introduction

Foreign aid aims at promoting efficiency and equality in poor countries. Yet econometric studies and casual empiricism suggest that the donor community in many cases has not been very successful in achieving this ambition, Mosley et al. (1987), Krueger et al. (1989), Cassen et al. (1994), and Boone (1995). One explanation for the lack of success may be found in what Buchanan (1975) has called the Samaritan's dilemma. This dilemma is closely related to the free rider incentives faced by potential contributors to a public good. In the context of altruistic transfers, the recipient's welfare enters the objective function of the altruist. Hence, the recipient's consumption is a non-rival good. The recipient free rides on the altruism of the donor through choice of consumption profile over time. By increasing consumption today, the recipient becomes poorer tomorrow and thereby induces greater transfers from the donor.

Alternatively, the Samaritan's dilemma may be seen as a rent-seeking problem, where the transfers constitute the rent and the choice of consumption profile the rent-seeking action taken by the recipient. Either way, strategic behaviour by the recipient leads to a Pareto inferior outcome: both parties could be made better off if the donor could credibly commit to a transfer of a given size, see Lindbeck and Weibull (1988). Assume for instance that the donor could present a once and for all gift equal in size to the one obtained by the recipient through strategic behaviour. In this case, the recipient has no incentive to distort his consumption profile over time. A smoother consumption profile improves recipient welfare. In addition, since donor consumption is unchanged and recipient welfare is a public good, the welfare of the donor is also improved. Donor commitment, then, has lead to a Pareto improvement.

The Samaritan's dilemma shows how strategic behaviour by potential recipients crowds out the beneficial effects of aid. However, while presenting the recipient with a fait accompli may solve the incentive problems in the donor-recipient relationship, this donor

strategy might increase incentive problems on a lower level. Consider a rent-seeking economy as described in Pedersen (1995) where income distribution is determined by political competition amongst influential groups. As long as the political competition requires real resources, such as lobbying, the Nash equilibrium will involve economic waste, with income distribution determined by the distribution of political skills. Pedersen shows that foreign transfers of a given size, even if allocated to the poorer groups, will end up being distributed according to political influence, part of it being dissipated due to an increase in rent-seeking.[1] Given that the rent-seeking economy described above gives a fair picture of many aid recipient countries, commitment of aid offers no solution to inefficiency and inequality.

The present paper demonstrates that this pessimistic theoretical conclusion may be reversed by an appropriate design of transfers. In fact, foreign aid may eliminate rent-seeking and inequality in the model-economy. When aid is offered to a rent-seeking society, the structure of the problem changes character from the Samaritan's dilemma to the issues raised by the rotten kid theorem, see Becker (1974, 1976). While the Samaritan's dilemma is related to free riding incentives in contributions to a public good, the rotten kid theorem is related to the concept of negative externalities. Egoistic family members, i.e., the rotten kids, have access to actions which increase personal utility but reduce family income, due to for instance negative externalities associated with these actions. The rotten kid theorem states that transfers from an altruistic «father» may eliminate such asocial behaviour from the egoistic recipients. This family-setup is quite similar to the rent-seeking economy sketched above. To see this, let the political pressure groups represent the rotten kids and rent-

[1]Empirical evidence indicates that rent-seeking and rent dissipation are serious problems in the Third world. For instance, according to Mohammad og Whalley (1984), yearly rent-seeking losses in India are between 30% and 45% of GDP. For a survey of the rent-seeking litereature focussing on rent dissipation, see Nitzan (1993).

seeking their asocial behavior. The foreign donor plays the role of the altruistic father and society's income replaces family income.

The intuition behind the rotten kid theorem is rather simple. Equalization of welfare in a social group through altruistic transfers implies that each individual in that group receives a share of the social income. Individuals will therefore refrain from activities which harm social income. The rotten kid theorem is based on two important assumptions. First, transfers must be operational to all egoistic agents, which means that the aid budget must be of a certain minimum size. Second, the transfers should be offered in a discretionary manner. That is, transfers should be flexible enough to serve as a «punishment» for asocial behaviour, see Bruce and Waldman (1990) on altruism and time consistency. Note that a reduction in transfers is not a punishment in the ordinary sense of the word, since it is simply the optimal response from an altruistic donor to inequalities amongst the recipients. In the case of foreign aid, discretion in practice means implementing small-scale and short-term projects. Note the contrast with the central policy implication of donor commitment in the Samaritan's dilemma problem.

Applying Becker's theory to questions of foreign aid gives rise to two important problems of aggregation. First, the donor must be able to observe the welfare of the pressure groups. While the rotten kid theorem may work well in a nuclear family context where the altruistic father is relatively well informed about the well being of his siblings, information about the well being of recipient groups in a foreign country is clearly harder to come by for the donor. In order to make the information problems less acute, the donor agencies in this model are assumed to operate on a local basis, call it a «regional» basis. For many kinds of aid projects and organizations, perhaps particularly for NGOs, this seems to be a realistic assumption. Each region consists of a number of communities which form effective political pressure groups, perhaps under the leadership of a tribal chieftain. Through geographical proximity, the donor agencies are assumed to be able to

observe the standard of living of a tribe. Since utility is a function of consumption alone and perfectly observable to the donor, the problems of transferable utility and asymmetric information stressed by Bergstrom (1989) in his critique of the rotten kid theorem are therefore assumed to be solved.

The second aggregation problem is that the donor must be able to target individual communities with its aid. In the family case, gifts can be bestowed upon each sibling. Aid, on the other hand, is frequently in the form of projects which benefit large sections of the economy, such as infrastructure improvements. Unless aid can be made group-specific, it will not be able to function as a punishment device for asocial behaviour, and the rotten kid theorem cannot be applied. In this paper, it will simply be assumed that the donors have access to small scale projects which produce excludable goods and services, making targeting of the welfare of communities feasible. This being the case, aid-projects can be made group-specific and be used to «tax» rent-seeking activities.

The model is based on Pedersen (1995), but differs from it in two important respects. First, the present study allows for a number of politically autonomous regions. Second, and most importantly, it focuses on the timing of aid relative to the rent-seeking activities of recipient groups, while Pedersen only considers the commitment case. The paper is organized as follows. Section 1 presents the rent-seeking model without aid and without political interaction between the regions, call it the autarky scenario. Section 2 introduces aid, 2.1 discussing the consequences of aid given under commitment, and 2.2 aid given under discretion. Extending the model, section 3 relaxes the autarky assumption and introduces interactions between donors and regions. Section 4 concludes.

1. The autarky model without aid

The recipient country consists of a number of regions, each inhabited by N groups, or tribes, denoted by subscript $i = 1,2,\ldots,N$. The initial

endowment of resources in a region is assumed to be fixed and given by $Y = \sum_{i=1}^{N} Y_i$. There is no migration between regions, which may be due to for instance cultural barriers or difficulties in gaining access to scarce land in another region. There might however be trade between the regions, in which case Y includes the gains from trade.

The regions are headed by local governments responsible for the implementation of economic policies. As argued by Grindle (1991), the implementation stage plays a particularly important role in poor countries, where policy making often is a very closed process. Due to the inability or, for the sake of political stability, unwillingness of the central government to control local governments, economic policy in these countries to a large extent is determined on a local level. In the following, the central government will simply be left out of the analysis, allowing us to focus on the more important interaction between local pressure groups and local governments.

Define autarky as the situation where the authority of each local government is limited to a single region. For simplicity, economic policy in this model deals only with questions of income distribution. The local governments are assumed to have access to lump-sum instruments of taxation, which in the autarky case implies that for each region $\sum_{i=1}^{N} T_i = 0$. In practice, redistribution should be interpreted as a broad range of policies where the costs are borne by the region as a whole but where the benefits accrue to a limited number of groups, examples being allocation of credit and land, or public sector employment. Of course, these policies might themselves be distortionary. The assumption of lump sum redistribution of income abstracts from such costs and is therefore likely to underrate the wasteful effects of rent-seeking. On the other hand, the model may exaggerate the effects of rent-seeking on income distribution. With the availability of lump sum redistribution, the tax base and therefore the rent (R) in this scenario consists of the entire initial endowment

within regions, i.e., $R = Y$.[2] As will become evident below, this means that income distribution is uniquely determined by political strength. Introducing limitations in the tax base due to for instance administrative constraints would modify this result, but not alter the qualitative results of the model.

For redistributive policies to be meaningful in the present context, local governments cannot be indifferent to income distribution between important groups in its region. For instance, inequality may trigger social unrest which in turn threatens the political survival of the incumbent local leadership. The local politicians therefore act according to a welfare function which exhibits decreasing marginal rate of substitution between any pair of consumption levels. For simplicity, the welfare function is of the following log-linear kind:[3]

$$W = \gamma_1 \ln(C_1) + \gamma_2 \ln(C_2), \ldots, \gamma_N \ln(C_N) \qquad (1)$$

where C_i is consumption and γ_i the welfare weight of group i, with $\sum_{i=1}^{N} \gamma_i = 1$ in each region. The groups are assumed to be equally large, so that C_i can be interpreted as group consumption. In addition, we abstract from income inequalities within groups. Taking account of differences in group size and within group inequalities would however not affect the main insights from the model. The above welfare function does not necessarily include the welfare of all groups in a

---

[2] An alternative formulation would be to define Y as an exogenously given budget at the disposal of the local government, to be allocated between the groups in the region.

[3]The log-linear function represents a monotone transformation of the Cobb-Douglas function. Hence, the elasticity of marginal utility is minus unity. Choosing a CES-formulation would not add much to the analysis. Generally, the lower the absolute value of the elasticity of marginal utility, the lower is income inequality in equilibrium, and vice versa.

region. Groups which are politically powerless are assigned a welfare weight equal to zero and therefore excluded from the competition for special treatment. In this model such marginalized groups are likely to achieve only a subsistence level of consumption.

The implementation of the tax and transfer policy implicitely defined by (1) is affected by local pressure groups through activities we shall call «lobbying» or «rent-seeking». Bureaucrats and local politicians may be motivated by both economic and social goals in their responsiveness to such activities. For instance, pressure groups may offer bribes or other favors in return for special treatment, and the ability and willingness to accomodate the interests of influential groups may boost the social prestige of policy implementors. Political influence is operationalized by endogenising the welfare weights, making them a function of rent-seeking. The influence functions determining the welfare weights are of logit form, similar to the probability functions in the seminal contribution of Tullock (1980). Let q_i measure the rent-seeking efficiency, or political skill, of group i. To simplify the algebra, assume constant returns to lobbying outlays, B. The influence functions can then be described as:

$$\gamma_i = \frac{q_i B_i}{\sum_{i=1}^{N} q_i B_i} \qquad \text{for} \qquad \sum_{i=1}^{N} q_i B_i \neq 0$$

and (2)

$$\gamma_1 = \gamma_2 = \ldots = \gamma_N \qquad \text{for} \qquad \sum_{i=1}^{N} q_i B_i = 0$$

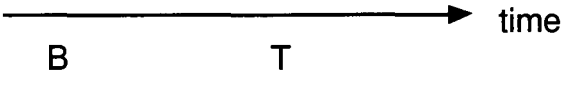
The larger is group i's share of total political influence, the larger will its welfare weight be. If no resources are invested in the political game, equal welfare weights will be assigned to the groups. In this model, then, any inequality is due to rent-seeking. In addition to lobbying in the traditional sense, rent-seeking may involve activities such as strikes, demonstrations and military buildup. These rent-seeking activities require real resources and therefore involve economic waste.

It would be straightforward to model these activities as bribes. In that case, of course, there would be no rent-dissipation, only a transfer of income from the groups to the local politicians. Such transfers may however be undesirable for society for purely ethical reasons, or for instance because bribe-money frequently ends up in foreign bank accounts. In the absence of aid, the budget constraint of group i is:

$$C_i = Y_i - T_i - B_i \tag{3}$$

In this static model, each group seeks to maximize its consumption, with rent-seeking as the only control variable. Problems of collective action within the groups are not considered. The groups move simultaneously, which implies that political influence is determined as a Nash equilibrium. The sequence of moves is illustrated in the figure below.

Figure 1.



First, the groups simultaneously and non-cooperatively choose levels of lobbying. Second, the tax policy is implemented. Government tax policy is found by substituting (3) into (1) and maximizing with respect to T_i, taking the welfare weights as given. The first order conditions are

$$T_i = Y_i - B_i - \gamma_i(R - B) \tag{4}$$

Inserting (4) into (3), the tax policy implies that each group ends up consuming a share of regional income, net of rent-seeking outlays, equal to its welfare weight:

$$C_i = \gamma_i(R - B) \tag{5}$$

The groups know the government's tax rule and understand the influence function (2). Maximizing (5) with respect to their rent-seeking outlays, Nash equilibrium is characterized by:

$$B_i = (1-\gamma_i)(R-B) \tag{6}$$

which implicitely defines the reaction functions of the groups. From (6) it is easily seen that total rent-seeking in a region is given by:

$$B = \frac{(N-1)R}{N} \tag{7}$$

where the region's resource endowment is the rent and $(N-1)/N$ is the degree of rent dissipation. Clearly, rent dissipation is an increasing function of the number of players. In the limiting case, when the number of players is sufficiently large, complete rent dissipation takes place.[4] Using (3)-(7), taxation of group i equals:

$$T_i = Y_i - \frac{R}{N} \tag{8}$$

which implies that the region's resource endowment is equally distributed between the groups. Since the extent of rent-seeking is inversely related to political skill, however, consumption across groups will vary. Plugging (6) into (2), the relative consumption and rent-seeking levels between any two groups within a region can be expressed as:

---

[4] It is reasonable to assume that there exists a subsistence level of consumption which defines an upper limit to the extent of rent-seeking and hence the number of groups involved in this game.

$$\frac{\frac{\gamma_j}{1-\gamma_j}}{\frac{\gamma_k}{1-\gamma_k}} = \frac{q_j}{q_k} \qquad (9)$$

Evidently, $q_j > q_k \Rightarrow \gamma_j > \gamma_k$ which from (5) implies $C_j > C_k$ and from (6) $B_j < B_k$, and vice versa. In other words, the politically more skilful group will enjoy a higher level of consumption and spend less resources on rent-seeking than the politically weaker group. Intuitively, the tax policy of the government implies that part of the rent-seeking outlays of a group will be paid for by rival groups. In a sense, such expenses become «tax deductible». This makes it profitable for the weak groups to compensate the low «quality» of their lobbying with greater quantity. This reasoning also accounts for the fact that rent dissipation in this model is independent of the distribution of political skill, a result which should be contrasted with Tullock (1980) where the extent of rent dissipation is a decreasing function of the degree of asymmetry between the contestants.

2. Autarky and aid

In the autarky scenario, each donor agency, like the local government, operates in only one region. In other words, their aid budgets are region specific. For simplicity, assume there is only one donor in each region, an assumption which will be relaxed in section 3.1. The donors are assumed to share the government's distaste for inequality, but they are immune to political pressure from the recipient groups. The objective function of a donor is therefore identical to (1) with $\gamma_1 = \gamma_2 = \ldots = \gamma_N$. This means that aid should be allocated according to the Rawlsian maximin principle, priority always being placed on the poorer group. The aid budget of each donor agency, $A = \sum_{i=1}^{N} A_i$, is fixed, which means that Samaritan's dilemma-type problems do not arise. In addition, $A_i \geq 0$ for all i. In other words, donors are not able to tax groups in the recipient country directly. Note however that a reduction

in transfers to a specific group may have the same effect as an increase in the taxation of that group. This fact is at the core of the rotten kid theorem, derived below. The budget constraint of group i including aid is:

$$C_i = Y_i - T_i - B_i + A_i \tag{10}$$

As noted above, the aid allocation rule is given by the objective function of the donor. Aid policy therefore reduces to a question of timing: Should the donor seek to equalize consumption levels ex ante or ex post relative to local redistribution? In the «commitment»-scenario, aid allocation takes place prior to the rent-seeking and policy implementation phase, while in the «discretion»-scenario, the allocation of aid is made after the implementation of local redistribution. The sequence of moves in the two scenarios is illustrated in figures 2 and 3.

Figure 2. Commitment

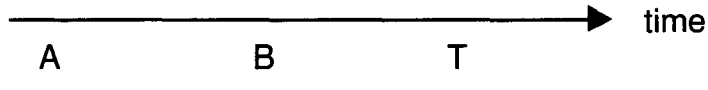
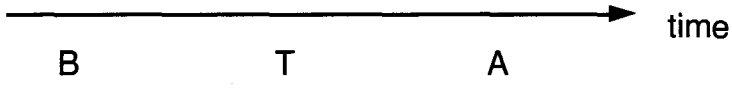


Figure 3. Discretion



Although not modelled explicitely, we can think of the difference in timing as caused by rigidity in the political system, due to for instance slow responsiveness of the pressure groups to aid inflows. If it takes time for the groups to organize their lobbying efforts, the donor agencies may be able to «surprise» the groups through the

implementation of short term projects. This is the discretion case. Long term aid projects, on the other hand, are more likely to be subject to rent-seeking. The pressure groups here have time to observe the projects and organize their lobbying efforts in order to gain a share of the surplus generated by the projects. This is the commitment case.

2.1 Commitment

In this scenario, aid is commited to a project for a number of years. Taking account of this type of aid in the rent-seeking model above is straightforward, since such projects simply add to the rent in a region, i.e., $R = Y + A$. The results regarding consumption, rent-seeking and rent dissipation carry through, while the equilibrium taxation of group i needs a slight modification:

$$T_i = Y_i + A_i - \frac{Y + A}{N} \tag{11}$$

Two things should be noted with respect to aid in this scenario. First, by increasing R aid increases rent-seeking, as is evident from (6). In fact, (7) shows that more than half of the aid budget will be dissipated, the degree of rent dissipation increasing in the number of pressure groups. Second, (11) demonstrates that the initial distribution of aid between groups does not matter. For a given aid budget, any increase in aid to one group will be perfectly crowded out by an increase in taxation. At the end of the day income distribution is uniquely determined by political influence. This scenario of donor commitment therefore cannot be said to achieve the main goals of aid, namely to foster efficiency and equality.

2.2 Discretion

Short term aid projects may be flexible enough to avoid rent-seeking, i.e., $R = Y$. This means that in the discretion scenario the donors are able to target specific groups with their aid, irrespective of the groups'

political strength. The objective function of group i in (5) should therefore be reformulated as:

$$C_i = \gamma_i(R - B) + A_i \tag{12}$$

Assume for now that the aid budget is sufficiently large so that all groups within a region always qualify for aid transfers. According to the maximin aid-allocation rule, this implies that consumption within a region is equalized. The aid allocation to group i which equalizes ex post regional consumption is:

$$A_i = \frac{R - B + A}{N} - \gamma_i(R - B) \tag{13}$$

which inserted into (12) yields

$$C_i = \frac{R - B + A}{N} \tag{14}$$

Equations (13) and (14) state the almost trivial result that when aid compensates for any difference in income distribution caused by political influence, each group ends up consuming equal shares of regional income. The implications of this result are however far from unimportant. Indeed, ex post equalization of consumption through discretionary aid policy has eliminated the incentives for rent-seeking. Since the only effect of rent-seeking is to reduce regional income which of course is harmful to each group, $B = 0$ in (13) and (14). By eliminating rent-seeking and equalizing regional consumption across groups, aid given under discretion has been extremely successful in fulfilling its goals. This is an application of the rotten kid theorem on foreign aid.

As noted earlier, the rotten kid theorem rests on the assumption that all groups qualify for aid. Clearly, this requirement need not always hold. Consider the figures below which summarize the

discussion so far and illustrate the consequences of a «too small» aid budget.

Figure 4a. No aid

		G 2 P	G 2 A
G 1	P	6, 6	0, 8
	A	8, 0	3, 3

Figure 4b. Commitment, aid=8

		G 2 P	G 2 A
G 1	P	10, 10	0, 16
	A	16, 0	5, 5

Figure 4c. Discretion, aid=8

		G 2 P	G 2 A
G 1	P	10, 10	8, 8
	A	8, 8	7, 7

Figure 4d. Discretion, aid=5

		G 2 P	G 2 A
G 1	P	8.5, 8.5	5, 8
	A	8, 5	5.5, 5.5

Notes: G 1 = Group 1, G 2 = Group 2, P = Passive, A = Active

Figure 4a presents the rent-seeking game without aid for a region consisting of two groups, 1 and 2, each with a strategy space {P, A} in terms of lobbying. Regional resources equal 12 units. If the groups play {P, P}, i.e., refrain from lobbying, they share the regional resources equally and there is no rent dissipation. Assuming for simplicity that the groups are equally skillful in lobbying, the {A, A} case leads again to a sharing of resources, but in accordance with (7) there is a fifty percent rent dissipation. In the cases where only one party is involved in lobbying, the active party gains while the passive party looses everything. The dominant strategy of both players is {A, A}, and the Nash equilibrium is therefore of the Prisoner's dilemma type.

In figure 4b eight units of aid are given under commitment. This adds to the rent in the economy and does not change the outcome of the game. The dominant strategy is still {A, A}. Figure 4c presents the case where the eight units of aid are given under discretion. In this case, aid always befalls the poorer group and is shared equally when the groups are equally well off. Since this aid budget is sufficiently large to

equalize consumption for all combinations of the players' strategies, the outcome of the game changes from a Prisoner's dilemma to {P, P}. In figure 4d aid is still allocated in a discretionary way, but the aid budget is now reduced to only five units which is not sufficient to equalize consumption levels for all strategy combinations. There are two Nash equilibria in pure strategies in this game, {A, A} and {P, P}. Game theory does not provide an answer to which of the equilibria the players will coordinate on, see for instance Fudenberg and Tirole (1993). An even smaller aid budget of for instance three units would lead to {A, A} as the unique Nash equilibrium, which illustrates that a very small aid budget may not be sufficient to eliminate rent-seeking.

The more uneven is the distribution of political influence between the groups, the greater is the gap in the payoffs due to lobbying from the stronger party, and therefore the larger must the aid budget be in order to achieve ex post consumption equalization. If the aid budget cannot be increased so as to eliminate the rent-seeking equilibrium, political reform which somehow strengthens the political influence of the weaker groups should be encouraged before offering aid. In other words, when political influence between groups is too disparate relative to the aid budget, political reform should preceed aid. The donors can contribute to improving the political skill of the poorer groups by measures such as literacy campaigns, education, primary health services, free press, and so on. A more equal distribution of political skill reduces the size of the aid budget required for the rotten kid theorem to hold. Note however that even if the aid budget is too small to eliminate rent-seeking and political reform is a lengthy and costly process, discretion in this model is always superior to commitment as a donor strategy. This is because small scale aid projects are less likely to be subject to rent-seeking and more likely to have a positive impact on equality.

3. Extensions

3.1 More than one donor in a region

So far there has been only one donor offering aid to the recipient region. What happens if there are more than one donor? If all donors care about all groups in a region, nothing new is added to the model since only the total aid budget is increased.[5] It is possible to show, however, that for the rotten kid theorem to hold, universal altruism is not needed. What is required is overlapping preferences, in the sense that at least one group's welfare enters the objective function of more than one donor. For concreteness, assume that there are three groups, 1, 2 and 3, and two donors, a and b. Donor a cares about groups 1 and 2, and donor b cares about groups 2 and 3. The poverty orientation of aid implies that each donor seeks to equalize the consumption of the groups he or she cares about. We know that for the rotten kid theorem to hold, the aid budget of each donor must be large enough to equalize the consumption level of its recipient groups. But this implies an equalization of consumption between all groups. The proof is straightforward. Donor a achieves $C_1 = C_2$, and donor b achieves $C_2 = C_3$, which together implies that $C_1 = C_2 = C_3$. Since consumption is equalized across the board, all groups will refrain from taking part in the competition for political favors. The important point to stress here is that this first-best outcome can be achieved even though the groups belong to different «families», to use Becker's terminology. If the donors have overlapping preferences, and their aid budgets are large enough to create equalization of consumption in each subgroup-family, the rotten kids will act as if they themselves were altruists.

3.2 Interregional transfers

This section deals with the problems that arise when donor altruism is not overlapping. Imagine a situation where in autarky the donor in

---

[5] Note that free riding on the donor level is not an issue here since aid budgets are assumed to be exogenous.

each region is able to equalize consumption between the groups and thereby eliminate rent-seeking and inequality. Now let there be a reform which merges the political authority of some regions. Call the merged regions «states», and let economic policy be defined on a state basis. The autarky assumption is now relaxed since transfers of resources between regions within the state may take place. As long as the donor agencies operating in separate regions do not merge, however, this political reform may lead to economic waste and inequality. Consider a state consisting of two regions, α and β. In the autarky case consumption is assumed to be equalized by donors a and b. Let C_α represent consumption of a group in region α and C_β ditto in region β. With subscript α and β denoting regional affiliation and $T_{\alpha\beta}$ representing transfers of resources from region α to β, and similarly $T_{\beta\alpha}$ from β to α, (14) must now be rewritten as:

$$C_\alpha = \frac{R_\alpha - B_\alpha + A_\alpha - T_{\alpha\beta}}{N_\alpha}$$

$$C_\beta = \frac{R_\beta - B_\beta + A_\beta - T_{\beta\alpha}}{N_\beta} \qquad (15)$$

The first order conditions are:

$$\frac{\partial C_\alpha}{\partial B_\alpha} = 0 \Rightarrow -\frac{\partial T_{\alpha\beta}}{\partial B_\alpha} = 1$$

$$\frac{\partial C_\beta}{\partial B_\beta} = 0 \Rightarrow -\frac{\partial T_{\beta\alpha}}{\partial B_\beta} = 1 \qquad (16)$$

which simply state that lobbying is worth while as long as the increase in transfers to a region, or reduction in taxation, exceeds the marginal rent-seeking costs, equal to unity. This is no different from the rent-seeking game with commited aid presented in section 2.1, and the

results from that analysis will carry through. With two regions in the state, half of the state's access to resources, including aid, will be dissipated. Differences in consumption levels in the two regions will be fully accounted for by difference in political skill.

The policy implication for the donors is quite clear. If economic policy leads to a transfer of resources across regions, the donors responsible for these regions should also merge. Given that their aid budgets are sufficiently large and given their ability to implement the short term projects characterizing the discretion scenario, the merged donor agencies will be able to implement the rotten kid theorem. Intuitively, interregional transfers of resources due to rent-seeking must be offset by interregional transfers of aid punishing such wasteful activities. If regions are tied together by both economic policy and foreign aid, these regions in practice constitute a single region, and the analysis from 2.2 can be directly applied.

4. Conclusions

This rent-seeking model shows that aid efficiency may depend in a rather dramatic way on the characteristics of the transfer. Since part of the value added from long term development projects will be dissipated and the remainder allocated according to political strength, aid efficiency can be said to be low. Short-term and small-scale projects, however, have a much higher chance of success. This is because flexibility in aid projects may lead to the results implied by the rotten kid theorem, eliminating the problems of rent-seeking and inequality.

Although the rotten kid theorem thus applied to the area of international aid transfers produces some very optimistic theoretical results with respect to aid efficiency, the results should be interpreted as tendencies rather than accurate descriptions or predictions of reality. As stressed in the introduction, problems of asymmetric information between donor and pressure groups may be large and the donor may lack the ability to design projects which target the welfare of individual groups. In addition, the aid budget must be sufficiently large to create

an interior solution. The idea that the donor agencies by frequent reoptimization of their project portefolio are able to punish groups aquiring resources through rent-seeking does however seem plausible. Moreover, even if rent-seeking is not entirely eliminated, both economic waste and inequality is likely to be smaller when aid is offered under discretion rather than commitment. Aid efficiency in the discretion case can therefore be expected to be relatively high. In addition to stressing discretion, the above mentioned problems of aggregation and insufficient aid budgets suggest two policy implications for donors dealing with a rent-seeking society. First, the donor agencies should operate on a local basis in order to mitigate the information problems. Second, if the aid budget is not sufficiently large and cannot be increased, aid should be preceeded by a political reform which bridges the gap in political skill between the groups.

Further research needs to be done in several areas related to this article. On a theoretical level very little has been done on the incentive effects of foreign aid. Indeed, the surveys dealing with the consequences of aid cited in the introduction do not mention it as a problem at all. Clearly, there is also a need for further theoretical contributions to our understanding of how government works. The influence function approach popular in the rent-seeking literature and used in this paper are convenient shortcuts for describing very complicated processes. More detailed descriptions of the interaction between legislators and pressure groups or civil society in general may produce important new insights. On an empirical level, credible assessments of the consequences of aid are obviously called for. Of particular importance to this paper is whether small scale projects have performed better than large scale projects. Moreover, to the degree that NGOs posess superior information about the conditions in recipient countries relative to government controlled donor agencies, NGOs are in a better position to implement the rotten kid theorem and realize a higher aid efficiency.

References

Becker, G.S. (1974). «A theory of social interactions,» *Journal of Political Economy* 82 6: 1063-1093

Becker, G.S. (1976). «Altruism, egoism, and genetic fitness: Economics and sociobiology,» *Journal of Economic Literature*: 817-826

Bergstrom, T. C. (1989). «A fresh look at the rotten kid theoem,» *Journal of Political Economy* 97: 1138-1159

Boone, P. (1995). «Politics and the effectiveness of aid,» CEP-Discussion Paper no. 272, Center for Economic Performance, London

Bruce, N. and M.Waldman (1990). «The rotten kid theorem meets the Samaritan's dilemma,» *Quarterly Journal of Economics*: 155-165

Buchanan, J.M. (1975). «The Samaritan's dilemma,» pp. 71-85 in Phelps, E. S., ed. *Altruism, morality and economic theory*, Russel Sage Foundation

Buchanan, J. M. et al. (1980). «Toward a theory of the rent-seeking society,» College Station: Texas A&M University Press

Cassen. R. et al. (1994). *Does aid work?*, Second editin, Oxford, Clarendon Press

Fudenberg, D. and J. Tirole (1993). *Game theory*, The MIT Press

Grindle, M. S. (1991). «The new political economy: Positive economics and negative politics,» in Meier, G.M., ed., (1991). *Politics and policy making in developing countries*, ICS Press, San Fransisco

Krueger, A. O. et al., eds. (1989). *Aid and development*, Baltimore and London, John Hopkins University Press

Lindbeck, A. and J.W. Weibull (1988). «Altruism and time consistency: The economics of fait accompli,» *Journal of Political Economy* 96 4: 1165-1182

Mohammad, S. and J. Whalley (1984). «Rent-seeking in India: Its costs and policy significance,» *Kyklos* 37: 387-413

Mosley, P. et al. (1987). «Aid, the public sector and the markets in Less Developed Countries,» *The Economic Journal* 97: 616-641

Nitzan, S. (1993). «Modelling rent-seeking contests,» *European Journal of Political Economy* 10: 41-60

Pedersen, K.R. (1995). «Rent-seeking, political influence and inequality - a simple analytical example,» *Public Choice* 82: 281-305

Tullock, G. (1980). «Efficient rent-seeking,» in Buchanan et al. (1980), op. cit.