BOARD OF CONTRIBUTORS COLUMN

The Journal's Board of Contributors is composed of five distinguished scholars who periodically contribute commentaries or articles on a wide range of subjects which are of timely interest to regional and urban specialists.

REGIONAL GROWTH REVISITED

HARRY W. RICHARDSON*

I

About two years ago I wrote a book titled *Regional Growth Theory*[1] (RGT for short) which has aroused some strong reactions, both for and against. This paper provides me with a good opportunity to take a critical look at the analysis, and to see whether it can be taken any further. There are several pretexts for such introspection. First, some of the reviewers misunderstood the argument, and to answer them would clear up some important points. Second, the mystery of the regional growth process is such a vital question for spatial analysts that it is hardly necessary to justify giving more time to it. Third, I have cherished—and still cherish—a wild ambition to develop a theoretical framework for spatial growth analysis that is independent of level of spatial disaggregation, stage of economic development and institutional environment. Realistically, this goal is likely to remain frustrated. Nevertheless, there are at least two directions worth a little exploration. One is to extend the analysis downwards to the intraurban level, touched upon but not developed in RGT. Such a step might help to reconcile theories of regional development with urban development, and reinforce the claim that this is one field rather than two. The other is to examine whether the model is capable of application to the developing world. The search for a general theory that can straddle the three worlds is a just and worthy crusade. Western economists must be rather tired of the third world jibe that they travel with capitalist models in their pockets. In any event, the problems of regional development are much more serious in the developing world than in either North America or Western Europe, and the argument that theory should be responsive to policy needs has much in its favour.

II

The bones of the argument in RGT can be summed up in a few sentences. The weaknesses of most of the work in regional growth theory are its neglect of space and distance, or their treatment in a trivial fashion, and its heavy reliance on the price mechanism, particularly for factors of production, as a spatial allocator. If space is introduced intraregional growth determinants become at least as important as the interregional. The spatial distribution of population and

*Professor of Regional and Urban Economics, University of Pittsburgh.

economic activity reflects the net impact of opposing sets of forces—those of concentration (agglomeration) and those of dispersion. Growth rates vary with location and over time because the relative strength of agglomeration and dispersion factors alters over space and intertemporally. Accordingly, a satisfactory theory of regional (spatial?) development must assign an important role to these factors. RGT stresses agglomeration economies, a multi-dimensional concept much broader than that of economies of scale, as the main force promoting concentration. Conversely, friction costs favour dispersion while immobile locational constants impose constraints upon agglomeration. Also, the emergence of diseconomies will tend to slow down the rate of agglomeration. Locational preferences are a significant non-economic but entirely rational variable. In the general case, locational preferences for households perpetuate patterns of dispersion while locational preferences for firms tend to reinforce agglomeration in the major metropolitan centres. Historically, the growth path for nations tends to conform to predictable spatial patterns: initial polarisation in one or two regions that becomes cumulative in the most rapid industrialisation phase; subsequent diffusion of development into other regions; spatial concentration within each region in a limited set of major urban centres; and spatial readjustment within these urban areas, as economic and social change eventually promotes decentralisation. These spatial trends are comprehensible only in terms of variations in the relative strength and incidence of forces of agglomeration and dispersion. These ideas are made specific in the form of a regional growth model which is potentially operational but is not tested in RGT.

III

RGT has been criticised on several grounds.[3] They include:

(i) the study attacks a false version of the neoclassical model (a neoclassical 'straw man');

(ii) the locational variables used in RGT can be accommodated within the neoclassical framework;

(iii) the approach is eclectic;

(iv) the model is not tested;

(v) some of its variables are badly specified, cumbersome or difficult to understand;

(vi) the model is concerned with total rather than per capita growth;

(vii) the model overemphasises agglomeration economies and urban size;

(viii) the model lacks generality because, as specified in contrast to its verbal statement, it does not deal with the possibility of reversal of spatial trends, but applies best to the early development phase of initial concentration;

(ix) there is no acknowledgment of the fact that interest in regional economics in the United States has declined in recent years;

(x) the study skates over the problem of defining regions.

Some of these points are easily dealt with. In regard to (x), although the regional delimitation problem was discussed in RGT, the aim was to develop a spatial theory of development stressing agglomeration and dispersion tendencies that are independent of regional boundary definitions. (ix) is mere ethnocentrism, since interest is, if anything, growing worldwide, and the aim of any respectable

theory should be to shed light on a broader spectrum of institutional environments than that of the United States. As for (iii), eclecticism is an advantage in regional science since progress depends not only upon the thought of many individuals but on the contributions of many different disciplines. The reasons for (iv) were explained in detail in RGT, though as pointed out below (see pp. 9-11) the original stand has been modified. Since the model contained the rate of growth of population (labour supply) as a variable, (vi) is easily accommodated via a mechanistic reformulation. I confess to (v): the capital stock density variable is a clumsy and probably non-operational method of taking account of the intraregional spatial distribution of investment, and a substitute needs to be found; locational preferences are measured better by (100 per cent *minus* the percentage of people born in the region but living in other regions) as suggested in RGT (p. 200) and adopted in the empirical analysis discussed below rather than by a length of residence term as used in the RGT model. Criticism (viii) is important. The objective of RGT was to explain the *variation* in the relative strength of agglomeration and dispersion forces over space *and* time. Insights were offered on this point in the verbal explanation in RGT, but the analysis was not carried over into the formal model. Much more work needs to be done. Of course, introducing the spatial dimension makes dynamic problems much more complex. The critical question is how to identify precisely when and why the process of spatial polarisation in the national economy gives way to one of interregional diffusion.

Since (i) and (ii) are related criticisms and can be treated together, this leaves two: (i + ii) and (vii). Whether RGT gives too much attention to agglomeration economies (vii) can be resolved only by empirical testing. This is not an immediate answer in view of the obstacles in the way of measurement of agglomeration economies. The equation of agglomeration economies with size is to misunderstand the arguments of RGT. Agglomeration economies refer to economies that induce people and activities to cluster together rather than to the economies *resulting from* agglomeration. Although some agglomeration economies are size-related, e. g. the provision of high-level urban services that require a large population threshold, it is more accurate to regard increasing size as the result rather than the cause of agglomeration economies. Several important externalities, both positive and negative, are not size-related, e.g. climate, social mix, industrial pollution.[3] Also, the *net* agglomeration pull reflects the outcome of several, identifiable agglomeration economies. Even if each could be assumed a function of size, the mix will vary among locations in a manner that is independent of size. Furthermore, this mix will include diseconomies which are even less a function of size than economies. In addition, the distance between centres is an important element in agglomeration economies, a factor made explicit by von Böventer (1975) in his division of total agglomeration effects into *intra*city and *inter*city components, the latter measuring spillover agglomeration economies due to proximity to other large centres. This refinement allows the prediction that the minimum critical size for a centre from the growth potential point of view is lower for cities close to bigger agglomerations than for cities located in sparsely populated regions.

One aspect of (vii) which is closely linked with criticisms made under (ii) is whether RGT makes a false distinction between agglomeration economies and factor returns, and at the same time overemphasises the former. This criticism has

been made from two extreme standpoints. One is to argue that 'the chain of causation runs from innovations and market sizes and high productivities to price and return differentials' and that empirically there is so much multicollinearity between agglomeration economies and factor returns that the latter should be dropped from a regional growth model. The other is to suggest that agglomeration economies are merely one element in the profitability variable considered by the profit-maximising location decisionmaker. This is consistent with the traditional neoclassical view that externalities are minor irritations that foul up the price mechanism, but because they are minor can be kept under control. RGT included both agglomeration economies *and* factor returns within the model. This may have been a mistake, indicating a lack of courage in one's convictions and/or an underestimation of the identification problem. However, there is an analytical distinction between 'agglomeration economies' and related concepts such as 'economies of scale' or 'externalities'. The spatial element is crucial. The concept of agglomeration economies is complex and multidimensional, referring to all sets of forces that induce households, firms and public agencies to locate near to each other in space. This is *the* plot of regional economics, not a footnote in the last chapter.

RGT has been most heavily criticised because it dared to attack the dominant ideology of the economics profession—the neoclassical framework. One strand in the argument is that RGT concentrates on a naive version of the neoclassical model. Another is that the neoclassical model is a general model capable of predicting divergence as well as convergence. The third and most serious is that neoclassicism is so flexible that it can cope with, and perhaps be enriched by, all the variables and hypotheses in RGT. Most of these issues are matters of judgment. If a naive neoclassical model was used, it is because this has dominated the literature. Also, care was taken to describe it as 'naive' and 'crude'. The generality of the neoclassical theory was admitted (the most important modification is to introduce multiple sectors with an export commodity), just as in cumulative causation theory the dominance of spread over backwash effects can predict convergence. Alternative theoretical frameworks very often represent not dissimilar ways of looking at the same phenomena. As assumptions are relaxed and refinements introduced, it becomes a question of taste and judgment as to whether it is better to stick with the original theory or to substitute a more appropriate if less palatable alternative.

I reject neoclassicism as a basis for explaining regional growth not because it is impossible to introduce sufficient adaptations and qualifications to allow a neoclassical theory to provide an approximate simulation of the process of regional development but because the neoclassical approach has certain implications which conflict with my perception of the determinants of spatial distributions. More specifically, objections to the neoclassical theories are: their reliance on the price mechanism as a spatial allocator of resources; emphasis of marginal adjustments, whereas spatial functions are discontinuous and locational changes usually mean inertia (no change) or a long-distance jump; the neglect, or at best trivial treatment, of space (e.g. spatial diffusion theory); the assumption that growth can be constructively analysed, even at the abstract level, in terms of an aggregate production function and a homogeneous capital stock; a predilection for, if not exclusive emphasis on, equilibrium solutions; a greater facility with

deterministic rather than probabilistic models; and the policy inference that regional inequalities can be satisfactorily dealt with by reducing market imperfections and by adjusting the price system via taxes and subsidies rather than by infrastructure strategies and comprehensive regional planning. It is possible that some defence can be constructed against some of these charges. But can they all be refuted? If not, in my judgment, some alternative theoretical framework must be developed. RGT was a modest first step in that direction.

The absorptive capacity of the neoclassical model can be exaggerated. Locational preferences are not merely a constraint on income (or profit) maximisation, but a contradiction of it. Agglomeration pull is not reflected to any major degree in prices, since many of the important agglomeration economies cannot even be quantified, not to mention monetised.[4] Spatial frictions cannot be comprehended solely or even primarily in terms of transport costs. Locational constants have a powerful influence on moulding the spatial structure of the economy, but their survival and subsequent growth owes far more to political pressures and social habits than to economic logic. The predictive capacity of the neoclassical model to cover the divergence case depends on introducing intermittent exogenous disturbances rather than on the internal mechanics of the model itself. These mechanics rely on negative feedbacks (e.g. rapid capital accumulation depressing capital yields which slow down the rate of capital accumulation) rather than the positive feedbacks that reinforce the circular spiral of a typical divergence model such as cumulative causation theory. Finally, the evolution of the space economy relies far more on chance and expectations, risk aversion, herd instincts (i.e. in the form of interdependent location decisions), locational inertia and on large-scale, indivisible and irreversible investment decisions than on *economically* rational responses to changes in the relative price structure. If all these factors can be embraced by the neoclassical framework, this is a very different form of neoclassicism from that taught in the classroom or written about in the journals.

IV

RGT expresses a purist fear of testing. Since then, I have overcome this quite justified reluctance and carried out some fairly crude empirical analysis (Richardson, 1974a). This required burying two nasty qualms. First, it was necessary to assume that worthwhile results could be obtained by tests of regional growth hypotheses at the *state* level. No regional scientist would maintain that the state was an ideal economic region, but it has a unity because certain policies apply over its territory and its residents are bound together by common concerns. Also, most of the data are available only at the state level. Second, the absence of capital stock and capital yield data meant running the risk of incorrectly specified models. This is a major blow since whether high capital yields occur in rich or poor regions is at the heart of most controversies about regional growth.

Gross State Product (GSP) was the growth measure used.[5] Tests using quasi-neoclassical variables alone yielded poor results. Agglomeration economy variables performed much better, but were still not satisfactory; some variables lacked significance, and an agglomeration economies equation explained only 50 per cent of the variance in state growth rates (see Richardson, 1974a, for more

details). Much better results were obtained by stepwise regression procedures, which in effect implied testing a 'hybrid' model. Each variable in the equations could be justified in terms of theory or, at least, on the basis of plausible and rational propostions about the determinants of regional growth. The equation shown in Table 1 contains some interesting results.

Table 1

DETERMINANTS OF RATE OF GROWTH IN GROSS STATE PRODUCT, 1955-64

Growth rate of gross state product, 1955-64 (% per annum) = 15.59 + .0172*
(2.10)

Urbanisation − .2158* Unemployment − .0015** Wage − .0034** Per Capita
(2.31) (3.47) (5.03)

Income − .0200** Population Density + .0814** Federal Spending + .2410**
(5.04) (3.22) (9.57)

Tourist and Recreation Expenditures − .4001** Regional Growth − .0028*
(2.95) (2.29)

National Urban Hierarchy Rank− .6153** Growth Rate of Income + .0080**
(3.14) (4.92)

Income Density + .0576** Locational Preferences+ .0676** Migration Attraction.
(4.69) (8.81)

R^2 = 0.9185 d.f. = 34
** = significant at 0.01 level
* = significant at 0.05 level
t-values in parentheses

Perhaps surprisingly, there was little intercorrelation between unemployment, wages and income per capita. The traditional neoclassical wage-unemployment relationship does not appear to hold, presumably because of stickiness of wages and labour market imperfections. Wages and per capita income are not strongly related due to wide interstate differentials in the share of unearned income. There were no strong indications of multicollinearity elsewhere in the equation.

From the point of view of the RGT model, the results support the importance of agglomeration economy variables (urbanisation, income density) and of locational preferences. Also, the high degree of significance for tourist and recreation expenditures points to strong links between amenity resources and regional

growth. Moreover, differences in the spatial distribution of population and economic activity within states (e.g. urbanisation, population and income density, migration and locational preferences, the spatial contiguity of the 20 best predicted states) are associated with variations in state growth rates. This finding highlights the deficiencies of non-spatial models that have dominated much of regional growth analysis.

There are other implications for regional growth theory. The significance of the Federal spending variable suggests the need to include the public sector and public policy variables in regional growth models. Several successful variables (per capita income growth, locational preferences, and migration attraction) refer to past conditions in the state, and it is clear that current growth performance is not independent of historical influences. The inverse relationship between current growth and past income growth indicates that regional growth models should allow for the possibility of trend reversal. Considered as a whole, the findings suggest that the regional growth process is a much more complex and interrelated phenomenon than some of the simpler growth models imply.

Finally, two other results merit a brief comment. First, several variables show that the more backward states grew faster (wage, per capita income, past growth, population density, national urban hierarchy rank), strong evidence of convergence. Second, there was a negative relationship between the growth rate of a state and of its region. One inference might be that a reasonable size of region should be smaller than the Department of Commerce regions.

V

Despite its disparagement by some people as a crude empirical tool, the gravity model and other members of its family has considerable appeal as a framework for macrospatial theory. If appropriate exponential weights are applied to both the mass and distance variables, the model is responsive to variations in the relative strength of agglomeration and dispersion forces. Nevertheless, the gravity model loses much of its theoretical value by its reliance on univariate mass and distance measures and by burying the influence of complex polarisation and diffusion forces in an empirically derived exponent. The alternative technique of income potential mapping suffers from a different kind of aggregation by losing the influence of the individual agglomeration in a more or less continuous surface. To remedy these deficiencies we need a methodology that takes account of the multidimensional character of agglomeration economies (and diseconomies), the fact that the impact of each is reduced over space and to varying degrees, the complex nature of spatial frictions (also multidimensional), and of the possibility of measuring both the spatial range of an individual agglomeration and the relative pulls of several agglomerations on a particular point in space.

These criteria led me to develop the concept of *agglomeration potential* (Richardson, 1974b). The agglomeration potential of centre j on location i, ${}^{i}Z^{j}$, can be represented by the expression

$$ {}^{i}Z^{j} = f\left[\frac{\sum_{n} A_{n}^{j}}{\sum_{x} d_{ij_x}}\right] \qquad \text{additive form} $$

or by

$$ {}^{i}Z^{j} = f\left[\frac{\pi_{n} A_{n}^{j}}{\pi_{x} d_{ij_x}}\right] \qquad \text{multiplicative form} $$

where A_n^j is the value of agglomeration economy n in centre j and d_{ij_x} is the distance between i and j measured by variable x (e.g. transport costs, travel time, social distance). Thus, agglomeration potential depends upon disaggregation of the agglomeration function into 1, 2, ..., n agglomeration economy (or diseconomy) variables and of distance (or spatial frictions) into n+1, n+2, ..., x components. The choice between the multiplicative or additive version is determined by how certain important theoretical and empirical considerations are resolved.[6]

The most obvious applications of the agglomeration potential concept are in evaluating the relative pull of alternative centres j, k, l, ...etc. at i via estimates of ${}^{i}Z^{j}$, ${}^{i}Z^{k}$, ${}^{i}Z^{l}$, etc. or by measuring the overall agglomeration potential of centre j on surrounding space. If it could be assumed that spatial influence extended concentrically,[7] this would be given by $2\pi r_j \int^{r} Z^j(d)\,\partial d$ where r is that radius from j where $Z^j(d) = 0$. The concept is also useful in explanations of the macrospatial distribution, in household migration and plant relocation models, and in hierarchical diffusion and city size distribution theory.

To claim more generality, the agglomeration potential model would need extending in at least three directions. How far can it help to explain intrametropolitan agglomeration? Does the concept have any relevance for developing countries, and in particular can it shed light on the spatial distributions prevailing there? As stated, the concept deals only with variations in agglomeration and dispersion over space. To be helpful to regional growth analysis, it must be dynamicised to accommodate variations over time. How can this be done?

How to treat distance at the mass is a familiar but knotty problem in all potential models. Sometimes the mass is treated as a spaceless point; frequently, the mass is assigned an arbitrary radius over which potential is assumed constant; or the problem may be attacked via empirical estimation. The multidimensional character of both the agglomeration effect and distance in the agglomeration potential model permits a more sophisticated treatment. For instance, the inclusion of agglomeration diseconomies and the possible variation for each agglomeration component in distance-decay functions may help to explain changes in intraurban spatial structure. For example, many diseconomies (e.g. congestion) decay very rapidly with distance outside the CBD, but may have

disproportionate quantitative weight within the CBD itself. The result may be very low values for agglomeration potential within the CBD but much higher values in the suburbs or on the outskirts of metropolitan areas. In a general spatial model this hypothesis is consistent with the strength of decentralisation forces. Another relevant feature is that some elements of distance, of which travel time is perhaps the prime example, may be much more important within cities than outside them. The weight of travel time in the measurement of effective distance may help to account for the development of large intraurban clusters (subcentres) of like and unlike economic activities.

The primate city size distributions endemic in developing countries (especially Latin America and South East Asia) reflect very clearly the distribution of agglomeration potentials. The concentration of agglomeration economies in the national metropolis and other leading cities is so overwhelming that the polarisation of migration flows from rural areas to the primate city is inevitable. The presence of major diseconomies in these cities does little to repel migrants, partly because these diseconomies receive less weight in the agglomeration function, partly because information flows about 'public bads' decay much more rapidly over distance than information about jobs and opportunities. In those countries with secondary cities, these cities generate some agglomeration economies but their agglomeration potentials are low because of high spatial frictions: their hinterlands frequently lack an efficient transport network, uncertainty costs are high and their backward social spaces transmit information poorly. Nevertheless, in some cases these secondary cities are becoming more competitive in the attraction of migrants and industry as spatial frictions are reduced as a by-product of economic development and social change.[8] The resulting increasing concentration of people and economic activities has a cumulative impact on agglomeration potential, thereby attracting yet more migrants. This leads in to the third and most important modification of agglomeration potential theory— its extension to spatial dynamics.[9]

The reconciliation between spatial and intertemporal variations in the relative strength of agglomeration and dispersion tendencies via the agglomeration potential concept simultaneously achieves consistency between macrospatial theory and the locational behaviour of individual decisionmakers (households) and firms). Abstracting from interurban differentials in natural increase and changes in participation rates, it may be assumed that city size adjustment takes place primarily via the migration of households and firms. The distribution of migrant destinations from a specific location will reflect relative agglomeration potentials as perceived at that location. The results of these migration decisions, however, will be to alter these relative Zs because *some* components in the total agglomeration effect are functions of city size. At first sight, this might suggest cumulative migration trends in which the largest cities with the highest Zs consistently gain at the expense of other centres in the national urban system. This may happen, particularly in certain phases of the development process, but it is not inevitable. Other forces may come into play, such as pressure on urban housing and services, dampening effects on income and links between migration rates and metropolitan unemployment. These slow down the increase in agglomeration potential, partly because some agglomeration economies and diseconomies vary non-linearly with city size, partly because some are unrelated to size, and partly via changes in weights.

Furthermore, in developed countries the predominant pattern of migration flows is intermetropolitan (Alonso, 1971). Thus, some urban areas may lose migrants, affecting their agglomeration pull and changing their Z values. Although the typical pattern may be declining Zs in cities experiencing net outmigration and increasing Zs in net inmigration cities, this may not be universal for reasons already suggested. Moreover, agglomeration potentials may change quite suddenly and non-marginally as a result of a variety of influences— changes in the location of government-induced employment (whether direct in the form of government jobs or indirect via contracts to private industry), shifts in tastes e.g. the rise of resort cities, the strength of information flows— both accurate and inaccurate— and migration interdependence (i.e. the herd instinct), e.g. the postwar California trail.

Although these generalisations are woefully non-specific, they illustrate the simultaneous interdependence between spatial variations in agglomeration and dispersion forces and their time paths. Changes in the spatial distribution of population and economic activity induce changes in agglomeration potentials over time, and these changing agglomeration potentials induce further spatial movements. Also, it is not too difficult in a simultaneous model of this kind to explain how over time increasing spatial concentration may give way to a more diffused pattern of spatial development. The answer lies in changes in the structure and composition of the agglomeration potential function and its relative distribution. More specifically, these changes include: (i) an increase or reduction in the *number* of agglomeration or distance variables; (ii) non-linearities in how agglomeration economies (and diseconomies) vary with city size; (iii) changes in the values of agglomeration economies that are independent of size; (iv) changes in the agglomeration and/or distance weights due to changes in tastes, technology or social attitudes; (v) improvements (or within cities possibly deterioration) in transport connections, communications and information flows, cultural and linguistic differences and other elements affecting the degree of spatial friction.[10]

VI

Although the agglomeration potential approach was developed subsequently to the model in RGT, it is not clear to me whether it is superior or not. I suspect that it is more satisfying from a conceptual point of view, and that it may be easier to manipulate in formal models. On the other hand, the problems of implementation are immense, and the RGT model is much closer to being empirically tested. However, does it lack generality? Considering its key elements—agglomeration economies, locational preferences, locational constants, and measures of connectivity (transport and communications costs)—are these useful for the analysis of intraurban growth and for understanding spatial development trends in developing or socialist countries? Few would argue that the neoclassical models could pass these tests.[11]

The urban economics literature contains several analyses that examine how agglomeration and dispersion forces affect urban spatial structure. The standard treatment is to show how spatial residential equilibrium is obtained in a model where economies of scale of CBD production are the main force inducing agglomeration and the sole dispersing factor is transport costs narrowly defined as

commuting costs to the CBD.[12] Apart from the restrictive treatment of agglomeration economies and spatial frictions, the models abstract from spatial reality by retaining the familiar assumption of a one-dimensional city. As yet, this framework has not been used to analyse growth situations. Moreover, the assumption that agglomeration economies are created only in the CBD is so much at variance with the structure of modern cities that a more complex approach is needed. As one of the theorists of this school admits: 'the single production node is clearly unsatisfactory and what is really needed is a theory of subsidiary production nodes. It seems that real progress in developing the latter must await a stronger theory of the gains from spatial concentration' (Riley, 1974, p. 248).

The claim of RGT to provide the basis for a general spatial theory is reinforced by the fact that its key concepts assist an understanding of metropolitan spatial structure and, more important, shed light on the determinants and form of urban growth. The recognition that agglomeration economies are multidimensional, embracing centripetal influences far broader than production scale economies, and can be created at many locations—even intraurban locations since some economies have a spatial range shorter than that of the city as a whole—opens up possibilities for analysing multicentric cities[13] and for explaining shifts in relative agglomeration pulls between competing centres (especially between the CBD and suburban subcentres). These latter shifts have a major impact on patterns of urban growth, and may also affect the rate of aggregate urban growth since the peak agglomeration economies within the metropolitan area (wherever they occur) largely determine its success in interurban competition for mobile economic activities. Locational preferences also enter into the analysis, to some extent as a constraint on locational shifts out of the CBD for activities which have lost the economic rationale for a central location, but primarily as an explanation of changing residential location patterns and of the expansion (or decline and disappearance) of individual neighbourhoods.[14]

To specify more precisely how these variables affect intraurban growth requires a formal model. How can such a model be developed since the dominant theory is that of a monocentric city in which all jobs are located in the CBD surrounded by a residential ring?[15] This question is difficult to answer, but a possible solution is the development of discrete rather than continuous models in which the city is divided up into a number of areal sections, perhaps a minimum of fifty. MacKinnon (1974) has recently proposed a procedure for dealing theoretically with a city of this kind, though his approach has not yet demonstrated its capability of handling the reality of multiple employment centres. Nevertheless, expectations are high.

VII

Although the institutional environment assumed in RGT is that of the developed world (particularly the United States and Western Europe), the basic model is intended to be applicable to developing countries. The hypotheses examined in Chapter 5 under the title of 'Towards a General Theory of Spatial Development' are believed to be universal, and the analysis of the initial phase of spatial concentration there is based as much on theorising about present-day developing economies as on the historical experience of developed countries. There

are several *prima facie* grounds for justifying the relevance of the RGT approach for understanding spatial trends in the third world: the inclusion of political and social factors as important influences on changes in the spatial distribution of population and economic activities; the emphasis on agglomeration economies which is consistent with explanations of primate city size distributions; the role given to locational constants which is useful for accounting for the irregular spatial distributions (e.g. intensive coastal development) prevalent in many developing countries; the importance assigned to transportation and communication networks and to the evolution of the national urban hierarchy as determinants of the rate and pattern of spatial diffusion; and the stress on locational preferences, not so much as a constraint on migration from poor to rich regions as in developed countries but rather as a major factor keeping business firms and professional elites in the national metropolis and other leading cities.

The crucial significance of locational constants in moulding the long-run development of the spatial structure of the economy is particularly noticeable in many of today's developing countries. Drawing upon Friedmann's (1966) dictum that regional economic growth is externally induced, it is apparent that the current spatial imbalances in so many areas in the developing world (e.g. Latin America) can be explained primarily by the historical process of colonialism. The largest metropolis is very often coastal, and its growth was due to the fact that the country was originally opened up for development by colonialisation from overseas. The international functions of such a city may persist indefinitely, certainly for long after the passing of the colonialist phase. The repercussions affect interregional levels of development as well as the skewness of the city size distribution. This is because the scale and spatial spread of hinterland development are functions of city size.

If location constants determine initial spatial conditions and constrain future spatial patterns, the specifics of the subsequent development path require bringing agglomeration economies and locational preferences into the analysis. The cumulative character of the agglomeration economies created in the national metropolis and the weakness and sub-threshold levels of agglomeration economies in other cities are at the heart of explanations of the 'internal colonialist' structure found in so many developing countries today. The secondary cities lack countervailing power, and consequently labour and capital flows polarise towards the primate city. The theories of spatial domination needed to explain this phenomenon have an economic as well as political and social dimensions, but the concepts of agglomeration economies and agglomeration pull are more relevant than the equilibrium framework and marginal price adjustments of neoclassical models. The weak or non-existent data base, the higher risks of investing at untested locations and the differential market sizes of the centre and the periphery increase the level of uncertainty in location decisions in developing countries. In practice, uncertainty reinforces agglomeration (Webber, 1972). Whether the lower degree of uncertainty in big city locations is an agglomeration economy or whether it is better treated as a risk-aversion element in locational preferences is a semantic point, but in either case it downgrades the relevance of neoclassical theory. This is due not to the inability of neoclassical models to include risk and uncertainty variables (of course they can) but to the resulting destruction of their predictive power.

Locational preferences of business firms are too complex a phenomenon to be absorbed in a cost-benefit model of relocation that is consistent with the neoclassical approach. Some factors, such as the familiar overestimation of returns at the centre relative to the periphery (Friedmann, 1966; Hirschman, 1958; Alsonso, 1968), can be accommodated, but others cannot. For example, how does one attach a price to the need of some types of firms to be close to the centres of economic and administrative decisionmaking, in view of the indivisibilities in central government? Also, the psychic welfare gains that accrue to middle class business elites from metropolitan life styles bear little direct relation to the operational efficiency of the plants and enterprises that they manage.

Finally, few would quarrel with the hypothesis that in developing countries the spatial diffusion of technological and social change is a major determinant of the rate of development outside the core regions. RGT attempted to deal with this by stressing connectivity, the national urban hierarchy (especially the size of regional metropolises) and the social structure of underdeveloped regions. It is noticeable that, almost without exception, economists have treated the question of innovation diffusion without mentioning the frictions of space (e.g. Mansfield, 1968).

VIII

Empirical evidence from centrally planned economies supports the view that agglomeration economies and locational preferences strongly influence spatial growth rates.[16] This is in spite of Koropeckyj's argument that neoclassical theory is most appropriate for analysis of regional growth in centrally planned economies. This paradox is explained by the fact that central planners have more control over the location of factors of production, particularly capital—the mobility of which is very sticky in mixed economies. If potential returns really were higher in backward regions, central planners would be able to discount risks and uncertainty and invest there. In fact, to take the example of the Soviet Union, there is no evidence that the planners have behaved in this way. On the contrary, investment policy has strongly favoured the regions containing the major urban agglomerations.[17] Empirical studies have shown that the marginal productivity of capital in industry in the USSR's million-plus cities is more than double its level in towns of less than 50,000 (Holubnychy). Disparities in levels of economic development (e.g. gross industrial output per capita, capital per worker) and in welfare indicators (e.g. consumption) have widened between regions in the long run. Very recently, there has been some convergence, but it has been slow and intermittent.

Locational preferences also have an impact on spatial efficiency in the Soviet Union. In the less developed regions of the Central Asian republics, for example, attempts to promote industrial development in the cities have been frustrated by the unwillingness of the rural and poorly educated indigenous Turkic population to migrate to the urban areas. As a result, the rate of industrial and urban growth in these regions depends very much upon the ability to induce in-migration from the European republics (Whitehouse).

Finally, the stated goals of USSR spatial policy emphasise both agglomeration and the need for dispersion. Although the implicit conflict between these objectives is ignored in official policy statements, the potential dichotomy between

agglomeration and dispersion as principles of spatial development can be traced back to the writings of Lenin and Engels (Hamilton). While Lenin stressed the benefits of concentrating production in large integrated industrial complexes, particularly near energy and raw material sources, Engels emphasised the importance of achieving a more spatially even distribution of industry and of eliminating urban-rural differentials. It may be possible to reconcile partially this apparent contradiction because certain activities (e.g. heavy industries) gain from spatial concentration whereas others (e.g. some services, light footloose industries) may be dispersed at minimum sacrifice in costs.

A short paper cannot do justice to the hypothesis that the RGT model is applicable to developing and/or centrally planned economies. To evaluate this would require major research efforts and the study of many individual countries. Nevertheless, it is believed that the generalisations offered here support two propositions. First, the 'agglomeration economies—locational preferences' regional growth model, which was developed to explain spatial growth trends in a wide variety of cross-cultural settings, attains a degree of universality. Second, it is a richer and more flexible theory than the neoclassical model. Even if the latter can deal with such variables as agglomeration economies, spatial frictions and locational preferences by subsuming them in price and return differentials, this at the expense of trivialising space and over-simplifying its influence on regional growth.

IX

This paper is a mélange of concept-clarifying statements, verbal theorising, crude testing and casual empiricism. It is probably insufficiently technical for most tastes. Many would prefer to start by developing a model that can be manipulated with available tools and to see how far it can take them. I chose to approach the problem the other way around—to explore what theory should be trying to do, and to seek for consensus on its main features. To formalise a more rigorous theoretical model and to undertake serious hypothesis testing are subsequent steps. There are major bottlenecks in both directions. The construction of an economic theory type of model will be difficult because of the inability of mathematics to handle the discontinuities of space easily, while the main alternative—a simulation model—is not only costly in time and resources but hard to develop conceptually, e.g. the need for inclusion of a hierarchy of several regional and sub-regional levels. Hypothesis testing requires much more work in data collection, particularly on the spatial distribution of the capital stock and on returns to capital at different locations as well as on the measurement of agglomeration economies. These are tasks calling for many skills. The interdependence of problems within regional studies is paralleled by a need for interdependence among regional analysts if these problems are to be solved.

FOOTNOTES

1. Richardson (1973a).
2. To avoid personalisation in the debate, critics are not mentioned by name.
3. This was a central argument in Richardson (1973b).
4. On the other hand, there is a school of thought that regards higher wages in big cities as compensation payments for social costs (Hoch, 1972; Tolley, 1974).
5. The data were developed by L'Esperance and Nestel (1967). The series have been extended more recently by Cohen and Maeshiro (1973).

6. See Richardson (1974b) for discussion of this problem.
7. This assumption is unlikely to hold because space is non-homogeneous economically and socially.
8. This is similar to Pedersen's (1970) argument that hierarchical diffusion of innovations tends to replace general spatial diffusion during the process of economic development.
9. *En passant*, I admit to an awareness that the hypothesis that the distribution of city sizes will reflect the distribution of agglomeration potentials at particular points in space (specifically the locations of would-be migrants) comes dangerously close to a truism. This is part of the price paid for generalising the theory. Except in the rare counterintuitive cases, general theories are frequently not far from statements of semi-obvious truths.
10. In developing countries the reduction of spatial frictions may be the most important step since agglomeration pulls are ineffective unless they can exert themselves over space. Agglomeration potential must have a minimum critical value, and cities smaller than the national metropolis may not achieve this because of the steepness of the distance-decay function.
11. The work in the so-called 'new urban economics' is very much in the neoclassical tradition, e.g. its reliance on Cobb-Douglas production functions. The literature has concentrated on the problem of equilibrium residential densities and the related questions of rent structure and housing-roads land use allocations. It has produced interesting results such as how to correct for externalities (e.g. traffic congestion). However, there are two crucial weaknesses. First, the sole locational characteristic is distance from the CBD, therby making the city one-dimensional and permitting continuous functions more easily handled mathematically. Second, all the models have hitherto been static rather than dynamic. For elaboration of these arguments see Mills and MacKinnon (1973) and Richardson (1973c).
12. For examples see Lave (1970) and Dixit (1973).
13. Historically-determined locational constants have an important role in accounting for the location of subcentres, e.g. an old town or village centre absorbed by the spatial spread of the metropolis, a suburban education or office complex, the juxtaposition of multiple mode transportation terminals, etc.
14. This implies that 'environmental area—neighbourhood preference' residential location models provide a better explanation of residential location behaviour than more narrowly economic models such as that emphasising the trade-off between housing and commuting costs.
15. 'It is somewhat ironic that economists are becoming proficient at building models in which all employment is in the CBD and all housing outside it, when cities look less and less like that paradigm' (Mills and MacKinnon, 1973, p. 600).
16. The writers quoted in this section all contributed papers to Bandera and Melnyk (1973).
17. However, supported by studies of the relationships between locational efficiency and city size in Poland and the Soviet Union, the 1971-5 Five Year Plan has seen some shift towards promoting growth in small and medium-size cities.

REFERENCES

W. Alonso (1968), 'Industrial Location and Regional Policy in Economic Development', Institute of Urban and Regional Development, UCLA, Berkeley, WP 74.

W. Alonso (1971), 'The System of Intermetropolitan Population Flows', Institute of Urban and Regional Development, UCLA, Berkeley, WP 155.

V. N. Bandera and Z. L. Melnyk, eds. (1973), *The Soviet Economy in Regional Perspective,* Praeger, New York.

E. G. von Böventer (1975), 'Regional Growth Theory', *Urban Studies,* 12.

J. Cohen and A. Maeshiro (1973), 'Testing the Monetary Hypothesis at the State Level', mimeo., Department of Economics, University of Pittsburgh.

A. Dixit (1973), 'The Optimum Factory Town', *Bell Journal of Economics and Management Science,* 4, 637-51.

W. L'Esperance and G. Nestel (1967), 'A Comparative Analysis of Gross State Products, 1955-64', *Bulletin of Business Research,* Ohio State University, July, 4-6.

J. Friedmann (1966), *Regional Development Policy: A Case Study of Venezuela,* MIT Press, Cambridge, Mass.

A. O. Hirschman (1958), *The Strategy of Economic Development,* Yale UP, New Haven, Conn.

I. Hoch (1972), 'Income and City Size', *Urban Studies,* 9, 299-328.

L. B. Lave (1970), 'Congestion and Urban Location', *Papers and Proceedings, Regional Science Association,* 25, 133-49.

J. MacKinnon (1974), 'Urban General Equilibrium Models and Simplicial Search Algorithms', *Journal of Urban Economics,* 1, 161-83.

E. Mansfield (1968), *The Economics of Technical Change,* W. W. Norton, New York.

E.S. Mills and J. MacKinnon (1973), 'Notes on the New Urban Economics', *Bell Journal of Economics and Management Science,* 4, 593-601.

P. O. Pedersen (1970), 'Innovation Diffusion within and between National Urban Systems', *Geographical Analysis,* 2, 203-54.

H. W. Richardson (1973a), *Regional Growth Theory,* Macmillan, London; Wiley, New York.

H. W. Richardson (1973b), *The Economics of Urban Size,* Saxon House, Lexington Books, Lexington, Mass.

H. W. Richardson (1973c), 'A Comment on Some Uses of Mathematical Models in Urban Economics', *Urban Studies,* 10, 259-66.

H. W. Richardson (1974a), 'Empirical Aspects of Regional Growth in the United States', *Annals of Regional Science,* 8, 8-23.

H. W. Richardson (1974b), 'Agglomeration Potential: A Generalisation of the Income Potential Concept,' *Journal of Regional Science,* 14: 325-326.

J. G. Riley (1974), 'Optimal Residential Density and Road Transportation', *Journal of Urban Economics,* 1, 230-49.

G. S. Tolley (1974), 'The Welfare Economics of City Bigness', *Journal of Urban Economics,* 1, 324-45.

M. J. Webber (1972), *Impact of Uncertainty on Location,* MIT Press, Cambridge, Mass.