

UC Berkeley

IURD Working Paper Series

Title

Equity And Its Relation To Efficiency In Urbanization

Permalink

<https://escholarship.org/uc/item/2tt231jn>

Author

Alonso, William

Publication Date

1968-07-01

Peer reviewed

EQUITY AND ITS RELATION TO EFFICIENCY IN URBANIZATION

**William Alonso
Professor of Regional Planning
University of California, Berkeley**

July 1968

Working Paper No. 78

**The work reflected in this paper was done under a grant from the
Economic Development Administration of the U.S. Department of Commerce.**

All Rights Reserved

INTRODUCTION

This paper divides into two parts. The first is a discussion of the goals of national policies for regional development, focusing primarily on the conceptual and operational difficulties of dealing with the goal of equity. The second part is a paradigm of the interplay between the goals of efficiency and equity in interregional migration. Many simplifications are made to highlight this interrelation, which has been the source of considerable confusion.

NATIONAL GOALS FOR REGIONAL POLICY, WITH SPECIAL ATTENTION GIVEN TO EQUITY

National regional policy may respond to many goals. Most commonly, in addressing itself to the issues of urbanization and depressed areas, policy focuses on the goals of national economic growth and of the distribution of income and unemployment. Recent academic literature has concentrated on these, usually under the names of efficiency and equity. But it must be recognized that several other goals for regional policy are commonly encountered. These include the occupation of territory for defense purposes,¹ the occupation of frontier regions to create national consciousness, identification, pride, and purpose,² and

¹Instances of this are the policies of India with respect to its Chinese and Pakistani borders, Israel's settlement policy, Thailand's Northeast development efforts, and many well-known instances in the history of the United States.

²Among the best known of these are the contemporary efforts of Brazil, Peru, and Venezuela to occupy and develop their interiors.

policies aimed at civil stability or political gain.³ Related to this last, but ultimately distinct, are policies directed to supporting or weakening the solidarity of ethnic or cultural groups, usually but not always in response to the tension between the national interest in socio-cultural unity and the groups' desire to maintain their solidarity and promote their interests as a group.⁴ We shall return to this issue below in the discussion of the concept of equity.

This paper confines itself to the relation of the goals of growth and income distribution, and tries to clarify their interplay. Public debate and academic discourse have had a great deal of trouble dealing cleanly with this interrelation. For instance, one often meets statements that population movement from the countryside to urban centers should be discouraged for two reasons: First, the urban centers are subject to diminishing returns as the result of congestion and like factors and, consequently, further movements of population to these centers is inefficient. Second, the cities are already rich and powerful, whereas the countryside is being drained of its best young people by this migration, thus increasing the inequality among regions. Although both of these positions are often held simultaneously (albeit seldom stated with precision), a moment's thought will raise doubts as to whether both

³ Perhaps the most interesting current instance of this is the common experience of most nations that have tried to use growth poles for regional development. Although the logic behind this strategy is the concentration of energy and investment, political pressures, usually associated with the territorial representation embodied in the legislative branch, lead almost invariably toward the proliferation of the centers.

⁴ Instances of this include the tragic experience of tribal conflict in African nations, the Iranian handling of the Kurdish question, and Spain's enduring separatist Basque and Catalan movements. Needless to say, this issue in many different forms has become a central one in America's racial situation.

can be true simultaneously. We shall see that although one statement does not necessarily contradict the other, both can be true only in unlikely circumstances.

The goal of efficiency is, most simply, a goal of national economic growth. It is commonly accepted both at the level of political statements and of technical analysis. Slight variations exist as to how to measure performance. The most easily understood and commonly used criterion is the rate of growth of gross national product. More technical measurements use the sum of the discounted values of future levels of consumption, but these distinctions are not fundamental to the argument in these pages. The goal is generally understood and accepted, and very specific statements are possible about performance which, in one form or another, are intelligible to most people and thus affect political decisions.

Equity does not enjoy such clarity. Public policy accepts the desirability of assisting the less fortunate, but there is no commonly accepted basis of conventional or scientific wisdom on which to base operational judgments of need or performance. It is extraordinary that there has been so little technical discussion of a concept so central to political economy. In ordinary usage, equity is the quality of being equal or fair. The idea, in some form, is at the root of the concept of social justice. But it is not mere equality. "To each according to his needs" implies that these needs may differ. If a man has greater capacity for enjoyment than his brother, is it more equitable to give them the same income or the same enjoyment? Our concept of what is fair is an evolving one, as we slowly forge an economic bill of rights consisting of guaranteed income, access to education, compensation to victims

victims of discrimination, crime, sickness, etc. Similarly, our idea of who is our brother is an evolving one. American democracy long considered the slave as a non-person. The contemporary Negro movement in this country demands parity with white America, not with Haiti or black Africa. Equity is intuitively accepted as a goal by people and governments, but it remains full of ambiguity and problems of meaning and technical measurement, especially when dealing with regional issues. Although the paradigm presented later will revert to strong simplifications, a brief discussion of some of the more technical issues may be useful.

The geographic scale of regional definition is crucial to the evaluation of inequality. It is well known that a fine-grained regionalization will result in higher measures of inequality than a coarse-grained regionalization. In other words, the same index of inequality for the United States computed by state will be lower than if computed by county because much of the inter-county variation (i.e., inequality) is averaged out when the larger unit is used. The solution to this problem is not to use smaller and smaller regional units, for the logical limit of this is to have one-person regions, and such an index would be that for the national population as a whole and would not be a regional index. When a regional measure is used, it is because it is believed that there is a meaningful spatial association of the phenomenon. If there is no such association, or if the association is not meaningful for the purpose at hand, there is no particular sense in using a regional index.

In brief, a regional measure of inequality or concentration depends on the variable of regional scale, and a meaningful measure

should be based on divisions of the territory which reflect the spatial structure of the socio-economic system.⁵ Consider a possible case in which certain depressed regions can develop only by internal specialization which increases the inequalities within the region. National equalization of regional incomes would hide overall divergence for the national population if regional considerations were ignored. National interest in equity might well accept this regional convergence in spite of a-regional divergence, especially if increases in intra-regional inequality are the result of some of the people in the poor regions being better off without worsening the condition of others.

A similar problem results from the customary consideration of equity in terms of relative numbers, such as rate of unemployment. Areas which are to receive aid are frequently identified by such criteria. But comparison of the average income of a high-income area (such as the New York metropolitan area) with that of a low-income area (such as one of the Southern states) may hide the fact that there may be more poor people in the prosperous area than in the poor one.⁶ This type of problem does not arise in the measurement of efficiency because in that case

⁵This problem has begun to receive some attention, particularly among geographers, in a slightly different formulation: that of spatial intercorrelation. It is the two-dimensional equivalent of the problem of serial correlation in time series. The use of smaller and smaller regions in computing the index runs into the problem of ignoring the tendency of a small area to be like its neighbors or, more technically, it ignores the problem of spatial autocorrelation. A recent illustration of this problem is to be found in K. E. and A. F. Taeuber, Negroes in Cities (Chicago: Aldine Publishing Co., 1965). The authors compute indices of segregation for American cities, using the city block as their unit. Indices of segregation are indices of inequality by another name. The deficiency of this procedure is that a city with many small Negro districts might receive the same index score as a city with one massive ghetto.

⁶A very interesting treatment of this question is found in G. Davies, "Regional Underemployment, Labour Availability, and Redeployment," Oxford Economic Papers (March, 1967).

the performances of subdivisions of the national territory add up to national performance. In the case of equity, measurement is relational and therefore not additive in a straightforward sense, so that the measure will be extremely sensitive to the scale and manner of regionalization.

The definition of equality depends on the terms of reference. Thus, a low-income region may specialize in low-wage industries but have higher than average wages for those industries. Measured simply on the basis of per capita income, that region would appear disadvantaged. But, if its industrial composition is taken into account, the opposite conclusion would follow. Comparative statements usually carry a heavy baggage of implicit assumptions as to the basis of comparison. The previous example pointed to the effects of composition, but similar problems arise in the definition or perception of the universe within which we make our comparisons. Thus, Appalachia appears low-income within the United States, but would be high-income if it were considered within a context of the Western Hemisphere.

Indices of inequality, concentration, localization, and segregation are really the same indices by other names. There are surprisingly few techniques that have been used in the construction of these indices. Among the principal ones are the relative mean deviation, the standard deviation, the Lorenz and Gini indices and, most recently, measures of entropy.⁷ They all share the same underlying logic: an ex ante expectation is compared to the observed value for each unit, and the difference or ratio between expected and observed values by units is combined

⁷ See H. Theil, Economics and Information Theory (Amsterdam: North Holland Publishing Co., 1967). This work is of particular interest because it is based on information theory and therefore may be used to relate expectation to observed values explicitly. However, Theil limits himself to the mean as the ex ante expectation.

according to some formula. The literature has attended almost exclusively to the choice among combinatorial formulae, and the expected value has been taken as a matter of course as the mean of the universe's distribution.⁸ Very little theoretical work has been done on the setting of expected values, yet this is the key to the meaning of the index. The importance of how expectations are defined was shown in the example above of the low-income, high-wage region.

In territorial or regional indices, because of the spatial autocorrelation already mentioned, the use of the national mean as the expected value is only one possible approach, and not necessarily a good one in every case. It presumes that we make no use of information of regional location (or spatial association), and its use in geographic work is equivalent to ignoring the information of the sequence of observations in the analysis of a time series. This is fine for some purposes but deceptive for others. Techniques based on spatial association, such as potential mapping, map generalization, and geographic spectral analysis, hold great promise as alternative ways of generating sets of expected values for which useful statements can be made, linking explicitly the level of inequality with regional scale.

These techniques have a further interest. Generating an expectations surface by, in effect, running some version of a two-dimensional moving average over the observed values provides a controllable regional scale (the range or some measure of dispersion of the weighting function used) without dividing the national territory into discrete regions. This feature is particularly attractive, since a conventional division

⁸The Gini index proves to be very subtle, relating expected value not to the mean, as do the other measures, but rather to the weighted distribution of values in the rest of the universe as a probability function.

into discrete regions will suffer from a necessary arbitrariness in the setting of boundaries. Human interaction seldom respects hard boundaries, and inequality and concentration are most often studied as problems of lack of social or economic integration. These techniques are all based on some form of a distance decay function which simulates the friction of space. When the rate of decay is nil, we are saying that distance plays no role. In this case, the expected value by these techniques becomes the mean, which is therefore a limiting special case.

An interesting development is the increasing governmental practice of setting expectations on normative rather than statistical grounds. Thus, we have definitions of poverty for families based on their lacking an income which is set not from the distribution of incomes but from a far broader societal frame of reference. Needy regions are normally defined by their proportion of needy people. While it is obvious that such bench marks are necessarily arbitrary and often poorly constructed, there is sense in measures of equity or inequity based on some concept of what ought to be. Equity is the quality of being equal or fair. Those measures based on statistical distributions measure what is, and take equity to mean formal equality. The normative measures tell us how close we come to our concept of fairness, to what we expect of ourselves. Still, however the expected values are set, thereafter in general the index measure is constructed in the usual way by relative deviations, etc.

While the measures of efficiency have become sophisticated in dealing with the time dimension by the use of discounting procedures, equity seems to be difficult enough to measure at one point in time, and few have ventured into the problem of intertemporal comparisons. How much inequality in the short run can be tolerated for greater equality

in the long run?⁹ The much-debated problem of trade-offs between efficiency and equity, and their millenary reduction to a common metric which might enable us to combine them into a common measure, must await development of some theory to discount inequality streams.

Thus far most of the discussion has dealt with the equity problem from the point of view of income inequality, and this is the common approach of economists. Politicians and others have usually spoken of it as if it had to do with high rates of unemployment. Recent work has shown that both are right: U.S. counties characterized by low incomes are usually not the same counties characterized by high unemployment.¹⁰ Yet both types of area are troubled and the proper object of equity policy. Thus equality and inequality turn out to be multidimensional. We have only arbitrary ways of relating these two dimensions of inequality. Our operating policies have used rough-and-ready formulae that combine these two for the definition of areas that are in need of aid, usually combining a third dimension of sustained outmigration. A single, all-purpose measure based on sound theory is nowhere in sight.

A fundamental problem remains untouched. Out of its foundations in the eighteenth-century rationalism of Locke and Bentham, economics

⁹H. Peter Rydell, in the only treatment of this problem of which I am aware, suggests treating each component region as concerned exclusively with its own efficiency goal, and combining their discounted consumption levels into an equity index which would thus be interregional and intertemporal. It may be objected that this procedure does not take sufficient account of the perception of inequality at the time. Thus, two discounted income streams may be equal, but incomes in the two regions may be unequal at virtually all times and this may be perceived as inequality. Alternatively, the suggested procedure seems to depend on the starting date of the comparison.

¹⁰Area Welfare: Eligibility for Development Assistance, Report prepared for the Area Redevelopment Administration of the U.S. Department of Commerce (Cambridge, Mass.: Regional and Urban Planning Implementation, Inc., May, 1965), Vol. 1, Chap. VII.

views the world atomistically, as if the level of satisfaction of an individual related exclusively to his own well-being and was unaffected by the circumstances and companionship of his fellows. On the other hand, it is common for the discussion outside of technical journals to take what may be termed an organic view of groups and regions. Some of this is clearly a semantic trap, which may be termed "the geographic fallacy," whereby territories are personalized and endowed with feelings.¹¹ For instance, it is perfectly possible for the per capita income of a depressed region to drop, but for all individuals in it to be better off. The conditions are that the depressed region be subject to diminishing returns and that the higher-income members leave for other regions where their skills are better rewarded. Since the region is, in a sense, overpopulated (diminishing returns), the leaving of some raises the incomes of those who remain. Both those who leave and those who stay are better off, but many will argue that the region has declined because its average income has dropped.¹² The classic economist's position is that this is an excellent development, since everyone is better off. But, in a sense

¹¹For an excellent discussion, see L. Winnick, "Place Prosperity vs. People Prosperity: Welfare Considerations in the Geographic Redistribution of Economic Activity," in Essays in Urban Land Economics, Real Estate Research Program, University of California at Los Angeles, 1966. The same issue is addressed in the U.S. Supreme Court ruling (in Baker vs. Carr) that "legislators represent people, not acres or trees. Legislators are elected by voters, not farms or cities or economic interests. . . . The weight of a citizen's vote cannot be made to depend on where he lives." Here the Court comes on the side of atomism, reflecting the eighteenth-century rationalism of the Constitution of the United States.

¹²The most common forms of the geographic fallacy seem to relate to our urban problems. Simplified statements about the decline of our cities ignore the fact that populations move across municipal boundaries (suburbanize), and compare measurements that refer to constant areas to shifting populations. Conversely, many of the positive evaluations of earlier attempts at urban renewal were based on keeping the same geographic base, but ignoring the change of population in the cleared area.

that I have never seen well articulated, many seem to feel that this represents a decline of a people, although it may represent improved conditions for persons. Pride, self-image, a valued way of life, a sense of identity with a people or a cultural tradition are reflected in this perception, which opposes itself to the rationalism of economists.

This viewpoint presents itself in many guises. Limiting ourselves to the contemporary scene in the United States, we find expressions of it not only in the public discussion of depressed areas, but in some of the propositions of black power, in the cries of ethnic or cultural communities threatened by urban clearance, and in the language of those concerned with Indian reservations. Economics has had little to offer here.¹³ Sociology may provide a more useful framework in terms of its classic distinction between mechanistic and organic solidarity (Gesellschaft and Gemeinschaft). The descendants of Adam Smith, dealing only in questions of a contract society, use conceptual instruments more suited to a modern, developed society and therefore miss some of the considerations that apply to the less developed societies (organic solidarity) which characterize many distressed areas.¹⁴ In the economist's language, it

¹³The proposition that the utility function of individuals can be made to include as independent variables the levels of consumption of others, weighted by the degree of concern of the individual for these others, does not seem to do justice to the question.

¹⁴The concepts of Bellassa and Friedmann about economic integration are particularly relevant here (B. Bellassa, The Theory of Economic Integration [New York: R. D. Irwin, 1961]; J.R.P. Friedmann, Regional Development Policy [Cambridge: M.I.T. Press, 1966]). Lack of integration is an irrational immobility of factors in spatial or sectoral terms, judged by the standards of rationalism. In that economic development may be viewed as the rationalization of production and consumption, it represents a shift from organic to mechanistic solidarity (from tradition to contract) and much of the problem of distinguishing between depressed areas and their populations may be viewed as a conflict between the two types of solidarity. Analysis and proposals framed entirely in terms of one or the other frame of reference will obviously be insufficient for consensus.

might be said that the organic solidarity of a region or a people is a merit good, but this does not advance us a great deal.

Finally, any consideration that will help frame decisions which affect the lives of people must be intelligible to more than a small fraternity of experts. Since equity is relational, its definition and measurements are quite naturally abstruse. Statements about the root mean square deviation or about the complement of entropy are unlikely to move political support or to engage the hearts of men. There is a lesson to be learnt from the effectiveness which the simplified measures of performance of the efficiency goal (principally, rate of growth of the economy) have had in mobilizing support for this objective. In spite of their statistical limitations, two measures of equity stand out as being understandable and therefore effective guides to political action. One is the use of normative bench marks discussed above: the number or proportion above or below some standard such as poverty level or unemployment rate. The other is rarer, but very promising. It is the relative weighted mean deviation from the mean in its one-tailed form, or the relative deviation. In the cases of unemployment, income, or racial distribution, it measures the number of workers, cents per dollar, or people in the population that would have to be moved in order to equalize conditions. Non-technical persons can easily understand a statement that, whereas some years ago we would have had to move X cents per dollar from the rich to the poor to equalize material welfare, we now have to move Y cents per dollar. To influence and inform policy, it is frequently important to present findings in measures such as this, which have vividness and intelligibility.

A PARADIGM OF EFFICIENCY AND EQUITY IN THE MIGRATION TO CITIES

The interplay between considerations of efficiency and equity is fairly simple if considered systematically. Yet there is enough complexity to give rise to considerable confusion, and at times equivocation, among scholars and national and international agencies. The reasons for this confusion seem to be two-fold. On the one hand, where a particular policy favors one goal over the other, its advocates are tempted, in seeking support, to represent it as serving both. On the other hand, the semantic shorthand by which we use the name of some piece of territory to speak of a population has led to untold confusion because people can move, whereas real estate cannot. We have referred to this above under the label of the geographic fallacy. It must be admitted, however, that the following attempt at clarification ignores many of the considerations raised in the preceding pages.

We shall conceive of a simplified country composed of two realms, one advanced and prosperous, the other depressed or backward.¹⁵ We pose the question of a national government which is evaluating, from the point of view of its goals of efficiency and equity, a possible move by a worker from a poorer area (periphery) to a city (center). To keep matters simple, we ignore movements of capital and, for the time being, we do not distinguish among types of migrants.

Figure 1 embodies the core of the argument. On the vertical axis we measure the marginal product of labor in the city, MP_c , and on the horizontal axis we measure the marginal product of labor in the periphery,

¹⁵ On this we follow many precedents, notably A. Hirschman's distinction between North and South (Strategy of Economic Development [New Haven: Yale University Press, 1958]) and J. Friedmann's center and periphery (op. cit.), whose terminology we borrow.

Key

+ = increase
- = decrease

MP_c = marginal urban product

MP_p = marginal rural product

EF = efficiency goal performance

EQ = equity goal performance

y_c = mean urban income

y_p = mean rural income

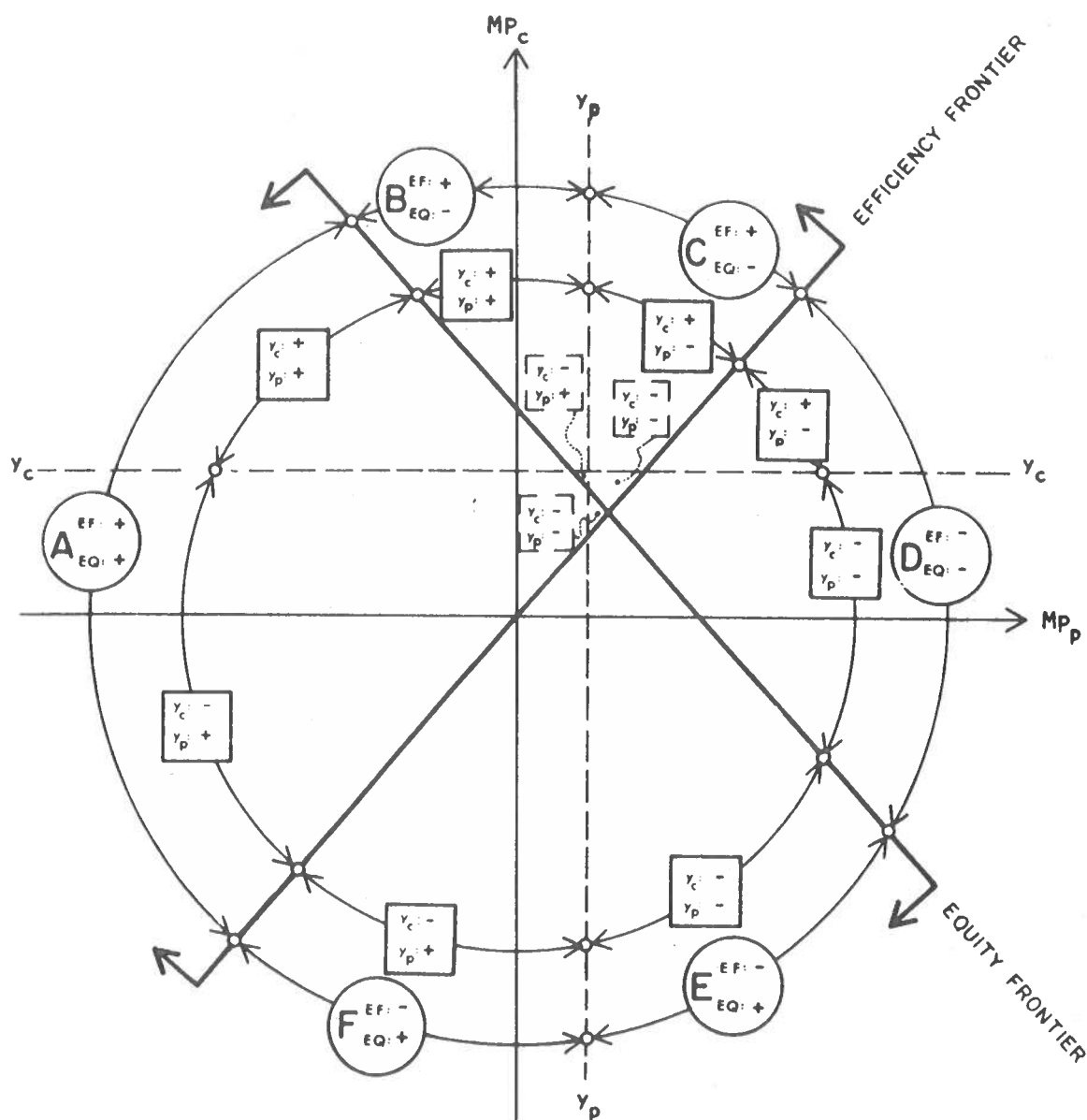


FIGURE 1. EFFECTS OF A MIGRANT'S MOVING FROM THE RURAL AREA TO THE CITY

MP_p .¹⁶ We extend both axes beyond the zero point to admit of negative values. A point on the graph represents the objective consequences of the movement of an individual in terms of marginal product, and the national government may evaluate the consequences in terms of its objectives.

Clearly, if marginal product of a worker is greater in the center than in the periphery, the contemplated migration will increase national product and serve the goal of efficiency. We can therefore draw a 45-degree line sloping upward through the origin, which we call an efficiency frontier, such that any point above the line represents circumstances where the efficiency goal would be served by the move. This obtains even in the negative quadrant. Under conditions of labor surplus both in the center and the periphery, as sometimes happens in developing countries, surplus workers may have a negative effect on product both in the city and in the countryside, and the national goal of efficiency would best be served by the man's being in the region in which his presence does the least harm.

On the same graph are plotted dashed lines y_c and y_p , which represent the per capita incomes¹⁷ in the center and the periphery, respectively, before the move. It will be noted that in the illustration, as is true

¹⁶ Marginal product may be thought of as increases of value added in a region or changes in the contribution to regional product from a marginal change in labor force at that location. It should be stressed that this does not necessarily mean that this marginal product is equivalent to the wages of the marginal worker. Even should institutional conditions result in wages being equal to marginal product within the firm, the contribution to regional product takes into account the net effect of external economies and diseconomies of the increase in population, many of which do not enter into the internal calculations of the firm.

¹⁷ For verbal simplicity, we shall equate regional per capita product with regional per capita income, and use the words "product" and "income" interchangeably.

in all countries that I know of, urban per capita income, y_c , is greater than rural per capita income, y_p . If the point which represents the marginal product contribution of the workers is above the y_c line, marginal urban product is greater than per capita urban income (a condition of increasing returns to scale for the city), and the move will result in a rise in per capita urban income.¹⁸ Similarly, if the point is to the right of y_p , the periphery also enjoys increasing returns to scale and the departure of the individual will lower rural per capita income. In such a case it is obvious that the equity goal will not be served, since the rich region becomes richer and the poor poorer.

At this point we can reduce the laborious analysis by plotting an equity frontier, which is a line such that to one side of it the equity goal is served, and to the other it is not. The line has been calculated on the "cents per dollar to be moved" criterion of equity,¹⁹ and takes into account the shift of population. Most alternative definitions of equity will also yield a straight line, differing slightly in slope, and passing through the intersection of the y_c and y_p lines or just under it. Figure 2 shows the equity frontier for five alternative

¹⁸We are assuming, of course, that there are no transfer payments among regions.

¹⁹The equation for the constant relative deviation is:

$$MP_p = \frac{(p_p y_p + p_c y_c)^2 - MP_c (p_p y_p [p_p + p_c] + p_p y_p + p_c y_c)}{p_p y_p (p_p + p_c) - p_p y_p - p_c y_c},$$

where the notation is as in the diagram, and p_c and p_p are the populations of the center and the periphery, respectively, before the move.

Key

- A. Constant difference in per capita regional products.
- B. Constant standard deviation in per capita regional products.
- C. Constant dollars to be moved to equalize per capita incomes.
- D. Constant ratio of per capita incomes.
- E. Constant relative deviation (cents per dollar of GNP to be moved to equalize per capita incomes).

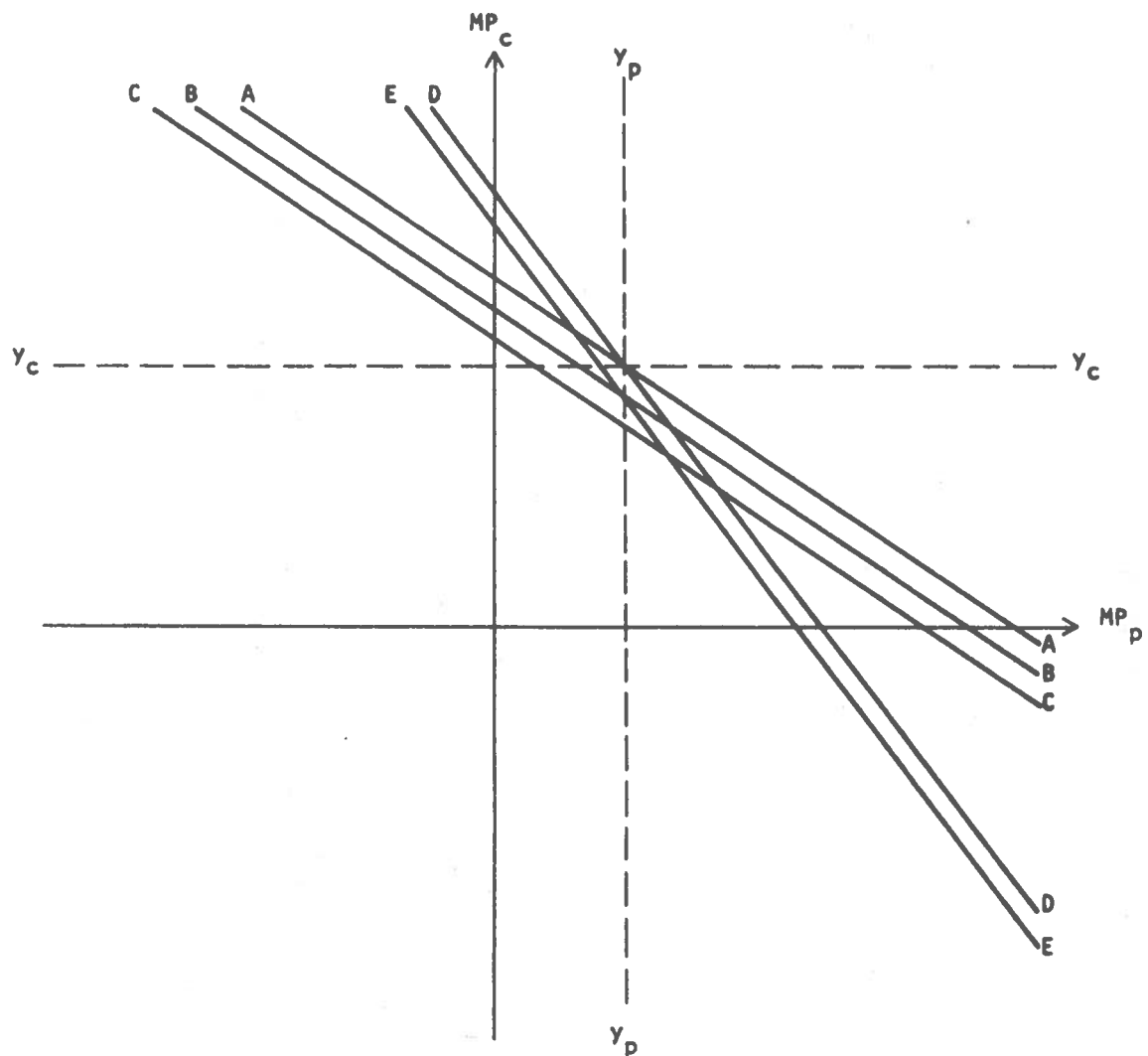


FIGURE 2. FIVE EQUITY FRONTIERS

definitions.²⁰ In all cases, if the point that represents the marginal product of the individual at both locations is below the equity frontier, the move will increase equity. The argument that follows is not affected by the measure of equity which may be chosen.

The equity and efficiency frontiers, together with the lines y_c and y_p , define a number of sectors and subsectors. Each sector has been labeled by a letter, A through F, and in each sector it is shown whether efficiency and equity are increasing or decreasing. In addition, for each of the subsectors we have indicated in small rectangular boxes the direction of change of the per capita incomes.

Sector A represents a condition in which the move would improve both equity and efficiency, and there can be no doubt that the move is desirable. The upper subsector corresponds to a situation in which the urban realm enjoys increasing returns to scale, so that further increases of urban population result in higher levels of welfare in the cities. This implies that the economies of agglomeration exceed the diseconomies. On the other hand, this subsector corresponds to a condition of rural overpopulation (as, for instance, in cases of regions undergoing rationalization of agricultural production or rapid rates of natural population growth). Those who argue that we should move people to jobs imply that we are in this sector. The lower subsector corresponds to a case of declining returns to labor in the city, resulting in a drop in average urban income, but of even greater overpopulation in the countryside. In other words, the upper subsector may be viewed as a situation in which

²⁰ It will be noted that the three absolute measures of difference, A, B, and C, share the same inclination, while the relative measures, D and E, have a slightly different one. This is an indication of the structural similarity of the relative mean deviation to the ratio of per capita income, which is, of course, a highly intelligible measure but is limited to the comparison of pairs of regions. The "cents per dollar to be moved" measure can be used for any number of regions.

the move has a very positive consequence, whereas the lower subsector permits only making the best of a bad situation. Such conditions are sadly imaginable where economic development is flagging.

Just to the right of the y_p line there is a small subsector of A which is quite interesting. The average incomes of both city and periphery drop, but total average income for the nation increases and the goal of equity as well as that of efficiency is served. This represents a circumstance in which the returns to labor in the city are decreasing while those to rural labor are increasing, but where the returns to urban labor are still larger than to rural labor. Still, in a sense it is surprising that the efficiency goal should be served when the mean incomes in both regions are dropping. This is the result of the shift of population from the poorer to the richer region. In general, as can be seen from the diagram, this apparent paradox will occur in other cases where marginal product is very near to average product; that is say, in cases where both regions are close to constant returns to scale.

Sector D is close to being symmetrically opposed to sector A. Throughout sector D, marginal returns to labor are higher in the periphery than at the center, and therefore efficiency is served by not moving. Similarly, through this sector the move would also reduce national equity. In the upper subsector of D, city incomes would go up and rural incomes would go down. Both the city and periphery can use more labor, but there is a higher marginal return to rural manpower. In the lower subsector, the city is too big, in the sense of diminishing returns (and at the very bottom of the subsector, below the MP_p axis, too big in the sense of an absolute decline in urban product), while the rural sector still enjoys increasing returns. One might even argue that if a nation finds

itself in sector D, as many believe they do, it might try to encourage migration from the cities to the rural areas. Worldwide experience shows, however, that it is very hard to move people back to the farms once they are in the cities. The frequent statements to which we referred, that urban-to-rural migration is both inefficient and leads to greater inequality, can be true only if national circumstances are represented by a point in this sector. These circumstances may be typified as (1) rural production would increase more than proportionately with increases in population, and (2) urban returns to population growth are lower than rural returns. The first proposition would be true only in rare cases.

Sectors C and F represent circumstances of true conflict between the two goals. If the nation finds itself in sector C, migration will increase national product but will also increase inequality. Over the upper part of the subsector, both city and countryside have increasing returns to labor, but marginal product is greater in the city. In the lower part of the sector there is a small area where we again meet the apparent paradox of a drop in the per capita income in both regions, but a rise in national per capita income. The explanation of the paradox has already been mentioned. Sector F is a sad circumstance in which to be. The nation is overpopulated. Both city and countryside have declining returns, and over most of this sector (to the left of MP_c) the surplus bodies actually lower absolute production. In this plight, the most efficient solution results in an increase of inequality, and there is genuine conflict between the two goals. The choice in both sectors C and F must be made by whatever political processes are available.

Sectors B and E are perhaps the most interesting. Sector B appears to satisfy the efficiency criterion, but to deny that of equity.

On closer examination it proves to be a Pareto optimum. Efficiency is served, of course, and though equity is not, every one is better off than before. This is not a quirk of the particular index used, but general to all measures of equity. In practice it represents a situation of surplus labor in the periphery and increasing returns to the urban realm. This may be taken to be a frequent situation. If we say that permitting or encouraging migration is the national policy, its consequences are to raise incomes in the cities and the rural areas, but the cities grow sufficiently faster in per capita income that there is an increase in inequality. Consider the opposite policy, that of restricting the migration. The regions are more equal, but not only is national growth lower, the per capita income in the rural region is also lower (as is the per capita income in the urban region) than it would be if migration were permitted. This situation may be widespread if dynamic manufacturing and service sectors in the cities result in increasing returns to scale and if structural transformations are taking place in the rural primary sector, resulting in labor surplus. The best documented case of this condition is that of Southern Italy, where early policy called for an equalization of incomes between North and South, but where experience prevailed in setting the more modest objective of raising Southern incomes while recognizing that Northern incomes might rise faster.²¹

Sector E, by contrast, might be called an invidious Pareto situation. While the contemplated migration might serve the goal of equity

²¹A particularly clear account of this shift in objectives is found in H. Chenery, "Development Policies for Southern Italy," Quarterly Journal of Economics (November, 1962). Of course, in the reality of the Italian case, many other factors were operating, such as the movement of capital, skilled labor, and entrepreneurship. Nonetheless, the basic shift resulted from the realization that a Pareto optimum might be possible which violated the prima facie concept of equality.

at the expense of that of efficiency, in fact everyone would be worse off after the move since the per capita income in both regions would be lower. It would take a rather bitter concern for formal equity to insist on migration under these circumstances. If a migration from the city to the countryside were possible, its result would be an increase in efficiency, and, whereas everyone would be better off, formal inequality would increase. It must be noted that the conditions for this quandary are that marginal returns to labor in the rural areas be very high and that marginal returns to labor in the cities be very low.

Throughout this analysis we have considered the migration of a worker as if all workers were alike. Clearly, in any nation people with different skills and attitudes will place differently in the diagram. For instance, a very good farmer might place in the right-hand side of the diagram, and if he were unable to adapt to city life, in the lower right. A skilled machinist might place on the upper part of the diagram. A poet, a physicist, a laborer, an accountant, a medicine man would all place differently. Similarly, the consequences of an individual's move are not likely to be so certain that they can be represented by a point. Rather, the diversity of individuals, even within a class, and the uncertainty of prediction would result in an area or probability surface on the diagram, and the conclusions would have to be stated probabilistically, especially if the area fell into more than one sector.

WHERE DO WE STAND: CONSEQUENCES FOR POLICY

The paradigm presented in these pages is only a systematic classification for evaluating the consequences of the act of migration with respect to the goals of efficiency and equity. The remaining question

is the important one, and it is one of fact: where do we stand in the diagram? Considering the importance of this issue and the radical policies (usually opposed to urbanization) that have been advocated and often espoused in Europe, in the socialist countries, and in the developing nations, it is indeed remarkable that there is so little knowledge of fact concerning the marginal product of populations in regions. Rather, it would seem that the flood of pronouncements on this issue by politicians, economists, aesthetes, and assorted urbanologists has created a sort of self-reinforcing conventional wisdom supported primarily by a well-polished rhetoric which has discouraged not just empirical inquiry but even the examination of its logical basis. Without knowing (or asking) whether the urban or rural realms are subject to increasing or decreasing returns, most countries have adopted at least the rhetoric of policies of decentralization and the advancement of backward regions. The necessary factual assumption is that cities are inefficient (diminishing returns), and that there are many unexploited opportunities in the periphery (increasing returns), placing them in sector E or the lower part of sector D. A conscientious search of the literature reveals no empirical basis for this opinion.

This is not the place to review the slender evidence which exists. It may be expected that circumstances will differ from country to country and from period to period. Neither is the argument being made that the policy of having no policy be adopted while years are spent on research which may prove inconclusive. Rather, we should get on with research on a vital policy issue and frame policies more consistent with current estimates of economic reality.

