An Economic Analysis of Internal Migration in Brazil

Gian S. Sahota*

Vanderbilt University, U.S.A., and IPE, São Paulo, Brazil

1. Introduction

Migration's role in economic development is widely recognized. The heart of the economic development problem is generally considered to lie in a gradual reallocation of the labor force from the agricultural sector to the industrial sector (see, for example, Lewis, 1954). Since urban fertility is unlikely to exceed rural fertility, the shift of population must obviously come about through migration. The factors underlying labor migration (especially from the agricultural areas to the industrial areas) need to be isolated. Such an analysis, on the one hand, will increase our understanding of the mechanism of labor adjustment and thus of an important aspect of the process of economic development. On the other hand, it can serve as a basis for policies concerning employment, antipoverty, and other economic matters.

Why do people leave their homes? Is the typical migrant rational? Do migrants respond to economic incentives, or are they mainly actuated by non-economic factors? Do migrants represent random samples of the population, or do they belong to select groups? Do unhindered movements of people tend to eliminate interregional differences in income? Do responses to given stimuli differ among age groups? These are some of the questions which this study seeks to answer.

The present study is confined to an analysis of *interstate* migration in Brazil. Brazil is a vast country with wide disparities of levels of living, economic structure, and growth rates among different states. If a sample

* In developing this study, I benefited from discussions with Affonso C. Pastore, Carlos A. Rocca, and William O. Thweatt. I am also grateful to the referees of this *Journal*, whose detailed comments and suggestions led to substantial refinement of this paper. I wish to make acknowledgments to IPE, University of São Paulo, and the Graduate Program of Economic Development and the Graduate Center for Latin American Studies, Vanderbilt University, for meeting the research costs. Thanks are also due to USAID, under whose auspices I visited Brazil, without which this study would not have been possible.

consists of *international* migration, the determination of various response-stimuli relationships may be hampered by several economic and non-economic barriers to migration, such as the visa problem, foreign exchange difficulties, controls on foreign travel by home governments, racial antagonism, cultural differences, and so on. At the other extreme, if one confines the inquiry to migration among the communities of a small country or a state within a country, one might be handicapped by a lack of heterogeneity of migration factors, insufficiency of the variability of observations, and the absence of most of those economic and non-economic factors that are represented by geographical distance. The Brazilian economy, in this respect, provides an interesting example of a country that offers the best of both worlds for a study of migration. Thus, on the one hand, there are no visa problems or racial barriers among the states. On the other hand, the country is so vast and the economic disparities are so wide that, compared with the internal migration in this subcontinent, the international migration among, for example, the countries of Europe might look like a local phenomenon.

In Section 2, I review very briefly three hypotheses or theories of migration. The data are defined in Section 3. The variables expected to explain migration are discussed in Section 4. The empirical results are derived from two models: a single-equation model estimated in Section 5 and a simultaneous-equation model estimated in Section 6. A brief comparison with the existing works and the conclusions of this study are given in the final section.

2. A Brief Survey of the Theories of Internal Migration

Of the various hypotheses purporting to explain migration, three (though not entirely distinct) may be briefly mentioned. One of these is based on the neoclassical theory of investment and has appeared largely in the recent works of the Chicago School, in particular Schultz and Sjaastad.[1] Internal migration is put in a framework of costs and returns of investment in human capital. The returns—rather, expected returns—consist of the differentials in the income streams accruing to the migrant from the (expected) better opportunities. Costs are comprised of money costs in the form of an increased expenditure on food, lodging, and transportation; non-money costs, which include income foregone during the period spent on traveling, searching for, and learning a new job; and psychic costs, such as homesickness, acclimatization strains, and so on. Thus, substantial differences in current earnings in two different places may continue to exist without inducing significant migration because the costs of migration may

[1] A summary of these works may be found in a special issue of the *Journal of Political Economy* (Vol. LXX, No. 5, Pt. 2, 1962). In particular, see Schultz (1962) and Sjaastad (1961, 1962).

(actually, or in the estimation of the prospective migrant) exceed the observed differences in earnings. Older workers may be less disposed to move simply because the payoff period is shorter; the expected lifetime earnings are smaller for them than for younger workers. Moreover, as Sjaastad has emphasized, investment in migration is usually accompanied by supplementary investments. "Cognizance of and attention to these additional investments offer a promising clue to observed immobility in the face of large differentials in current earnings" (Sjaastad, 1962, p. 93). According to this approach, investment in migration, like other investments in human beings (for example, education), has high payoff and stimulates growth.

An alternative approach to migration may, from the point of view of pinpointing a single center of major contributions, be associated with the Harvard School. In particular, it permeates the works of Simon Kuznets. The best single source for work along these lines (and related works) is the multivolume study, *Population Redistribution and Economic Growth—United States, 1870–1950*.[2]

This approach delineates the relationship between internal migration and economic development in terms of the selectivity of people. The migrants come from select groups. They are the dynamic risk-taking beings who have high capacity to detach themselves from the traditional surroundings and adapt themselves to the unfamiliar environment. They are stimulated to move to the centers that offer better economic opportunities. The resulting redistribution of population promotes subsequent growth and induces further migration of the select individuals. The jobward-migrating components of the population and the indicated interaction effects, according to Kuznets, may be among the most productive from the standpoint of economic growth. "Their [migrants'] very rootlessness may promote adjustment to new opportunities" (Kuznets, 1957, p. 3). Schultz and Sjaastad would not deny the selectivity hypothesis in migration. They would, however, explain the migration of the select people, for example, the younger, the better educated, or more enterprising people, by the expectancy of higher income differentials or longer payoff periods of the streams of future income, rather than merely saying that particular select groups have a higher propensity to migrate.

A third explanation of internal migration comes from the late nineteenth- and the early twentieth-century English economists, notably Ravenstein (1885, 1889) and Redford (1926). Their studies are still widely quoted. The forces that determine migration were called "the laws of migration" by Ravenstein. In substance, this explanation is simply in terms of the "push" and "pull" factors. People are pushed from rural areas by factors such as outmoded land-tenure systems, unfavorable terms of trade, wide dispersion

[2] In particular, see the papers by Kuznets and Thomas (volumes of 1957, 1960, 1964).

of property and income, pressure of rural poverty in general, and so on. Migration may occur despite the fact that the migrants may only be able to drift into petty jobs like those of a porter, shoe shine, or underemployed casual laborer, rather than remunerative employment in modern sectors. Similarly, the "pull" factors, apart from the employment, education, and other opportunities, may also consist simply of the "bright lights" of the towns, even though improvements may be taking place in the countryside. According to this hypothesis, therefore, migrants are not necessarily motivated by economic (cost-and-return) calculations. Yet, while the English explanation of the motivations underlying internal migration materially differs from the American explanations, the consequences are similar. Thus, according to Ravenstein, "Migration means life and progress; a sedentary population stagnation" (1889, p. 288).

It should be clear from the brief survey of the theories of internal migration that not only the consequences but also most of the explanatory variables in these approaches are the same. The differences arise mainly in emphasis and interpretation.

Rather than assuming one of these approaches in advance, this study in part blends and in part tests all of them. It differs from others in the formulation of the model, the specification and estimation of the functions, the kind of data used, and in the study of an underdeveloped economy of special interest. The basic migration equation is formulated in the costs-and-returns framework of Schultz and Sjaastad. The selectivity and the interaction effects of migration and economic development of Kuznets are included by allowing migration and wage rates to be determined endogenously in a simultaneous system, by analyzing subsamples of different age groups of migrants, using education by age groups and by sectors as the explanatory variables, and so on. As already remarked, these variables are also very relevant in the capital-theoretic approach of the Chicago School—only the interpretation and emphasis, in part, differ. Finally, the "push" and "pull" factors are tested, very roughly, by analyzing the effects of variables, such as the dispersion of incomes, urbanization, and so forth, which cannot be entirely called economic. In addition, the residual effects are measured by the use of the region dummy variables which, though not efficacious in separating the (residual) economic effects from the non-economic ones, at least allow for their combined effects, thereby sparing the rest from being biased on that account.

3. The Data

The model describing migrants' behavior in Brazil cannot be formulated out of context with the data available. The major source of data on internal migration is the census of population. The latest census for which internal migration data are available is the 1950 census, which was therefore used.

(The relevant data for the 1960 census have not been published or even processed as of September, 1967.) The best information available on migration is the distribution of population by age and sex and by state of birth and state of residence on the day of enumeration of data. It is thus a snapshot of the level at a point of time, rather than a flow measure. As such, it lumps together temporary migrants and permanent migrants, old migrants and new migrants, and so on.[3] The data are obviously not ideal, but are the only ones available. The available evidence, however, indicates that in the interstate migration in Brazil the seasonal elements are insignificant. Besides, as will become apparent in the succeeding pages, the migration function estimated here includes mostly the explanatory variables which reflect relatively permanent rather than transitory influences. The major exceptions are the rates of growth of incomes and, to some extent, wage rates. More on this later.

Several of the explanatory variables used in this study, for example, education and urbanization, are also available in the population census. The remaining variables come from diverse sources, almost all of which are reported in IBGE, *Anúario Estatístico do Brasil* (annual).

The regional wage rates were calculated by dividing the salaries and wages paid to production workers in manufacturing by the average number of production workers. The data on education are available by age groups and by sectors (agricultural and non-agricultural) for individual states. The index of education used in this study consists of the aggregate number of primary, high school, and college male graduates in the given age group or economic sector as a percentage of the total male population in the corresponding age group or sector. This index was considered to be preferable to the others available from the census of population. The dispersion of incomes was calculated, for the rural sector only, from the size holdings of cultivable land, under the assumption that rural incomes are closely associated with the ownership of cultivable land. The relevant data for the urban centers are not available. Farm establishments by size holding are available, from the census of agriculture, for five categories only. A few of the sketches of the distribution of farm establishments gave little indication of these being unimodal. Hence, no standard measures like the Lorenz curve or Gini index were employed. Instead, a rough index of the dispersion of landed property was prepared by taking the ratio of the average size holdings of the lowest 20 per cent of farm establishments to the weighted regional average of the remaining 80 per cent of the establishments. This index was used as a proxy for the dispersion of incomes in the respective states.

[3] This is an unfortunate fact of the form in which the data are available—not an infrequent experience of the researchers engaged in empirical studies of the underdeveloped countries. The inclusion of an additional question in the census enumeration, for example, on "the number of years of residence in the destination state," would have made the migration study a considerably more interesting one.

For this study, data were collected for nineteen of the states (or "regions") of Brazil. The map (Fig. 1) shows the state boundaries, density of population, and per capita income of various states. The observations on the dependent variable, migration, consist of the number of adult males born in state i and residing in state j. There are, thus 342 ($=19 \times 18$)

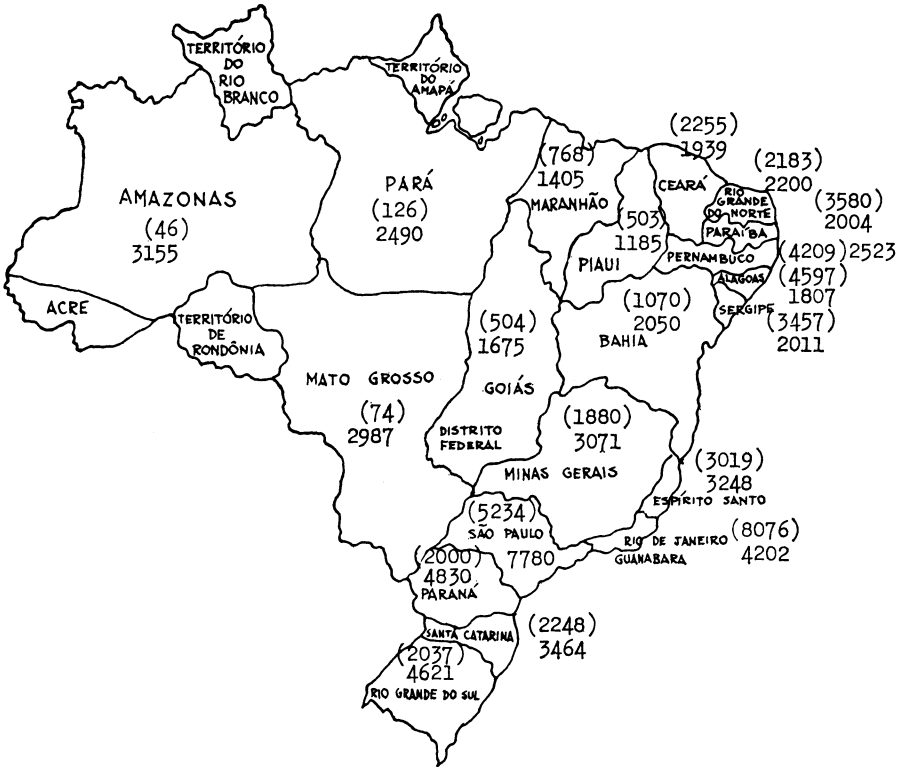

FIG. 1.—Sketch of the states of Brazil with per capita incomes and densities, The first figure (the value in parentheses) for each state studied in this paper refers to density of population per 100 square kilometers in 1960, and the second figure (the value without parentheses), refers to (estimated) per capita income, of 1960 in 1949 cruzeiros.

observations in the sample. The important variables of the model are briefly discussed below.

4. The Variables

The Migration Variable

The study is confined to adult males. Migration is studied by age groups: the migrants in the age group 15–29 years (to be denoted as "young"

migrants) and the migrants in the age group 30–59 (to be denoted as "middle-aged" migrants). The proportion of the people that migrated in the age group 15–29 to the total population in that group is slightly higher than the corresponding proportion pertaining to the age group 30–59. More important than the comparison of the numbers of migrants in the two age groups, however, is the contrast between the responses of the two age groups to given stimuli. Data on the explanatory variables are also available by age groups. Hence, the breakdown by age groups (along with education, etc.) is presumed to shed some light on the demographic selectivity.

The most important explanatory variables tried in this study are the *wage rate* (also treated as an endogenous variable), *education*, *urbanization*, *density* of population, *geographic distance*, and the *level*, the *rate of growth*, and the *dispersion of income*, both in the origin region and the destination region.

Wage Rate

Internal migration may result from and influence geographical differences in the productivity of labor. The differences are usually reflected in wage differences. The wage rate, therefore, is assumed to be the most important economic variable influencing migration. An individual's decision to migrate is presumably not based on average income but on the income he can earn, given his training and dexterity. The majority of the migrants are generally considered to come from rural areas to urban areas, or from smaller towns to larger ones. The most relevant expected income in the destination region to a migrant seems to be the prevailing wage of the workers. Due to the lack of data on migration by occupations, it was assumed that the relative position of an occupation in the wage structure is nearly the same in all regions, even though levels may differ, and that the average wage rate in the manufacturing industries is the most relevant one for the purpose in hand.

Here, two observations are in order. First, the living costs may differ between regions. Since a majority of the migrants are supposed to come from the (low cost-of-living) agricultural regions to the (high cost-of-living) industrial regions, the observed money wage rates may overstate the real wage rates. To the extent that is so, the coefficient of migration with respect to wage rate will be underestimated. The coefficient of wage rate (when used as a dependent variable) with respect to its explanatory variables, on the other hand, will be overestimated. Due to the lack of reliable cost-of-living data at the state level for the census year under reference, it was thought that an arbitrary correction of the observed wage rates to take account of cost-of-living differences might introduce an even greater bias. Accordingly, no such correction was attempted. Second, perhaps the

expected wage rate rather than, or in addition to, the current wage rate would be the more appropriate explanatory variable. However, the form in which the data on the migration variable are available (see above) would not lend itself to appropriate lag models. Hence, only the current-period wage rate was used in this study.

Education

Education is an important variable that may account for systematic differences in individual responses. The effects of the so-called brain drain of educated individuals from the less developed countries to the relatively more developed ones, particularly the United States, have recently been brought into the limelight (see, for example, H. G. Johnson, 1964).

The same (dissatisfaction) hypothesis may also be maintained with regard to internal migration: Educated individuals are regarded as relatively more mobile and adaptable and also brighter and more alert to changing opportunities. The increased earning capacity per se of the educated individuals is widely acknowledged. What, however, is not always recognized is the increased capacity and option of an educated individual *to learn still more*[4] (and perhaps also to explore more). From this it does not necessarily follow, however, that the migration rate will be relatively higher among the educated individuals. For, in general, high-income regions are also the high-education regions. Since the educated individuals of the origin region have to compete with the educated individuals of the destination region, *ceteris paribus*, a high level of education in the destination region may serve as a deterrent to the educated immigrants. Moreover, the average incomes in such destination regions may overstate the wage rates available to the less educated migrants, while the average incomes in the origin region may understate the wage rate obtainable by the more educated individuals. Mere social status may provide a psychic income. A priori, therefore, it is not possible to predict the influence of education one way or the other. We shall wait to see what the empirical results reveal.

Urbanization

Cities are considered to be the dynamic centers that, through history, have been the cradles of civilization, progress, and revolutions. Dynamic individuals may be attracted to these centers. Moreover, the urban centers offer superior educational opportunities, more amenities, wider contacts, and other benefits. At the same time, according to the findings of one researcher, the rural-urban migration may often be only a fraction of the

[4] See, for example, Schultz (1962) and Gisser (1965).

urban-urban migration, and the rural migrants may move to big urban centers by stages (Herrick, 1965). Acclimation to a money economy and superior media of information about jobs may be among the causes of the latter phenomena. During unemployment periods, there may even be a reverse movement from the cities to the villages. Thus, the urbanization effect may go either way. On the whole, however, we would expect a net pull effect of urban centers on the rural workers.

Two variables were tried as the indexes of urbanization: (*a*) the proportion of population living in cities of 5,000 people or more, and (*b*) the proportion of state income originating from manufactures as a proxy for industrial-urban centers.

Density

Density may attract migrants, or it may serve as a "push" factor. It may affect migration both for the origin region and the destination region. The push factor may be due to the pressure of population. On the other side, initial density might have partly resulted from earlier immigration. If so, old migrants may attract new migrants simply because the latter are more aware of the advantages of migrating and may be assured of aid and information about jobs. This might start a snowball effect. Density may serve as a proxy for this snowball effect.

Geographic Distance

Distance is a variable that may represent more than mere economic costs. The information and knowledge about jobs is closely related to distance. Differences in language, dialects, dress, food, social customs, and cultural moves are generally accentuated by distance. Uncertainty about job opportunities may increase with distance. In the absence of a fully specified migration function, therefore, distance may serve as a proxy variable for the omitted factors. In this study, however, I shall be using the dummy variables for regions or groups of regions which, rather than distance, are expected to take care of most of the effects of the omitted variables. The distance variable is, therefore, hypothesized to account mainly (but not solely) for the economic *costs*—which, in addition to the out-of-pocket money costs incurred in traveling, and so forth, by the migrant and his family, should include income foregone during the transfer period. It is predicted, therefore, that distance will be negatively correlated with migration.

The Level, the Rate of Growth, and the Dispersion of Per Capita Income

Per capita income may serve as a substitute for wage rate. In some of the regressions of this study, per capita income was tried as the income foregone in the origin region, and wage rate as the expected income in the

destination region. This assumption was based on the possible prevalence of the joint-family system in the rural areas and single-family system in the urban areas. *The rate of growth of income*, net of the rate of growth of population, was regarded as an index of the creation of employment opportunities (there being no data on the rate of unemployment). Finally, the *dispersion of income* among households may influence migration. For instance, for any given level of average income, high dispersion of income—usually associated with higher percentages of the population receiving incomes well below average—may, *ceteris paribus*, mean high rates of migration. This may be accentuated in times of adversity (such as those due to crop failures or adverse terms of trade) when, for example, the push factors may impinge harder upon regions with a higher degree of income dispersion than those with a relatively lower degree of income dispersion. As a matter of fact, in many underdeveloped countries, land reforms which distribute land more equally among the peasant families are usually considered to result, other things being equal, in a relatively contented and stable peasantry with attachment to their homes and discouragement to migration. Clearly, this variable can be included in the push-pull category.

Capital gains that may be foregone through migration from agriculture are often suggested as being a relevant variable for migration study (D. G. Johnson, 1963). This factor should act as a disincentive to migration as, due to high inflation in Brazil, real estate is considered among the safest and most highly lucrative investments. No data are available to account for the effects of this factor in Brazil.

5. The Single-Equation Model and Substantive Results

Recall that the measure of the interstate migration is the number of adult males born in region i (the origin region) and residing in region j (the destination region). All variables pertain to the year 1950 unless otherwise stated. The variables are defined below:

M_{ij} = absolute number of migrants from region i to region j; M_i = total migration from region i; $m_{ij} = M_{ij}/M_i$.

w_i, w_j = wage rate per production worker per year in cruzeiros in region i, region j.

y_i, y_j = per capita income in cruzeiros in region i, region j.

Y_{mi}, Y_{mj} = income originating in manufacturing in region i, region j.

δ_i, δ_j = a measure of the dispersion of income in region i, region j.

g_i, g_j = two-year rate of growth of the state's per capita income, 1949–51, in region i, region j.

E_{1i}, E_{1j} = percentage of educated people in the age group 15–29 years in region i, region j.
E_{2i}, E_{2j} = percentage of educated people in the age group 30–59 years in region i, region j.
E_{3i}, E_{3j} = percentage of educated people in the agricultural sector in region i, region j.
E_{4i}, E_{4j} = percentage of educated people in the non-agricultural sector in region i, region j.
U_i, U_j = index of urbanization as measured by the percentage of population in cities with population of 5,000 or over in region i, region j.
d_i, d_j = density of population per square kilometer in region i, region j.
D_{ij} = geographic distance in kilometers from the capital city of region i to the capital city of region j.

Starting with the costs-and-returns approach, potential returns (R) to a migrant from an origin region i to a destination region j may be approximated by

$$R_{ij} = R(w_j; E_j; g_j; \ldots), \tag{1}$$

where E_j is to be interpreted as the earning potential in the destination region specific to the average education attainment of the migrants from region i. The other variables are to be interpreted likewise.

Costs of migration (C) may be expressed as

$$C_{ij} = C(w_i; E_i; g_i; D_{ij}; \ldots), \tag{2}$$

where w_i is the origin-region wage income foregone, E_i is the earning potential in the origin region specific to the average education attainment of the migrants from region i, and so on for other variables.

The complete migration function, then, is

$$M_{ij} = M(R_{ij}, C_{ij}) = M(w_j; w_i; E_j; E_i; g_j; g_i; D_{ij}; \ldots). \tag{3}$$

As stated earlier, even though three broad theoretical approaches to migration have been distinguished, not only the effects of migration in all of them are recognized to be similar, but most of the explanatory variables also are common. For example, education is common to all but is supposed to work through higher potential lifetime earnings according to the Chicago School approach; through selectivity, *inter alia*, according to the Harvard approach; and through the push and pull of employment opportunities, and so forth, according to the English approach. Where possible, particular influences will be distinguished by the introduction into equation (3) of additional variables, for example, income dispersion for push-pull factors, and so on.

With a view to determining the algebraic form of the migration function, several alternative constructions were tested. The double-log form gave the best fit (as judged from the standard errors of prediction converted to comparable measure), except that the distance variable acquired a higher level of statistical significance when entered as the square root of absolute kilometers. Likewise, a number of other specifications of the variable forms were also tested. These are discussed in a footnote below.[5] The resulting migration function with a few of the key variables, therefore, is of the following form

$$M_{ij} = w_j^{a_1} w_i^{a_2} E_j^{a_3} E_i^{a_4} g_j^{a_5} g_i^{a_6} U_j^{a_7} U_i^{a_8} \exp\left(a_9\sqrt{D_{ij}} + a_{00} + \sum_{i=2}^{I} a_{0i}\theta_i + v_{ij}\right), \tag{4}$$

[5] In particular, the following forms of the dependent variable were tested:

M_{ij} = the absolute number of migrants from region i to region j, giving an over-all migrant responsiveness.

$M_{ij}/M_i = m_{ij}$, that is, the number of migrants to region j from region i as a percentage of total migration from region i, which is supposed, in general, to perform as a relatively more appropriate "allocation variable."

$M_{ij}/n_i = (M_{ij}/M_i)\,(M_i/n_i)$, where n_i stands for the total population of state i, M_{ij} stands for m_{ij} weighted by the proportion of population migrating from region i where m_{ij} is an unweighted category. (Sjaastad examined the form $M_i/M_{ij} = 1/m_{ij}$ as the dependent variable. I did not try it.)

Likewise, the following three forms of the explanatory variables were tested, namely, considering the education variable, E, for illustration,

$$m_{ij} = (E_j/E_i)^{\alpha_1} \ldots, \quad \text{(i)}$$
$$m_{ij} = E_j^{\alpha_2}/E_i^{\beta_2} \ldots, \quad \text{(ii)}$$
$$m_{ij} = (E_j - E_i)^{\alpha_3} \ldots. \quad \text{(iii)}$$

It may be remarked that form (i) hypothesizes that migrants respond to relative differences in the variables and that the elasticities of migration with respect to a given variable in the origin region and the destination region are equal and opposite in sign. The imposition of an identical elasticity in the two regions i and j will be referred to as the "homogeneity restriction." In this form, the variable series E_i and E_j, and so on, will be converted to one series E_i/E_j, and so on. A minor gain in such a specification is in using up a lesser number of degrees of freedom and a possible reduction of multicollinearity. Form (ii) assumes that migrants' responses to identical percentage changes in a given variable in the origin region and the destination region can differ, that is, that the elasticities of a variable in the origin region and the destination region can be unequal. When $|\alpha_2| = |\beta_2|$, (ii) will be identical to (i). Form (iii) implies that migrants respond to absolute, rather than relative, differences in the variables in the two regions. The results of the variable forms (i) and (ii) only are reported in the text.

It may be of interest to note that in his empirical study, Sjaastad used only the j variables and omitted the i variables altogether (1961). The effects of the omitted i variables, in his study, will either appear in the general constant (when none of the i variables is correlated with any of the j variables), or in the coefficients of the j variables (or the error terms) and so bias the coefficients (when some of the i variables are correlated with the j variables). Since most of the variables are likely to change in the same direction in most of the states, the coefficients of the j variables will have a tendency to be biased downward.

where v_{ij} is a disturbance term assumed to have zero mean and constant variance and uncorrelated with the explanatory variables. The θ_i are the region dummy variables that take the value of one for the region concerned and zero for others. The symbol "exp" stands for "exponential." The above model will be estimated by the single-equation least-squares

TABLE 1

INTERNAL MIGRATION IN BRAZIL, AGE GROUP 15–29 YEARS (DEPENDENT VARIABLE: THE LOG OF THE NUMBER OF MIGRANTS, $\text{Log}_e M_{ij}$) *

Coefficient of (1)	Regression 1 (2)	Regression 2 (3)	Regression 3 (4)
Constant	1.35	0.13	−11.70
$\text{Log}_e d_j$	1.35 (2.99)[5.00]	...	...
$\text{Log}_e d_i$	−0.02 (− 0.08)[4.99]	...	...
$\text{Log}_e w_j$	1.81 (2.94)[8.60]	2.08 (3.84)[8.60]	...
$\text{Log}_e w_i$	0.24 (0.60)[8.61]	...	...
$\text{Log}_e y_j$	...	...	2.25 (4.03)[7.87]
$\text{Log}_e y_i$	...	−1.53 (− 3.65)[7.88]	−0.77 (− 1.90)[7.87]
$\text{Log}_e (Y_{mj}/Y_j)$	...	...	0.46 (− 0.03)[2.62]
$\text{Log}_e (Y_{mi}/Y_i)$	...	...	−0.01 (− 0.03)[2.62]
$\text{Log}_e \delta_j$	...	...	0.05 (0.94)[0.43]
$\text{Log}_e \delta_i$	...	...	−0.13 (− 1.17)[0.45]
$\text{Log}_e E_{(1,2)j}$	0.26 (0.57)[2.61]	...	0.37 (0.90)[2.56]
$\text{Log}_e E_{(1,2)i}$	−0.71 (− 3.17)[2.61]	...	−0.95 (− 4.28)[2.56]
$\text{Log}_e E_{3j}$	...	0.80 (0.40)[1.39]	...
$\text{Log}_e E_{3i}$	...	−0.35 (− 1.64)[1.39]	...
$\text{Log}_e E_{4j}$	...	0.41 (1.92)[3.33]	...
$\text{Log}_e E_{4i}$	...	−1.56 (− 2.61)[3.33]	...
$\text{Log}_e g_j$	1.28 (1.14)[2.32]	...	...
$\text{Log}_e g_i$	−3.38 (− 3.97)[2.33]	...	...
U_j	−1.00 (− 1.37)[0.29]	...	...
$\text{Log}_e U_j$	...	−0.09 (− 0.55)[1.33]	...
U_i	2.50 (2.59)[0.30]	...	...
$\text{Log}_e U_i$	...	0.69 (2.80)[1.34]	...
Dummy 1	−0.18 (− 0.26)[0.50]	−1.62 (− 2.62)[0.05]	−2.16 (− 2.75)[0.05]
Dummy 2	−0.33 (− 0.87)[0.37]	−0.13 (− 0.31)[0.37]	−0.84 (− 1.95)[0.37]
Dummy 3	−0.82 (− 2.45)[0.47]	−0.70 (− 2.02)[0.47]	−1.46 (− 3.58)[0.47]
$\sqrt{D_{ij}}$	−0.117(−14.18)[38.68]	−0.111(−18.61)[38.68]	−0.116(−14.07)[38.68]
S.E.E.	1.73	1.73	1.76
R^2	0.520	0.514	0.497
d.f.	327	328	329

* The numbers in parentheses are t-ratios of the respective coefficients preceding these numbers. The figures in square brackets are the mean values. The mean value of $\log_e M_{ij} = 4.87$ and $\log_e M_i = 9.89$. The symbol i stands for the origin state, and j for the destination state. Due to the limitations of computer capacity at the University of São Paulo, the dummy variables were entered for certain groups of i states rather than for each of the i states. Thus, Dummy 1 stands for the state of Amazon. Dummy 2 is comprised of Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, and Alagoas. Dummy 3 includes Sergipe, Bahia, Minas Gerais, East Santos, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, and Rio Grande do Sul. The dummy for Mato Grosso and Goiás was suppressed to serve as a basis for comparison and to preserve non-singularity of the variable matrix in the presence of the general constant term. S.E.E. = standard error of estimate; R^2 = the coefficient of determination; and d.f. = degrees of freedom.

The mean values (in logs or otherwise, as the case may be) were measured in the units given below:

d	= the number of persons per square kilometer;
w	= wage rate per man per year in cruzeiros (1950 prices);
y	= per capita income in cruzeiros (1950 prices);
Y_m/Y	= percentage of income originating in manufacturing;
δ	= the ratio of average land holding of the farm establishments in the lowest 20 per cent size holdings divided by the weighted average size holding of the remaining 80 per cent;
g	= two-year rate of growth of income (1949–51) net of the rate of growth of population for the corresponding period;
U	= proportion of population residing in towns of 5,000 or over;
Dummies	= the number of observations for the dummy concerned divided by the total sample (342 observations) (absolute values, not logs);
$\sqrt{D}$	= the square root of road distance in kilometers from the capital of state i to that of state j;
E	= the number of primary, high school, and college male graduates in the given age group or sector as a percentage of total males in the corresponding age group or sector.

method. The simultaneity problem will be tackled in the succeeding section.

The matrix of intercorrelations among the variables indicates a disturbingly high collinearity between education and wage rate. This may result

TABLE 2

INTERNAL MIGRATION IN BRAZIL, AGE GROUP 30–59 YEARS (DEPENDENT VARIABLE: THE LOG OF THE NUMBER OF MIGRANTS, $\text{LOG}_e M_{ij}$) *

Coefficient of (1)	Regression 4 (2)	Regression 5 (3)	Regression 6 (4)
Constant	1.44	−0.70	−12.80
$\text{Log}_e d_j$	0.93 (2.23)	...	...
$\text{Log}_e d_i$	0.18 (0.61)	...	...
$\text{Log}_e w_j$	1.05 (1.84)	1.76 (3.36)	...
$\text{Log}_e w_i$	−1.20 (− 2.69)	...	...
$\text{Log}_e y_j$	...	...	1.82 (3.44)
$\text{Log}_e y_i$	...	−1.70 (− 4.18)	−1.69 (− 3.83)
$\text{Log}_e Y_{mj}$	...	...	0.59 (2.73)
$\text{Log}_e Y_{mi}$	...	...	−0.25 (− 0.79)
$\text{Log}_e \delta_j$	...	...	...
$\text{Log}_e \delta_i$	...	...	...
$\text{Log}_e E_{(1,2)j}$	0.87 (2.06)	...	0.72 (1.89)
$\text{Log}_e E_{(1,2)i}$	−1.63 (− 5.95)	...	−1.80 (− 7.25)
$\text{Log}_e E_{3j}$	...	0.11 (0.62)	...
$\text{Log}_e E_{3i}$	...	−0.48 (− 2.32)	...
$\text{Log}_e E_{4j}$	...	0.47 (2.28)	...
$\text{Log}_e E_{4i}$	...	−1.82 (− 3.14)	...
$\text{Log}_e g_j$	0.08 (0.08)	...	...
$\text{Log}_e g_i$	−4.70 (− 5.07)	...	...
U_j	−0.00003(− 0.39)	...	...
$\text{Log}_e U_j$	...	0.017(0.11)	...
U_i	0.0003 (3.58)	...	...
$\text{Log}_e U_i$	...	0.91 (3.81)	...
Dummy 1	−0.55 (− 0.82)	−1.88 (− 3.14)	−1.79 (− 2.21)
Dummy 2	−0.36 (− 0.86)	−0.50 (− 1.24)	−0.77 (− 1.61)
Dummy 3	−0.90 (− 2.80)	−0.87 (− 2.63)	−1.27 (− 3.08)
$\sqrt{D_{ij}}$	−0.112 (−14.40)	−0.106(−13.39)	−0.111(−14.16)
S.E.E.	1.63	1.67	1.66
R^2	0.571	0.547	0.549
d.f.	327	328	331

* The mean value of $\log_e M_{ij}$ is 5.237 and $\log_e M_i$ is 10.06. For other notes, see the footnote to Table 1.

in large standard errors (low degree of precision) of the coefficients concerned, and it may become difficult to obtain the estimates of the independent (though not necessarily the joint) effects of these variables. In interpreting the coefficients of education and earnings, therefore, such damaging effects of multicollinearity will be kept in mind.

The regression results are presented in Tables 1, 2, and 3. Tables 1 and 2 give alternative regression results for the age groups 15–29 and 30–59, respectively, with $\log_e M_{ij}$ as the dependent variable. Table 3 presents a

few results under the homogeneity restriction (as defined in footnote 5), using M_{ij}/M_i as the dependent variable.

Most of the coefficients are significantly different from zero at the conventional levels and have the right and predicted signs. Distance comes out with a negative coefficient in all of the regressions, signifying that it is a definite deterrent to migration. The "young" migrants appear to concentrate slightly more on a smaller number of destination states than the

TABLE 3

INTERNAL MIGRATION IN BRAZIL (DEPENDENT VARIABLE: $\text{LOG}_e\, m_{ij} = \text{LOG}_e\, [M_{ij}/M_i]$) *

Coefficient of (1)	Age Group 15–29 Regression 7 (2)	Age Group 30–59 Regression 8 (3)
Constant	0.29	0.31
$\text{Log}_e \tilde{d}$	0.160 (5.26)	0.159 (5.28)
$\text{Log}_e \tilde{Y}_m$	0.053 (2.13)	0.050 (2.02)
$\text{Log}_e \tilde{w}$	0.081 (2.19)	0.067 (1.81)
$\text{Log}_e \tilde{y}$	0.280 (7.17)	0.312 (7.67)
$\text{Log}_e \tilde{E}_{(1,2)}$	−0.082 (− 3.46)	−0.177 (− 4.21)
$\text{Log}_e \tilde{E}_3$	0.017 (1.06)	0.059 (2.69)
$\text{Log}_e \tilde{E}_4$	−0.049 (− 2.54)	−0.036 (− 1.87)
$\text{Log}_e \tilde{g}$	0.314 (10.53)	0.335 (11.34)
$\text{Log}_e \tilde{U}$	0.0006(0.73)	0.0007(0.91)
Dummy 1	−0.320 (−12.12)	−0.334 (−12.89)
Dummy 2	−0.0008(− 0.18)	0.0002(0.04)
Dummy 3	0.138 (1.95)	0.122 (1.74)
$\sqrt{D_{ij}}$	−0.033 (− 3.03)	−0.028 (− 2.59)
S.E.E.	0.150	0.148
R^2	0.569	0.577
d.f.	327	327

* The variables with tilde (~) stand for the value of the variable concerned for region j relative to that in region i, for example, $\tilde{d} = d_j/d_i$.

"middle-aged" migrants. Thus, the logarithmic average allocations (m_{ij}) for the age groups 15–29 and 30–59 are 5.02 and 4.82, respectively. I shall interpret the results briefly for each of the variables.

Density (d) is a strong pull factor in the destination region with an elasticity of 1.35 for the "young" migrants and 0.93 for the "middle-aged" migrants. The elasticity coefficient is significantly different from zero at the 1 per cent level in all of the regressions. When the homogeneity restriction is imposed (Regressions 7 and 8), the elasticities for the two age groups become almost identical. Note that the dependent variable in Regressions 7 and 8 is the allocation of migration (M_{ij}/M_i) rather than the absolute number of migrants M_{ij}. The lower value of elasticity in these regressions, therefore, is due to the change in units of measurement.

Density of the origin region seems to have no effect on migration. For the "young" group, the density coefficient even has the wrong sign, but it is not significantly different from zero in any of the regressions.

The *earnings* variable (wage rate [w] or per capita income [y]) is of special interest here. It is significant in both the origin region and the destination region and has the right and expected sign. The elasticity of migration with respect to earnings in the destination region is approximately 2 for the "young" group and approximately 1 for the "middle-aged" group. (Compare also the coefficients of $\tilde{w}$ in Regressions 7 and 8 of Table 3.) In the case of the latter age group, equivalent changes in earnings in the origin region and the destination region almost cancel out (see Regressions 4, 5, and 6). The "young" migrants, on the other hand, appear to be attracted more by a rise in earnings in the destination region than a similar rise in the origin region (observe Regressions 1, 2, and 3).

The computed disparity of responses between the two age groups appears, in part, to be due to the already noted collinearity between education and wage rate. Thus, the coefficient of education in region j is consistently lower for the "young" group than that for the "middle-aged" group, indicating (as is usually the case under multicollinearity) that part of the weight of the education variable in the case of the "young" group has perhaps been tilted in favor of the wage variable. This may also be checked by comparing the sums of the coefficients of education and wage rate—the sums of coefficients being a better predictor when there is multicollinearity (provided the same multicollinearity continues to exist in the period of prediction). These sums for the two age groups are much closer to each other than the corresponding individual coefficients are. The overall results of the regressions of Tables 1 and 2 give the impression that, in the case of the "young" group, the existence of intercorrelation between wage rate and education has possibly led to an overestimate of the coefficient of wage rate and (correspondingly) an underestimate of the coefficient of education.

The above is only one possibility. The other interpretation is that the "young" migrants are definitely more sensitive to wage differentials than are the older migrants. This is supported by the alternative computations presented in Table 3—in which multicollinearity is supposed to be reduced as a result of the homogeneity restriction—where both of the coefficients of education and earnings are higher for the "young" group than those for the "middle-aged" group. The magnitudes of the sums of the coefficients of education and earnings, just discussed, also lend support to this interpretation. Thus, even though these sums are closer to each other than are the corresponding individual coefficients, the sum of these coefficients for the "young" group is consistently higher than that in the corresponding regressions for the "middle-aged" group.

The computed higher sensitivity of the "young" migrants to wage differentials is susceptible to alternative explanations under the three approaches discussed in the beginning of this study. Under the "selectivity" hypothesis, for instance, it may be interpreted to mean that the young people are more adventurous relative to their older counterparts. It seems more logical to infer, however, that the greater responsiveness of younger workers is in anticipation of a longer payoff period for them than for the older workers.

Dispersion of income (δ) is significant neither in the origin region nor in the destination region. It will be recalled that the index of the dispersion of income prepared from the size distribution of agricultural land is, at best, very crude. A possible explanation for this result may, thus, be found in the untrustworthiness of the data used for this variable or the inappropriateness of the proxy variable. Since the presence of such a variable will affect not only estimates of its own coefficient but others as well, this variable was omitted from all but one regression.

Industrialization gives a disappointing result, in the unrestricted equations of Tables 1 and 2. It is a significant variable in the case of the "middle-aged" group and that, too, for the destination region only. In the regressions presented in Tables 1 and 2, the industrialization variable was entered along with the variable of per capita income (y). These two variables, however, are intercorrelated. As such, it may be possible that the y variable has "robbed" the industrialization variable. Additional regressions (not reported here) with the industrialization variable only (dropping the y variable), however, did not improve the effect of industrialization significantly. The possible explanation appears to be that industrialization per se is not particularly attractive to migrants. Why should it be, with its smoke-stack chimneys and dusty and hot jobs, anyway? Whatever attraction there is in this variable as such is perhaps largely due to the other factors that are associated with industrialization (and not taken care of separately), such as the existence in the industrial centers of old relatives and acquaintances who may be the disseminators of information and who may be helpful in other ways to the newcomers, at least in the beginning. This may, of course, create a snowball-like effect.

Education, on the whole, is a significant variable. Three measures of education were tried. These, it may be recalled, are education by age group ($E_{1,2}$), education in the agricultural sector (E_3), and education in the urban sector (E_4). Considering only the *signs* of the coefficients, education comes out, by and large, as an attraction to migrants in the destination region and a deterrent to migration in the origin region. For the "young" group, however, only two of the nine coefficients of education are significantly different from zero at the 5 per cent level (Tables 1 and 3). It is quite likely, as already discussed, that these low significance levels have been caused by a strong collinearity between education and wage rate. In the case of the

"middle-aged" group, on the other hand, all of the nine coefficients are significant (Tables 2 and 3). The education elasticities of Table 2 indicate that the deterrent effect of the origin-region education is from twice to four times as strong as the attractiveness of the destination-region education. In particular, a 1 per cent rise in the destination-region education of the "middle-aged" group increases migration (of the same age group) by about 0.0075 per cent. An equivalent rise in the origin-region education deters migration, on an average, by about 1.72 per cent. The corresponding elasticities for the destination and origin regions for E_3 and E_4 are 0.11 and -0.48, respectively, for rural education, and 0.47 and -1.82, respectively, for urban education. The rate of increase of the destination-region education, thus, has to be roughly four times as high as that in the origin region to cancel out the sedentary effect of the latter. No wonder that, in the restricted regressions of Table 3, the education coefficient is persistently negative.

In a recent paper, T. W. Schultz (forthcoming) has advanced a hypothesis of which the present results appear to be a direct test. Schultz asks the question: "Are large farms necessarily inefficient?" His answer is couched in terms of entrepreneurial talent: Large farms require highly competent entrepreneurs; there are undoubtedly a considerable number of entrepreneurs who are equal to the task of managing large complex Chilean farms, but they are in short supply. Schultz derives his hypothesis in the following words:

> It is a puzzle: why should the supply of this entrepreneurial talent be so small? To get at this puzzle, I would advance the following hypothesis: Individuals who have the necessary entrepreneurial talent to manage a large farm can command a level of income which is normally sufficient to acquire first rate schooling for their children and other social services, none of which are available to families who reside on large farms. Outside of the larger cities, Chile—unlike California—has no first-rate schools, either elementary or secondary, and none of the social services that middle and high income families want and can afford. Thus in Chile it means a divided family; for the wife and children must live in a city. This arrangement on the part of the family is unacceptable to most individuals (husbands) who have the necessary competence to manage a large farm [pp. 7–8].

The negative coefficient of local education in the present study (assuming a higher level of local education *ipso facto* implies higher-quality education) is, thus, quite consistent with Schultz's hypothesis.

Apart from its direct effects, education may also have certain indirect effects on migration. These, apart from higher earnings, may include such things as the information and awareness of opportunities at home and

elsewhere, prestige and snob appeal at home, and so on. Some of these effects might have been picked up by the dummy or distance variables. Moreover, due to a high collinearity with wage rate, the possibility of the equilibrium values of the education parameter being tilted one way or the other due to the inclusion or exclusion of some extreme observations in another sample cannot be dismissed. Since education is likely to have a trend component in all the states, while it may be difficult to realize a 4:1 disparity in the rate of growth of education in different states, the most we can conclude from the present regressions is that, on the whole, the *direct net* effects of education on migration are not likely to be substantial.

The coefficient of the *rate of growth of income* of the destination region has the right sign (though not significantly so) in the unrestricted regressions of Tables 1 and 2. The growth variable of the origin region, on the other hand, has the right and expected sign and is highly significant for both of the age groups (Tables 1 and 2). In the restricted regressions of Table 3, the (relative) growth variable has the right and expected sign and significantly so.

Due to the unfortunate fact that the dependent variable relates to migrants who migrated at various times in the past, a single growth rate—a two-year growth rate here—will not be a very apt explanatory variable unless the growth rates have continued to be of an equivalent order of magnitude during the past several years. Furthermore, there may be a lag in the dissemination of information, a lag in the response to the stimulus provided by growth rates, and so on. In view of this, further trials were made by using the lag values as well as the five-year average values of the growth variable. In these latter regressions, the significance level of the coefficient of destination region, indeed, goes up, but could scarcely be stretched to be significant even at the 10 per cent level.

The *urbanization* coefficient, too, has the wrong sign. As the results indicate, the origin-region urbanization, *ceteris paribus*, encourages migration, while destination-urbanization deters migration. The urbanization coefficient is significant for the origin region for both of the age groups. It is, however, not significant for the destination region. The results are partially in agreement with those of Herrick (1965) for Chile. Herrick found that the city-to-city migration is much the more dominant phenomenon in comparison with the rural-urban migration. In part, however, my conclusion is in conflict with Herrick's, in that in his study the bigger the city, the stronger the pull. For instance, in the past three decades, the capital city of Santiago, with almost one-third of Chile's population, has shown a kind of gravitational pull for the migrants. In my study, other things remaining constant, the urban people seem to be anxious to emigrate, but where they want to go is not indicated by the urbanization variable by itself. In a way, this conclusion appears to support the proposition made in Section 2 to the effect that the urban people are relatively

more mobile. Their mobility, however, appears to be explainable more by economic factors, such as earnings, than by the non-economic pulls and pushes, such as urbanization per se.

About half of the *dummy* variables are significantly different from zero, indicating that the specification of the function is improved by the inclusion of the dummies. Taking an overview of various regression coefficients, it is clear that, vis-à-vis the region of Mato Grosso and Goiás (the omitted region), every other region has an attraction for its own residents not to emigrate. Out of the three regions represented by the three dummies, contrary to the common view, the Amazon people (represented by Dummy 1), *ceteris paribus*, appear to be least anxious to emigrate. The central-southern region (Dummy 3) comes next in its sedentary effect, and the northeastern region last. This result may be of some interest to the sociologists and anthropologists, since the forces lying behind the dummy variables may be attributed to any of the cultural and social and other push and pull forces, and perhaps also the ethnic factors. It must be stated, however, that such an interpretation will at best be conjectural in that, first, not all dummies are significant; and second, there can still be some economic factors, costs, and benefits excluded from the model whose effects have not been captured by the set of explanatory variables actually used.

The *distance* variable is highly significant. It has the negative, and expected sign, and so is a deterrent to migration. The regression coefficient is approximately 0.11 in all of the regressions of Tables 1 and 2. Recall that the dependent variable is the log of the number of migrants while distance, D, is entered as $\sqrt{D}$. The mean distance between the (capitals of the) regions is about 1,500 kilometers. The point-elasticity coefficient at this mean is about 0.79. It appears that a 10 per cent (or 150-kilometer) increase in distance deters migration by as much as about 8 per cent. We have seen above that the elasticity coefficient of wage rate is 1.81 for the "young" group. The 1950 mean wage rate was 453 cruzeiros per month. Accordingly, we can surmise that the inducement to migration of an increase in wage rate of 10 per cent of 45 cruzeiros per month (in 1950 prices) in the destination region will be canceled out by an increase in distance of about 22 per cent or 330 kilometers.

The traveling cost of 330 kilometers by road in 1950 was around 75 cruzeiros, which from the 45-cruzeiro differential in wage rate could be retrieved in a little over one and one-half months in the case of an unmarried person and in about seven months in the case of a married couple with two children. Hence, the *direct* money cost of moving does not appear to be the main barrier. The indirect money costs, such as income foregone during the migration period, the higher rentals in the urban areas, and similar costs, however, may be substantial, but perhaps still not so prohibitive as the distance coefficient indicates. In order to have a broad view of the latter

costs, I made some rough estimates which are presented in a footnote (below).[6] According to these estimates, for an average migrant from Pernambuco to São Paulo (about an average migration distance in Brazil), all initial monetary and non-monetary costs of migration can be recouped from the differential earnings in about ten months in the case of an unmarried migrant and in about thirteen months in the case of a family of four members. The traveling costs alone can be recouped in about one month by a single migrant and about five months by a migrant with family. These costs are not prohibitive, but neither are they paltry, especially for a migrant from the low-income strata for whom the risk premium (including the lender's risk and the borrower's risk) may be quite

[6] Consider an individual migrating from Pernambuco to São Paulo. The relevant monetary and non-monetary costs and returns, in 1950 cruzeiros, using some guesswork for some of the items, and neglecting differences in prices and taxes and biases in the measurement of the state incomes, are computed below:

	Pernambuco	São Paulo
1. Calculated monthly wage	587	1,196
2. Monthly per capita income	252	778
3. Differential wage (from row 1)	...	+609
4. Monthly rent and utilities, assuming a rent cost of 10% of the local wage in the rural Pernambuco and a rent of 20% of wage in the urban São Paulo	59	239
5. Monthly wage differential net of rent (row 3 − row 4)	...	+429
6. Traveling cost by bus for a distance of 2,212 kilometers between the state capitals of Pernambuco and São Paulo:		
a) For the migrant alone	...	502
b) For a family of four members	...	2,010
7. Income foregone at the rate of the Pernambuco wage rate for an assumed period of three months (the time assumed to be spent on the preparation, traveling, and job searching)	...	1,761
8. Three-months rent, assuming no free accommodation from relatives or friends	...	717
9. Capital cost:		
a) 6*a* + 7 + 8	...	2,981
b) 6*b* + 7 + 8	...	4,488
10. The recoupment period using a 14% rate of interest (the normal real rate of interest in rural Brazil), assuming all money was borrowed in the beginning of migration:		
a) Corresponding to 10(*a*)	...	10 months
b) Corresponding to 10(*b*)	...	13 months

Corresponding to the non-pecuniary costs of migration, there may also be non-pecuniary returns, especially in the urban centers, for example, better public services, shorter working hours, better educational facilities for children, and so on. This point has been emphasized by Gupta (1966). To the extent these returns exist, the recoupment period calculated above will be overestimated. No allowance has, however, been made for this factor. For one thing, this factor is impossible to measure. Moreover, it may be canceled out against the relatively higher prices and taxes in the urban areas, which also have been ignored.

a bit higher than that indicated by the market rate of interest. In any case, the out-of-pocket costs are not so ignorable as they would appear in the conclusions of the empirical studies by Sjaastad (1961) for the United States and Beals, Levy, and Moses (1967) for Ghana. In addition, as mentioned earlier, psychic costs are also a factor to be reckoned with in any cost-return analysis of migration. Thus, while Beals, and at one place Sjaastad also (1961), looked for the explanation of a sharp decline in migration as geographical distance increases in the cultural and social differences—the left-out variables—this study tends to restore the importance of pecuniary costs of moving in explaining the negative coefficient of distance.

Coming to the regressions under the homogeneity restriction (Table 3), we find that the coefficients of all of the variables except $\tilde{E}_3$ for age group 15–29 are significantly different from zero. Education and migration appear to be negatively related. The remaining variables are all positively related with migration. Another feature of these regressions to be noted is that the significance level of Dummy 1 (the Amazon region) goes up considerably, while that of $\sqrt{D_{ij}}$ goes down substantially, indicating some relationship between distance and the Amazon dummy. It appears that when the homogeneity restriction is imposed, that is, when the changes at home are given the same weight in the people's decisions to migrate as are changes in distant places, the location of a region in relation to the destination region gains in importance at the expense of the distance variable. Moreover, urbanization and industrialization have, in this case, the right signs.

6. The Simultaneous-Equation Model and Substantive Results

In the foregoing model, no account was taken of a possible simultaneity among the variables. In particular, wage rate may be affected by migration even within one year. If so, the coefficients estimated by the single-equation least-squares method will not be unbiased. In order to allow for these interaction effects, therefore, a two-equation model, using the key variables, was also formulated as follows:

$$m_{ij} = \tilde{w}^{a_1}\tilde{d}^{a_2}\tilde{U}^{a_3}\tilde{g}^{a_4}\tilde{E}^{a_5}\exp\left(a_6\sqrt{D_{ij}} + a_{00} + \sum_{i=2}^{I} a_{0i}\theta_i + \eta_1\right)$$

(the migration supply equation); (5)

$$\tilde{w} = m_{ij}^{b_1}\tilde{d}^{b_2}\tilde{U}^{b_3}\tilde{g}^{b_4}E_j^{b_5}\exp\left(\sum_{j=2}^{I} b_{0j}\theta_j + b_{00} + \eta_2\right)$$

(the migration demand equation) (6)

Supply of migrants = demand for migrants. (7)

It may be noted that, in this model, partly as a result of the knowledge gained from the preceding section and partly with a view to reducing computational work, a pairwise homogeneity condition has been imposed on the corresponding variables of the pairs of regions i and j, except E_j. It was assumed that the entire incidence of the traveling cost falls on the migrant, so the distance variable appears in the supply equation only. On the other side, insofar as education is concerned, the wage offers were supposed to be geared to the local levels of education only; hence, only E_j appears in the demand equation. The dummy variables are so specified as to remove the origin-region effects in the supply equation and the destination-region effects in the demand equation. The variables with tilde and the dummies were defined earlier. The η's are the disturbance terms which are assumed to be uncorrelated with the exogenous variables and among themselves.

TABLE 4

STRUCTURAL PARAMETERS COMPUTED BY TWO-STAGE LEAST-SQUARES METHOD

Symbol for Coefficient (1)	Coefficient of (2)	Age Group 15–29 Regression 9 (3)	Age Group 30–59 Regression 10 (4)
	Dependent Variable: $\text{Log}_e m_{ij}$		
a_{00}	Constant	0.477	0.257
a_1	$\text{Log}_e \tilde{w}$	0.146(2.56)	0.111(1.89)
a_2	$\text{Log}_e \tilde{d}$	0.069(2.73)	0.108(2.91)
a_3	$\text{Log}_e \tilde{U}$	0.096(1.37)	0.125(1.29)
a_4	$\text{Log}_e \tilde{g}$	0.038(3.19)	0.013(6.12)
a_5	$\text{Log}_e E_i$	−0.030(−1.99)	−0.036(−2.27)
a_6	$\sqrt{D_{ij}}$	−0.013(−1.75)	−0.014(−1.85)
a_{01}	Dummy 1	−0.063(−1.35)	−0.070(−1.50)
a_{02}	Dummy 2	−0.093(−2.98)	−0.115(−3.37)
a_{03}	Dummy 3	0.011(0.45)	−0.004(−0.18)
η_1	S.E.E.	0.127	0.106
...	R^2	0.478	0.451
...	d.f.	332	332
	Dependent Variable: $\text{Log}_e \tilde{w}$		
b_{00}	Constant	0.362	0.018
b_1	$\text{Log}_e m_{ij}$	−1.969(−2.19)	−1.842 (−2.03)
b_2	$\text{Log}_e \tilde{d}$	0.026(1.98)	0.017 (4.15)
b_3	$\text{Log}_e \tilde{U}$	0.295(2.43)	0.322 (2.51)
b_4	$\text{Log}_e \tilde{g}$	0.075(1.35)	0.0733(1.83)
b_5	$\text{Log}_e E_j$	−0.055(−2.04)	−0.015 (−1.70)
b_{01}	Dummy 1	0.411(3.87)	0.391 (4.17)
b_{02}	Dummy 2	0.039(0.59)	0.225 (3.23)
b_{03}	Dummy 3	0.057(0.95)	0.173 (3.12)
η_2	S.E.E.	0.529	0.534
...	R^2	0.767	0.786
...	d.f.	333	333

It will be seen that the model is *logically* complete, as there are two equations and two endogenous variables, m_{ij} and $\tilde{w}$. But, *statistically*, the model is overidentified on the basis of coefficient restrictions. Accordingly, it will be estimated by the two-stage least-squares method.

The results of this estimation are presented in Table 4. I shall concentrate mainly on the supply function. It may be observed that the elasticities of migration with respect to the computed values of relative wage rate ($\tilde{w}$) are higher and those of the growth rate ($\tilde{g}$) lower in this model than the corresponding elasticities of the single-equation model. This appears to be a natural concomitant of simultaneity between wage rate and migration. The fit seems to have improved, as the standard errors of estimate (S.E.E.) are significantly lower in Table 4 than the corresponding errors in earlier computations. The coefficients have the right and expected signs, except perhaps education (for which, in any case, no sign was predicted). There is positive response to relative increase in density, wage rate, urbanization, and growth rate. Urbanization does, at last, acquire the right sign, and in about half of the cases emerges as a significant variable in attracting migrants. The magnitude and the significance level of the income-growth variable declines somewhat, but the coefficient still remains highly significant. The elasticity of migration with respect to wage rate continues to be about one-third higher for the "young" group than that for the "middle-aged" group. Distance remains a deterrent to migration, though the distance coefficients are relatively lower in this model than those obtained in the previous model. On the whole, the magnitudes of the corresponding responses of the two age groups appear to have come somewhat closer to each other as compared with those observed in the single-equation model.

7. Final Assessment, Conclusions, and Policy Implications

It will be of some interest to compare the foregoing results with the results of the existing studies. The present results are not strictly comparable to those of the existing studies, since the model as well as the variables of this study differ in important ways from those used by others. Also, there are other problems inevitably involved in such comparisons. Yet the enterprise seems to be worth the effort. Three studies bearing closely on the present problem, known to have been done before, are: the study of interstate migration in the United States by Sjaastad (1961); the study of inter-regional migration in Ghana by Beals *et al.* (1967); and the study of urban migration (from the outlying districts to the capital city of Santiago) in Chile by Herrick (1965). Herrick does not use regression analysis, without which it is difficult to hold the effects of some variables constant in order to study the effect of other variables. Hence, Herrick's results can at best be compared with the signs of the coefficients of this study. As to the studies of Sjaastad and Beals, I have already noted, while referring to

their methodologies at the relevant points, that Sjaastad tests the destination variables only and neglects the origin variables, while Beals tests fewer variables than I have, uses a somewhat different form of the dependent variable, works with only thirty-six observations against 342 of this study, and differs from this study in certain other minor aspects.

We have just seen—in this study as in the studies of Sjaastad and Beals—that distance sharply reduces migration. However, while, according to Sjaastad and Beals, the high ratio of the elasticity of distance to that of earnings is inconsistent with the hypothesis that distance proxies for only the money costs of migration, the evidence gleaned from this study is *not* inconsistent with the hypothesis that distance proxies to a significant extent for the money costs of migration.

Sjaastad averaged the coefficients of twelve regressions for twelve states and derived an elasticity of 1.52 for income and −0.84 for distance. The corresponding elasticity coefficients of this study range between 1.05 and 1.81 for wage rate and 0.60 to 0.80 for distance—surprisingly close to the U.S. elasticities. In Beals's study, the absolute magnitudes of these coefficients are about twice as high as ours, while the relative values are close to ours. In the above two studies, the rate of change of income for the destination state and population (which can be compared to the density variable of this study) were both significant and of the expected sign. By and large, the same is the case in the present study.

Sjaastad found a significant response of migrants to the region-income distribution (as computed from the income tax returns). Herrick purported to show that land reform emphasizing family-sized farms would retard rural migration. Beals did not try any such variable. In the present study, no such relationship could be established, though the cause of this could very well be the unrepresentativeness of the data on income distribution, computed as they were from the size holdings of cultivated land.

Our conclusion with regard to urbanization is partially in agreement and partially in conflict with the corresponding conclusion of Herrick's.[7] Herrick found that "a far greater proportion [of migration] came to Santiago from other cities . . . [and] rural-urban migration was only a miniscule fraction of the total movement toward Santiago" (1965, p. 103). The findings of this study do show a relatively greater tendency on the part of the urban people to emigrate, but the destinations are not necessarily the bigger and bigger cities. Rather, the direction of migration seems to be determined more by the economic factors such as earnings than by urbanization per se. A policy implication of this result is that a conscious decentralization of industrial and employment centers will not in itself be inimical to migration. Sjaastad's skill factor and Herrick's education factor stimulate migration. Beals's education variable retards migration in both

[7] Sjaastad did not test this variable.

the destination region and the origin region. The model of this study appears to support the above findings with regard to the destination region but not with regard to the origin region in which education seems to retard migration.

In summary, among the major findings of this study is the fact that internal migration in Brazil is highly responsive to earning differentials. Distance is a strong deterrent to migration, which to a significant extent can be attributed to the direct (monetary and non-monetary) costs of moving, though associated costs and investments are also important. There is evidence that education attracts migrants both in the origin region and the destination region. Moreover, it certainly promotes migration to the extent that it influences other variables conducive to migration. Economic costs and returns appear, on the whole, to dominate the behavior of migrants, though some relevance of the non-economic "push" and "pull" factors is not denied. There is evidence that urbanization and industrialization promote mobility, but only a shaky evidence that these characteristics in themselves attract migrants. Accordingly, a policy of industrial and employment decentralization does not seem to prove detrimental to migration. Density acquires a gravitation-like pull on migrants. Among the possible causes of this pull, in combination with the industrialization variable, can be a previous presence of relatives and friends in the destination regions, who may be the source of information and temporary aid. The rate of growth of income, employed to serve as a proxy for the expansion of income and employment, though shaky in some regressions, turns out, on the whole, to be a significant variable.

Significant, though not dramatic, differences in responses to various stimuli by migrants in different age groups are discernible. This result, combined with the observed simultaneity among earnings and migration, lends support to Kuznets' selectivity and interaction approach as well as to the capital-theoretic approach of Schultz and Sjaastad. On the whole, there is a strong conformity of the migration function to the neo-classical costs-and-returns approach, which has served to modify several of the impressions regarding the theories of migration. Finally, the quantitative estimates of the migration effects of various factors, a reasonably good fit of the migration function, and the high significance levels of most of the parameters may serve as policy guides for the planning of economic development and human resources in Brazil as well as in other underdeveloped countries.

Some of the policy implications following from this study for Brazil may be briefly indicated.

First-rate schooling in the rural areas will induce the entrepreneurial farmers not to migrate to cities but to stay in agriculture and develop scientific farming. Decentralization of industry and employment centers will not be detrimental to migration, and urbanization per se will not

necessarily attract migrants. On the contrary, decentralization will aid in lowering the rate of increase of congestion and slumming of old cities, as well as reducing the geographical distance between the springs of migrants and the sinks of migrants. The high-density states are likely to continue growing denser, unless the density element is countered by the creation of immigration incentives in the lower-density states. These incentives may include subsidies or tax concessions to industry, so that the wage rates and employment opportunities are increased in the low-density regions; more publicity about the new employment centers; and other similar measures. Better roads and economical long-distance bus/train services will lower the cost barrier of geographical distance. Finally, although at any given time there will usually exist some dispersion of average earnings among different regions due, for instance, to different occupational structures and the concomitant equalizing differences and differences arising from non-competing groups (Friedman, 1962, chap. xii), given an appropriate educational and industrial location policy, internal migration in Brazil appears to be responsive enough to earnings differentials as to substantially reduce the present interregional differences of incomes.

References

Beals, Ralph E., Levy, M. B., and Moses, L. N. "Rationality and Migration in Ghana," *Rev. Econ. and Statis.*, XLIX, No. 4 (November, 1967), 480–86.

Friedman, Milton. *Price Theory: A Provisional Text.* Chicago: Univ. of Chicago Press, 1962.

Gisser, Micha. "Schooling and the Farm Problem," *Econometrica*, XXXIII, No. 2 (July, 1965), 582–92.

Gupta, Suraj B. "The Economics of International Migration: A Theoretical Analysis." Univ. of Western Ontario, Dept. Econ., Faculty Seminar, Paper No. 1, 1966 (mimeographed).

Herrick, Bruce H. *Urban Migration and Economic Development in Chile.* Cambridge, Mass.: M.I.T. Press, 1965.

Johnson, D. Gale. "Efficiency and Welfare Implications of U.S. Agricultural Policy," *J. Farm Econ.*, XLV, No. 2 (May, 1963), 338–48.

Johnson, Harry G. "The Economics of Brain Drain," *Sixth Annual Seminar on Canadian-American Relations* (Ottawa, Canada) (1964), pp. 37–50. Also published in *Minerva* (Canada), III, No. 3 (Spring, 1965), 299–311.

Kuznets, Simon, Thomas, Dorothy, *et al. Population Redistribution and Economic Growth—United States, 1870–1950.* 3 vols. Philadelphia: American Philosophical Soc., 1957, 1960, 1964.

Lewis, W. Arthur. "Economic Development with Unlimited Supplies of Labor," *Manchester School*, XXII, No. 2 (May, 1954), 139–91.

Ravenstein, E. G. "The Laws of Migration," *J. Royal Statis. Soc.*, XLVIII, No. 2 (June, 1885), 167–227; ———. *Ibid.*, LII, No. 2 (June, 1889), 198–99.

Redford, Arthur. *Labor Migration in England, 1800–1850.* Manchester: Univ. Press, 1926.

Schultz, Theodore W. "Reflections on Investment in Man," *J.P.E.*, LXX, No. 5, Part 2 (October, 1962), 1–9.

———. "An Endeavor to Clarify the Economic Components Underlying Chilean Agriculture," in Inter-American Development Bank. *Economic Growth Theory and the Profitability of Farming in Latin America* (forthcoming).

Sjaastad, Larry A. "Income and Migration in the United States." Unpublished Ph.D. dissertation, Univ. of Chicago, 1961.

———. "The Costs and Returns of Human Migration," *J.P.E.*, Vol. LXX, No. 5, Part 2 (October, 1962).