Place-to-Place Migration in Colombia*

Gary S. Fields
Cornell University

I. Introduction

This paper builds upon earlier work of mine which explored the determinants of population migration in Colombia.¹ As before, the basic proposition is that areas' economic opportunities play a central role in determining the spatial allocation of the population.

My earlier paper used published data from the 1973 Colombian Census of Population to establish that the rates of net lifetime migration into Colombia's 23 provinces (or, as they are known in Colombia, "departments") are associated with those areas' labor market conditions. The present paper uses unpublished data for 12 zones (six regions, rural and urban segments of each) to analyze the causes of place-to-place migration flows (lifetime and recent) for the population as a whole and for eight specific demographic groups (four educational categories and the two sexes).

Several hypotheses are formulated and tested. The results give strong support to the economic model of migration, overall and in explaining differences in migration behavior among demographic groups. For Colombia this is particularly useful since some past work has been interpreted to the contrary.²

II. Place-to-Place Migration Models

Place-to-place migration studies analyze the rate of gross population flow between pairs of origins and destinations. The economic model

* Research on this paper was begun at Yale University and continued at Cornell University. The Rockefeller and Ford Foundations contributed financial support. Ruth Daniel, Gary Moss, and Se-Il Park did the computer work; Deborah Nivison did the typing. Helena Ribe and Jeffrey Williamson offered helpful comments on an earlier draft. I am pleased to acknowledge the assistance of these institutions and individuals.

¹ Gary S. Fields, "Lifetime Migration in Colombia: Tests of the Expected Income Hypothesis," *Population and Development Review* 5 (June 1979): 247-66.

² For a review of the literature on migration in Colombia, see Helena Ribe, "Income of Migrants Relative to Non-Migrants in Colombia" (Ph.D. diss., Yale University, Department of Economics, 1979).

of place-to-place migration is that even though various individuals face different economic opportunities in various locations and some have different preferences for one location relative to another, the *rate* of migration between i and j is systematically and positively related to the general attractiveness of labor market conditions in the destination (E_j), negatively related to the economic attractiveness of the origin (E_i), and negatively related to the cost of moving from i to j (C_{ij}):

$$M_{ij} = f(E_i, E_j, C_{ij}) , \quad f_1 < 0 , \quad f_2 > 0 , \quad f_3 < 0 . \quad (1)$$

Typically, E_i and E_j are approximated by average income, employment probability, etc., and C_{ij} by distance. The remainder of this section explores the specific functional form of the place-to-place model (1).

Symmetrical Models

The starting point for place-to-place migration research is a simple symmetric specification based on the rate of return to investment in migration. If we summarize the relationship between the benefits and costs of migration by an internal rate of return r , then

$$M_{ij} = f(r) , \quad f' > 0 . \quad (2)$$

As with other human investments, such as education, the rate of return on migration may be approximated by either the difference or the ratio between economic conditions in origin and destination:

$$M_{ij} = f(E_j - E_i, C_{ij}) , \quad f_1 > 0 , \quad f_2 < 0 \quad (3)$$

and

$$M_{ij} = f(E_j/E_i, C_{ij}) , \quad f_1 > 0 , \quad f_2 < 0 . \quad (4)$$

The difference form is appropriate when *none* of the costs are opportunity costs, the ratio form when *all* are opportunity costs. This model looks very much like simple human investment models as they are applied in other fields, such as education and occupational choice.

By its very nature, the symmetrical model holds that the “push” of unfavorable economic conditions in the origin and the “pull” of favorable economic conditions in a possible destination are equally strong. A one-unit (or percentage) increase in one of the components of E_i is assumed to have the same effect on migration as a one-unit (or percentage) decrease in the corresponding component of E_j , and the elements of E_i and E_j are the same. Thus, the symmetrical model

does not allow for the possibility that labor market conditions in origin and destination may have differential effects from one another. To allow for this, we may formulate various types of asymmetrical models.

Asymmetrical Models

One of the most elementary criticisms of the symmetrical model is that the model seems to assume perfect information in labor markets. If information were perfect, potential migrants might be as receptive to changing labor market conditions in a distant origin as they would be to changes of the same magnitude in their present locations. However, once we recognize that individuals know more about labor market opportunities in their present locations than in far away places, we are led to expect a particular form of asymmetry,

$$M_{ij} = f(E_i, E_j, C_{ij})f_1 < 0, \quad f_2 > 0, \quad f_3 < 0, \quad |f_1| > |f_2|, \quad (5)$$

whereby it is hypothesized that changes in origin economic conditions have a larger effect on migration than a like-sized change in destination conditions.

Origin economic conditions might be thought to dominate destination conditions for another reason: push from the land. Lipton,³ for example, argues that this is particularly strong for relatively poorly educated workers. This is another reason why the pattern hypothesized in equation (5) might arise.

Asymmetrical models based on imperfect information or push from the land have a problem. In several studies of advanced economies (though not up to now for Colombia), empirical tests have been conducted and the results go the other way. That is, economic conditions in the destination are consistently found to outperform those in the origin.⁴ Several ad hoc explanations for this particular pattern or asymmetry have been proposed.

One line of thinking holds that capital market imperfections may strongly impede mobility. By this argument, superior economic conditions in the origin increase the ability of potential migrants to finance profitable moves. Hence, the better are origin conditions, the greater the number of people who move out.

³ Michael Lipton, *Why Poor People Stay Poor: Urban Bias in World Development* (Cambridge, Mass.: Harvard University Press, 1977).

⁴ I am using the term "outperform" here in the crude sense of higher *t* statistics in a linear model: $M_{ij} = \sum_m \beta_i^m E_i^m + \sum_m \gamma_j^m E_j^m + C_{ij}$. In this equation, the imperfect information hypothesis leads us to expect larger regression coefficients for the *i* variables than for the *j* variables, i.e., $|\beta_i^m| > |\gamma_j^m|$ for all *m*. Also, since in a linear model for any given element of E_i and E_j the standard errors of the regression coefficients are identical, the imperfect information model would lead us to expect greater statistical significance of origin variables as compared with the corresponding destination ones.

Another possible reason for this same effect is that migration itself may be desirable as a consumption good. By this argument, in high-income origins, the income effect leads to greater consumption of most goods, including migration. Once again, relatively favorable origin economic conditions may result in more out-migration.

A third possible explanation relies on aggregation error. We may distinguish between (i) currently employed workers and (ii) persons who are currently unemployed or out of the labor force. The latter group is necessarily "in the job market," so average labor market conditions in the origin may be a very good proxy for job opportunities if they remain in their present location. However, currently employed persons, because they are working in particular jobs with particular wages, fringe benefits, etc., face job opportunities which are proxied only imperfectly by average conditions in the origin. Since a very large number of potential migrants are currently employed, we would by this argument expect the function with weak effects of origin conditions to dominate the function with strong origin effects.

We thus have several explanations for the observed asymmetry. What they have in common is that favorable economic conditions at origin increase the means for moving but reduce the incentives. Although the relative strength of these effects cannot be predicted a priori, they together lead us to expect only a weak correlation between origin economic conditions and the rate of out-migration. At destination no problem of offset arises, since changes in destination conditions affect incentives only.

Logarithmic Specifications and Logistic Models

Despite their differences, the models considered so far have in common the assumption that origin and destination economic conditions enter linearly in the migration function. A new literature, built around an alternative assumption and resulting in logarithmic rather than linear estimation, has arisen in recent years and shows signs of substantial promise. This literature recognizes that the migration decision is inherently a choice between a finite number of mutually exclusive discrete alternatives. As such, it is amenable to analysis by the polytomous logistic model, developed in economics by McFadden and applied to the migration decision by Schultz.⁵

The logistic model holds that an individual's decision to locate (or

⁵ Daniel McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. Paul Zarembka (New York: Academic Press, 1974); and "Quantal Choice Analysis: A Survey," *Annals of Economic and Social Measurement* 5, no. 4 (1976): 363-90. T. Paul Schultz, "A Conditional Logit Model of Internal Migration: Venezuelan Lifetime Migration within Educational Strata," Discussion Paper no. 266 (New Haven, Conn.: Yale University, Economic Growth Center, September 1977).

relocate) in place j given that he now lives in i depends on a linear combination Z_{ij} of origin and destination conditions in the following specific way:

$$P_{ij} = e^{Z_{ij}} / \sum_j e^{Z_{ij}} \quad (6a)$$

where

$$\sum_j P_{ij} = 1 \quad (6b)$$

for all i . For a variety of reasons noted by Schultz, Z_{ij} is thought to be a linear function in the *logarithms* of the origin and destination conditions X_i and X_j and the distance D between i and j :

$$Z_{ij} = \alpha + \sum_m \beta_m \ln X_{mi} + \sum_m \gamma_m \ln X_{mj} + \delta \ln D_{ij} .^6 \quad (7)$$

Combining (6) and (7), we obtain the general form:

$$\ln (P_{ij}/P_{ii}) = \alpha + \sum_m \tilde{\beta} \ln X_{mi} + \sum_m \tilde{\gamma} \ln X_{mj} + \tilde{\delta} \ln D_{ij} , \quad (8)$$

the tildes ($\sim$) indicating transformations of the respective coefficients of (6) and (7). Since the variation in P_{ij} is undoubtedly much greater than the variation in P_{ii} , we might regard P_{ii} as roughly constant across labor markets. This assumption justifies approximating equation (8) by double logarithmic estimation of the function:

$$\ln M_{ij} = f(\ln X_i, \ln X_j, \ln D_{ij}) . \quad (9)$$

Empirical estimates of this function are presented below.

III. Data and Variables

The data for this study are taken from the 1973 Colombian Census of Population. A 4% sample of the questionnaires was made available by the National Statistical Office (DANE) in computer-readable form.

⁶ The reasons offered by Schultz for preferring the logarithmic form of Z_{ij} are: (1) the expected wage hypothesis posits multiplicative interactions between wage rates and employment rates, which are easily specified logarithmically; (2) the ratio of expected incomes approximates the rate of return to migration in the case where opportunity costs are the most important costs of migrating; (3) specified in this way, the logistic model is comparable to non-logistic models in double-logarithmic form; (4) in empirical research on migration in Venezuela, the logarithmic form of Z_{ij} explained a larger share of the variance than other forms.

This public-use sample, known in Colombia as the "Muestra de avance," contains 777,800 individuals.

The Colombian census sample offers exceptionally fine detail. Answers to the following locational questions were obtained: Where do you live now? Where did you live last? How long ago did you move from the last place? Where were you born? Each locational variable was coded down to the level of the municipality. The urban or rural character of the place of residence was also ascertained.

The availability of data on a very large sample of individuals rather than geographic aggregates, the multiplicity of questions on location, and the richness of geographic detail offer an unprecedented opportunity for disaggregated research using the place-to-place model. Data limitations have impeded past researchers working on other LDCs. Even in the United States, which often is held up as the model of data availability, the smallest unit for which we have place-to-place migration information is nine large census regions.

In what follows for Colombia, two concepts of "migrant" are used. One is *lifetime migrant*: someone whose birthplace differs from current residence. The other is *recent migrant*, defined as an individual who has moved within the last five years. These concepts are used to construct three migration variables. *MIGRATE* is the rate of gross lifetime migration into an area. Its numerator is the number of persons now living in that location whose birthplace was some other municipality. Its denominator is the area population. *RLFMIGIJ* is the rate (per 100) of lifetime migration between zone *i* and zone *j*. Its numerator is the number born in *i* and now living in *j*. Its denominator is the number born in *i*. Finally, we have *RRECMIGIJ*, which is the rate of recent migration between *i* and *j*. Its numerator is the number who lived in *i* 5 years previously. Only persons over the age of 10 are included in these rates.

The geographic areas defined for this study take account of both the region of location and its rural or urban character. I divided the country into six regions according to departments.⁷ The breakdown is given in table 1. Each region was in turn divided into rural and urban sections. The 12 resulting areas are called zones. For this study of place-to-place migration, the 12 zones are both origins and destinations. There are thus 144 place-to-place migration flows (including nonmigration). Of these flows, 1/4 are rural-urban observations, 1/4 are urban-rural, 1/4 are rural-rural, and 1/4 are urban-urban.

The distance between one zone and another (*DISTIJ*) is, to a certain extent, arbitrary since the zones are not only big (the usual

⁷ Colombia is divided into 23 departments. A department is a geographic unit rather like an American state though with less administrative autonomy.

TABLE 1
REGIONAL GROUPINGS USED IN PLACE-TO-PLACE MIGRATION STUDY

Region	Departments Included
Region I	Córdoba, Bolívar, Atlántico, La Guajira, Magdalena, Sucre, César
Region II	Antioquia, Caldas, Risaralda, Quindío
Region III	Chocó, Valle, Cauca, Nariño
Region IV	Huila, Tolima, Meta
Region V	Boyacá, N. de Santander, Santander
Region VI	Bogotá, Cundinamarca

problem) but also not topographically closed.⁸ The details of the distance calculations are available upon request.

The economic variables used to explain migration flows are the zone's income and employment opportunities. The income variables used are the average income in the origin and destination (*AVGYI* and *AVGYJ*, respectively). These were calculated from census respondents' answers to the question: "What was your income in pesos last month?" The employment variables used are the employment rates in the zones of origin and destination (*EMPLRI* and *EMPLRJ*, respectively). Following the usual definitions, we have (i) the employed: those who worked in the preceding week or who did not work but had a job; (ii) the unemployed: those who were looking for a job and either had worked before or were looking for a job for the first time; (iii) the labor force: $i + ii$; (iv) employment rate: i/iii .

The demographic characteristics of an area's population also play an important role in determining migration behavior. As people age, their likelihood of moving declines. Better-educated individuals exhibit higher migration rates. In Colombia, as in other Latin American countries, women are more likely to migrate than men. One way to include these factors would be to add origin-specific demographic factors to the basic place-to-place model:

$$M_{ij} = f(E_i, E_j, C_{ij}, DEMOG_i), \quad (10)$$

where $DEMOG_i$ is a vector of demographic characteristics of the population in the origin i . A straight linear specification such as

$$M_{ij} = \alpha + \beta_1 E_i + \beta_2 E_j + \beta_3 C_{ij} + \beta_4 DEMOG_i + \epsilon \quad (11)$$

would assume that demographic factors shift only the intercept of the migration function but not the slope, that is, that each demographic

⁸ That is, movement from one part of the zone to another part of the same zone may take you out of the zone.

group's migration rate is changed by the same amount by an extra dollar of destination income, by a 1% higher employment rate, or by a 1-kilometer difference in distance. These assumptions are implausible, since we have reason to believe that these demographic groups differ in their responsiveness to economic stimuli. In recognition of this belief, we might include a number of interactive variables in (11). Alternatively, and more conveniently, we could stratify the sample and run separate models for each demographic group. That is the course followed below. The stratifying variables used are of two types: (1) sex: males' and females' migration functions are estimated separately; (2) education: the sample is stratified into four education groups—none, primary (some or complete), secondary (some or complete), higher (some or complete).

One variable that is usually included in migration functions has not received mention: size of population in origin and destination. The omission of population variables is not an oversight; they are excluded for econometric reasons. Suppose, as adherents of the gravity model would have us believe, that the statistical function linking migration on the one hand and economic and demographic variables on the other is

$$M_{ij} = k(POP_i^\alpha POP_j^\beta / DIST_{ij}^\gamma)(E_j^\delta / E_i^\varepsilon), \quad (12)$$

where k is a constant, POP_i and POP_j are origin and destination population, respectively, E_i , E_j , and $DIST_{ij}$ are as before, and the Greek letters are parameters.⁹ Equation (12) can be rewritten in logarithmic form as

$$\log M_{ij} = \log k + \alpha \log POP_i + \beta \log POP_j - \gamma \log DIST_{ij} + \delta \log E_j - \varepsilon \log E_i. \quad (13)$$

A simple linear regression of the variables in (13) using Ordinary Least Squares (OLS) will *not* work properly, though it has been done, because all regression coefficients would suffer from simultaneous equations bias due to endogeneity of population.¹⁰

⁹ This form comes from Ira Lowry, *Migration and Metropolitan Growth: Two Analytical Methods* (Los Angeles: University of California Press, 1966).

¹⁰ Simultaneous-equations bias arises because the populations in origin and destination are themselves determined by previous migration. If the same underlying process that determines present migration also determined past migration (as is assumed by adherents of both gravity models and economic models of migration), places that had experienced higher-than-expected in-migration in the past would (i) experience higher than-expected in-migration at present and (ii) have larger populations at present. Factors i and ii together lead us to expect a positive correlation between the error in the regression equation and the levels of the independent variables. When there is such a correlation, OLS estimates are biased.

To avoid biased estimates of the migration functions, the choice is either (a) to estimate a complete system which also explains past migration, or (b) estimate a reduced form from which population variables are omitted. Option a, being more complete, is preferred in principle; but since it is impossible with Colombian data, it must be rejected in practice. Option b provides less information (population effects are not estimated), but what estimates it does provide (on the effects of origin and destination economic conditions and distance) are unbiased. Option b is the only feasible alternative which is justified econometrically. Because it is better to have good information on a limited set of relationships than poor information on a larger set, population in origin and destination are excluded from further consideration.

IV. Empirical Results for the Full Sample

Tabular Evidence

Table 2 presents some basic data for the 12 geographic zones. These zones represent the urban or rural sections of the department groupings indicated in table 1. For mnemonic purposes, the zones are referred to by the name of the principal city in each.

Six observations may be made: (i) When a migrant is defined as someone who lives in a different municipality from the one in which he or she was born, the rate of in-migration across various geographic zones ranges from .23 to .59. The migration rates are higher into urban zones than into rural zones. (ii) Average incomes in the several zones range from 394 to 2,231 pesos per month. Within each region, average incomes are always higher (by a factor of three or four to one) in the

TABLE 2
BASIC DATA FOR 12 ZONES IN COLOMBIA

Zone	Rate of In-Migration (MIGRATE)	Average Income in Pesos (AVGY)	Employment Rate (EMPLRATE)
1. Barranquilla, urban	.39	1628	.97
2. Barranquilla, rural	.23	665	.99
3. Medellín, urban	.58	1818	.97
4. Medellín, rural	.34	687	1.00
5. Cali, urban	.59	1631	.97
6. Cali, rural	.28	453	.99
7. Neiva, urban	.56	1555	.98
8. Neiva, rural	.36	596	1.00
9. Bucaramanga, urban	.56	1519	.98
10. Bucaramanga, rural	.23	394	.99
11. Bogotá, urban	.63	2231	.97
12. Bogotá, rural	.27	545	1.00

SOURCE.—Muestra de Avance, Censo de Población, 1973.

NOTE.—For definitions of zones and variables, see text.

urban segment than in the rural segment. (iii) The employment rates (as a percentage of labor force) range from .97 to 1.00. Within each region, the employment rate is higher in rural than in urban areas. (iv) Zones with higher incomes show higher in-migration rates ($r = +.92$). (v) Zones with higher employment rates exhibit lower in-migration rates ($r = -.85$). (vi) High-income zones have low employment rates ($r = -.95$).

Some of these observations are as expected and some are not. The higher migration rates into urban areas reflect the net urbanization of the Colombian society and are as expected. Also, it is not surprising that urban areas have higher incomes than rural areas. I see a cause and effect relationship between these two observations: higher average income in urban areas causes rural-to-urban migration.

At first, the negative correlation between in-migration rate and employment rate might seem surprising. Following Harris and Todaro, we might expect to find potential migrants being drawn into high employment probability areas.¹¹ In a multivariate relationship, the expected correlation between rate of in-migration and probability of employment should be positive after controlling for income. But this may not show up in a simple correlation or even in OLS estimation of a single migration equation, since we also expect that higher income in an area causes a higher unemployment rate there because of an inflow of job seekers. Indeed, the correlation between income and employment rate is negative. This explains why the simple bivariate correlation between in-migration rate and employment rate is also negative—income is a very strong intervening variable.

One feature of these data is very hard to explain: the unexpectedly high employment rates which imply corresponding unemployment rates of just 2% on average. These rates of unemployment are substantially lower than the rates of 10%–25% issued by DANE and used in Colombia by the Corporación Centro Regional de Población.¹²

My first inclination upon encountering these rates was to dismiss them as a statistical aberration and to discard the variable. But upon further reflection, I think otherwise. Is it not reasonable to expect that in a poor country like Colombia 98% of the people who were economically active in the week preceding the census had some kind of job? Or put differently, how many could afford to have done no work for an entire week? An answer of 2% seems at least as plausible to me as

¹¹ John Harris and Michael Todaro, "Migration, Unemployment, and Development," *American Economic Review* 60 (March 1970): 126–42.

¹² The unemployment rates I calculated match DANE's unadjusted unemployment rates; these rates are unpublished. DANE, however, publishes "adjusted" unemployment rates. I have tried without success to obtain a definitive statement of why the rates were adjusted at all and how the adjustments were done. I have been able to learn only that the unadjusted rates appeared "too low" so they were raised.

an answer of 10% or more. Also, we find higher unemployment rates in urban than in rural areas. Is it not reasonable to expect that the poorer areas of the country would have even less open unemployment for the same reason (inability to afford open unemployment for a whole week), and thus rural unemployment rates would be lower than urban rates?

Regression Results

This section presents the results of fitting the place-to-place migration model to the full sample (i.e., without stratifying by sex or education). The geographic unit is the zone (12 zones, 144 place-to-place observations). The hypotheses are that high-income zones will exhibit higher in-migration rates and possibly lower out-migration rates than do low-income zones; that zones with higher employment rates have higher rates of in-migration and lower rates of out-migration than other zones *ceteris paribus*; and that zones that are further from one another have less migration between them.

These hypotheses are tested using four separate functional forms. The first functional form takes the rate of lifetime migration (*RLFMIGIJ*) as the dependent variable and regresses it on the distance between *I* and *J* (*DISTIJ*), the average incomes in *I* and *J* (*AVGYI* and *AVGYJ*), and the employment rates in *I* and *J* (*EMPLRI* and *EMPLRJ*). The second regression instead uses as the dependent variable the rate of recent migration (*RRECMIGIJ*). Regressions 3 and 4 are double-log versions of 1 and 2, respectively. All variables are as defined in Section III.

The results are presented in table 3. Taken together, regressions 1–4 exhibit a consistent effect on migration only for distance. In those equations the economic variables are not impressive. Only three of the 16 regression coefficients involving economic variables are statistically significant, and one of these has the wrong sign. At first, this would seem to cast doubt on the relevance of the economic model of migration in the Colombia context.¹³

We should not hasten to discard the economic model too quickly. As I shall now show, the particular specifications in regressions 1–4 involve too little theory applied to too many variables.

The expected-income model of migration leads to the expectation that an equilibrium configuration is one where high-income zones attract migrants seeking the high-income jobs, which in turn causes higher unemployment. That is, besides the relationships in regressions 1–4 which are of the form

¹³ Michael Todaro (*Internal Migration in Developing Countries* [Geneva: International Labour Office, 1976]) reports that past researchers working on other countries also experienced difficulty with the expected income variables.

TABLE 3
PLACE-TO-PLACE MIGRATION REGRESSIONS FULL SAMPLE, VARIOUS FUNCTIONAL FORMS

FUNCTIONAL FORM	DEPENDENT VARIABLE	INDEPENDENT VARIABLES						R^2	SEE
		Constant	<i>DISTIJ</i> or Its log	<i>AVGYI</i> or Its log	<i>AVGYJ</i> or Its log	<i>EMPLRI</i> or Its log	<i>EMPLRJ</i> or Its log		
1. Linear	<i>RLFMIGIJ</i>	1516.7	-.0510* (.0050)	-.0123* (.0062)	-.0081 (.0062)	-6.898* (3.293)	-8.009* (3.293)	.455	13.34
2. Linear	<i>RRECMIGIJ</i>	106.6	-.0075* (.0010)	-.0017 (.0013)	.0006 (.0013)	-.4164 (.6668)	-.6096 (.6668)	.381	2.70
3. Log-log	<i>LOGRLFMIGIJ</i>	138.3	-.7703* (.0525)	-.1969 (.4463)	.5968 (.4463)	-3.418 (22.62)	-26.34 (22.62)	.659	1.10
4. Log-log	<i>LOGRRECMIGIJ</i>	24.3	-.4189* (.0487)	.1736 (.4147)	.0940 (.4147)	10.18 (21.02)	-15.44 (21.02)	.372	1.02
5. Linear	<i>RLFMIGIJ</i>	18.87	-.0465* (.0049)	.0001 (.0019)	.0062* (.0019)	...	...	.416	13.71
6. Linear	<i>RRECMIGIJ</i>	3.47	-.0071* (.0010)	-.0010* (.0004)	.0017* (.0004)	...	...	.375	2.69
7. Log-log	<i>LOGRLFMIGIJ</i>	2.07	-.7681* (.0523)	-.1336 (.1528)	1.0849* (.1528)	...	...	.656	1.09
8. Log-log	<i>LOGRRECMIGIJ</i>	-.57	-.4186* (.0485)	-.0151 (.1417)	.3803* (.1417)	...	...	.369	1.01
9. Linear	<i>RLFMIGIJ</i>	24.39	-.0464* (.0051)	1.1314 ^a (.9967)		...	...	.375	14.14
10. Linear	<i>RRECMIGIJ</i>		-.0071* (.0010)	.9288 ^a (.1932)		...	...	.349	2.74
11. Log-log	<i>LOGRLFMIGIJ</i>		-.7680* (.0555)	.6092 ^a (.1149)		...	...	.608	1.16
12. Log-log	<i>LOGRRECMIGIJ</i>	1.94	-.4186* (.0488)	.1977 ^a (.1010)		...	...	.354	1.02

* Regression coefficient is statistically significant at least at the 95% level; numbers in parentheses = SEs.

^a Ratio *AVGYJI*.

$$MIGIJ = \alpha + \beta_1 DISTIJ + \beta_2 AVGYI + \beta_3 AVGYJ + \beta_4 EMPLRI + \beta_5 EMPLRJ + \varepsilon, \quad (14)$$

there are also functional relationships of the form

$$EMPLRI = \gamma + \delta MIGIJ + \eta \quad (15)$$

and

$$EMPLRJ + \gamma' + \delta' MIGIJ + \eta'.$$

Substituting (15) into (14) yields

$$MIGIJ = \alpha + \beta_1 DISTIJ + \beta_2 AVGYI + \beta_3 AVGYJ + \beta_4(\gamma + \delta MIGIJ + \eta) + \beta_5(\gamma' + \delta' MIGIJ + \eta') + \varepsilon. \quad (16)$$

One problem which appears in (16) is that the right-hand-side variables are functionally related to the dependent variable. This creates simultaneous-equations bias if OLS is used. A second problem with (16) is the likelihood of severe multicollinearity between the income and employment variables, which can be inferred from the strong simple correlation coefficient between *AVGY* and *EMPLR* (-0.95). The bias problem leads to an understatement of the relationship (as measured by the regression coefficient) between average income in a zone and the rate of migration to or from it. The multicollinearity problem leads to less reliable point estimates of the regression coefficients, needlessly large standard errors associated with them, and unwarranted rejection of hypotheses which ordinarily would be accepted as correct.

Fortunately, to deal with these two problems, the solution is the same: *drop EMPLRI and EMPLRJ out of the regressions*. The way this solves the simultaneous-equations problem is to create a reduced-form structure, which is estimable without bias by OLS. It deals with the multicollinearity problem by removing one of the collinear variables so that the effect of the other can be estimated more accurately.

The resultant estimates are presented in regressions 5–8. In each case, higher average income in a destination is found to exert a statistically significant attraction on migrants. The estimated values are all quite plausible, for example, regression 7 implies that a 1% increase in average income in a destination will lead to a 1.08% increase in lifetime in-migration.

Comparing the two sets of regression estimates (regressions 1–4 with regressions 5–8) shows that each regression coefficient on *AVGYJ* is substantially larger in the latter set and each standard error sub-

stantially smaller. Thus, the second set of specifications largely averts the econometric problems inherent in the first set. As one indication of just how much better it is to drop the employment variables out, note that the standard errors of estimate (SEE) are lower (by about 1%) when fewer variables are used; or put differently, the econometric problems in regressions 1–4 were severe enough to distort the regression estimates to such a degree that the resultant regression line fit the data less well than a line estimated from less data.

One other finding bears mention. The results indicate marked asymmetry. In regressions 5–8, the estimated coefficient on origin income is smaller than the coefficient on destination income, and in only one of these four regressions does higher origin income appear to discourage out-migration. This suggests that the factors considered in Section II which produce asymmetrical results in other countries may be operating in Colombia as well.

What if we had instead specified a symmetric model? Compare regressions 9–12, which are based on income ratios, with regressions 5–8, where origin and destination income are estimated separately. If we had only the results from 9–12, the economic model of migration would have looked “better” (to put it crudely, as gauged by the presence of an asterisk on all but one regression coefficient); but the asymmetry would not have been found. Now that we have found this asymmetry in the Colombian data, future research into migration decision making is needed to understand why it is there.

V. Empirical Results for Eight Subsamples

We know from past research that women in Colombia migrate at higher rates than men and that better-educated workers migrate at higher rates than less educated workers. Why these groups’ migration rates differ is poorly understood.¹⁴ This section presents a partial explanation.

The hypotheses on education differentials, for persons of either sex, are: (a) the average propensity to migrate rises with education; (b) the marginal propensity to migrate rises with education; (c) the gain from migration rises with education. The reasons for hypothesizing that better-educated workers have higher average migration propensities are that some people migrate in order to acquire an education while others migrate in order to obtain the jobs appropriate to highly educated people. The reasons for hypothesizing that better-educated workers have higher marginal propensities to migrate are both socio-

¹⁴ Todaro lists the relationship between education and migration as one of the high-priority areas for research. As he puts it: “Although it is well known that more education increases the propensity of an individual to migrate, we are still unclear as to how much of this increased propensity can be explained solely by economic factors . . . and how much is the result of the impact of education on a rural individual’s ‘world outlook.’ ”

logical and economic. The former include such things as education's role in broadening one's horizons, creating an awareness of opportunities, instilling a willingness to take risks, and inculcating a desire for modern things. A more economic explanation is that better-educated workers migrate at higher rates because they have more to gain: absolute income differences may be greater across regions for better-educated workers than for less educated ones.

The hypotheses on sex differentials, for persons of any given educational attainment, are: (a) the average propensity to migrate is higher for females; (b) the marginal propensity to migrate is higher for females; (c) the gain from migration is lower for females. A general characterization of migration in Latin America is that females have higher average and marginal propensities to migrate than do males. In Colombia, as in other Latin American countries, it is common for teenage girls and young women to migrate, leaving the men behind. Many Colombian women take jobs as personal service workers, especially in domestic service; few Colombian men would dare do so. Thus, for a given monetary gain in income, women would be expected to migrate at higher rates; and for a given change in monetary incentives, women would be more responsive than would men. On the other hand, it is also argued that discrimination against women, cultural differences in males' and females' roles in Colombian society, and parents' preferences for investing in education and skills of their sons rather than their daughters, all act to hold down women's incomes and to reduce the gain from migration for women below the gain for men.

The simple tabular evidence in table 4 suffices to test parts *a* and *c* of these hypotheses.

First, the average propensities: for both males and females, the place-to-place migration rate increases monotonically with education; it is four times as high for those with higher education as for those with none. Within an education group, females' migration rates are higher

TABLE 4
DATA ON EIGHT SUBSAMPLES, 12 GEOGRAPHIC ZONES, COLOMBIAN CENSUS

Subsample	Migration Rate (<i>RRECMIGIJ</i>)	Mean Income	Standard Deviation of Income
1. Males, no education	1.10	619	189
2. Males, primary education	1.42	900	355
3. Males, secondary education	3.19	2072	610
4. Males, higher education	4.85	6818	1997
5. Females, no education	1.25	360	109
6. Females, primary education	1.75	554	173
7. Females, secondary education ..	3.41	1528	263
8. Females, higher education	4.90	2801	997

than males' rates. Note that sex-specific differences in migration rates across education groups are much larger than education-specific differences in migration rates across sexes.

Now for the gains from migrating: the standard deviation of income across zones is 11 times as high for males with higher education as for males with no education; for females, the ratio is nine to one. What this means is that a random move by a highly educated male would add 11 times as much to his income (in pesos) as would a random move by an uneducated male; for females, the gain is 9 times as many pesos. These ratios are very similar to the ratios of income means. Thus, though the absolute gains are much greater for those with more education, the percentage gains in income from making a move are quite similar for the various educational groups. Differences in absolute gains are also found in comparisons by sex. At any given level of education, females have less to gain absolutely from making a move than do males; for example, the standard deviation of income across zones for persons with primary education is 900 pesos for males and only 554 pesos for females. However, females' incomes are also lower than the incomes of males in comparable education categories, so the relative gains from migrating are not very different.

To test the hypotheses about the marginal propensities to migrate for various education and sex groups, we turn to regression analysis. These regressions parallel those presented above for the full sample. For each of the eight subsamples defined in table 4, separate migration functions were run relating the migration rate of persons in the i th group to group-specific economic conditions in origin and destination:

$$MIGIJ^i = \alpha^i + \beta_1^i DISTIJ + \beta_2^i AVGYI^i + \beta_3^i AVGYJ^i + \beta_4^i EMPLRI^i + \beta_5^i EMPLRJ^i + \epsilon^i. \quad (17)$$

Note the group-specific superscripts for all relevant variables. Equation (17) includes the employment rates as well as the income variables and distance. These employment rate variables were found to cause estimation problems for the full sample. To avoid similar problems with the subsamples, $EMPLRI^i$ and $EMPLRJ^i$ were then dropped and a second set of regressions was then estimated for each subsample. Table 5 presents the results of double logarithmic estimation of these two sets of regressions for recent migration of the eight subsamples (*LO-GRRECMIGIJ*).

Here are the principal findings from the regression analysis:

i) Econometric difficulties arise for nearly every subsample in the first set of regressions: average incomes are seldom significant and employment rate variables have the wrong sign more often than not.

TABLE 5

PLACE-TO-PLACE MIGRATION REGRESSIONS, EIGHT SUBSAMPLES, ALTERNATIVE FUNCTIONAL FORMS
(Dependent Variable = Log of Rate of Recent Migration *LOGRRECMIGIJ*)

Subsample	Constant	Log of <i>DISTLI</i>	Log of <i>AVGYI</i>	Log of <i>AVGYJ</i>	Log of <i>EMPLRI</i>	Log of <i>EMPLRJ</i>	<i>R</i> ²	SEE
Employment Probability Included								
1. Males, no education	−63.84	−.43* (.05)	.12 (.42)	.27 (.42)	6.81 (10.87)	6.95 (10.87)	.366	1.02
2. Males, primary education	82.56	−.44* (.05)	.12 (.47)	−.53 (.47)	7.89 (14.60)	−24.86 (14.60)	.391	1.01
3. Males, secondary education	182.80	−.35* (.05)	−.32 (.59)	.01 (.59)	8.89 (12.80)	−47.82 (12.80)	.453	1.07
4. Males, higher education	62.23	−.20* (.06)	−.46 (.30)	.58* (.30)	12.07 (7.27)	−25.51 (7.27)	.221	1.20
5. Females, no education	71.11	−.41* (.05)	−.20 (.25)	.11 (.25)	−.08 (14.45)	−14.89 (14.45)	.346	1.03
6. Females, primary education	81.32	−.43* (.05)	−.29 (.28)	.40 (.28)	−8.85 (12.77)	−8.59 (12.77)	.334	1.10
7. Females, secondary education	115.45	−.34* (.06)	−1.12* (.60)	2.21* (.60)	5.36 (7.83)	−31.86* (7.83)	.341	1.18
8. Females, higher education	−.168	−.13* (.06)	.087* (.044)	.103* (.044)	−1.27 (.89)	1.56* (.89)	.120	1.16
Employment Probability Excluded								
1. Males, no education	1.98	−.43* (.05)	−.09 (.25)	.06 (.25)	...	...	.363	1.01
2. Males, primary education	1.34	−.44* (.05)	−.11 (.19)	.21 (.19)	...	...	.377	1.01
3. Males, secondary education	−7.74	−.35* (.05)	−.69* (.29)	1.95* (.29)	...	...	.396	1.12
4. Males, higher education	−.67	−.20* (.06)	−.58* (.31)	.83* (.31)	...	...	.136	1.25
5. Females, no education	1.58	−.41* (.05)	−.20 (.23)	.23 (.23)	...	...	.341	1.03
6. Females, primary education	.19	−.42* (.05)	−.21 (.25)	.48* (.25)	...	...	.329	1.10
7. Females, secondary education	−6.28	−.34* (.05)	−1.13* (.63)	2.24* (.63)	...	...	.260	1.24
8. Females, higher education	1.18	−.13* (.06)	−.08* (.04)	.09* (.04)	...	...	.088	1.18

* Regression coefficient is statistically significant at least at the 95% level; numbers in parentheses = SEs.

Apparently, this is because of the same simultaneity and multicollinearity problems that plagued the regressions for the full sample. These problems may be lessened by eliminating the employment rate variables. Hence the first set of results, which include the employment rates, should be given little weight and the second set, which excludes employment rates, should be used instead.

ii) In the second set of results, the economic model of migration works for most subgroups: higher average income at destination significantly attracts migrants in five of the eight subsamples.

iii) The asymmetric results persist in the subsamples. For seven of the eight subsamples, the estimated effect of destination income in migration is larger in absolute value, often significantly so, than the corresponding effect of origin income.

iv) As educational attainment increases up to the secondary level, the estimated effects of origin and destination income both get larger in absolute value. That is, better-educated groups exhibit more responsiveness to regional differences in income than do persons with less education. However, this effect is diminished for both sexes at higher education, presumably because of the high propensity of highly educated individuals to migrate in any case. The patterns encountered in the lower educational ranges, which include the great bulk of the Colombian people, confirm the hypothesis that, on the whole, education raises the marginal propensity to migrate in response to economic opportunity.

v) The proportion of variance explained declines precipitously at higher education. This may be for genuine behavioral reasons—that persons with more education tend to move from rural to urban areas anyhow—or because the explanatory variables for the highly educated group are measured less precisely.¹⁵

vi) For each education group, females are more responsive to geographic differences in economic opportunity than are males. (Compare the coefficients on the income variables in subsample 5 with those in subsample 1, subsample 6 with subsample 2, etc.) This confirms the hypothesis that the marginal propensity to migrate is higher for females than for males.

vii) For each education group, the proportion of variance explained is higher for males than for females. Probably this is because Colombian women migrate more often for noneconomic motives (e.g., marriage).

The results of the demographic breakdowns may be summarized by an “incentive effect” and a “behavior effect.” The incentive effect

¹⁵ Taken together, the six rural zones had only 81 males and 13 females in the higher-education category, or an average of just 13 males and 2 females per zone. With so few cases, the migration and average income variables must certainly contain considerable errors in measurement.

arises when one group faces a wider range of opportunities than another. The behavior effect refers to the greater willingness of one group to move in response to given opportunities. For the education differentials, these effects reinforce one another and lead better-educated workers to migrate more. But for the sex differentials, the two effects act in opposing directions, causing the differences between males' and females' migration rates to be smaller than would be predicted from either component alone.

VI. Conclusion

This paper has explored the determinants of place-to-place migration in Colombia using an economic model of migration behavior. Six specific hypotheses have been formulated in the course of this paper. These hypotheses, and the test results, are:

Hypothesis 1. High income zones have higher in-migration rates and lower out-migration rates than do low income zones. This hypothesis is partially confirmed. For the population as a whole and for nearly all demographic subgroups, high-income zones are found to have higher in-migration rates than low-income zones. However, by the usual statistical criteria, it cannot be said with a high degree of confidence that higher origin income retards out-migration on balance; null results are the rule.

Hypothesis 2. Zones with more stable employment have higher rates of in-migration and lower rates of out-migration than do other zones ceteris paribus. The available data do not permit econometrically valid tests of this hypothesis since the ceteris cannot be held paribus, though econometrically invalid tests can be performed. The hypothesis is neither confirmed nor refuted; it has not been subjected to rigorous test because of data limitations.

Hypothesis 3. Zones that are further from one another have less migration between them. This hypothesis is confirmed in every test for all groups.

Hypothesis 4. For persons of either sex (a) the average propensity to migrate rises with education; (b) the marginal propensity to migrate rises with education; (c) the gain from migration rises with education. All parts of this hypothesis are confirmed.

Hypothesis 5. For persons of any given educational attainment (a) the average propensity to migrate is higher for females; (b) the marginal propensity to migrate is higher for females; (c) the gain from migration is lower for females. All parts of this hypothesis are confirmed.

Hypothesis 6. The large disparities in migration rates across education groups and the smaller disparities in migration rates by sex can be explained by differences among demographic groups in incen-

tives to migrate and differences in behavior even for the same incentives. The incentive effect and behavior effect work in the same direction to cause better-educated workers to migrate at higher rates, whereas they offset one another and result in smaller differences between males and females. Hence, the hypothesis is confirmed.

Overall, the results sustain the empirical validity of the economic model of migration in the Colombian context, for both sexes and for various educational groups. In addition, these data suggest some valuable insights into the workings of the Colombian labor market. For one thing, the open unemployment rate is a poor measure of the goodness or badness of labor market conditions. This is not a novel conclusion, coming as it does a decade after the ILO report on Colombia which first proposed an income criterion for evaluating employment conditions, but it is sometimes forgotten.¹⁶ Second, the large income differentials encountered suggest that wage flexibility has not come close to equilibrating labor markets spatially. One is hard pressed to defend in the Colombian context the notion that wages are determined by supply and demand. Third, these patterns are consistent with the view that past migration may have been caused by large rural-urban income differentials, just as the economic model of migration would predict, with unemployment differentials arising as a response. And finally, these patterns suggest that the Colombian economy may have drawn near a Harris-Todaro equilibrium: interregional inequality, rural-urban migration, and urban unemployment and underemployment all persisting *ad infinitum*. Whether such an equilibrium actually has been reached remains to be seen.

¹⁶ International Labour Office, *Towards Full Employment: A Programme for Colombia* (Geneva: International Labour Office, 1970).