This work is licensed under a
Creative Commons
Attribution – NonCommercial - NoDerivs 3.0 Licence.

To view a copy of the licence please see:
http://creativecommons.org/licenses/by-nc-nd/3.0/
ETHNICITY IN KENYAN RURAL-URBAN MIGRATION:
A TEST OF THE INFORMATION HYPOTHESIS

by

Hillard Huntington

August 1973

Any views expressed in this paper are those of the author. They should not be interpreted as reflecting views of the Institute for Development Studies or of the University of Nairobi.
ETNICITY IN KENYAN RURAL-URBAN MIGRATION:  
A TEST OF THE INFORMATION HYPOTHESIS

by

Hillard Huntington

August 1973

ABSTRACT

The paper examines the importance of information flows in determining the pattern of rural-urban migration in Kenya. The similarity between the ethnic compositions of the rural and urban areas is used to approximate the level of information linking the two areas. Empirical models incorporating this variable provide a more powerful explanation of migration rates, while at the same time causing distance, which is usually hailed as an important force, to be an insignificant determinant of migration. Secondary conclusions include: (a) urban and rural income elasticities appear to have different magnitudes, and (b) the expansion in the urban modern sector does not contribute to the explanation of migration rates.
Any attempt to explain migration in Kenya must contend with the ecological as well as cultural diversity within the country. Ecological conditions are often incorporated into economic models under the label of rural incomes. In these models, the migrant is hypothesized to respond to the differences between destination and source incomes as well as inversely to the transportation costs, as measured by distance. The second factor—the ethnic-similarity between host and sender populations—has not been explicitly introduced into empirical estimates of the migration decision. Although this variable has a non-economic flavor, ethnicity is highly relevant to the economic decision of seeking employment in African towns. Kinsmen are often the primary source of information about employment in both the modern and the small-scale sectors of the urban economy.

The introduction of ethnicity into the present model provides some important insights into an old variable—distance. Although distance is frequently used to represent transportation costs, it consistently explains more of the variation in cross-section migration rates than the other costs or benefits of the decision, specifically the source and destination income levels. Many have acknowledged that distance may also be incorporating other influences, e.g., cultural differences in inter-regional migration in Ghana studied by Beals et al. However, seldom has it been suggested that these influences are the main source of distance's sacred role in the migration decision.

1. However, most income measures used by economists to approximate rural income do not do justice either to the ecological conditions or the opportunity cost of labor. See Hymer & Rasnick. 2.
2. One of the few exceptions is Nelson. Those familiar with his work will notice that it was an important stimulus to the present study.
This study investigates the distance variable in Kenyan rural-urban migration by separating the effects of ethnic diversity from that of transportation costs. It will be shown that once the similarity of ethnic composition between two areas is held constant, distance becomes an insignificant variable. Moreover, the replacement of distance by the ethnic-similarity measures contributes substantially to the explanatory power of the regression model.

The empirical investigation begins with a test of the simple model that variations in migration rates are determined by variations in the following variables: the levels of urban and rural incomes and the distance between rural and urban areas. Migration rates in this study will be defined as the level of migration between rural area, \( i \), and urban area, \( j \), divided by the product of rural and urban populations, or \( m_{ij} = \frac{n_{ij}}{n_i n_j} \). This particular form of the dependent variable attempts to control for the influences of population size on the level of migration, thereby making the economic relationships more readily discernible. For example, the random chance that migration will be greater occurs not only as the rural population increases but also as the destination population is larger. One would expect that the level of migration into a section of Nairobi would be less than that into the entire city, independent of any economic variations.

The urban-income level is expected to be directly related to migration rates. The "pull" effect of this variable has been so frequently noted, that no additional comment on it is needed. However, the rural-income level has a much more ambiguous role. Higher rural incomes/substitution effect, which raises the attractiveness of remaining in the rural area. At the same time, they engender an income effect that could very likely encourage more outmigration. The strength of this income effect would depend upon the spread of the cash economy among other factors. Increased monetization of the rural area could increase migration to the towns by: (a) providing the rural residents with job experience that would help them to obtain urban employment, (b) placing more of the residents' home production and consumption activities into a market system and thereby reduce the costs of migration (both resource and
psychological), and (c) creating a demand for market goods and investments that requires income from outside the rural economy. Higher rural incomes also allow expenditures on education, which is a decided benefit in obtaining urban employment. In an attempt to isolate the impact of schooling, an education variable measuring the level of primary schooling as a percentage of origin population was added to this first run. Finally, greater distances are expected to reduce migration.

The regressions were estimated in double-log form for migration flows between six rural provinces in Kenya and the eight towns that were studied by Rempel (13). The regression equation, where the variables are natural logs, is:

$$ m_{ij} = a_0 + a_1 Y_j + a_2 Y_i + a_3 H_i + a_4 D_{ij} $$

where, $m_{ij}$ = migration rates (defined above) for 1964-68, $Y_j$ = urban modern-sector earnings in j, 1964, $Y_i$ = rural monetary income in i, 1964, $H_i$ = primary school enrollment as a percentage of province population, 1969, $D_{ij}$ = distance between area, i, and town, j.

The regression results are presented in column (a) of the table on the next page. All coefficients (elasticities) are significant at least at the 95% level. (Only the rural income one falls below a 99% level of significance). The substitution effect of the rural income appears to predominate in the cross-section, as the coefficient is significantly negative. However, the magnitude of the rural income elasticity is considerably smaller than that for urban incomes, which demonstrates the conflicting income and substitution

1. See Hymer & Resnick (5) for an analysis of changing resource costs.
2. Berg (2) discusses this form of an income effect with reference to the concept of backward-bending supply curves for labor in Africa.
3. The correlation between schooling and rural incomes is a mild .54.
TABLE = Regression Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-34.16**</td>
<td>-31.12**</td>
<td>-31.21**</td>
<td>-31.57**</td>
</tr>
<tr>
<td></td>
<td>(4.94)</td>
<td>(5.26)</td>
<td>(5.29)</td>
<td>(5.26)</td>
</tr>
<tr>
<td>Urban Income</td>
<td>5.91**</td>
<td>5.56**</td>
<td>5.46**</td>
<td>5.57**</td>
</tr>
<tr>
<td></td>
<td>(4.53)</td>
<td>(5.01)</td>
<td>(4.96)</td>
<td>(4.92)</td>
</tr>
<tr>
<td>Rural Income</td>
<td>-0.90*</td>
<td>-1.10**</td>
<td>-1.04**</td>
<td>-1.05**</td>
</tr>
<tr>
<td></td>
<td>(2.07)</td>
<td>(2.94)</td>
<td>(2.84)</td>
<td>(2.82)</td>
</tr>
<tr>
<td>Primary Education</td>
<td>2.11**</td>
<td>1.29</td>
<td>1.15</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>(2.95)</td>
<td>(2.01)</td>
<td>(1.85)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.89**</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnic - Similarity</td>
<td>0.80**</td>
<td>0.93**</td>
<td>0.93**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.79)</td>
<td>(6.06)</td>
<td>(6.02)</td>
<td></td>
</tr>
<tr>
<td>Employment Expansion</td>
<td></td>
<td></td>
<td></td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.53)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.56</td>
<td>.69</td>
<td>.69</td>
<td>.69</td>
</tr>
</tbody>
</table>

# Negative because it is a log of a Fraction.
* Significant at 95% level.
** Significant at 99% level.
Numbers in parenthesis are t-values.
effects - of the former. Equation (a) explains over one half of
the variation in migration rates.

The second regression equation introduces a variable intended
to demonstrate the role of information in the process. Presumably,
distance incorporates at least two major components: fewer
information linkages and greater transportation costs. The first can
be separated from the other through introducing a variable that will
hold constant the importance of these linkages. This variable is the
similarity in ethnic compositions between rural and urban areas.

If information and ethnic-similarity are to be important,
several conditions must be assumed to exist. (1) Information is
scarce and hence an important component in the migration decision.
(2) The more information that flows between two areas, the more widely
dispersed it will be over the population of the emigration area,
i.e., information is not concentrated in a few households.
(3) Information is channeled within an ethnic group from one area to
another but not between groups.

1. The urban-income coefficients reported were for earnings
estimates corrected for the uneven distribution of non-African
employees in the different towns. (See Appendix A). These
income estimates tended not to fluctuate greatly from town to town.
Since migration rates demonstrated considerably variation, the
income elasticity appears to be quite high. Regressions were also
run using uncorrected modern-sector incomes, which appear to
overstate the differences in African incomes between the larger and
smaller towns. The elasticity for these estimates were
considerably lower (and significant) and probably represent
reasonably well a lower limit to the magnitude of the income
elasticity. These were approximately twice the magnitude of the
rural-income elasticity and thus support the contention that the two
have unequal magnitudes. This was also true for a second rural
income measure which included some unpublished data on rural,
nonagricultural industries in addition to agricultural income.
If these conditions exist, then the number of potential urban-rural contacts will approximate the variations in information flows and hence contribute substantially to the explanation of migration rates.

The number of potential contacts for one individual in the rural area would be the number of people from his ethnic group that are now residing in the town. The contacts for one ethnic group can be calculated by multiplying the number of group members in the rural area by that in the urban area. The potential contacts for all groups would simply be the summation of these products for each group, or:

\[ C_{ij} = \sum_{t=1}^{T} P_{it} P_{jt} \]

where, \( P_{it} \) = population of rural area, \( i \), in ethnic group, \( t \),
\( P_{jt} \) = population of urban area, \( j \), in ethnic group, \( t \),
and \( t = 1, \ldots, T \) ethnic groups.

Migration rates would then be a function of these potential contacts divided by the total populations in both areas,

\[ M_{ij} = \frac{C_{ij}}{P_{i} P_{j}} \]

Equation (3) notes that the ethnicity concept can be computed by multiplying the proportions of each area's population (rural and urban) belonging to a particular ethnic group in 1962, the products being summed over \( T \) ethnic groups.

1. Of course, rural-urban contacts provide many other services. Remitted urban earnings, the psychological benefits of being around one's friends and relatives, and economic support for job-searching are some of the more important services that encourage the urbanization process in certain towns.
So far, little has been said about the kinds of information that ethnic group members can transmit to each other. At one level, the information about employment is of a general nature, i.e., "conditions look good (or bad) in Nairobi." This general type of information will tend to amplify whatever the response to the economic conditions will be. If the conditions are good, sufficient information will encourage migration. In this way, ethnic-similarity and economic conditions are complements and influence migration through a multiplicative effect. To be encouraged, a potential migrant needs both the income incentives and information; it is not an either-or situation. Nelson (12) concludes that this primarily explains why double-log regressions for migration are a superior form to arithmetic ones.

However, information also incorporates a sizeable specific element. Urban contacts also provide information on specific jobs available; how to obtain them, etc. Moreover, a rural resident from a particular ethnic group probably realizes that his opportunities for employment are greater in some towns than in others regardless of the over-all employment picture. Hence, this specific information encourages migration when the general conditions may not favor it otherwise. In this manner, migration rates and the information linkages defined by the ethnic groups should be positively correlated.

The results of the second regression equation are presented in column (b) of the earlier table. In this regression, rural income shows a significant substitution effect, whereas some of the "pull" of the urban income elasticity has been incorporated into the ethnic-similarity measure. The most remarkable result, however, is the insignificance of the distance variable. Once ethnic differences are accounted for, distance does not contribute to the regression results. When distance is replaced by ethnic-similarity (column (c) in the table), the explanatory power of the regression equation remains at its considerably higher magnitude ($r^2 = .69$ as compared to .56 in the first regression). Only the education variable remains insignificant at the 95% level.
A final regression equation explaining migration rates was executed using urban employment expansion as an additional independent variable. The use of this variable for migration in some developed countries has been criticized as being an almost definitional relationship. A majority of geographical migrants in the United States already have a new job in the destination area before they move. Hence, most migration will not occur in the absence of new jobs, unless many residents of the destination area are losing theirs at the same time. However, the situation in a country like Kenya is much different. Most migrants do not have employment waiting for them and often anticipate periods of unemployment or marginal employment while they seek an urban job (Gugler (3), Todaro (15). Moreover, any measure of employment expansion in Kenya will exclude the numerous opportunities in the non-enumerated or small-scale sector. It would appear, then, that the migration-employment expansion relationship in Kenya would not have the definitional connotations that its counterpart has in some developed countries.

The construction of an employment-creation measure followed, as closely as was possible, that suggested by Todaro's model of migration and urban unemployment (15). Todaro hypothesized that the relevant ratio was the number of new jobs divided by the number of job-seekers already in the city. It was not possible to approximate this latter number for all towns so the total population in the town under consideration was used as a proxy. If unemployment rates in the towns were at all similar, total population would move with the absolute size of the pool of aspirants to modern-sector employment. Regression equation (d) shows that the employment expansion in the urban modern sector had an insignificant (even negative) effect on migration rates. (A simple percentage increase in employment also failed to be a significant explanatory variable). There are several possible explanations. First, the measure may not truly reflect the chances of obtaining employment because of the inability to approximate the number of job-seekers. Although possible, it does not appear probable. The towns most deviant from the expected pattern appear to be the two larger and higher-income towns, which show too small an increase in employment to "justify"
the migrants that they have attracted. The Harris-Todaro general-equilibrium condition (4) would predict that these towns had a higher rate of joblessness.¹ Thus, the expected stock of job-seekers, or the rate times the urban population, would increase more rapidly than the urban population alone as the analysis shifted from the smaller to the larger towns. This would lower the job-opportunity index in the large towns and thus drive even further the regression estimates away from the predicted pattern. In short, the job-creation rate measure, if it has a serious bias, would tend to favor the Todaro hypothesis.

A second explanation may lie in the nature of the information provided by ethnic groups. If it were to include a sizeable specific element — i.e., a person moves to a town where his tribesmen have been successful in the past — then the response to variations in the job-creation rate may not be that sharp. It may be that those who are most responsive to these variations are those whose main source of information is not the ethnic linkages, e.g., secondary school leavers. As this alternative source of information spreads through more of the population, important changes in the nature of the rural-urban migration process will undoubtedly emerge.

A final explanation would be that, in addition to modern-sector employment, the rural population responds to opportunities in the small-scale sector as well. This could be because they view changes in this sector as either: (a) providing additional

¹ The condition is: \( Y = \frac{N}{S} Y_u \), where \( Y_r \) and \( Y_u \) are the rural and urban incomes, respectively, and \( u \) and \( N/S \) is the employment rate. If rural income is constant and urban income is greater for these towns, then the employment rate will be less. In other words, with a homogenous rural area (as in their model), higher incomes encourage more job-seekers.
urban opportunities that have a higher remuneration than rural work, or (b) affecting the costs of urban job-search in the modern sector. In either case, this argument would lead one to conclude that there is no systematic relationship between modern and small-scale opportunities for urban Kenya as a whole. Different towns will have dissimilar experiences in this respect.

The ethnic-similarity index is undoubtedly influenced by past migration. To the extent that this previous migration responded to approximately the same economic conditions that govern present migration, this variable may not be completely independent. For example, Nairobi may have had the highest urban income during both migration periods, thereby making past migration (and, hence, ethnic-similarity) a function of urban income. A fifth regression (not reported in the table) was estimated with ethnic-similarity as the dependent variable and the income levels and distance as explanatory ones. Only distance is a significant force in determining ethnic similarity, and it has been dropped from regression (c).

1. The I.L.O. (6) has suggested that the remuneration from a number of small-scale opportunities in the urban areas exceeds that from rural work.

2. The correlation between ethnic-similarity and past migration is .64.
This analysis has shown that attempts to estimate the income elasticities of migration may be seriously impaired by a misspecification of the supply function. In the particular case of Kenyan rural-urban migration, ethnic-similarity has proved to be an important variable, which has frequently been neglected in statistical studies. From a more general perspective, the results suggest that distance may not be able to incorporate completely the differences in information flows between source and destination areas. In these cases, additional information variables are needed.

Several secondary conclusions also arise from the analysis. First, it appears incorrect to view the urban-rural income differential as being one policy parameter. The regressions presented above have demonstrated that urban "pull" and rural "push" operate with dissimilar magnitudes. This has been attributed primarily to the opposing substitution and income effects of the latter. Second, little evidence was apparent for viewing the employment creation in the urban modern sector alone as being a significant incentive for migration.

1. It is also possible that migration will influence the income levels. I have tried to avoid this simultaneous-equation problem by choosing the income variables at the beginning of the migration period.
APPENDIX A: THE DATA

Statistical information on migration in African countries is notoriously scarce and inadequate. Hence, any conclusions arising from the data should be made with considerable qualification. In Kenya, some data on migration flows and income levels can be used by making a few adjustments. Although these estimates are not always ideal, I believe that the several conclusions derived from the analysis appear rather clearly despite these problems.

Migration: Migration levels were estimated from the I.L.O. report's figures on urban population growth and Rempel's sample of urban migrants. The I.L.O. report listed the annual growth rate in the African population for each major town (6, p.23). From these, the annual growth in numbers was established for each town, and an index was developed using Eldoret, which had the smallest increase, as a base equal to 1. This index was deflated by the implicit weights used in Rempel's survey (13, p.24); the new weights equalled the I.L.O. ones divided by Rempel's. Migration flows between provinces and the towns were then estimated by multiplying Rempel's observations on urban place of residence by rural province of birth (14, p.23) by the appropriate weight for each town. Finally, each migration flow was divided by the respective rural and urban population in 1962 (source: 11, Vol.I, and 9, p.15). One difficulty with these estimates is that no correction was made for: (1) different rates of return migration from each of the towns, or (2) the natural population increase in each town. Both of these could affect the relationship between the growth of an urban population and the gross immigration rate, which is the variable we wish to approximate, ideally.

Urban Income: Urban income is essentially earnings in 1964 in the modern sector of each town divided by employment (7, p.19 & 31). However, these figures include non-African employment as well. Consequently, an effort was made to express these income levels as earnings per African labor unit, or:

\[ Y_{u} = Y \cdot N_{A} \]  

where \( Y = \text{total earnings} \)  
and \( N_{A} = \text{total African labor units} \).

The divisor is a summation of African, Asian, and European employees, where each group is weighted by its average income relative to the African average.
earnings. Then,

\[ Y_u = \sum_{i=1}^{3} w_i N_i, \]

where \( w_i \) = earnings by race index

There are several steps in this calculation. (1) Average 1964 earnings by race for all private industry & commerce and public service employees were found in the 1972 Statistical Abstract (9, p.210 & 219). These figures were expressed as an index, where a base value of 1.00 was assigned to average African earnings. (2) The 1964 total employment in each town was distributed between the three races, according to their shares in 1971 employment (9, p.217). The 1971 base was dictated by necessity and is certainly not ideal. (3) The resultant employment by race was multiplied by the appropriate earnings index to produce total African labor units for each town. (4) Total earnings in each town were divided by total African units to yield average urban income for the African population.

This measure produces two major biases, which operate in opposing directions. The weighing of non-African employment by the average earnings of all races for the entire country may not reflect the wage structure existing in some cities. Non-African employment tends to be concentrated in the larger towns, where incomes in general are likely to be higher. Hence, \( \omega_i \) in the formula may be relating large-town non-African earnings to all urban earnings for Africans. If this is the case, \( \omega_i \) will be too high a correction for the smaller towns, whose incomes may be understated. On the other hand, \( N_i \) in the formula represents the 1971 structure of employment by race, which would undervalue the 1964 participation rates of non-Africans in the smaller towns. Hence, this adjustment would overstate the African income in these smaller towns. It is not possible to discern which bias is the stronger of the two.

**Rural Income:** Total rural income is comprised of three parts: (a) 1964 modern sector earnings in each province (7, p.28) minus the total earnings in the major towns, (b) 1965 value of crops on small-holder farms (10, p.23), and (c) the 1964 grade stock valued at the 1964 export price (10, p.103). The average provincial rural income was calculated by dividing total income by rural adult males (or total provincial adult males minus the adult males in major towns). No attempt was made to include the large
farms because the Africanization of large-scale agriculture was only in its initial stages in 1964. (Large-farm employment is included in the modern-sector employment figures).

**Distance:** Distance was defined as the miles between a town and the center of each Province's most-densely populated area.

**Ethnic-Similarity:** The calculations were explained in the text. They included any ethnic groups whose total population throughout Kenya exceeded 100,000. Occasionally, two or more groups were combined, e.g., Embu-Neru, Mijikenda-Taita, and the Kalenjin-speaking groups. This was done because, as the number of separate groups increased, the ethnic-similarity measure approached zero. The source of the data was the 1962 Census report (11, Vol. I & II). Due to boundary changes since 1962 the ethnic compositions of the rural areas were computed at the district level and then summed for each province. Some errors are involved in changing the 1962 boundaries into present ones. However, the ethnic-similarity measure is not very sensitive to the minor errors involved in this step.

**Education:** The province figures used were a weighted average of primary-school enrollment by district (6, 78-79).

**Employment expansion:** Modern-sector employment over time in each town was given in (7, 8). I used the period 1964-68.

**Ranking of observations for certain-variables:**

**Urban income: corrected**
- Nairobi, Mombasa, Kisumu, Nakuru, Nyeri, Thika, Nanyuki, Eldoret.

**Urban income: uncorrected**
- Nairobi, Mombasa, Nakuru, Kisumu, Nyeri, Thika, Eldoret, Nanyuki.

**Rural income:**
- Central, Rift Valley, Coast, Eastern, Western, Nyanza.

**Employment-expansion:**
- Kisumu, Nakuru, Nairobi, Thika, Nyeri, Eldoret & Nanyuki, Mombasa.
Working Paper No. 123 represents one part of my research project on rural-urban migration in Kenya. I wish to outline very briefly what my entire project includes in hopes that it might reduce the chances of research duplication.

I have been investigating the effect of information and the expansion of a money economy in rural areas on the supply of urban labor. Most of my work has focused on aggregate indicators of these forces at the district level. It contrasts sharply with previous work on rural-urban migration in this country, which has tended to emphasize the urban dynamics of the process (e.g., Todaro, Rempel, Johnson, Tobin, and Wasow among others).

The study has essentially three parts. First, a resource allocation model has been developed to explain under what conditions the expansion of the money economy in rural areas could coincide with more outmigration. This essentially involves an extension of the Boserup (The Conditions of Agricultural Growth) and Flinn-Resnick (5) models where a rural economy has a considerable non-market sector.

Second, rural province-to-town migration functions have been estimated using growth estimates from the I.L.O. report (6) and migration flows from Rempel’s survey (13, 14). The purpose of this part is to identify the relative importance of information flows, as given by the distribution of ethnic groups, in addition to testing several differences and the urban modern sector employment expansion. (See Working Paper No. 123).

And third, the analysis will be conducted at the district level to examine migration from some 30 rural areas (districts or combination of districts) into the two urban districts – Nairobi and Mombasa. With the aid of some unpublished information, I have attempted to reconstruct the boundary changes in the intercensal period so that migration flows during the 1962-69 period could be approximated. I shall be examining these flows with respect to a number of variables reflecting rural conditions in the districts – cash employment, cash income, amenity facilities, schooling enrollments, population density, and of course, ethnic composition, among other factors. Although most of the district indicators are either in crude or untabulated form, I do plan to make them as well as the migration flows available as soon as possible.
REFERENCES CITED


