PREDICTING MIGRATION IN KANSAS: AN EVALUATION OF ALTERNATIVE PROCEDURES FOR SELECTING PREDICTORS AND MODELS

by

KENNEY HAROLD BARB

B. S., Kansas State University, 1966

42-6074

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Sociology and Anthropology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

Approved by:

Cornelia B. Flora
Major Professor

Acknowledgements

I would like to thank my major professor, Dr. Cornelia Flora, for her assistance in completing this project. The advice and insights of Dr. James Miley and Dr. Richard J. Greechie were also beneficial and stimulating.

Most importantly I would like to thank my wife, Marilyn, for her patience, understanding, and encouragement. Without her substantial contributions, this project would have been immeasurably more difficult.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I

THE AREA OF INVESTIGATION AND THE PROBLEM

Migration as a Field of Study

Migration is of importance as a field of study because it is one of only three ways in which populations change their size and composition. Fertility and mortality are the additional two means by which such changes can occur. In the industrialized, developed nations like the United States, migration has the most immediate and dramatic impact. This is true because of the rapidity with which large numbers of persons may be affected by wide variation and sudden changes in migration rates. Conversely the death rate in such countries is low and unchanging, or declining only slightly. Birth rates are low and declining and fluctuate gradually when compared to internal migration rates. These are characteristics of populations in the "late transitional" stage of the demographic transition developed by Frank W. Notestein (Bogue, 1969: 56).

In areas with relatively stable death and birth rates, migration is thus the major component of population change. If migration is selective of people or areas with particular demographic, social, and economic characteristics, not only the size of populations in given areas is changed but also its composition. Involved in changes in the size and composition of populations, migration is seen as both a cause and effect of social change.

Generalizations about migration as a source of social change have been commonplace for millennia. Becker and Barnes (1961) cite evidence of the well-recognized interrelationship of migration, cultural contact, and mental mobility by thinkers as early as Herodotus, Thucydides, Plato, Aristotle, and others. In a section dealing with the Greco-Roman mobility and mentality Becker and Barnes point out that migration received one of its earliest explicit treatments at the hands of Thucydides (471-400 B.C.) who wrote as follows:

> ...Hellas appears to have had no stable sedantary population until a comparatively recent date, and to have been subject in earlier times to migrations, in which populations were easily dislodged from their homes under pressure from some more numerous body of intruders. There was no trade and no security of intercourse by sea or land. Each community lived at a subsistence level by its own local production, without accumulation capital or investing it in the land, since none could foresee when the next invader would deprive them of their homes, which they had not yet learned to fortify. They also took it for granted that their bare daily bread would be as easy to gain in one place as in another... For those reasons they migrated readily and therefore did not develop great manpower or armaments. The richest territories... were particularly exposed to changes of population. The fertility of the soil produced accumulations of power which resulted in ruinous civil disorders, and at the same time these countries were more eagerly coveted by foreigners... (p. 135)

Apparent in the foregoing passage is a trend toward geographic determinism, a concept which proved not to be wholly correct. The important thing, however, is that various reasons or explanations were offered to account for migration as well as comments about its results. The point illustrated by the above evidence is that migration's causes and effects are certainly not a newly recognized problem area to be dealt with. Despite its long known position of importance in relationship to social structure, social scientists are still relatively limited in the degree of precision with which they may make factual statements about who will migrate and the causes and effects of such a movement.

The failure of social scientists to handle migration effectively is not a result of their lack of diligent attention to the subject. Migration is a complex subject involving many variables. The perspectives from which the student of migration may proceed are varied; seldom can general laws of migration be determined which will endure over changes in time and locale. The problems associated with varied perspectives and multiple variables that hinder the social scientist in the study of migration are certainly not unique or insurmountable. In fact, these are the types of situations social scientists expect to deal with due to the nature of their subject matter. More important in hindering the students of migration in developing simple, inclusive generalizations about migration is that until recently very little data was available from which one could study migration, especially internal migration. This is not intended to imply that as soon as data is available insights about migration will become self evident, but rather that such data is an essential prerequisite to more illuminating inquiry.

International migration generally received the attention of record keepers earlier than did the internal mobility of populations. In the United States official records of immigration have been kept since 1820. A federal agency began to maintain records on immigration from the United States in 1904 (Shryock and Siegel, 1971: 20). The first census to ask a question from which internal migration studies could be derived was carried out in 1850. The 1850 census asked for state of birth of the native population. However it was not until the 1940 census that questions were asked on residence at a fixed date in the past. The data from the 1940 census provided material for analysis of migration that was gathered from direct questions about mobility. In addition to the direct questions method of obtaining information about migration there is also a second method available. Estimates of net migration can be made using various residual methods. The residual method considers the difference between the total change in population during a period and the change due to natural increase to be the result of net migration.

The availability of these two types of data from which migration may be studied consequently determined the types of migration research which has developed in the United States. One school of research considers the volume, direction, and differential characteristics of migrants based primarily on place of residence at a specific past time or state of birth. Reliance of this school on recall data subjects it to the shortcomings of such information, which is a disadvantage, but not a serious one when considered in comparison with the great contributions such data provides in the study of differential migration.

The second school of research employs data from the residual method. The perspective focuses on "net migration as a component of area population change. The unit of analysis is the ecological area and its characteristics, rather than groupings of migrants" (Stinner and DeJong, 1969: 456). Variations in social, demographic, and economic variables are seen as stimulating or retarding migration.

It should be noted that there is another way of obtaining data on the mobility of a population. This method is known generally as a direct registration method. The United States does not maintain the records necessary to obtain such data although Scandinavian and other European countries as well as a few East Asian countries do maintain such population registers. The population register requires that all persons who move transfer a record from one local registry office to another. This provides a universal, continuous record of changes of residences from which migration statistics can be compiled.

Several miscellaneous sources of migration information exist which have been used with varying degrees of success depending upon the scope of the study and the rigor with which the study was carried out. The Social Security Administration in the United States administers the Old Age and Survivors Insurance program for the government. The records kept by the agency are a source of data from which migration can be approximated. These records provide a non-census continuous registration system. Donald Bogue made a pioneering study with data available from such sources for the States of Ohio and Michigan in which he tried to measure job mobility as well as geographic mobility (Bogue, 1952).

Shryock and Siegel (1971) point out that the commercially produced "city directories" available in most sizeable cities in the United States have been used for sources of data on mobility, with the study of Norristown, Pennsylvania by Goldstein (1958) being the classic example.

Given the nature of the problem which was pursued in the present study, the residual method was employed. The specific advantages and shortcomings of using the residual method will be discussed in a later section dealing with the data collection and net migration rates.

The Problem

The purpose of this study was to make some initial investigations into the feasibility of developing a model to predict net migration rates for Kansas counties. The focus of the study is on evaluation of ways to select variables for the model as well as on the type of model to be used. It was hoped that if such a model could be developed it would prove useful to city, county, and state planners in considering future needs for specific areas of the state. One aim of this project was to develop a simple, effective approach that would use data that was readily available or data that could be calculated at a nominal cost. Such data, consisting of characteristics of counties at a given point in time, would be used to predict the net rate of migration to or from a county over an ensuing time period.

The decade from 1960-1970 saw a wide range of variation in migration for Kansas Counties (Table 1). Two counties, Saline and Geary, had out migration rates of over 30 percent. Twelve counties had out migration rates between 20-30 percent; 48 counties ranged between 10 to 20 percent; and 29 counties experienced between 0.1 to 10.0 percent out migration. One county, Finney, had no net migration. Thirteen counties experienced net in migration, but only four of these were over 10 percent. Johnson county had the highest net in migration for the state with +34.3. Of the state's 105 counties, 91 experienced net out migration, one county had no net migration, and only 13 counties had net in migration.

Table 1. Net Migration Rates for Kansas Counties, 1960-1970

20% And Over Net In-Migration

Johnson	+34.3
Douglas	+20.6

10% to 20% Net In-Migration

Riley	+19.6
Lyon	+11.3

0.1% to 10% Net In-Migration

Sherman	+7.1
Haskell	+7.0
Crawford	+2.1
Leavenworth	+1.9
Jefferson	+1.8
Osage	+1.1
Ellis	+0.2
Jackson	+0.1
Wichita	+0.1

0% Net Migration

Finney	0.0

0.1% to 10% Net Out-Migration

Franklin	-1.2
Harvey	-1.9
McPherson	-2.4
Linn	-2.6
Ford	-3.2
Wallace	-4.3
Grant	-4.4
Bourbon	-4.5
Pottawatomie	-4.6
Shawnee	-4.8
Cherokee	-4.8
Gray	-4.9
Waubaunsee	-5.0
Labette	-5.4
Neosho	-5.6
Reno	-5.7
Morton	-5.7
Scott	-6.1
Miami	-6.6
Clay	-6.7
Butler	-6.9
Cloud	-7.7
Ottawa	-8.1
Anderson	-8.2
Woodson	-8.3
Allen	-8.4
Thomas	-8.7
Coffey	-9.5
Doniphan	-9.9

10% to 20% Net Out-Migration

Sumner	-10.0
Marion	-10.0
Dickinson	-10.3
Cowley	-10.5
Stanton	-10.8
Brown	-10.9
Keary	-11.0
Republic	-11.6
Kiowa	-12.1
Wilson	-12.1
Cheyanne	-12.2
Chase	-12.2
Montgomery	-12.3
Mitchell	-12.3
Wyandotte	-12.5
Smith	-12.5
Norton	-12.5
Phillips	-12.6
Sedgwick	-12.8
Edwards	-12.8
Atchison	-13.0
Morris	-13.2
Jewell	-13.5
Nemaha	-13.6
Kingman	-13.8
Rice	-14.1
Washington	-14.5
Osborne	-15.3
Barton	-15.6
Stevens	-15.8
Elk	-15.8
Marshall	-16.0
Lincoln	-16.0
Logan	-16.1
Ness	-16.4
Clark	-16.4
Decatur	-16.7
Gove	-16.8
Seward	-17.4
Meade	-17.6
Harper	-17.6
Greenwood	-17.6
Chautauqua	-17.6
Rush	-18.7
Sheridan	-18.9
Ellsworth	-19.3
Comanche	-19.4
Stafford	-19.5

20% to 30% Net Out-Migration

Pratt	-20.3
Pawnee	-20.3
Hodgeman	-21.0
Russell	-21.1
Lane	-21.6
Hamilton	-22.2
Greeley	-22.2
Barber	-22.2
Rawlins	-22.4
Graham	-22.6
Rooks	-27.1

30% and over Net Out-Migration

Saline	-30.1
Geary	-30.2

The Frame of Reference

The frame of reference on which this study is based is essentially a derivative of a theoretical perspective known as neo-orthodox ecology. In view of the fact that the term "ecology" has been subjected to multiple uses across the academic disciplines (Duncan, 1959a: 678-679) as well as having been given various interpretations by the popular movements to clean up the environment some clarification and specification is in order. For those persons interested in the historical development of social thought which provided a sound background and set the stage for the emergence of an ecological perspective several interesting sources exist. The work of Ibn Khaldun (Dawood, 1967), Jean Bodin (cf., Becker and Barnes: 309), Montesquieu (cf., Becker and Barnes: 410), and especially Durkheim (Simpson: 1947) should be consulted. The work of these persons preceded the existence of any ecological school of thought as such. However, the present day ecological perspective has inherited a strong legacy from the writings of the previously mentioned men. One of the best references illustrating the implications of the connecting ties is Leo F. Schnore's "Social Morphology and Human Ecology" (1965) in which he points out the Durkheimian legacy of human ecology.

The ecological perspective has evolved from its beginning in the work of Park and Burgess (cf., Park: 1936) at the University of Chicago in the 1920's into three distinct schools of thought. Theodorsen (1961) has referred to these as the socio-cultural approach, social area analysis, and the neo-orthodox perspective. As mentioned above the third one, the neo-orthodox perspective provides a frame of reference which justifies the assumptions, design, and methodological consideration used in this project.

The neo-orthodox perspective contains four components which have been labeled the "ecological complex" (Duncan and Schnore, 1959b: 136). The complex is made up of population, organization, environment and technology which have been referred to as P.O.E.T. for mnemonic purposes. Ecological analysis may be said to generally be concerned with the presence or absence of various forms of the four major components and with the consequences for the human collectivity of such presence or absence. It is a perspective which centers on the assumption that the ways of life are a function of the conditions of life. The "ways of life" may be translated as "social organization" thus, enabling one to characterize the ecological perspective as primarily being concerned with the study of social organization.

The development of the neo-orthodox position has been primarily the result of the work done by Otis Dudley Duncan, Leo F. Schnore, Jack Gibbs, and Walter Martin. The four men have provided a theoretical framework and also have generated a good deal of the empirical support for their perspective.

The general goal of the neo-orthodox ecologist is to study the social life of man as a phase of the ecosystem. The ecosystem is the interacting environmental and biotic system: Duncan points out that the most

> ...salient feature of man's evolution is the expansion of his niche in the ecosystem. This expansion has involved (1) increase in numbers sustained by (2) increasing human resourcefulness in extracting the requisite supplies of energies and materials from the environment and (3) an elaboration of the patterns of organization of the human collective efforts involved in this activity (Duncan, 1964: 40).

Contained in the above characterization are the four components referred to above as P.O.E.T. Duncan and Schnore conceptualize the components as being interelated in the ecosystem in order to bring out the ..."aspect of interrelatedness which some writers have identified as the most fundamental premise in ecological thinking." (Duncan and Schnore, 1959: 136).

Within this ecosystem the elementary unit of analysis from an ecological standpoint is a pattern of activity. The system is seen as an overlapping, interpenetrating series of activities organized in various fashions.

This emphasis on the unit of analysis as an activity leads to a position which causes much controversy over the neo-orthodox perspective. Namely, it leads to the exclusion of individual considerations in the perspective. "The ecologist is interested in the pattern of observable physical activity itself rather than the subjective expectations that individuals may entertain of their roles. Ecological analysis does not attempt to explain the individuals feelings of obligation, the stresses he suffers as a consequence of performing several roles simultaneously or sequentially, or his motivational syndromes when he is engaged in different sorts of activities" (Duncan and Schnore, 1959b: 137).

Jack Gibbs and Walter Martin present a logical structure for neo-orthodox ecology (Gibbs and Martin, 1959) which is essentially the same as that presented by Duncan and Schnore. Gibbs and Martin increase the cause for criticism for excluding individual behavior when they define the focus of ecology as investigating man's sustenance activities in the cultural setting yet "independent of cultural values and the individual motives of men" (Gibbs and Martin, 1959: 33).

By taking such a position the neo-orthodox ecologist has committed himself to an avenue of study which compels him to view the distinctive activities he is studying, their numbers and kinds, as properties of aggregates or populations. The theoretical rationale underlying such a decision to focus on the aggregate level of analysis and exclude cultural along with individual differences has been argued elsewhere (Willhelm and Sjoberg, 1960; and Alihan, 1938). It is not my intention to evaluate here the difficult and complex issues involved in what is in effect an argument stemming from philosophical arguments over the basic positions of naturalism and idealism.

In this project I distinguish between the theoretical assumptions of ecological theory and the theoretical model it has provided for testing purposes. It is not my purpose to test the "theory" underlying neo-orthodox ecology. It is my intention to use a theoretical model developed from it as a frame of reference which illustrates the importance of migration as studied in this project. I am taking the approach employed by the neo-orthodox perspective as a given, viable frame of reference from which to view reality, leaving the theoretical arguments to others. What is important to understand at this point is that ..."[a] frame of reference is a set of definitionally interrelated concepts, defining the theoretically relevant properties of an object" (Bailey and Mulchy, 1971: 38). As such it can not be right or wrong, it does not "explain", and in a sense it can be pre-theoretical in that theories which explain may arise from it. The important question to ask is what is the quality of data a given frame of reference produces and what types of conclusions can be drawn from such data. It is this question which has importance to the present study. Before considering the question raised above about the logic of ecological inference and the interpretation of aggregate data it will be useful to summarize the preceding section.

The ecological complex as defined by the neo-orthodox ecologists embraces four concepts which were recognized by previous thinkers as relevant to the study of social organization. As such the ecological complex is an interdependent set of variables. A change in one component affects the other components. No one variable need be considered as an independent or dependent variable exclusively. The neo-orthodox perspective studies the adaptation of man by analyzing patterns of activities man is engaged in. As such the perspective studies populations or aggregates in spatially delimited areas which are assumed to have some distinctive unit characteristics. Thus studies from this perspective exclude individual level analyses and cultural variations.

The perspective furnishes the present study with a frame of reference which allows population to be seen as a dependent variable. Specifically, migration rate as a component of population change is seen as the activity pattern to be investigated as a dependent variable. Differential net migration rates are indexes of the adjustment or adaptation a population makes to the other components of the ecosystem. As such, investigations can be made of the impact of variations in community structure or organization on the net migration rate. In summary, social, economic, and demographic characteristics of the units being analyzed are indicators of variations which operate in the ecosystem to determine net migration rates. The frame of reference developed from the neo-orthodox perspective offers several points relevant to the following research: (1) the purpose of the societal level of analysis and the resulting need for the use of aggregate data is pointed out; (2) the exclusion of individual and cultural variations are justified; (3) the nature of the interelatedness of the concepts employed is shown to be considered most beneficially when thought of in ecosystem terms. The ecosystem is seen as cutting across various levels of society and thus entailing a macro level approach to the study of society; and (4) the sociological ancestry of the perspective can be shown to be sound and the current status of the perspective addresses itself to problems which are sociological in nature.

The question posed previously with reference to the quality of the data and the logic of ecological inference is in need of some discussion in light of its use in the present project. The data used to characterize the ecological units (counties) in the present project was aggregate data. Aggregate data is that which is given a unit characteristic although it is derived from a distribution of individual attributes or behaviors. The problem for some persons has centered around the debate over what kind of inferences one can make on the basis of aggregate data.

Ecological correlations employing aggregate data were the subject of strong criticism from W. S. Robinson in his now famous paper, "Ecological Correlation and the Behavior of Individuals" (Robinson, 1950). Robinson pointed out the ecological fallacy in attempting to make inferences about individual properties or behavior based on ecological correlations between group or aggregate characteristics. His example illustrated that the ecological correlation which showed the illiteracy rate to vary directly with percent Negro conveyed the false impression that individual Negroes were more often illiterate than non-Negroes. Robinson showed such reasoning to be fallacious and it seems fair to say that today everyone agrees an ecological correlation is not the same as an individual correlation.

Robinson's paper can essentially be seen as an attack on the interpretation and use of the ecological correlation rather than the validity of correlation itself. It should be kept in mind that it is not the purpose of this paper to provide explanations of why individuals may move. Rather the purpose is to identify characteristics of territories which account for net migration rates. In this case the net migration rate is also a characteristic of a territory as opposed to an individual property; which it can not be due to its definition.

This is not to deny that men's decisions influence whether, when, and where they move. It can be said however that empirical generalizations can be reached without considering such decision-making. Such generalizations are not seen as being fully explanatory of all the forces surrounding migration, but rather as being suggestive of possible partial answers. The aspect of reality which the neo-orthodox perspective captures should not be seen as an either-or type of proposition, but rather as being complementary to and augmenting the knowledge which other types of perspectives reveal.

Before moving on it should be pointed out, lest the neo-orthodox perspective seem too maligned, that the type of ecological fallacy Robinson pointed out can work in reverse: It is just as likely for persons to err in attempting to generalize from individual behavior to group or collective relationships. It is quite possible here to confuse the relationships among individuals with those governing aggregates. The essential point about either type of fallacy, be it ecological or "individualistic" (Alker, 1969: 78), is that one recognizes the possible errors in attempting to use such data and adjust his inferences accordingly.

Review of the Literature

It is appropriate to review at this time some major studies which have examined of migration from a perspective similar to the one employed in this project. This will give some background on the problem and illustrate more clearly the nature of the problem and the procedures which have been used in analyzing net migration rates.

Models similar to the ones employed in this study have been used in previous attempts to establish ecological correlation between area's characteristics and net migration rates. Some of the more notable findings are discussed below. The independent variables selected for this study were chosen in light of the following research.

"The study of population components of population growth has long been performed for the major units of nations: states. However the detailed analysis of growth in terms of particular communities dates only from the 1950" (Bogue, 1969: 776). Donald Bogue did one of the first analyses of the entire nation in terms of its detailed parts when he studied net migration and natural increase over the 1940-50 decade (Bogue, 1957). The units of analyses in Bogue's study were economic regions and subregions of the states. In nonmetropolitan areas where the net migration was for the most part outward, Bogue computed correlations which indicated out-migration occurred most often in agricultural areas of low level of living or with large concentrations of nonwhites. The rates of out-migrations were less in areas with some manufacturing but there was so little manufacturing growth in the nonmetropolitan areas that change in manufacturing had no effect on retarding the outflow. Bogue used changes in variables over the decade to account for variation in rates and hence his project was somewhat different from the one this paper entails. Nonetheless the correlations below offer potential areas of inquiry by showing the relationship of the given variables with nonmetropolitan net migration rates:

Variable	Correlation With Non-Metropolitan Net Migration Rate 1940-50
Percent change, total population	.917*
Percent change, rural-nonfarm population	.548*
Percent change, rural-farm population	.488*
Percent in agriculture, 1950	-.519*
Percent nonwhite, 1950	-.286*
Percent change, nonwhite, 1940-1950	.393*
Percent in manufacturing, 1950	.279*
Median income, 1949	.645*
Farm operation level of living index	.584*
Rural-nonfarm level of living index	.533*
Percent change in retail sales, 1939-1947	.082
Infant mortality rate, 1939-1940	-.162

*Significant at the 95 per cent confidence level (Bogue, 1969: 778).

The approach Bogue used in this study focuses on an <u>ex post facto</u> analysis of net rates. This is in contrast to the focus of this paper which takes characteristics at the beginning of a period and attempts to predict migration rates. Nonetheless studies like Bogue's are still useful to this project in that it illustrates interrelationships between characteristics of areas and net migration rates. For such studies to be of use in the model employed here, variables were evaluated in terms of the potential effect they may have over the next decade.

Before considering studies which have used the same approach that this project employs it will be useful to review the findings of other types of migration studies. From this review it is possible to gain insight into some potentially fruitful types of variables which might be used, as well as theoretical understandings of why they are useful.

Gladys K. Bowles and James D. Tarver provide one of the most illuminating and complete studies of net migration in their work, Net Migration of the Population, 1950-60, by Age, Sex, and Color (1966), the principal findings of which are summarized in chart from in Bogue (1969: 784).

Some of their findings which follow suggested variables which were used in this study. On the whole there was a difference in 1950-60 net rates of migration depending upon the degree of urbanism. Counties less than 50 percent urban lost population generally and the losses were more severe for completely rural or only slightly urbanized counties. This finding led to one of the control classifications used in this study, which is based on percent of county that is rural.

Tarver and Bowles also found a heavier exodus from counties where median family income was lowest and high in migration where family income was greatest. This encouraged the researcher to code "% of families under $2,500" and "over $10,000", as well as "median income" as possible relevant indicators of net rates.

The age differential in migration rates was made quite obvious in the Bowles findings. Rates were the highest in the age categories between 18 and 34 years of age. In view of the fact that migration is selective with reference to age categories the dependent variable in this study was viewed not only as a net migration rate but was also broken down into age-sex specific rates.

The U.S. Census for 1960 published as part of a special series characterizing the population a section dealing with migration. Mobility for States and State Economic Areas (1960) is a source of many relevant indicators of migration rates. For example from Table 6 of the above report it is quite obvious that the rates of migration vary directly with the level of education obtained. From such knowledge one concludes that the "% of the population over 25 with a high school education or more" might be a measure to look at in attempting to predict migration rates. "Median school years completed" is also another measure which could be relevant.

It should be noted that the generalization above holds for persons in the ages 25-34. Hamilton found in his studies of rural to urban migration that while for younger ages migration tends to be selective of the better educated, the reverse is true at the older ages. Older people who leave rural areas tend to be less well educated than those who remain (Hamilton, 1959).

It is a well documented generalization (cf. U.S. Census of Population, 1960a, Mobility for States and State Economic

THIS BOOK CONTAINS NUMEROUS PAGES WITH MULTIPLE PENCIL AND/OR PEN MARKS THROUGHOUT THE TEXT.

THIS IS THE BEST IMAGE AVAILABLE.

Areas, Table 9) that migration is selective of occupational categories. Professionals and white collar workers are more mobile than blue collars and laborers. Farmers and farm laborers tend to comprise a small percent of all migrants. Non-farm laborers also tend to be a relatively immobile group. Persons not in the labor force have generally high mobility rates. These facts suggest that percent of persons employed in various types of occupations as well as the percent of people in the labor force are of interest in attempting to predict migration.

As a result of studies of differential migration like those discussed above, it is well known that variation in age, income, sex, marital or household status, occupation, and education affect migration. Given such knowledge some people have taken an approach toward predicting migration rates which assumes that if characteristics of populations in given areas can be measured in a way which indicates the relative occurrence of such characteristics, then one can predict migration rates from such knowledge.

Working from such a position James D. Tarver developed a prototype in the late 1950's for predicting the movement of people during a given period of time. "According to the formulations of this archtype, migration among the various subdivisions of states, or of the United States is a corollary of the distinctive demographic, economic, and social structure existing therein" (Tarver, 1961: 207). In Tarver's first test of the model, he collected various data at a state level from the 1940 census. On the basis of preliminary regression analysis he classified the data into social, economic, and demographic categories. It was his purpose to use this data as independent variables in predicting net migration for the 1940-50 decade. The first results achieved were remarkable: Tarver accounted for 98 percent, 87 percent, and 95 percent, respectively, of the variations in the net migration rates of whites, non-whites and the total population. It is Tarver's contention that variables shown to be this powerful in accounting for variation of rates can be measured at the beginning of a period and used to predict migration for a given interval.

Tarver illustrated the amount of variance the economic, social, and demographic characteristics accounted for independently as well as interdependently. Independently they were found to account for 72 percent of the white rate's variation while interdependently they added another 26 percent to the variation accounted for. Tarver concluded that while the evidence discovered was prima facie support for the model he did not bar the remote possibility of spurious relations among the variables. He pointed out that if this existed it would limit the models ability to predict during another period.

Tarver discovered that for white rates the economic variable ("percent of employed persons working in construction industry") was the most significant accounting for 69 percent of the variation. The two social variables, "median school years by persons 25 and over" and "percent of persons 25 and over completing four or more years of high school" accounted for 41 percent of the variation. The demographic variables by themselves accounted for 20 percent of the variation. "Percent of population under 20 years of age" and "percent of population white" were the two demographic variables which Tarver used.

Tarver was encouraged enough with his first results on the state level that he applied a similar model to county level data for 1950 in an attempt to account for the 1950-60 net migration rates (1965a: 3). A "multiple covariance model" was used and accounted for over 30 percent of the variability in county net rates. Tarver used an additive model in this study which employed one independent variable, census division, and six covariables. In the interest of clarifying Tarver's usage of "independent variable" and "covariables" Hamilton has pointed out ..."It would be simpler to say that of the seven independent variables six are continuous (cardinal) and one is nominal. Also "census division" as handled mathematically for regression purposes is actually seven nominal dichotomies - which some writers refer to (somewhat peculiarly) as dummy variates" (Hamilton, 1965: 17).

Five variables (census division, percent nonwhite, median family income, percent under 20 years of age, and percent employed in construction) explained the most appreciable amount of the variability in mobility rates. Two variables, "median years of school completed" and "percent completing four or more years of high school" were not significant. Tarver inferred that his model may not be as effective for county level data as it is for state level data in view of his poorer predicting power. He did not test for interaction effects in his initial analysis of the data.

Hamilton's discussion (1965: 13) of Tarver's work led to a rejoinder (1965b: 17) by Tarver, the essence of which was a test of the interactive effects of the relevant variables. By testing for the interactive affects in an expanded model Tarver accounted for about 66 percent of the variability which was more than twice that accounted for in the original small model. These results can be seen as evidence that perhaps the county level data is more powerful than Tarver originally expected. It also illustrated the importance of considering variables taken simultaneously rather than independently.

William F. Stinner and Gordon DeJong (1968: 455) have applied an ecological model similar to Tarver's to an analysis of Negro migration from the South. The data they collected was from a county level unit of analysis and net migration rates for age cohorts were the dependent variables. Their aim was to derive a configuration of factors which would statistically predict the nature and extent of Negro migration.

The specific social, demographic, and economic variables employed in the study are not as important to this project as were those of Tarver since Stinner and DeJong specifically selected their variables with reference to their effect on Negro rates. The important finding to emerge from their study was that the ability of the ecological model to predict migration rates was highest in the young ages and declined steadily into the older ages.

The previous use of such ecological models with multiple regression analysis has left several questions partially unanswered. One purpose of this paper is to address itself to some of these questions. One, what is the best way to select the most useful variables given the myriad to choose from? Two, is the county unit a profitable unit of analysis from such a perspective or is it so responsive to sudden changes in structure variates that initial characteristics lose their explanatory power over the normal intercensal period of a decade? Three, are the interactive effects of social, demographic, and economic variables useful enough to warrant further study? Four, is it a characteristic of such models that they decline in explanatory power as the age specific net migration rates are drawn from older categories? And fifth, can such a model be used in Kansas to effectively predict migration and if so what are the important considerations in applying such a model? With the above considerations in mind it is now time to turn to the design and procedure employed here to answer such questions.

CHAPTER II

DESIGN AND PROCEDURE OF THE STUDY

An Overview

This study was an attempt to evaluate methods of locating independent variables which can be used to predict county net migration rates. Twenty-four independent variables were selected from the 1960 census of Kansas on the basis of their potential explanatory value as illustrated in prior research. These were taken as characteristics of Kansas counties in 1960 and tested for their ability to account for variation in net migration rates over the ensuing decade, 1960-1970.

The project attempts to locate the variables which are the most reliable indicators of changes in migration rates. Two methods are used to select the variables to be run in a regression model. A factor analysis was performed on the twenty-four independent variables. Three variables emerged as indicative of discrete factors and were run as independent variables in an attempt to account for the variation in migration rates. A preliminary step-wise regression analysis was also performed in an effort to isolate those items with the most ability to predict migration rates. The three best items resulting from this analysis were also run in the regression model. Both the factor analysis and the step-wise regression technique were employed in selecting the independent variables so that a comparison of the two techniques could be made.

The regression model used employed two basic equations. Both equations were run with both sets of independent variables selected. This allowed a comparison not only of the effectiveness of the step-wise versus factor analysis method of selecting variables, but also a comparison of the two regression models.

It was hypothesized that the independent variables would vary in their efficiency in predicting migration depending upon the degree of rurality of various counties. Consequently the 105 counties were divided into those which were from 0-50% rural, 51-99% rural, and 100% rural. This was accomplished by using the residual of the "percent urban" for each county as taken from Table 2, item 9, in the 1967 County and City Data Book (1967).

It was also hypothesized that there were different forces at work in counties which experienced net in-migration and those that experienced net out migration. Consequently the counties were also divided in this fashion and run separately in the models. A total of six categories resulted with the additional one being "all counties" taken together and run in the model. These six categories were the result of an attempt to increase the likelihood of accurate prediction as well as offer potential insight into the phenomena in question.

In addition to using the net migration rate as the dependent variable age-sex rates were also considered as dependent variables. Rates for males and females in two age categories were selected for the test. The rates for persons who were 10-14 years of age in 1960 and hence became 20-24 over the decade were chosen as one important category. This was done because of the large numbers of persons in such age categories who migrate. This age category is seen as a very important time in a persons life as he or she is involved in the post high school, post adolescent adjustment period. It was deemed to be of interest to determine what structural forces may be at work on such persons.

The age category 50-54 in 1960 was also chosen as a dependent variable for investigation. This was done to allow comparison of the effectiveness of the variables selected in predicting youthful versus elderly migration.

Thus, in summary of the overall design, there were two methods used to choose two sets of three independent variables. Each set of the independent variables were run in a two equation regression model. The equations were run on each distinct classification of Kansas counties in an attempt to account for variation in a total of five dependent variables.

The mathematical equations used in the regression model were set up as below where Y is the dependent variable and X_1, $X_2 \ldots X_n$ are the independent variables. The first equation was a simple additive model: $Y_{123} = U + B_1X_1 + B_2X_2 +$

$B_3X_3 + \varepsilon$. The second equation was also an additive model, but with the addition of an interactive effect: $Y_{1-7} = U + B_1(X_1) + B_2(X_2) + B_3(X_3) + B_4(X_1X_2) + B_5(X_1X_3) + B_6(X_2X_3) + B_7(X_1X_2X_3) + \varepsilon$. It should be noted that the four additional variables in the second model are created from the original variables. As such they are interaction terms by definition in this paper. They are not tests of interaction effects as thought of in the traditional schema for the analysis of covariance. As was indicated in an earlier section such a model had been noted to increase efficiency in accounting for migration rates. It is obvious that with the addition of any variable to an equation in a regression model we will not explain less of the variance. That is, any additional variable even if it has only a random effect cannot decrease the amount of variance accounted for in the three term equation, but it can reduce the level of probability. The problem then becomes not whether the second model with the interactive effects explains more of the variance, which it will do in almost every case, but whether it explains significantly more of the variance.

It is the precise problem raised above which requires some comment here. If the second equation involving the expanded model with the "interaction" terms is significantly better statistically speaking, what does this mean logically and substantively? This is a problem which has been addressed by Hamilton (1965: 13). The issue revolves around the interpretation and use of the results. The aim here is to distinguish between "explanation" and "significance". The purpose of this project is to develop a prediction model with the ability to forecast probable net migration rates by finding variables which account for the variation in the rates. If this can be accomplished, this project will be deemed successful. However it is apparent that anything that works, in this case the prediction model, is somewhat limited if it is not able to be understood logically. Consequently if the full model proves to be a substantially better predictor, it is appropriate that explanations be offered as to why it is so. Consequently I offered what may be referred to as "soft" explanations which were judged relevant to a theoretical explanation. This is an addition to the "hard" statistical significance which emerged.

As Hamilton (1965: 14) has pointed out, if one desires only a "pragmatic prediction equation", the procedure of the two models outlined above is quite appropriate and orthodox. This was in part my aim. Consequently when examining the findings, the reader should be aware of the connotation of words like "explain" and "account for".

The advantage of such a model employing characteristics of counties at the beginning of a census period is that it allows one to take the corresponding characteristics at a given date and predict migration over a future period. This requires that one assume the migratory trends remain approximately constant during the following period. Models which employ measures of change in characteristics which occur during a census period do not provide comparable predictive ability. Such models cannot be operationalized until the measures of change in the selected variables are available, and these are only available at the end of the period.

The Independent Variables

Below is a list of the original twenty-four independent variables which were coded for the project. This list was then used to choose the variables considered in the study.

1. % of total county population that is employed in agriculture hereafter referred to as (% agric.).
2. % of total county population that is employed in construction (% const.).
3. % of total county population that is employed in manufacturing (% mfg.).
4. % of males 14 and over employed in the labor force.
5. % of population that is 18 years of age or under (% 18 or <).
6. Effective fertility ratio (fer. ratio).
7. Median family income (med. inc.).
8. % families under $3,000 income.
9. % families over $10,000 income.
10. Median age of the population.
11. % of persons 25 years of age and over who have completed high school or more education (% 25>HS +).

12. Median school years completed by persons 25 years old or over.

13. % of housing units that are owner occupied (% own. occ.).

14. % of population that is rural-farm (% RF).

15. Farm operator family level of living index (farm op. lev. of liv.).

16. % of civilian labor force that is unemployed (% unemp.).

17. % of commercial farms under $2,500 in value of products sold (%<$2,500 VPS).

18. % of commercial farms over $10,000 in value of products sold (%>$10,000 VPS).

19. % of personal income from government disbursements (% PI gov't.).

20. % of population residing in state of birth.

21. % of total population that is between 7 and 13 yrs. of age and enrolled in school (% 7-13 in sch.).

22. % of males who are 14 and over and married (% married).

23. % of population that is employed as white collar workers (% WC.).

24. % of population that is rural-farm or rural non-farm (sum of two) (% RF + RNF).

The variables were coded and punched by this researcher rounding off to the nearest tenth of a percent. All of the variables were either taken directly from or calculated from the 1962 and 1967 County and City Data Book which is published by the Bureau of the Census or from the 1960 Census of Population - Kansas, Volume One, Part Eighteen (1960b).

Selecting the Independent Variables

Twenty-four independent variables were selected on the basis of prior research in which they had been found to be of importance in predicting variation in net migration rates. The variables were indicators of social, demographic, and economic characteristics of the 105 counties in Kansas.

The objective was to logically and systematically reduce this pool of twenty-four original variables to a pool of a more manageable size which could be used in a multiple regression model to account for variation in the net migration rate. Factor analysis was decided upon as one appropriate method to use in the preliminary analysis of the data.

Factor analysis is a technique of multivariate analysis which allows one to discover the structure of a multivariate universe of data. This is accomplished by revealing clusters or bundles of closely related elements contained within a matrix of correlation coefficients. Factor analysis starts from the premise that given "x" variables, in this case 24 original observations, the total variation in the matrix can be explained in terms of "y" factors (plus residual error). The number of factors, y, derived is less than the number of original variables, x. The objective then is to express as much variation as possible with as few factors as possible while minimizing the residual error term.

When the axes between the factors are rotated orthogonally, as they were in this research, the factors which emerge from the analysis are uncorrelated with one another. This technique allows for the isolation of factors which account for differing fractions of the total variation in the matrix. Each factor which results from orthogonal rotation may be considered to be unique in comparison with the other factors. This is not true of an alternative form of axes rotation known as oblique rotations where factors may share common variables.

A most important fact about factor analysis is that the results which are arrived at by such a technique are greatly influenced by the data fed into the analysis. If demographic variables alone are used one must certainly expect a demographic factor to emerge as highly significant. The factor analysis performed on the data for this project was essentially an analysis of social, economic, and demographic variables. Consequently, at least one factor of each type was expected. The purpose in running the analysis was to determine if there were any factors other than a social, economic, and demographic in the data. Secondly, the factor analysis yields information about the components of each factor and can reveal information about the structuring and interplay of the original universe of data. Thirdly, the analysis allows one to eliminate redundant variables and select the highly diagnostic variables within each factor.

An initial factor analysis was performed on twenty-four original variables. Three of the four factors which resulted were obvious indicators of the dimensions they were measuring. A rural factor, an S.E.S. factor, and age structure factor were evident. The fourth factor was not as clearly defined as the other three. "Per cent employed in construction" had the highest loading on the fourth factor with a -.65. In an attempt to sharpen the three evident factors and clarify the fourth a second analysis was done after eliminating variables which did not load highly ($\pm$.65) and those that were redundant. Percent males 14 and over who were married was eliminated on the basis of its highest loading (.61) as were "median age" (-.59) and "% 14 and over in the labor force" (.53). The "percent rural plus percent rural non-farm" variable was eliminated because it was seen as redundant with other variables measuring the same dimension. Two other variables, "%ag.", and "% rural farm", had higher loadings on the factor than did the "% rural farm plus % rural non-farm".

Twenty items were run in the second analysis. Again four factors resulted. The first factor which had been referred to as an S.E.S. indicator was beginning to appear as more of an education dimension. The second factor which previously appeared as a rural indicator now was beginning to look like a rural-farm or agricultural dimension. The age structure factor was showing a tendency toward a "youthfulness" dimension. The fourth factor was still not clearly defined. Again however percent employed in construction had the highest loading with a -.82.

Once again variables were eliminated which did not load highly on any factors, were redundant, or theoretically unsound. Fertility ratio (.59) and % unemployed (-.52) were eliminated on the basis of their highest loadings. Median school years completed by persons 25 and over (.64) was eliminated as a redundant item. Percent of persons 25 and over with high school or more (.65) had a slightly higher loading. The second variable was also deemed to be a more relevant variable theoretically, because it indicates the percent of persons who completed high school. Actually finishing high school is seen as more important than just an indicator of another year of school finished. Percent of persons residing in state of birth had its highest loading (.69) on the rural farm dimension which had several other items with much higher loadings on the dimension (% rural farm (.93); % ag. (.92). Thus % state of birth was seen as a poorer indicator of the dimension than other variables available. It was also recognized as potentially more the result of migration than a cause of it in the type of model employed herein. For these reasons the state of birth variable was eliminated. The "% farms<$2,500 V.P.S." variable was eliminated as it was recognized to be a polluted indicator of the poverty level it was originally hoped it would measure.

Due to large numbers of part time farmers who have sources of income other than their farm this indicator was seen as a poor measure of rural poverty level and hence was eliminated. Percent of housing units owner occupied was eliminated as a redundant, and poorer, measure of the educational dimension.

Fourteen variables remained and another factor analysis was performed. The first factor which emerged was the education factor. The highest loading was .797 on percent of persons twenty-five and over with a high school education.

The second factor was the rural-farm factor. Percent employed in agriculture had the highest loading on this factor with a -.939. The third factor had a .964 loading on percent employed in construction. There were no other high loadings on this factor and the item with the highest loading did not correlate with any other items. It is possible that percent employed in construction is a distinct item which might account for some variation that no other item would. However it was decided to eliminate this factor from consideration for use in the multiple regression model. This was done for a combination of reasons. The item was seen as a partial reciprocal of another variable already chosen, percent employed in agriculture. The construction item affected directly only a relatively small portion of the population when considered in comparison with other items like percent in agriculture. For instance the range of variation on the construction item was from 9.4% to 3.3% for a difference of 6.1%. By contrast the percent in agriculture varied from 53.6% to .8% or a difference of 52.8%.

There was also an interest in keeping the items selected for use in the multiple regression model to a minimum. This was necessary because given the sample size it was important to guard against too severely handicapping the model due to an increase in the degrees of freedom.

The final factor of the four which evolved in the last analysis was the age structure dimension. It was now apparent that this factor was indicative of a youthfulness dimension with its highest loading on percent under eighteen years of age (.795).

Given the nature of the data fed into the analysis it is not too surprising that three factors emerged which were indicative of an economic, social, and demographic component. The purpose of the analysis was to determine which item in each factor might best represent what that factor was suggesting. Therefore the item with the highest loading on each factor was selected to be representative of that factor. This is not to suggest that the items selected would in all future cases be stable indicators of the given factors they represent. The only way to check for such stability would be to test the same counties at two points in time or a set of different counties at one point in time. This however was not my purpose in the project. I used the technique as a device for selecting items with a further end product in mind. The use of the items in a multiple regression model is a partial test of the successfulness of the factor analytic method of selecting items. One could use the factor score of each county on each factor as an alternative to selecting the item with the highest loading, but in this project there was more interest in the specific variables.

The factors which resulted were characterized by items which loaded high on one factor and low on all remaining factors. This suggests that distinct measures were refined from the analysis. For this reason the items chosen to represent each factor are uncorrelated with each other. This should aid in maximizing their explanatory power in the regression model.

The items chosen for the regression model were: percent employed in agriculture, percent over twenty-five with high school or more, and percent under eighteen. Table 2 indicates the factors and item loading of each variable in the final analysis. Table 3 shows the correlation between the three items selected to represent each factor.

A second method of selecting variables to use in the regression model's two equations was used. A step-wise multiple regression was done using all the independent variables to predict the overall net rate. The step-wise procedure entered the independent variable which explained the most variance first. The second variable entered was the one which added the largest increment of variance explained to the equation. A separate step-wise regression analysis was done on all classification of counties. The first three variables entered in the equation were selected and run in the two equations for all the county classifications. Thus there were six sets of three variables each.

In summary, the selection of independent variables was accomplished in two ways. A factor analysis was performed on the independent variables for all counties. Three items were chosen and tested in a regression model run on all classifications of counties. Thus the factors which emerged from the analysis of all counties were taken to be appropriate for testing on all classifications of counties. There is reason to suspect that if a separate factor analysis had been done on each classification of counties different items may have emerged as representative of given factors. If the ability of the factors to predict in the class of "all counties" is noticeably better than in the other classes of counties, this suspicion will be confirmed.

The step-wise procedure was carried out with net migration rate as the dependent variable. The procedure was performed on all classification of counties yielding a different set of three independent predictors for each classification. These predictors were taken to be the best indicators of the age-specific rates also. This is a somewhat unwarranted assumption, but once again the test in the final regression

Table 2. Final Factor Analysis of Independent Variables

Item	Factor I	Factor II	Factor III	Factor IV
**% agri.	.080	-.939	-.087	.180
% const.	-.016	.074	.964	-.007
% mfg.	-.510	.717	-.128	-.024
**% 18 or under	.106	.032	.180	.795
Med. Inc.	.410	.716	-.045	.495
<3,000 fam. income	-.484	.650	-.012	-.486
>10,000 fam. income	.457	.501	-.063	.510
**% 25 > High School +	.797	.346	-.122	.008
% rur.-farm	-.110	-.934	-.103	.080
farm lev. of living	.695	-.031	.094	.502
% farms 10,000 VPS	.669	-.162	.037	.536
% pers. income from gov't	-.066	.270	.260	-.691
% pop. 7-13 and in school	.193	-.054	-.039	.754
% white collar	.195	.867	.068	-.206

* Factor I is education indicator

Factor II is agriculture; rural-farm indicator

Factor III is per cent in construction

Factor IV is age indicator of youthfulness

** Indicates the defining item chosen from each factor for regression model.

Table 3. Correlations of variables selected by the Factor Analysis

	% agri.	% 18 or under	% 25 > H.S. +
% agric.	-	-	-
% 18 or under	.12	-	-
% 25 > H.S. +	-.23	.16	-

will elucidate the relationship between the variables' ability to predict net rates and age-sex specific rates.

It can be seen from the procedure illustrated above that the only entirely sound comparison of the step-wise and factor analysis method of selecting the variables will necessarily be done on their respective ability to predict net migration for all counties. This is true since the factor analysis was not done on all classifications of counties and the step-wise was not done with age-sex specific rates as the dependent variable. Determining the consequences of each method of selection was thought to be of more interest.

Also an attempt was made to determine if the interactive effects of the second model added significantly to the power of the predictors. Possible explanations are offered as to why the items which were successful predictors of migration rates may have been so.

The Dependent Variables

Net migration is the balance between in migration and out migration and hence may be either a negative or positive figure. The net rates used in this research were taken from computations made at the Population Research Laboratory, Kansas State University and published under the title Migration in Kansas: Outmigration and Population Trends (Flora, et al., 1971). The vital statistics method was used in calculating the net rates. The vital statistics method is a simple way of estimating net migration rates employing the following equation:

$$M - (P_1 - P_o) - (B-D),$$

where

M = migration

P_1 = population at time one

P_o = population at a previous time

B = births

D = deaths

Population change during a period minus natural increase during that period is equal to net migration.

The vital statistics method depends upon the complete registration of births and deaths and also upon accurate population counts. Hence it is subject to the errors which may occur in such data.

The age-sex specific rates were obtained from the Kansas State University Population Research Laboratory also. The rates were computed using the forward survival rate method. "The basic formula for estimating net migration is

$$M_{x+t} = P^t_{x+t} - sP^o_x ,$$

where

x = an age or age group

t = the interval in years between censuses

P^o_x = the population aged x at the first census

P^{t}_{x+t} = the population at the next census at age x+t

S = the survival rate (Shryock and Siegel, 1971: 630)

Unit of Analysis

As has been indicated elsewhere the data drawn for this study was from county level statistics. The degree to which such a unit constitutes a legitimate ecological area varies. For the purposes of this paper, it should be recognized that the county is used as an approximation of such an area. The county as an administrative unit is well established and has universally familiar boundaries which makes it a practical unit to use. Migration thus consists of a change of residence across a county line. The size and shape of a county along with the geographic distribution of its population determine in part its net rate of migration.

For example, if a large part of the population lives near a county boundary the chances are increased that this county will have a higher rate of migration than a county in which the majority of the population lives closer to the geographic center of the county. The reason for this is that a short move in the first county may result in crossing a county boundary line. An extreme example would be simply a move across the street. By contrast in the second county on the average it would take a longer move to constitute a change in county residence.

In Kansas where this study was done the counties are relatively homogeneous in size and shape with most approximating a rectangle. It was assumed for the purpose of this study that the effect of county boundary implications would be uniform across the State and have a random effect. However such possible sources of influence on rates should be kept in mind in interpreting the results of studies using net rates on a county level.

It should also be pointed out that as the size of the unit of analysis increases the rate of migration decreases. Thus one cannot legitimately compare net rates based on different size units of analysis. For example, within county rates are greater than between county rates and the same holds for states and regions when used as units of inquiry. The results of this study should be interpreted in light of the above limitations which, while not ideal, do allow one to continue to do research. With the above limitations in mind let us move to an analysis of the findings.

CHAPTER III

PRESENTATION AND INTERPRETATIONS OF THE FINDINGS

Mean Migration Rates

Before beginning a discussion of the findings for each set of counties some prefatory remarks may be useful for setting the general background in which the study took place. The state of Kansas as a whole experienced a net out migration rate for the 1960-1970 decade of -10.3. (See Table 4). The age-sex specific rates show us that younger persons had higher out migration rates (male 10-14 (20-24), -28.9; female 10-14 (20-24), -32.6) than the state mean. In contrast the older ages categories have small positive rates (male 50-54, +0.6; female 50-54, +1.1).

Table 4 also shows that none of the classifications of counties had net in rates with the exception of the classification which considered counties with net "in" rates of migration. Also evident is the fact that only two sets of counties experienced net in rates for the young age specific rates. One of these was again the classification which considered only counties with net in rates. The notable exception then was the set of counties which were 0-50% rural or therefore "urban". This set of counties experienced a mean male rate for ages 10-14 (20-24) of +25.7. The mean female rate for ages 10-14 (20-24) was +6.3. In contrast the counties which were 51-99% rural experienced high net

Table 4. Statistical Means of Migration Rates for Kansas Counties by Types of County, 1960-1970

		Net	Male		Female	
			10-14 (20-24)	50-54 (60-64)	10-14 (20-24)	50-54 (60-64)
Mean Rates:						
All Counties	(N=105)	-10.3	-28.9	+0.6	-32.6	+1.1
0-50% Rural	(N=35)	-6.1	+25.7	-1.2	+6.3	-0.3
51-99% Rural	(N=25)	-11.8	-50.8	+0.1	-47.5	+0.1
100% Rural	(N=45)	-12.7	-59.2	+2.2	-54.5	+2.7
Net "In"	(N=14)	+5.8	+87.1	+0.4	+39.7	+2.2
Net "Out"	(N=91)	-12.7	-46.7	+0.6	-43.7	+0.9

out rates in these two categories, -50.8 and -47.5 respectively. The same pattern held in the counties 100% rural where the young male rate was -59.2 and the young female rate was -54.5.

It is interesting that the 0-50% rural or "urban" counties was the only set which experienced a negative mean rate of net migration in the older age-sex categories (male 50-54 (60-64) -1.2; female 50-54 (60-64) -0.3). The rest of the sets of counties had positive mean rates for the older categories although they were small ranging from a high of +2.7 (females in all rural counties) to a low of +0.1 females in 51-99% rural counties).

In summary, Table 4 illustrates that overall the state suffered net out migration over the decade. The largest portion of this net out rate was due to young people leaving rural counties. The most urban counties did have a positive mean rate of in migration in the young categories. Of the counties which experienced net in rates the largest proportion of the increase came about as a result of high rates of in migration in the young male category (+87.1). This category was constituted by the persons who were 10-14 years of age at the beginning of the decade and hence became 20-24 years of age during the decade. There is evident within the state a rural to urban trend consisting largely of young people. Also apparent is the fact that young people are leaving the state altogether as evidenced by the -28.9 (males 10-14; 20-24) and the -32.6 (females 10-14; 20-24) mean rates for all counties. It is now appropriate to turn to a more detailed analysis of some of these differences in net migration rates.

All Counties Results

As pointed out above, for a strict comparison of the methods used to select the variables it is necessary to consider their ability to account for variation in the net rate for all counties. Table 5 summarizes the findings of the analysis. It is apparent that the variables selected by the factor analysis were not as successful as the variables chosen in the step-wise procedure. The former account for only 9 and 21 percent of the variance in the restricted and full models. By contrast the step-wise variables account for 26 and 52 percent of the variance in the respective models. What is apparent about both sets of variables is that their power is significantly increased in the full model where all the interaction terms are used as independent predictors. The factor analysis variables more than double their amount of variance explained, increasing it to 21 percent and raising the significance level from .05 to .01. The step-wise full model almost doubles the size of R^2 making it a respectable 52 percent significant at the .01 level.

As was pointed out in Table 3 in the preceding chapter, the variables which resulted from the factor analysis were relatively uncorrelated with one another. The same was true of the variables selected by the step-wise procedure. As Table 7 illustrates the factor analysis variables correlate as follows: % ag. and % 18 <, .12; % ag. and % 25 > HS +, -23; % 18 < and % 25 HS +, 16. The step-wise variables are: % married and % white collar, -.04; % own occ. and % w.c., -.19; % married and % own occ., .14. This is desirable because the power of the regression model is maximized when uncorrelated independent variables are used. In this situation each variable is capturing a different aspect of the variance. This also makes any interaction which may occur between the variables theoretically more interesting. That is, if two items measuring different aspects of reality combine to increase the significance of the prediction equation this can be more fruitful than two theoretically similar items combining to explain more variance. The likelihood of the latter type of occurrence is more remote also since the two items which are theoretically similar would be more likely to be highly correlated with one another.

The ability of the step-wise variables to account for more variation is a function of their correlation with the dependent variable. Percent white collar has the highest "r" (.429) with the dependent net rate, Y, of any of the six variables under consideration. The two other step-wise variables correlate .119 (% housing units owned or occupied) and -.161 (% married) with Y.

By contrast the factor analysis variables have zero-order correlations of -.285 (% agriculture), -.073 (% 18 and under) and .141 (% 25 > H.S. +) with Y. Blalock has pointed out it is desirable to have independent variables which correlate highly with the dependent variable in order to maximize the power of regression models. It should be added, however, before the reader assumes the factor analysis variables did not meet this criterion at all, that the "% agriculture" with its -.285 correlation was the third highest correlation of all the variables on the net rates. This indicates that not many of the variables correlated highly

Table 5. Multiple R's, R^2's, and F-Values for all Counties
Comparison of Sets of Variables and Types of Models

All Counties
N=105

			Male		Female	
		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R	.300	.450	.210	.560	.145
	R^2	9.0	20.3	4.4	31.4	2.1
% Ag.	F-Value	3.33	8.5*	N.S.	15.4*	N.S.
% 18 & Under	(DF: 3,101)					
% 25 HS +						
	Full Model:					
	R	.458	.783	.244	.828	.169
	R^2	21.0	61.3	6.0	68.6	2.8
	Gain	12.0	41.0	1.6	37.2	.7
	F-Value	3.68*	22.0*	N.S.	30.2*	N.S.
	(DF: 7,97)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R	.507	.652	.241	.706	.219
	R^2	25.7	42.5	5.8	49.8	4.8
% W.C.	F-Value	11.7*	24.7*	N.S.	33.4*	N.S.
% Own Occ.	(DF: 3,101)					
% Married						
	Full Model:					
	R	.719	.971	.338	.886	.265
	R^2	51.6	94.2	11.4	78.5	7.0
	Gain	24.1	51.7	5.6	28.7	2.2
	F-Value	14.8*	226.4*	N.S.	50.6*	N.S.

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

Table 6. Variable Names for Corresponding Variable's Numbers in Correlation Matrices in Tables 7, 8, and 9

1. % employed in agriculture
2. % employed in construction
3. % employed in manufacturing
4. % 18 years old and under
5. cumulative fertility ratio
6. median income
7. % families with income under $3,000
8. % families with income over $10,000
9. median age
10. % persons 25 years and over with high school education or more
11. % housing units that are owner occupied
12. farm operator level of living index
13. % unemployed
14. % farms under $2,500 in annual value of products sold
15. % farms over $10,000 in annual value of products sold
16. net migration rate
17. % of personal income from government disbursements
18. % of population residing in state of birth
19. % of males 14 and over who are married
20. % workers employed in white collar jobs
21. % of population that is rural farm and rural non-farm

Table 7. Zero Order Correlation Matrix of Variables for All Counties (above diagonal) and Out Counties (below diagonal).

Variable Numbers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	-	-.14	-.68	.12	.13	-.53	.50	-.25	.25	-.23	.28	.13	-.55	-.15	.28	-.28	-.36	.53	.01	-.81	.82
2	-.13	-	-.02	.07	.10	.01	-.01	.01	-.07	-.06	-.17	.06	.02	.02	-.01	.12	.17	.01	-.13	.12	-.18
3	-.68	-.04	-	-.13	-.18	.26	-.18	.10	.01	-.08	.10	-.31	.61	.29	-.39	.36	.21	-.37	.05	.49	-.60
4	.03	.09	-.10	-	.58	.42	-.45	.42	-.49	.16	-.17	.40	-.26	-.30	.47	-.07	-.36	-.04	.01	-.05	-.01
5	.04	.10	-.14	.54	-	.35	-.41	.36	-.56	.26	-.27	.50	-.30	-.35	.46	-.09	-.18	-.04	-.02	-.03	.06
6	-.54	.01	.20	.43	.40	-	-.95	.84	-.66	.55	-.41	.49	-.05	-.20	.38	.21	-.17	-.46	.07	.55	-.49
7	.52	-.18	-.16	-.45	-.43	-.97	-	-.73	.72	-.55	.53	-.60	.08	.31	-.47	-.12	.18	.43	.01	-.49	.48
8	-.22	.02	-.01	.44	.44	.81	-.75	-	-.57	.53	-.22	.44	-.21	-.22	.46	.29	-.23	-.41	.14	.45	-.27
9	.27	-.07	-.02	-.51	-.57	-.71	.73	-.66	-	-.43	.51	-.51	.18	.25	-.45	-.11	.09	.28	.15	-.30	.24
10	-.17	-.08	-.17	.16	.29	.47	-.51	.42	-.37	-	-.39	.44	-.15	-.36	.40	.14	-.02	-.30	-.02	.44	-.23
11	.38	-.20	.01	-.23	-.29	-.62	.64	-.48	.53	-.50	-	-.46	.01	.30	-.43	.12	-.21	.40	.14	-.19	.30
12	.06	.08	-.28	.32	.48	.56	-.61	.51	-.51	.47	-.45	-	-.47	-.53	.75	-.07	-.42	-.03	-.06	-.02	.04
13	-.57	.01	.63	-.20	-.27	-.03	.06	-.23	.16	-.13	-.01	-.45	-	.36	-.49	.07	.36	-.27	-.08	.34	-.46
14	-.09	.01	.25	-.28	-.36	-.28	.35	-.33	.25	-.43	.27	-.49	.35	-	-.67	.19	.19	-108	.08	.01	-.12
15	.20	.02	-.36	.42	.43	.47	-.51	.58	-.47	.46	-.43	.70	-.48	-.65	-	-.07	-.45	-.12	-.08	-.12	.20
16	-.13	.13	.35	-.06	-.17	-.08	.09	-.08	.14	-.30	.25	-.09	.06	.17	-.03	-	.20	-.26	-.16	.43	-.28
17	-.26	.16	.19	-.30	-.19	-.30	.29	-.36	.20	-.16	-.13	-.43	.37	.10	-.43	-.07	-	-.20	-.25	.35	-.24
18	.50	-.01	-.33	-.06	-.06	-.36	.38	-.28	.27	-.18	.51	-.03	-.30	-.06	-.16	-.07	-.07	-	-.01	-.55	.48
19	-.08	-.12	.04	-.09	-.04	.05	.01	.08	.09	.02	.01	-.12	-.06	.12	-.18	-.08	-.08	-.04	-	-.04	.09
20	-.81	.12	.48	.01	.06	.46	-.47	.29	-.29	.31	-.36	.02	.41	-.07	-.06	.09	.28	-.47	.03	-	-.75
21	.81	-.17	-.61	-.05	.01	-.47	.47	-.22	.22	-.16	.38	.03	-.49	-.11	.18	-.17	-.21	.45	.06	-.76	-

Table 8. Zero Order Correlation Matrix for 0-50% Rural Counties (above diagonal) and In Counties (below diagonal).

Variable Numbers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	-	.23	-.57	.24	.09	-.33	.25	-.25	.10	-.25	.01	.26	-.52	-.09	.35	-.16	-.31	.57	-.27	-.74	.59
2	-.21	-	-.33	.27	.49	.01	-.12	.06	-.36	.20	-.32	.41	-.39	-.32	.46	-.01	.12	-.06	-.12	-.03	.07
3	-.67	.07	-	-.21	-.34	.07	.10	.07	.36	-.25	.46	-.40	.54	.28	-.40	.15	-.10	-.29	.20	.23	-.29
4	.54	-.07	-.37	-	.76	.63	-.71	.52	-.55	.35	-.18	.71	-.41	-.25	.64	-.20	-.50	-.06	-.06	-.07	-.36
5	.70	.05	-.62	.85	-	.46	-.63	.37	-.74	.50	-.52	.68	-.37	-.32	.54	-.33	.01	-.23	-.06	.11	-.38
6	-.47	-.14	.49	.37	.05	-	-.92	.91	-.56	.71	.02	.47	-.35	-.13	.32	.23	-.31	-.41	.18	.53	-.70
7	.37	.19	-.21	-.45	-.17	-.91	-	-.73	.73	-.72	.27	-.68	.42	.32	-.50	-.03	.28	.29	-.05	-.37	.64
8	-.27	-.19	.44	.49	.15	.95	-.83	-	-.41	.70	.19	.38	-.39	-.11	.31	.39	-.23	-.46	.22	.61	-.59
9	-.01	.11	.46	-.45	-.40	-.40	.66	-.32	-	-.74	.63	-.67	.41	.39	-.50	-.04	-.16	.27	.24	-.28	.43
10	-.36	-.07	.23	.16	.03	.75	-.75	.74	-.58	-	-.34	.50	-.42	-.43	.40	.31	.21	-.42	-.14	.66	-.47
11	-.01	-.02	.51	.05	-.18	.17	.20	.26	.57	-.11	-	-.47	.10	.40	-.44	.34	-.41	.19	.17	.06	.06
12	.36	-.03	-.50	.75	.68	.30	-.58	.38	-.74	.42	-.50	-	-.62	-.54	.84	.02	-.30	.01	-.17	-.02	-.25
13	-.51	.13	.41	-.67	-.68	-.24	.29	-.30	.56	-.34	.08	-.59	-	.32	-.47	-.23	.30	-.28	-.04	.11	-.08
14	-.34	-.01	.48	-.41	-.43	-.01	.24	-.09	.64	-.30	.49	-.73	.46	-	-.67	.07	-.08	-.01	.23	-.08	.04
15	-.53	-.15	-.53	.71	.67	.13	-.35	.28	-.54	.23	-.44	.90	-.57	-.82	-	-.01	-.24	-.10	-.30	-.09	-.19
16	-.68	-.03	.58	-.23	-.44	.70	-.59	.67	-.22	.84	.06	-.01	.06	.01	-.12	-	.03	-.18	-.11	.46	-.01
17	-.56	.16	.21	-.67	-.37	-.02	.01	-.22	.07	.13	-.35	-.43	.39	.36	-.52	.31	-	-.45	-.18	.37	.11
18	.62	.09	-.59	.07	.16	-.78	.71	-.70	.17	-.70	.07	-.05	-.10	-.05	.03	-.77	-.48	-	-.10	-.63	.57
19	.29	-.05	.24	.60	.29	.35	-.16	.55	.27	.01	.62	.21	-.19	.05	.31	-.01	-.74	.02	-	.08	-.09
20	-.87	.09	.61	-.24	-.45	.73	-.61	.62	-.21	.69	.11	-.09	.22	.12	-.25	.86	.36	-.78	-.04	-	-.57
21	.87	-.16	-.54	.31	.51	-.54	.47	-.38	.25	-.47	.07	.11	-.20	-.06	.25	-.64	-.30	.54	.20	-.85	-

Table 9. Zero Order Correlation Matrix for 50-99% Rural Counties (above diagonal) and 100% Rural Counties (below diagonal)

Variable Numbers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	-	.20	-.55	.13	.02	-.83	.77	-.73	.42	-.24	.60	-.18	-.24	-.11	-.06	-.22	-.30	.43	.20	-.58	.61
2	-.30	-	-.31	.14	-.28	-.22	.26	.05	.08	.01	.20	-.23	-.33	.11	-.21	.08	-.04	.27	.02	-.03	-.21
3	-.26	.21	-	.03	.25	.37	-.27	.27	-.21	-.21	-.17	-.12	.43	.12	-.08	.59	.24	-.25	.01	.08	-.18
4	.15	-.75	-.41	-	.27	-.06	.01	-.10	-.13	-.36	.42	-.27	.02	-.17	-.03	.02	-.03	.08	.09	-.09	.10
5	.15	.07	-.34	.79	-	.05	-.07	-.06	-.14	.09	.21	.02	-.12	-.33	-.01	.13	.04	-.03	.01	.04	.11
6	-.23	-.06	-.31	.80	.53	-	-.97	.87	-.51	.41	-.70	.49	.06	-.08	.37	-.11	.19	-.28	-.17	.43	-.45
7	.18	.12	.35	-.80	-.54	-.97	-	-.79	.52	-.51	.61	-.57	-.03	.20	-.50	.17	-.06	.30	.09	-.48	.43
8	.03	-.07	-.31	.74	.49	.82	-.76	-	-.54	.26	-.67	.32	-.11	.02	.21	-.10	.19	-.26	-.28	.43	-.55
9	-.09	.07	.41	-.99	-.80	-.83	.84	-.75	-	-.11	.19	-.07	.12	-.13	.01	.03	.01	.16	.15	-.21	.05
10	-.06	-.31	-.29	.31	.18	.38	-.40	.34	-.30	-	-.27	.73	-.05	-.45	.56	-.34	-.28	.04	.28	.35	-.26
11	.01	.05	.37	-.67	-.37	-.77	.74	-.69	.67	-.44	-	-.47	-.22	-.10	-.28	.12	-.24	.44	.35	-.30	.25
12	.09	-.11	-.42	.68	.56	.66	-.72	.57	-.73	.34	-.61	-	-.10	-.67	.83	-.56	-.32	.05	.24	.19	-.05
13	-.24	.28	.52	-.42	-.33	-.33	.38	-.32	.44	-.15	.39	-.47	-	.27	-.01	.18	-.10	-.46	.12	-.05	-.03
14	-.09	.34	.44	-.55	-.42	-.56	.61	-.54	.59	-.34	.48	-.47	.42	-	-.66	.37	.33	-.31	-.35	-.07	-.04
15	.14	-.19	-.48	.74	.50	.71	-.73	.71	-.76	.44	-.76	.66	-.53	-.76	-	-.48	-.45	-.07	.40	.20	.02
16	.02	.24	.45	-.02	.01	-.05	.03	.02	.02	-.15	-.01	-.04	.14	.29	.05	-	.48	-.24	-.33	.12	-.20
17	-.16	.21	.69	-.78	-.52	-.71	.74	-.66	.79	-.49	.69	-.76	.49	.70	-.77	.32	-	.13	-.74	.08	-.06
18	.20	.13	.12	-.30	-.03	-.48	.47	-.50	.30	-.35	.64	-.30	.36	.25	-.55	-.19	.34	-	.16	-.42	.31
19	-.05	-.23	-.23	.16	.03	.35	-.34	.33	-.15	.56	-.41	.21	-.17	-.31	.28	-.15	-.36	-.35	-	-.15	.11
20	-.54	.08	.05	.04	-.11	.17	-.18	.13	.01	.15	.03	.03	-.03	-.16	.11	-.01	-.02	-.20	.23	-	-.53
21*																					

with the dependent variable (see Table 7).

A point to be made here should be remembered throughout the ensuing chapter. A high correlation between two variables may result from either one of two distinct sources known as causality and communality (Schuessler: 1969). A variable may be correlated with another because one affects the other (causality) or because they are measures of the same thing (communality). Since a correlation may result in either of these ways one must be cautious about the inferences made from simple correlations taken by themselves. In interpreting the correlations discussed in this paper the danger of confusing communality and causality as sources of correlations should be kept in mind.

In conclusion it can be said that the step-wise procedure for selecting variables is the better of the two in this case. What is of more interest, however, is whether the interaction or full models differ significantly from the restricted models.

The full models do differ significantly from the restricted models. However when one considers the improvement gained by the full model over the restricted model within each set of variables the evidence is inconclusive as to which set of variables is better. The factor analysis variables appear to be slightly better in that they more than double the R^2 from 9% to 21% for a 12% increase or 3% over the amount necessary to double the power. By contrast the step-wise variables go from 25.7 to 51.6 for an additional 24.1 or 1.6 percent short of the amount of increase necessary to double the R^2.

The essential point nonetheless is that the step-wise full model is markedly better than any of the others in terms of the amount of variance it accounts for. The interpretation of the variables included in the step-wise is that a high proportion of white collar workers is associated with attracting in-migrants. In some specific cases in Kansas it might be necessary to say that a high proportion of white collar workers will tend to lessen the out-migration such areas may expect to experience. This may be because a high proportion of white collar workers may be in "service" type institutions such as universities which attract persons or retain them. It is also possible that a high proportion of white collar workers indicates the presence of bureaucratic forms of organizations which tends to attract more white collar workers, who are among the most mobile of occupation groups. Also to be noted is the fact that "% white collar" is negatively correlated (-.813) with "% in agriculture". This suggests that the white collar indicator tells us that urban or rural non-farm areas are more likely than rural farm areas to experience in migration.

When speaking of areas being more likely to experience in-migration, it is essential to remember that in fact when we speak of a net rate we are speaking about a residual. We in actuality can not comment on the "gross rate" of "in" and "out" migration which has occurred. It is known that every migration stream operates in the face of a counter stream of migration. For example, it may take a gross rate of .40 out migration to accomplish a net rate of out migration of .20. This would be true if while X persons moved out of an area .5(X) persons moved into the same area. Another example of the gross "in-out" relationship to the "net" may prove useful. An area could experience no net migration yet be a highly mobile area with a changing population. This hypothetical case is unlikely but could occur if an area lost 40,000 "out" migrants but gained 40,000 "in" migrants. The net migration has been zero. This hopefully makes clear some of the implications for the effects of various variables in affecting the net rate.

In essence the variables may function in one of four ways, or more likely, in a combination of all four. The variables may increase or decrease the likelihood of "in" or "out" migration, they may "pull" people into the area or "push" them from the area with differing degrees of success or failure. When it is said that a characteristic has the affect of *encouraging net in migration* this does not mean that all areas with such characteristics will experience in-migration. It may mean that some of these areas experienced less out-migration than they otherwise would have had. This may be the result of other more powerful variables at work.

In other words to say a variable has a positive correlation with in-migration is not to imply all areas with such characteristics will experience "in" rates.

In looking at the other two variables which were considered in the step-wise equation, we may see several possible interpretations of their significance. The slight positive association of "% of housing units owner occupied" with in migration may be seen in a variety of ways. It is possible that the variable may be an indicator of a county wealth dimension which attracts migrants to areas which have a higher level of living. Or the indicator could be seen as a type of "permanance" or stability measure indicating that the more persons who own homes and hence have ties with the community, the less likely that area is to experience out migration.

The "% of males 14 and over who are married" variable has negative association with net in migration (-.161). This would suggest on the surface that people don't want to migrate to a place where there is not a large proportion of single people. This requires that one assume migration is somewhat rational and that there is knowledge of the place of destination and its characteristics. If this can be assumed one could hypothesize the bulk of migrants are young and single and hence move to where there are others like them. To engage in such a line of reasoning, however, disposes one to the possibility of committing the ecological fallacy. One is hypothesizing about individual characteristics which might have potential explanatory ability in accounting for ecological correlations. Such reasoning could also be seen in conflict with the findings in 1960 that a higher proportion of the married population was comprised of migrants than was the single population (Bogue, 1969: 768).

The % married variable is more likely a function of other types of characteristics about the area, the age structure for instance, which are more significant determinants of migration rates.

In moving from a consideration of the net rates to the age-sex specific rates some important differences in the models' ability to predict begin to emerge. A glance at Table 5 shows that both sets of variables are quite different in their abilities to predict the rates of the respective age categories. Both sets of variables are quite capable of accounting for both male and female rates in the young age category, 10-14 (20-24), but neither can produce significant results in the older category 50-54 (60-64) and account for only a meager proportion of the variance. The full model again is markedly better than the restricted model in both cases. The encouraging ability of the sets of variables to account for the young rates suggests that much of their success in handling the net rates derives from this source and that inclusion of older age categories weakens the ability to account for net rates.

Again the step-wise full model is the better of the two for all cases. The step-wise full model accounts for 94 percent of the variation in male rates age 10-14 and 79 percent of the variation in the same age for females. Remembering that the step-wise variables were selected on the basis of their ability to predict net rates raises the interesting question as to how much more powerful the method might be if variables were selected on their ability to handle age-sex specific rates and then run in the intersection full model.

Another interesting point which was noted in considering the net rates held true for the age specific rates. Once again the full model using the set of variables derived from factor analysis, while not as powerful as the step-wise full model, nonetheless resulted in a larger proportional amount of variance accounted for when both models were compared to their respective restricted models. This suggests that there is more significant interaction occurring between the factor-chosen variables than there is between the step-wise variables. This may be partially because the factor analysis isolated distinctive items which comprise three rather basic dimensions, the demographic, economic, and social. The separation of items achieved in the step-wise procedure is not as clear cut theoretically or statistically.

Out Counties Results

The same type of analysis that was performed on all the counties of Kansas was also performed after controlling for net in and out migration. The correlation matrix which resulted from the analysis is presented in Table 7 below the diagonal. The factor variables correlate as follows: % ag. with % 18 and under, .03; % ag. with % 25>HS or more, .17; and % 18 and under with % 25>HS or more, .16. The step-wise variables which emerged were % mfg., % own occupied, and % farms > $10,000 value products sold (V.P.S.). These variables generally correlated higher with each other than did the factor variables with the exception of % mfg. and % own occupied which correlated only .01. This is in comparison to the .03 between % agr. and % 18 and under which was the lowest correlation between the factor variables. The remaining correlations, % mfg. with % farms > $10,000 V.P.S. (-.36) and % farms > $10,000 V.P.S. with % own. occupied (-.43) were stronger than correlations between the factor variables.

Despite the fact that the step-wise variables don't appear to be as distinctive as the factor variables we may still expect them to be as successful in handling the variation in rates. This is because the step-wise variables correlate higher with the dependent variable than do the factor variables. The factor variables are all negatively correlated with net migration rate. Percent in agriculture is -.13, % 18 and under, -.06, and % 25>HS or more, -.30. Perhaps the most notable aspect of the three correlations is the change which occurred in the third one, % 25>HS or more. By removing only 14 counties from the analysis the correlation changed from a +.14 with all 105 counties to a -.30 with only 91 out counties considered. In the all county case the +.14 had been the highest positive "r" but in the out county case it is now the largest negative "r". This alerts us to two ideas. One, the fourteen counties (13 in migration and 1 with no net change) have a strong influence on the education variable. Two, areas with a higher proportion of persons with a high school education or more are less likely than counties with a lower proportion of persons with high school or more to experience high out migration rates. In the case of the out counties the negative correlation indicates that migration rates decrease, even though they are still all negative, as the independent variables increase. The step-wise variables correlate .35 (% mfg.) .25 (% own occupied), and -.03 (% farms > $10,000 V.P.S.) with the dependent variable, net out migration rate. Thus areas which have a high proportion of people employed in manufacturing or a high proportion of persons who own their home are more likely to experience less out migration than counties which do not exhibit these characteristics. Noting that % mfg.'s highest correlation with any of the other variables is a -.68 with % agric. it is possible to speculate that perhaps the manufacturing variable entered the step-wise procedure as an indicator of the more general difference in migration rates due to rural-urban differences. The small correlation of -.03 between % farms > $10,000 V.P.S. and net rate is not a significant correlation at the .10 level.

The results of the analysis of the 91 counties which had out rates are summarized in Table 10. It is evident that the power of both sets of variables and both models is somewhat reduced when applied only to counties with net out migration. The factor variables account for only 12% and 14% of the variation in Y in the restricted and full models respectively. The step-wise are somewhat better resulting in R^2's of 24% and 28% in the respective models. While still producing results which are generally significant at the .01 level the proportion of variance accounted for is decreased. In fairness to the factor analysis variables it should be recalled we are now applying variables in the model which emerged from an analysis of all counties. This may account for the set of variables decreased explanatory power in comparison to the step-wise procedure which was run on all classification of counties. The step-wise procedure nonetheless is somewhat disappointing in its ability to account for variations in the net rate by only achieving an R^2 of 28 percent in full model. The variables which emerged in this step-wise analysis (% mfg., % housing units owner occupied, and % farms over $10,000 V.P.S.) showed no appre-

Table 10. Multiple R's, R^2's, and F-Values for Counties with Net Out Migration. A Comparison of Sets of Variables and Types of Models

		Counties with Net "Out" Rates N=91				
			Male		Female	
		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R	.347	.387	.241	.624	.179
	R^2	12.0	15.0	5.8	39.0	3.0
% Ag.	F-Value	4.0	5.1*	N.S.	18.5*	N.S.
% 18 & Under	(DF: 3,87)					
% 25 & HS+	Full					
	Model:					
	R	.378	.526	.282	.704	.228
	R^2	14.3	27.6	8.0	50.0	5.2
	Gain in R^2	2.3	12.6	2.2	11.0	2.2
	F-Value	N.S.	4.5*	N.S.	11.7*	N.S.
	(DF: 7,83)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R	.492	.583	.366	.700	.256
	R^2	24.2	34.0	13.4	49.0	6.5
% Mfg.	F-Value	9.3*	14.9*	4.5*	27.9*	N.S.
% Own Occ.	(DF: 3,87)					
% 10,000 VPS	Full					
	Model:					
	R	.528	.693	.394	.735	.266
	R^2	27.9	48.0	15.5	54.1	7.1
	Gain in R^2	3.7	14.0	2.1	5.1	.6
	F-Value	4.6*	10.9*	2.2	14.0*	N.S.
	(DF: 7,83)					

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

ciable addition to the amount of R^2 (24.2 to 27.9 for a gain of 3.7) when used in the full model. Though they increased the amount of variance explained (as they must do by necessity) the increase in the number of variables resulted in smaller "F" ratios.

As was the case with all counties, the models in the out migration counties have difficulty accounting for variation in the older age specific categories. Six of the possible eight results were non-signfiicant with the exception being male rates 50-54 with the step-wise predictors. The two cases which were statistically significant did not, however, account for much variation in Y (13.4% and 15.5%). The ability of all variables in both models was consistently superior in the female 10-14 category with the step-wise full model achieving an R^2 of 54 percent significant at the .01 level. All in all the evidence derived from the models ability to account for variation in out county rates indicates that much of the success of the model with all counties must have derived from the variations in those counties with net in rates. Turning to an analysis of the in counties this fact is verified.

In Counties Results

The first thing which should be noted about the in counties is that they number only fourteen. The small sample size should cause interpreters of the results to be somewhat cautious in the types of generalizations derived from the findings. These results are algebraic relations which obtained in this case but the ability to contend that similar results would occur within the same size sample at a different point in time is less than the ability with which we may make such contentions in other cases with larger samples.

With the above point in mind a second one to be noted is that while with "all counties" and the "out counties" the age specific young rates yielded consistently higher R^2 and F values than did the net rates, such is not the case with the "in" counties. The net rate is more precisely accounted for than age-sex rates in both equations and with both sets of variables. While results for the younger rates for both sexes are still better than the results for older rates the results of the net rate is better than either of the age-sex specific rates. This occurrence is no doubt in part a result of the way in which the variables were selected which was discussed earlier. It should alert the reader to the fact that in the in counties there is a significant amount of migration which does not occur in the age categories we reviewed here.

The factor analysis variables were much better in the case of the "in counties" than in either the "all counties" class from which they were selected or the "out counties" class on which they were also run. This confirms our supposition that much of their success in the "all counties" case derived from the "in counties". Even though the variables account for 89% and 94% of the variation in net rates in the restricted and full models they are nonetheless superseded by the step-wise variables. The step-wise variables (% white collar, % 25 and over with HS +, and fertility ratio) account for 93% and 97% of the variation in the respective models (See Table 11).

Remembering that the step-wise variables were selected on the basis of their ability to handle net rates it is not too surprising that on the male-female young rates the step-wise variables are not as successful as the factor analysis variables. The factor analysis variables account for 60% and 88% of the variation in male, 10-14 rates in the restricted and full models, respectively. Both are significant at the .05 level in this case.

As Table 11 illustrates the same pattern holds for the female 10-14 rates with the factor variables explaining more variation and being significant at the .01 level while the step-wise achieve only an .05 level of significance. Again, as in the previous two sets of counties, none of the variables or equations are successful with older age-sex specific rates.

One of the variables from the factor analysis set, % 25 and over with high school or more, also emerged in the step-wise procedure as a good predictor. An examination of the correlation matrix in Table 8 indicates that the "% 25 and over with high school or more" variable has the

Table 11. Multiple R's, R^2's, and F-Values for Counties with Net "In" Migration. A Comparison of Sets of Variables and Types of Models

		Counties with Net "In" Rates N=14				
			Male		Female	
		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R_2	.941	.775	.338	.829	.372
	R	88.6	60.1	11.4	68.7	13.9
% Ag.	F-Value	25.9*	5.0	N.S.	7.3*	N.S.
% 18 & Under	(DF: 3,10)					
% 25 & HS+						
	Full					
	Model:					
	R_2	.968	.938	.712	.962	.910
	R	93.6	88.0	50.7	92.6	82.9
	Gain in R^2	5.0	27.9	39.3	23.9	69.0
	F-Value	12.6*	6.2	N.S.	10.7*	N.S.
	(DF: 7,6)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R_2	.963	.567	.388	.758	.448
	R	92.8	32.1	15.0	57.4	20.0
% W.C.	F-Value	42.9*	N.S.	N.S.	4.5	N.S.
% 25 & HS+	(DF: 3,10)					
Fer. Ratio						
	Full					
	Model:					
	R_2	.984	.775	.635	.930	.902
	R	96.8	60.0	40.3	86.4	81.4
	Gain in R^2	4.0	27.9	25.3	29.0	61.4
	F-Value	26.0*	N.S.	N.S.	5.5	N.S.
	(DF: 7,6)					

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

second highest correlation (.84) with the dependent variables, net in migration rate. This high correlation indicates in part why the factor chosen variables were so successful in this case and also why the "% 25 and over" variable emerged in the step-wise procedure. This evidence indicated that areas with a large proportion of the population well educated are in a favorable position to attract migrants. This is not very surprising since it can be shown that the four counties in Kansas which experienced the highest rates of in migration were university counties or else had a junior college or other advanced educational opportunities near by. Johnson county, one of the counties, is constituted by the suburbs of Kansas City which offer various junior colleges and vocational training schools as well as numerous white collar type jobs. Percent white collar was also one of the step-wise variables which emerged and was correlated .86 with net in migration rate. The three other counties Douglas, Riley, and Lyon are the home counties of Kansas University, Kansas State University, and Emporia State, respectively. In addition Lyon county also is home for the College of Emporia, while Riley houses Manhattan Bible College, and Douglas county has the Haskell Institute. These types of institutions not only employ a large proportion of persons who are highly educated but they also attract a large supply of people who desire to become so and hence have a high school education or more.

We also traditionally think of the white collar jobs going to persons with more years of formal schooling. Thus areas with a high proportion of white collar workers could be expected to have a high proportion of persons over 25 with high school or more. This supposition would be warranted in the case of the in counties under consideration as evidenced by the fact that "% white collar" correlated .69 with the % 25 and over" variable. This is larger than the .44 correlation which obtained between the two items in the all county analysis. What is occurring in the in counties is in part a result of the circular effect these two variables have on each other. They both work hand in hand to encourage not only an increase in each other but also in exerting a positive influence on net migration rates.

The third variable in the step-wise model was the fertility ratio which correlated -.44 with net rate. This variable should be seen in the same light as was the "% married" variable which was discussed earlier. Namely, the fact that an area is having fewer children per 1,000 women between 15 and 44 is not likely to be an explicit characteristic attracting migrants in comparison to other areas where the fertility ratio is higher. It is more likely that, similar to the "% married" variable, the fertility ratio is a function of other variables associated with high in migration rates. In the specific case we can hypothesize that the higher the level of education an area has the less likely it is to have a high proportion of large families. In this case we are dealing with areas where we have a high proportion of people 25 and over with high school or more. A check of Table 8 indicates however that "% 25 and over with high school or more" correlates only .02 with fertility ratio. This causes us to look elsewhere for a possible explanation (remembering the restrictions of the small sample size we are working with). It should also be remembered that the generalization may hold in the four counties mentioned (Douglas, Riley, Lyon, Johnson) but be counter balanced by the other 10 counties with in rates. We also can remember that in these four university counties the fertility ratio may be somewhat deflated due to the larger number of single women in the age 15-25 who are attending school. In effect I am saying the fertility ratio is more likely an effect of migration rather than a cause of it in a theoretical sense. Nonetheless it does emerge as a predictor in the step-wise procedure.

The factor analysis variables, % agric., % 18 and under, and % 25 and over with H.S. or more correlate -.68, -.23, and .84 respectively with net in rate. The % 25 and over with H.S. or more has already been discussed. The % agric. correlation -.68 may be seen as suggesting that areas with a high proportion of agriculture workers, rural areas, are not likely to experience in migration. In this case % agric. and % rural farm plus rural non-farm correlate .87. Thus it can be seen as part of the general rural to urban or suburban trend in migration.

The smaller negative correlation (-.23) of the third variable % 18 and under with net rates indicates that an older age structure favors in migration. In the case of these 14 counties this may result from the fact that in comparison with all migrants not many people under 18 have migrated into these counties. If this is true and most migrants are 18 or older then this variable could be a result rather than a cause of migration to an ecological area. It should also be noted that % 18 and under is correlated quite highly (.85) with the fertility ratio which also has a negative (-.44) correlation with net in rate and was discussed previously.

The in counties present a dilemma of sorts. They are the most interesting because there are so few of them. It would be valuable to know as precisely as possible what characteristics these counties had at the beginning of the decade 1960 which enabled them to experience the rarity of a net in migration rate. However, it is precisely because they are so rare in number that we must be cautious as to the types of generalizations we attempt to make about them. In the face of this possible shortcoming it is nonetheless fairly safe to say (based on the strength of correlations and R^2's) that two variables seem most important. The two are: % of white collar workers and % of persons 25 and over with high school or more. A high proportion of both or either one of these conditions is important for attracting migrants. This is seen as due in a large part to the influence of opportunities for advanced education and training. One can begin to get an intuitive feeling that Kansas Counties' biggest "industrial" asset may be an education system which prepares persons to leave the area permanently. The in counties may only be a stop-over for many migrants preparing themselves for later work elsewhere. Despite the possibility that a large portion of in migrants may be of the temporary type the growth and economic development of communities associated with such types of growth rates are nonetheless usually deemed advantageous. In summary % white collar workers and % 25 and over with H.S. or more are the best indicators derived from this study for accounting for variation in net in migration. This is seen not only as a result of the direct effect each has on bringing people to an area but also as partly due to the attendant characteristics which are associated with such indicators. The step-wise restricted model was the best predictor explaining 92.8% of the variation in Y. The full model explained 96.8% of the variation but yielded a smaller F-value. Despite the difference in the degrees of freedom associated with each model both were significant at the .01 level.

0-50% Rural Counties Results

Having reviewed the findings of all counties as well as the in and out counties I will now precede to an analysis of the remaining sets of counties which were classified according to their degree of rurality. The first set to be considered is the classification of counties which were only 0-50% rural and therefore can be thought of as "urban" counties. Thirty-five such counties were analyzed and produced the correlation matrix presented in Table 8 above the diagonal. The pattern of correlations which resulted between the factor variables was like that which occurred in all county cases. Two variables had negative correlations with net rates of migration (% ag., -.17; % 18 and under, -.20). The third variable, % 25 and over with high school or more, had a positive "r" of .31 with net migration rates.

The step-wise variables which emerged in the "urban" counties set were % white collar, fertility ratio, and % unemployed. These three variables had higher correlations (.47, -.33, and -.23, respectively) with the net rate than did the factor variables. Again % white collar emerged as the most highly correlated variable with Y. This was also the case in the "all" and "in" county analyses.

The regression analysis of net rates did not yield significant results at .05 level when performed with factor variables (See Table 12). This may be in part because none of independent variables correlated high enough with the dependent variable in this case to be significant at the .05 level. The critical value of "r" in this case, with degrees of freedom N-2 or thirty-three is .349 (Roscoe, 1969: 301). The highest "r" obtained was the .31 between the % 25>H.S. or more and the net migration rate.

The stepwise variables did yield significant results (Table 12) with the restricted model achieving an R^2 of 58 percent and the full model yielding an R^2 of 75 percent. Both were significant at the .01 level.

When attempting to account for variation in young age-sex specific rates the ability of the two sets of variables was reversed. In the case of the young male category the factor variables accounted for 52% and 67% of the variation. By contrast the step-wise yielded an R^2 of 13% and 22% respectively in the restricted and full models. The factor variable results were significant in the young male category at the .01 level while the step-wise variables were non-significant. The same pattern held for the young female rates (Table 12) except that the step-wise variables, while still not as good as the factor variables, did nonetheless, yield results significant at the .05 level as they accounted for 24% and 43% of the variation in the female 10-14 (20-24) rate.

An important comparison to be made at this time is between the results of the step-wise variables in the all county case and the "urban" case. It is noted that the

Table 12. Multiple R's, R^2's, and F-Values for Counties 0-50% Rural. Comparison of Sets of Variables and Types of Models

			Male		Female	
Counties 0-50% Rural N=35		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R	.457	.720	.266	.770	.467
	R^2	20.8	51.9	7.1	59.2	21.8
% Ag.	F-Value	N.S.	11.1*	N.S.	15.0*	N.S.
% 18 & Under	(DF: 3,31)					
% 25 HS+						
	Full					
	Model:					
	R	.507	.821	.448	.835	.631
	R^2	25.7	67.4	20.0	72.8	39.8
	Gain	4.9	15.5	12.9	13.6	18.0
	F-Value	N.S.	8.0*	N.S.	10.3*	2.5
	(DF: 7,27)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R	.759	.358	.517	.490	.634
	R^2	57.7	12.8	26.8	24.1	40.2
% W.C.	F-Value	14.1*	N.S.	3.7	3.3	6.9*
Fer. Ratio	(DF: 3,31)					
% Unemployed						
	Full					
	Model:					
	R	.868	.473	.651	.655	.719
	R^2	75.3	22.4	42.4	42.9	51.7
	Gain	17.6	9.6	15.6	18.8	11.5
	F-Value	11.8*	N.S.	2.8	2.8	4.1*
	(DF: 7,27)					

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

results are better in the "urban" case with the full model accounting for 75 percent of the variation as opposed to only 52 percent in the all counties case. This indicates that different forces may be at work in urban counties as opposed to all counties and that the step-wise procedure when controlled for rurality is fruitful in purifying the research. Future discussion will consider the same comparison for the other two classifications of counties, 51% - 99% rural and 100% rural.

It may prove beneficial to note that this classification of counties is one of only two in which the regression analysis proved to be useful at all in handling the older age-sex specific rates. It will also be remembered from previous discussion of mean rates at the outset of this chapter that this was the only set of counties in which the older mean rates were negative. The step-wise full model was the most efficient in accounting for variation by yielding an R^2 of 43% for males and 52% for females. Table 12 illustrates that five of the eight possible F-values for the older rates were significant at least at the .05 level. This is in marked contrast to previous results.

<u>51-99% Rural Counties Results</u>

Table 9 presents the correlation matrix (above the diagonal) for the analysis of the counties 51-99% rural. The step-wise variables and their correlations with Y for this set of counties were: % mfg. (.59), farm operator level of living index (.56), and % married (-.33). The step-wise variables in addition to being relatively highly correlated with Y were also uncorrelated with each other (% mfg. with farm operator level of living, -.12; % mfg. with % married, .01; and % married with farm level of living, .24). Such characteristics lead one to suspect that their results in the regression analysis may be quite good. By contrast the factor variables correlate with Y -.22 (% agric.), .02 (% 18 or under), -.34 (% 25>H.S. or more). The results of the regression analysis support the speculation based on the correlations. The factor variables were nonsignificant at the .05 level in attempting to account for variation in the net rate, Y. The step-wise variables however, accounted for 64% and 66% of the variation in the respective models (Table 13). In this case the restricted model yielded an F-value of 12.3, significant at the .01 level. The full model, also significant at the .01 level, did not show an appreciable addition to the size of R^2 adding only two percent and resulting in the smaller F ratio of 4.7.

The correlations of the step-wise variables indicate a high proportion of persons employed in manufacturing exerts a positive force encouraging in migration. Also apparent is the fact that as the farm operation level of living decreases so does the likelihood of having a high in migration rate. This may be the result of push or pull factors. As the index

Table 13. Multiple R's, R^2's, and F-Values for Counties 51-99% Rural. A Comparison of Sets of Variables and Types of Models

Counties 51-99% Rural
N=25

			Male		Female	
		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R	.471	.659	.417	.732	.336
	R^2	22.2	43.4	17.0	53.5	11.3
% Ag.	F-Value	N.S.	5.4*	N.S.	8.1*	N.S.
% 18 & Under	(DF: 3,21)					
% 25 HS+						
	Full					
	Model:					
	R	.621	.778	.709	.829	.577
	R^2	38.5	60.6	50.2	68.7	33.3
	Gain	16.3	17.2	33.2	15.2	22.0
	F-Value	N.S.	3.7	N.S.	5.3*	N.S.
	(DF: 7,17)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R	.799	.773	.560	.760	.471
	R^2	63.8	59.8	31.4	57.9	22.2
% Mfg.	F-Value	12.3*	10.4*	3.2	9.6*	N.S.
Farm Level of Living	(DF: 3,21)					
% Married						
	Full					
	Model:					
	R	.813	.925	.638	.782	.627
	R^2	66.1	85.6	40.7	61.2	39.3
	Gain	2.3	25.8	9.3	3.3	17.1
	F-Value	4.7*	14.4*	N.S.	3.8	N.S.
	(DF: 7,17)					

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

goes down this may indicate that farm economic conditions on the average are poorer than elsewhere. Consequently persons may be forced to give up small farms and leave the county. Seen another way one could propose the index is an indicator which as it goes down, illustrates the absence of a potential pull vector at work in the county. That is the level of living is not attractive enough to bring persons into an area for economic reasons.

Moving to a consideration of the age-sex specific rates in this set of counties it can be seen that once more the models can not account for variation in the older rates. Only one of the eight possible F-values is significant, that one being the male rate in the step-wise restricted model. Accounting for 31 percent of the variation, an F-value of 3.2 is achieved which is significant at the .05 level.

Better results are yielded in considering the young rates with the step-wise variables being the superior set. The full model step-wise variables achieve an R^2 of 86% with the male rates 10-14 (20-24). The resulting F-value of 14.4 is significant at the .01 level. The restricted model yields the highest F-value (9.6) for young female rates while accounting for 58% of the variation. This leads one to conclude that the interaction terms in the full model are not as important in affecting females as they are in the male case.

When the comparison of the 51-99% rural counties and the "all" counties set is made, as in the "urban" case, the controlled case yields better results. The all counties case had an R^2 of 52 percent. The 51-99% rural case has an R^2 of 66 percent. In moving to a consideration of the last classification of counties it is noted the same event occurs.

<u>100% Rural Counties Results</u>

The step-wise full model accounts for 64% of the variation in net migration rates in the counties which are completely rural (Table 14). This in comparison with the R^2 of 52% in the all counties analysis. For all three of the classifications based on degree of rurality different sets of variables emerged from the step-wise analysis which were superior in their respective cases to the variables used in the all counties analysis. This suggests the degree of rurality does determine what structural characteristics affect migration rates. It is also possible to speculate that even more rigorous controls on the urban-rural continuum might be useful in achieving better results.

Table 9 presents the correlation matrix resulting from the analysis of the 100% rural counties. It will be noted that only twenty variables were correlated with the twenty-first one, the sum of % rural farm and % rural non-farm, having been deleted. This was necessary because the matrix used was taken from the multiple regression analysis. Since all

Table 14. Multiple of R's, R^2's, and F-Values for Counties 100% Rural. A Comparison of Sets of Variables and Types of Models

Counties 100% Rural
N=45

			Male		Female	
		Net	10-14	50-54	10-14	50-54
Set I	Restricted					
Factor Analysis	Model:					
Variables:	R	.141	.418	.190	.374	.203
	R^2	2.0	17.5	3.6	14.0	4.1
% Ag.	F-Value	N.S.	2.9	N.S.	N.S.	N.S.
% 18 & Under	(DF: 3,41)					
% 25 HS+						
	Full					
	Model:					
	R	.382	.596	.261	.558	.322
	R^2	14.6	35.4	6.8	31.1	10.3
	Gain	12.6	17.9	3.2	17.1	6.2
	F-Value	N.S.	2.9	N.S.	N.S.	N.S.
	(DF: 7,37)					
Set II	Restricted					
Step-Wise	Model:					
Variables:	R	.739	.692	.515	.704	.489
	R^2	54.5	48.0	26.5	49.5	24.0
% Mfg.	F-Value	16.5*	12.6*	4.9*	13.4*	4.31*
% 10,000 VPS	DF: 3,41)					
% 2,500 VPS						
	Full					
	Model:					
	R	.800	.736	.639	.752	.525
	R^2	64.0	54.2	40.8	56.6	27.6
	Gain	9.5	6.2	14.3	7.1	3.6
	F-Value	9.4*	6.3*	3.63*	6.9*	N.S.
	(DF: 7,37)					

Note: F-Values presented only if significant at .05 or less.
* = Significant at .01.
N.S. = Non-significant at .05.

the counties had "100%" on the twenty-first variable due to the nature of the control classification the standard deviation was zero. The standard deviation is a statistic used in the regression analysis and the analysis with the computer program can not be accomplished when the standard deviation is zero.

As can be witnessed in Table 9 the factor variables correlate quite low with net migration rate (% ag., .02; % 18 or under, -.02; and % 25 H.S. or more, -.15). We expect based on the strength of these correlations that the factor variables may be unsuccessful in accounting for much variation in Y. The results of the analysis (Table 14) reveal that the expectations are correct as the factor variables are nonsignificant at the .05 level in both the restricted and full models.

By contrast the correlations of the step-wise variables with Y are more encouraging (% mfg., .45; % farms < $2,500 V.P.S., .29; and % farms > $10,000 V.P.S., .05). The step-wise models do indeed account for significant proportions of the variation (restricted, 55% and full, 64%).

Again, as with the counties which were 51-99% rural, % mfg. emerged in the step-wise as a significant indicator of variation in net rates. This suggests that in the rural or mostly rural a high proportion of persons engaged in manufacturing activities encourages less out migration or aids in retarding the outflow of persons. It will be remembered that both the all rural and mostly rural counties had negative mean rates of migration. It could be hypothesized that % mfg. is positively correlated with migration rates because it acts as a pull force bringing people into the community. This is no doubt partly true, but given the limitations of the residual net rate this can not be determined definitely.

The factor variables as mentioned with reference to predicting the net rate, were generally unsuccessful in handling the age-sex specific rates. Six of the possible eight F-values were nonsignificant at the .05 level. The only two which were significant were the male rates 10-14 (20-24) in the restricted and full model. Each model only accounted for 18% and 35%, respectively, of the variation in Y however.

The step-wise variables were more efficient in handling the age-sex specific rates. Seven of the eight F-values were significant. The only nonsignificant one was the female, 50-64 (60-64) rate (Table 14).

Summary of Findings

Table 15 provides a summary of the overall results of the two sets of variables when used in the restricted and full models. As can be seen the step-wise variables give better results on the average. They resulted in 42 significant findings compared to 28 for the factor variables.

Table 15. Summary of Results of Sets of Variables and Models

		Net Rate	Male 10-14	Male 50-54	Female 10-14	Female 50-54	Total
Set I Factor Analysis Variables:	Restricted Model:						
	Sig.	3	6	-	5	-	14
	Nonsig.	3	-	6	1	6	16
	Full Model:						
	Sig.	2	6	-	5	1	14
	Nonsig.	4	-	6	1	5	16
	Total						
	Sig.	5	12	0	10	1	28
	Nonsig.	7	0	12	2	11	32
Set II Step-Wise Variables:	Restricted Model:						
	Sig.	6	4	4	6	2	22
	Nonsig.	-	2	2	-	4	8
	Full Model:						
	Sig.	6	4	3	6	1	20
	Nonsig.	-	2	3	-	5	10
	Total						
	Sig.	12	8	7	12	3	42
	Nonsig.	0	4	5	0	9	18

There was no difference in the full and restricted models when the factor variables were used. Both models produced 14 significant findings at .05 level or better and 16 non-significant findings.

In the case of the step-wise variables the restricted model produced more significant findings (22) than did the full model (20). The difference was in the models ability to yield one more significant finding in both the male and female 50-54 (60-64) age category. The fact that the restricted model slightly exceeded the full model in the number of significant findings should not obscure the more important fact that the full model generally accounted for more of the variation in Y and resulted in F-values which were significant at the same level as the restricted model. The restricted model's ability to produce more significant findings was in part due to the more stringent requirements placed upon the full model by the degrees of freedom resulting from increasing the number of predictors while retaining the generally small sample sizes. The encouraging results derived from the full model lend empirical support to the implied hypothesis that migration rates are a function of social, economic, and demographic variables acting interdependently.

Table 15 also points out that both models and both sets of variables experienced some difficulty in accounting for significant variation in the older age-sex specific rates.

Of the 48 possible findings only 11 were significant. By contrast 42 of the possible 48 F-values for the younger age-sex rates were significant.

The differences in the models and variables in accounting for male-female rates in general were not outstanding. The step-wise variables were more efficient in the handling of young female rates than they were in the young male case. All twelve possible F-values were significant for the females while only 8 were so for the males.

In summary, considering the number of significant findings and also the R^2's achieved, the step-wise full model is the best of the four possible combinations of models and sets of variables.

CHAPTER IV

CONCLUSIONS

Several questions were posed at the end of Chapter One in this paper. It is appropriate to review these questions at this time to see if this project has provided answers to them. The first question was concerned with the determining the best way to select variables from those available. It would appear from this project that the step-wise procedure is better than the factor analysis method. However, it was noted in Chapter Three that the interaction between the factor variables sometimes increased the size of R^2 more than did the interaction (full model) between the step-wise variables. This suggests that future models employing interaction terms might find it useful to use factor analysis as a method of isolating distinct indicators. It should also be remembered that in this piece of research the factor analysis was performed on all counties. However the items which emerged were used as independent variables in all the classifications. This is recognized as not being entirely fair to the factor analysis. The fact that the items ability to handle the dependent variables varied as it did indicates even more so that different factors may have emerged if an analysis had been done separately on the classification of counties. Nonetheless in the all counties case where a sound comparison could be made the step-wise procedure was better than the factor analysis procedure in selecting items which in the full model would account for variation in net migration rates.

A second question dealt with whether the county is a profitable unit of analysis for an ecological model when dealing with rates obtained over a decade. Based on the results of this paper I contend that the county is a profitable unit of analysis for statistical purposes. Achieving R^2's of as high as .97 (in county, net rate) is quite respectable. It would appear to be useful to continue to use data pertaining to intercensal periods of a decade. One can surmise however that it would be even more profitable and might increase the power of the models by reducing the time span they are expected to project over. For example a five or one year period could be used by taking the characteristics of counties (or other areal units) at the beginning of a period and projecting over the shorter period. This project has illustrated that the period of a decade and the county unit can be used profitably, but does not contend that other time intervals (shorter) and other units (i.e. economic areas) might not be more profitable.

The third question asked in Chapter One was whether the interactive effects of social, demographic, and economic variables warranted further study in such models. The full model's ability to explain more of the variation in the findings in Chapter Three while achieving a level of significance comparable to the restricted model provides a positive answer to the question. Migration as a complex phenomena to be explained can be approached with the promise of achieving better results if one considers various aspects of the social milieu interdependently as opposed to separately.

A fourth question explored the ability of ecological models employing structural variates to explain older age-sex specific migration rates. This project supported the findings reviewed earlier of Stinner and DeJong which illustrated that ecological models are not as effective in accounting for migration in the older age categories. In conversations with Dr. Cornelia Flora and Dr. George Peters of Kansas State it was pointed out that older persons may not migrate for the same types of reasons younger people do. Older persons may not be as directly affected by systemic forces operating in ecological models such as employed in this project. While younger persons are affected by the characteristics of their habitat or potential habitat which bear directly on their future it is possible that older persons are more oriented toward recapturing the past. Thus young persons are affected by variables indicating economic or educational opportunities available in given areas. Conversely older persons may be more influenced in their migration decision by social psychological factors. Such an explanation offers a plausible reason why the ecological model may not be appropriate in attempting to account for older age-sex specific migration rates and should alert researchers to consider alternative procedures.

The fifth question raised asked whether an ecological model could be used to predict migration in Kansas and if so what were the important considerations in applying such a model. A partial answer to this question has been provided by this research. It is possible to isolate indicators of county characteristics at the beginning of an intercensal period account for varying proportions of the net migration rates which occur during the period. This research did not take the next logical step and attempt to use such indicators to predict migration rates for the 1970-1980 decade. Such a test could be done and would partially test one of the basic assumptions of this research, namely, that there indicators which are basic enough in nature that their influence will remain constant, or nearly so, over two decades or longer.

It also became apparent during the course of this project that the various control classifications used increased the models ability to account for variation in net and age-sex rates. Degree of rurality and the direction (in or out) of the net rate were important depending upon the type of area under consideration. One disadvantage of using such controls is the resulting decrease in sample size. If the control is theoretically sound, however, and the resulting classification has at least twenty-five to thirty counties in it the control can usually be justified by increasing the power of the models. The wide range in the various models ability to handle age specific rates underlines the importance of considering such rates independently which in effect is a control classification for the dependent variables and one that should not be overlooked in future research. The age specific rates not only are more profitable for statistical purposes but are also more useful and interesting in a theoretical sense.

When this research was completed several ideas emerged which take the form of suggestions for further research. These suggestions are in part a function of some questions this project attempted to answer, but did not fully do so and secondly the result of some questions the project itself raised.

An interesting point confirmed by the project centers on the ecological model's inability to handle the older age rates. In addition to the earlier discussion of the possible reason for this phenomena being involved with potential social psychological motives another aspect should be considered. It is possible that since the older rates had such a small range of variation that the items used in the regression analysis were not sensitive enough measures to account for such small differences in rates. Also, the items chosen in the step-wise procedure were selected on the basis of their ability to handle net rates. This raises a question as to whether different variables would have evolved if the age specific rates had been the dependent variable. To answer this question the procedure should be rerun with the younger and older rates as dependent variables. This would allow one to determine if any of the variables in the initial pool of twenty-four could account for variation in the older rates, and if so what are they and how they differ from the type of variables which emerge in the case where the younger rates are dependent variables. This type of procedure allows one to determine if it is the type of model being employed or if it is the nature of the variables which caused it to fail in the present project. If such a procedure yielded results which were still generally nonsignificant it would lend support to the conclusion that the type of social, demographic, and economic variables used in this model's analysis are not relevant measures of the forces affecting areas which pull or push older persons to or from such areas.

A second suggestion for future work in this area would be to perform the factor analysis on each set of counties or control classifications. This would be a more sound design than assuming as this project did, that the items which emerged from the one analysis of all counties would remain constant across all sets of counties.

The model used in the present project assumed a linear relationship between the variables. It is suggested that future research test for quadratic and/or cubic relationships. This could be done quite simply with the addition of a few basic statements to the computer program. After the best variables had been selected the test for quadratic can be built into the model by squaring the desired variables as opposed to multiplying them as was done in this research. The cubic test involves cubing the desired variables.

Before such research is undertaken it is essential that a computer program other than one used in this project be available to the researcher.

It was recognized before this project was undertaken that it would be most desirable to have a program which would include as part of the output the beta weights of the respective variables. Such knowledge would enable the researcher to determine the relative weight or importance of each variable's influence on the dependent variable. An alternative which would allow one to get at a similar type of information would be the standardized partial multiple correlation coefficient. Neither of these statistics were included as output from the program used in this research and consequently many desireable types of inferences which should be made about the results are not available.

Future work should include these statistics. By doing so the researcher can determine the respective amount of influence each variable has on the resulting R^2. This type of information becomes critical in models employing variables which are interaction terms or variables which are testing for quadratic or cubic relations.

In the absence of such informative statistics the researcher can only speculate about the probable influence of each variable based on its correlation with the dependent variable. Some attempts were made to do this type of speculation in this paper, but the results should be interpreted cautiously. It is recognized that much of the significance of the full model with its interaction terms is lost when one is unable to tell which interaction term was most significant. If such information were available it would stimulate clearer theoretical insights as to how and why the variables were affecting migration rates.

It is also recommended that more work be done using different areal units which may be more relevant than the county. Shorter periods of time should be considered also to determine if cutting down on the decade interval would increase the utility of the models.

Recognizing the limitations placed upon inferences based on net migration rates it is suggested that if time and resources were available it would be useful to examine gross rates of in and out migration. The use of in and out rates in place of the residual net rate would aid in clarifying the effects of the independent variables and would sharpen the conclusions which could be drawn about specific variables.

Given the nature of the results available from this project we are somewhat limited in the precision with which we may speak of reliable indicators or the best variables to emerge from the project. Before generalizing it should be remembered that different sets of variables emerged for each set of controls imposed on the county classifications. Thus if one wished to predict rates for sets of counties he should see the particular variables for the classification desired.

With reference to the factor analysis variables we can only speculate about the relative importance of each variable based on its correlation with dependent variable. This was done in Chapter Three's discussion of the findings.

When considering the step-wise variables we are basically in the position as above with the exception that we do know the order in which each variable entered the equation attempting to explain variation in net rates. We also know how many times each variable appeared as one of the three best indicators. Table 16 summarizes the variables which appeared, the number of times they appeared, their relative position in the set of three on each occurrence, and their average rank on all appearances.

It is clear that the economic variables were the most important variables which emerged. Percent white collar and % mfg. both appeared in three equations and both were the first variable entered on all occasions. Achieving an

Table 16. Step-wise Variables: Rank of each Occurrence, Number of Occurrences, and Average Rank

Variable	Rank of Each Occurrence		Number of Occurrences	Average Rank
% W.C.	1,1,1	(3)	3	1
% Mfg.	1,1,1	(3)	3	1
% Own. Occ.	2,2	(4)	2	2
Fertility Ratio	3,2	(5)	2	2.5
% Farms > $10,000 VPS	3,2	(5)	2	2.5
% Married	3,3	(6)	2	3
% 25 > H.S. +	2	(2)	1	2
Farm Op. Lev. of Liv.	2	(2)	1	2
% Unemployed	3	(3)	1	3
% Farms < $2,500 VPS	3	(3)	1	3

average rank of "one" is the best possible score and both variables had an average rank of "one". Percent of homes owner occupied was the third best variable appearing on two occasions as the second variable entered in the equation. This variable again underscores the importance of economic indicators.

The results summarized in Table 16 lead one to conclude that future research should continue to recognize the importance of economic determinants.

The previous suggestions have pertained to improvement on a similar design and methodology as employed in this project. It would also be well to consider the use of different designs and techniques in analyzing the same phenomena.

One of the most obvious alternatives or additions would be to perform a path analysis (Duncan, 1966: 1-16). Such an analysis, while not providing proof positive of causal relations, could nonetheless move toward confirming the possibility of causal links between the variables, thus reducing the amount of speculation made necessary in this project on the basis of simple correlation coefficients. Such a technique offers the potential to reduce errors resulting from confusing correlations due to communality and causality (Schuessler, 1969: 21-23).

A path analysis would also serve as a partial test of how appropriate the frame of reference used in this study is for such research. It will be remembered the ecological complex offers a system in which characteristics of populations (migration rates) can be seen as dependent variables. It would prove interesting to discover the degree to which such a view of migration as a dependent variable is warranted.

For the purposes of statistical analysis in the regression models used in this paper migration rates were seen as the result of variation in selected independent variables. This should not obscure a fact which the ecosystem concept is meant to underscore. Namely, that while for the purposes of analysis any of the ecosystem components may be seen as a dependent variable they are in fact interdependently related to each other and a degree of circular causality exists. Any interpretation of migration rates as a dependent variable would do well to keep such considerations in mind.

This project only scratched the surface of a fascinating and intriguing phenomenon of human behavior: migration. Migration should continue to be of interest to social scientists in its own right as a distinct area of study. It is also recognized that a multitude of other types of social research are complicated and made more difficult to carry out due to the effects of migration. Models employing an ecological perspective as well as designs to measure the social psychological underpinnings of migration would seem to warrant much more attention from demographers and social scientists specifically trained as sociologists. Such models taken separately, or more desireably, integrated with one another hold promise for elucidating the complex phenomenon of human migration.

REFERENCES

ALIHAN, M.A.
1939 Social Ecology: A Critical Analysis. New York: Columbia University Press.

ALKER, HAYWARD R., JR.
1969 "A Typology of Ecological Fallacies." Pp. 69-86 in Dogan, M. and Stein Rokkan (eds.), Quantitative Ecological Analysis in the Social Sciences. Cambridge, Mass.: M.I.T. Press.

BAILEY, KENNETH D., and PATRICK MULCAHY
1971 "Sociocultural Versus Neoclassical Ecology: A Contribution to the Problem of Scope in Sociology," Sociological Quarterly 13 (Winter): 37-48.

BECKER, HOWARD and HARRY ELMER BARNES
1961 Social Thought From Lore to Science, Vol. 1. New York: Dover Publications, Inc.

BOGUE, DONALD J.
1952 A Methodological Study of Migration and Labor Mobility in Michigan and Ohio in 1947. Oxford, Ohio: Scripps Foundation for Research in Population Problems.

1957 Components of Population Change, 1940-50: Estimates of Net Migration and Natural Increase for Each Standard Metropolitan Area and State Economic Area. Ohio: Scripps Foundation for Research in Population Distribution No. 12.

1969 Principles of Demography. New York: Wiley.

BOWLES, GLADYS K., and JAMES D. TARVER
1966 Net Migration of the Population, 1950-60, by Age, Sex, and Color, 2 vols. Washington, D.C.: U.S. Government Printing Office.

DAWOOD, N. J. (ED.)
1967 The Muqaddimah: An Introduction to History. Ibn Khaldun. New Jersey: Princeton University Press.

DUNCAN, OTIS DUDLEY
1959a "Human Ecology and Population Studies." Pp. 678-716 in Phillip M. Hauser and Otis Dudley Duncan (eds.), The Study of Population. Chicago: University of Chicago Press.

1959b "Cultural, Behavioral, and Ecological Perspectives in the Study of Social Organization." The American Journal of Sociology 65 (September): 132-146.

1964 "Social Organization and the Ecosystem." Pp. 36-82 in Robert E. L. Faris (ed.), Handbook of Modern Sociology. Chicago: Rank McNally.

1966 "Path Analysis: Sociological Examples." American Journal of Sociology 72 (July): 1-16.

FLORA, CORNELIA, et al.
1971 Migration in Kansas: Out Migration and Population Trends. Manhattan, Kansas: Agricultural Experiment Station, Kansas State University.

GIBBS, JACK P., and WALTER J. MARTIN
1959 "Toward a Theoretical System of Human Ecology." Pacific Sociological Review 2 (Spring): 29-36.

GOLDSTEIN, SIDNEY
1958 Patterns of Mobility, 1910-1950: The Norristown Study. Philadelphia: University of Pennsylvania Press.

HAMILTON, D. HORACE
1959 "Education Selectivity of Net Migration From The South." Social Forces 38 (October): 33-42.

1965 "County Net Migration Rates" Rural Sociology 30 (March): 13-17.

PARK, ROBERT E.
1936 "Human Ecology." American Journal of Sociology 42 (July): 1-15.

ROBINSON, W. S.
1950 "Ecological Correlations and the Behavior of Individuals." American Sociological Review 55 (June): 351-357.

ROSCOE, JOHN T.
1969 Fundamental Research Statistics for the Behavioral Sciences. Chicago: Holt, Rinehart, and Winston, Inc.

SCHNORE, LEO F.
1965 The Urban Scene. New York: The Free Press.

SCHUESSLER, KARL
1969 "Causality Versus Communality." Et. al. (Summer) Los Angeles, California: 21-23.

SHRYOCK, HENRY S., and JACOB S. SIEGEL
1971 The Methods and Materials of Demography. U.S. Bureau of the Census, U.S. Government Printing Office: Washington, D.C.

SIMPSON, GEORGE. (Trans.)
1947 The Division of Labor. Emile Durkheim Glencoe, Ill.: Free Press.

STINNER, WILLIAM F., and GORDON F. DEJONG.
1969 "Southern Negro Migration: Social and Economic Components of an Ecological Model." Demography 6 (November): 455-471.

TARVER, JAMES D.
1961 "Predicting Migration." Social Forces 39 (March): 207-213.

TARVER, JAMES D., and WILLIAM R. GURLEY
1965a "The Relationship of Selected Variables with County Net Migration Rates in the United States, 1950-1960." Rural Sociology 30 (March): 3-13.

1965b "Rejoinder." Rural Sociology 35 (March): 17-22.

THEODORSEN, GEORGE
1961 Studies in Human Ecology. New York: Harper and Row.

U.S. BUREAU OF CENSUS
1960a Mobility for States and State Economic Areas. Washington, D.C.: U.S. Government Printing Office.

1960b U.S. Census of Population - Kansas Vol. 1 Part 18. Washington, D.C.: U.S. Government Printing Office.

U.S. BUREAU OF THE CENSUS
1962 County and City Data Book, 1962. Washington, D.C.: U.S. Government Printing Office.

1967 County and City Data Book, 1967. Washington, D.C.: U.S. Government Printing Office.

WILLHELM, SIDNEY M., and GIDEON, SJOBERG
1969 "The Concept of the Ecological Complex: A Critique." American Journal of Economics and Sociology 23: 241-248.

APPENDIX 1

ESTIMATES OF MIGRATION RATES BY AGE AND SEX FOR KANSAS, 1960-1970

Age Cohort 1960		1970	Male Rate	Female Rate
0-4	to	10-14	-6.3	-7.1
5-9	to	15-19	-5.0	-3.4
10-14	to	20-24	-2.5	-5.5
15-19	to	25-29	-10.2	-10.9
20-24	to	30-34	-11.2	-8.0
25-29	to	35-39	-9.8	-7.5
30-34	to	40-44	-8.5	-7.1
35-39	to	45-49	-8.8	-7.6
40-44	to	50-54	-5.1	-3.7
45-49	to	55-59	-3.8	-3.5
50-54	to	60-64	-1.6	-0.3
55-59	to	65-69	-5.1	-1.7
60-64	to	70-74	-3.3	+0.8
65+	to	75+	-3.9	0.0

APPENDIX 2

ESTIMATES OF SELECTED MIGRATION RATES BY AGE, SEX AND COUNTY, 1960-1970

County	Age Cohort 1960 10-14 to 1970 20-24 Male	Female	1960 50-54 to 1970 60-64 Male	Female
Allen	-54.1	-44.9	8.3	-0.5
Anderson	-55.3	-57.8	0.6	2.7
Atchison	-3.2	-11.9	-8.0	-10.3
Barber	-64.8	-61.5	-1.7	8.2
Barton	-54.1	-40.3	-7.6	-6.6
Bourbon	-40.9	-25.4	5.4	11.9
Brown	-54.4	-55.3	8.5	11.2
Butler	-39.6	-27.7	1.2	3.9
Chase	-56.7	-55.0	-5.8	-11.5
Chautauque	-68.2	-62.8	-2.3	13.3
Cherokee	-42.6	-34.5	-0.2	5.9
Cheyenne	-66.5	-49.7	8.7	0.5
Clark	-72.3	-75.1	-1.3	16.5
Clay	-53.6	-50.5	8.1	4.8
Cloud	-46.7	-42.3	2.0	-0.7
Coffey	-57.5	-58.3	9.3	17.7
Comanche	-58.8	-62.9	-7.8	0.0
Cowley	-32.0	-23.4	-2.1	-3.9
Crawford	54.5	36.5	2.1	10.9
Decatur	-67.4	-58.6	1.3	-2.9
Dickinson	-38.8	-30.3	-9.6	-10.9
Doniphan	-43.4	-38.9	-8.9	-3.2
Douglas	306.0	248.8	-1.6	2.6
Edwards	-66.1	-62.1	2.6	5.9
Elk	-70.3	-60.4	-6.1	7.0
Ellis	92.6	80.9	0.3	0.0
Ellsworth	-67.2	-64.7	-5.9	-6.5
Finney	-27.8	-12.2	-10.9	-3.1
Ford	-22.5	-10.9	6.5	6.5
Franklin	-19.7	-7.1	3.1	3.5
Geary	373.7	112.0	-11.9	-8.7
Gove	-54.1	-57.4	12.9	-3.4
Graham	-62.2	-59.4	-8.2	-6.8
Grant	-44.7	-19.9	4.1	-10.7
Gray	-54.7	-36.2	13.8	8.9
Greeley	-63.8	-59.8	-4.8	-7.5
Greenwood	-60.8	-54.6	-0.6	-4.2
Hamilton	-67.4	-58.5	-0.9	7.9

Appendix 2 (Continued)

County	Age Cohort 1960 10-14 to 1970 20-24 Male	Female	1960 50-54 to 1970 60-64 Male	Female
Harper	-64.0	-59.7	4.1	2.6
Harvey	-58.3	-12.0	6.4	6.2
Haskell	-27.7	-20.4	8.2	15.4
Hodgeman	-67.3	-63.9	-12.7	-15.1
Jackson	-51.8	-38.2	0.3	-0.3
Jefferson	-50.1	-29.8	8.7	-5.9
Jewell	-66.3	-54.8	-2.0	10.1
Johnson	-22.1	9.6	6.0	10.0
Kearney	-49.4	-55.7	3.6	3.6
Kingman	-63.4	-60.6	6.4	-5.5
Kiowa	-57.3	-66.1	8.0	-1.7
Labette	-43.3	-33.0	2.2	-0.1
Lane	-66.1	-57.1	0.6	-1.8
Leavenworth	31.2	-18.2	-9.8	-4.1
Lincoln	-69.2	-64.8	-1.5	6.5
Linn	-57.1	-51.4	13.8	0.4
Logan	-63.5	-53.6	-2.8	-15.2
Lyon	156.0	156.1	1.7	2.1
Marion	-6.2	-5.7	-2.7	6.2
Marshall	-46.3	-44.7	4.2	3.6
McPherson	-65.5	-53.7	2.6	3.8
Meade	-60.0	-64.4	-2.6	-3.3
Miami	-50.2	-37.4	-6.8	-5.2
Mitchell	-58.1	-52.3	9.3	4.8
Montgomery	-50.9	-43.4	3.5	3.2
Morris	-58.7	-59.5	3.1	4.0
Morton	-52.2	-38.4	38.4	46.5
Nemaha	-65.7	-65.8	1.7	-1.6
Neosho	-43.9	-37.5	-1.1	0.9
Ness	-68.8	-70.0	8.9	9.6
Norton	-55.8	-54.1	-0.7	3.7
Osage	-44.2	-38.6	7.2	3.1
Osborne	-62.5	-66.3	5.8	4.7
Ottawa	-62.7	-47.6	16.4	8.6
Pawnee	-43.9	-43.3	-14.5	-11.0
Phillips	-56.0	-47.1	4.1	-3.0
Pottawatomie	-31.3	-29.7	11.4	8.8
Pratt	-55.8	-52.4	-8.6	5.3
Rawlins	-69.5	-63.6	-4.8	4.1
Reno	-23.0	-19.0	0.8	1.2
Republic	-67.1	-58.3	5.6	-1.6
Rice	-42.4	-42.2	-6.8	2.0

Appendix 2 (Continued)

County	Age Cohort 1960 10-14 to 1970 20-24 Male	Female	1960 50-54 to 1970 60-64 Male	Female
Riley	907.6	265.0	-1.9	-0.6
Rooks	-60.9	-55.5	-4.6	0.3
Rush	-68.1	-55.6	-13.3	-1.5
Russell	-66.5	-53.2	-8.6	-8.8
Saline	-28.8	-6.3	-10.4	-3.9
Scott	-48.0	-37.4	3.6	-2.5
Sedgwick	-5.1	3.4	-8.2	-6.7
Seward	-31.2	-17.6	-6.2	-18.7
Shawnee	25.8	27.5	1.6	3.0
Sheridan	-65.3	-64.5	0.9	-3.4
Sherman	-5.9	-24.5	5.7	6.5
Smith	-62.8	-64.4	-1.6	1.8
Stafford	-77.1	-68.5	-11.7	-6.5
Stanton	-49.2	-34.5	4.9	-13.1
Stevens	-50.7	-36.1	-6.8	-1.8
Summer	-51.3	-44.8	2.0	1.2
Thomas	-38.4	-36.2	10.9	9.5
Trego	-69.4	-68.3	-4.7	-8.4
Wabaunee	-53.6	-41.5	6.8	6.5
Wallace	-46.1	-54.9	11.7	6.5
Washington	-57.7	-62.8	-2.0	0.4
Wichita	-40.2	-16.9	-6.2	3.9
Wilson	-57.4	-52.5	2.9	-2.5
Woodson	-52.1	-61.1	-5.2	9.5
Wyandotte	-23.6	2.2	-7.8	-7.7

PREDICTING MIGRATION IN KANSAS: AN EVALUATION OF ALTERNATIVE PROCEDURES FOR SELECTING PREDICTORS AND MODELS

by

KENNEY HAROLD BARB

B. S., Kansas State University, 1966

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Sociology and Anthropology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

This study was carried out to evaluate two methods of selecting indicators to predict migration rates for Kansas counties. In addition to evaluating the methods of variable selection an effort was made to determine which of two regression models was better suited to account for variation in migration rates. The project was a preliminary inquiry into the feasibility of developing a model, or models, which could efficiently predict migration rates in Kansas.

Data collected from the 1960 U.S. Census of Population for Kansas was employed in an ecological model. Two methods were used to reduce a pool of twenty-four social, economic, and demographic variables to three variables. Factor analysis and step-wise multiple regression analysis were the techniques used to select the best three variables from the original pool.

The variables selected were tested in two regression models for their ability to account for net and age-sex specific migration rates over the 1960-1970 decade. The first regression model (restricted model) considered only the three variables selected as the best in each case. The second model (full model) employed the three original terms plus four more terms which were designed to test for "interaction effects" of the three original variables. The models allowed a test of the hypothesis that the interdependent effects of social, economic, and demographic characteristics are more efficacious than are their independent effects in accounting for variation in migration rates. The analysis was performed on all counties of Kansas as well as five control classifications based on direction of net migration in the county (in or out) and on degree of rurality.

The results of the analysis indicated that the step-wise regression technique is the better of the two techniques for selecting indicators. The full model was shown to be generally better than the restricted model. Economic indicators emerged as more reliable than the social or demographic indicators in accounting for variation in migration rates.

Indicators varied in their ability to handle classification of counties depending upon the direction of net migration and degree of rurality. The models were much more successful in handling younger age specific rates than they were in handling older age specific rates. No great differences emerged between the models ability to account for male-female rates but there were some slight differences.

This research lends support to the efficacy of employing ecological models to an analysis of migration rates in Kansas counties. However, the project underscores the fact that such models may not be entirely appropriate for and analysis of older age specific migration rates and offers possible explanations for the models' failure in such cases. The project was seen as capturing only one aspect of the complex phenomena of human migration. It is recognized that ecological models are complementary with and augment knowledge revealed by other designs which evaluate the social psychological underpinnings of migration.