Statistics Canada Statistique Canada

Occupational Composition of Canadian Migration

by Leroy O. Stone

Census Analytical Study

Occupational Composition of Canadian Migration

By Leroy O. Stone
with the assistance of
Susan Fletcher and
Frances Aubry

Published by authority of
The Minister of Industry, Trade and Commerce

February 1979
8-0003-503

Price: $1.40

Catalogue 99-752E

Ottawa

Version française de cette publication
disponible sur demande (n° 99-752F au catalogue).

FOREWORD

The Canadian censuses constitute a rich source of information about the condition of groups and communities of Canadians, extending over many years. It has proved to be worthwhile in Canada, as in some other countries, to supplement census statistical reports with analytical monographs on a number of selected topics. The 1931 Census was the basis of several valuable monographs, but, for various reasons, it was impossible to follow this precedent with a similar programme until 1961. The 1961 Census monographs received good public reception and have been cited repeatedly in numerous documents that deal with policy problems in diverse fields such as manpower, urbanization, income, the status of women and marketing. They were also of vital importance in the evaluation and improvement of the quality and relevance of Statistics Canada social and economic data. This successful experience led to the decision to continue the programme of census analytical studies. The present series of analyses is focused largely on the results of the 1971 Census.

The purpose of these studies is to provide a broad analysis of social and economic phenomena in Canada. Although the studies concentrate on the results of the 1971 Census, they are supplemented by data from several other sources. These reports are written in such a way that their main conclusions and supporting discussion can be understood by a general audience of concerned citizens and officials who often lack the resources needed to interpret and digest the rows of numbers that appear in census statistical bulletins. For these persons, interpretive texts that bring the dry statistics to life are a vital dimension of the dissemination of data from a census. Such texts are often the only means that concerned citizens and officials have to personally perceive benefits from the national investment in the census. This particular report is one of a series planned to be published concerning a variety of aspects of Canadian life, including income, language use, farming, family composition, migration, adjustment of immigrants, human fertility, labour force participation, housing, commuting and population distribution.

I should like to express my appreciation to the universities that have made it possible for members of their staff to contribute to this programme, to authors within Statistics Canada who have freely put forth extra effort outside office hours in preparing their studies, and to a number of other members of Statistics Canada staff who have given assistance. The Social Science Federation of Canada has been particularly helpful in the selection of authors for some of the studies, and in arranging for review of several manuscripts. In addition, thanks are extended to the various readers, experts in their fields, whose comments were of considerable assistance to the authors.

Although the monographs have been prepared at the request of and published by Statistics Canada, responsibility for the analyses and conclusions is that of the individual authors.

PETER G. KIRKHAM,

Chief Statistician of Canada.

PREFACE

This report is the second part of one of the series of 1971 Census Analytical Studies. The first part is entitled *The Frequency of Geographic Mobility in the Population of Canada* and is being published in a separate volume. The Census Analytical Studies, as well as the 1971 Census Profiles that have already been published, are intended to place into the public domain synthesized and interpreted census information. Although very large numbers of Canadians benefit only indirectly from these studies (through digests of the information prepared for mass communication and through the improvement of the basis of work by public and private policy analysts), they constitute a dimension of census data dissemination of the first order of importance. This observation arises from the fact that the vast majority of Canadians make direct use only of synthesized and/or digested information derived from a census.

This study should be viewed as a demographic analysis of an important aspect of the behaviour of Canadians – their geographic mobility. It concentrates upon variations in occupational and educational composition among different groups of migrants, rather than on the spatial pattern of residence change. In so doing it is addressed to at least three kinds of interests.

Firstly, many public officials and citizens who are concerned with policies and programmes in which the mobility of Canadians is an important factor should be able to obtain some benefits (in the area of improved knowledge) from reading Chapter 1 and the summaries of research findings that are reported in Chapter 4. The intended area of knowledge pertains to the degree and manner in which the occupational composition of a body of migrants is related to characteristics of the areas that serve as origins and destinations of the migration. It is hoped that this study will advance their knowledge, so that the assumptions of policies that relate to (or whose outcomes involve) the geographic mobility of Canadians can be more firmly grounded than formerly. Citizens concerned with the impact and evaluation of such policies should have an enhanced ability to assess the work and proposals of government, a matter of some significance in a democratic state. Some aspects of relevant government policies are cited briefly in the concluding summary section of Chapter 4.

Secondly, specialists and students who require somewhat intensive knowledge about the mobility of Canadians should be able to gain from the detailed models, statistical estimates and interpretations in Chapters 2 - 4, concrete information about the association of the occupational pattern of mobility with characteristics of the areas of origin and destination of migration. In the process of reviewing these materials, some of the specialists and students will likely be stimulated to explore on their own either questions that come to mind (or are sharpened) in the light of the discussion in the text or important analytical "loose ends" that remain at the end of this necessarily brief and exploratory work.

Thirdly, in a number of ways, this study contributes to the evaluation of the usefulness of 1971 Census data and to the planning of future censuses of Canada. In particular, certain limitations of the census data, for the purposes of analysing the socio-economic pattern of mobility in Canada, are clearly stated. Through this study, Statistics Canada is better able to help those who wish to use census data on migration. The use of basically inadequate data to address complex and important issues of knowledge is an ever-present problem in our society. Hopefully, this study illustrates ways of expanding the bounds of applicability of census migration data to practical questions.

Three factors have seriously delayed the publication of this report. Firstly, the achievement of a "clear" 1971 Census master file with respect to migration data occurred only near mid 1974. Secondly, the properties of the census data prompted the writer to opt for the adaptation and further development of a virtually brand new methodology for multivariate analysis. The explication of this development and the preparation of the related computer programmes were major time-consumers. (The resulting methodological advance is of major significance for the potential uses of Census of Canada data, however.) Thirdly, more urgent duties of a managerial nature continually interfered with the author's attention to this study, and as a result, the first draft (which was ready in late 1975) required serious and extensive revision in the latter half of 1976 and in early 1977. Despite the late publication, however, the value of this study should not be seriously impaired, because it deals with stable patterns in the geographic mobility of Canadians.

Many persons have made vital contributions to the completion of this report. The staff members of the Senior Advisor on Population Studies Division have been extremely supportive, Susan Fletcher (in the area of developing and implementing programming and statistical concepts that expanded the applicability of the general methodology, in editing the revised manuscript, and in drawing up data processing specifications), Frances Aubry (in programming census tabulation runs and in editing parts of the revised manuscript), and Marie-Claire Couture (in supervising the work of the clerical staff), and the clerical staff itself. Extremely valuable professional consultation was received from Stephen Fienberg (University of Minnesota), Leo Goodman (University of Chigago), Marvin McInnis (Queen's University) and James Simmons (University of Toronto). The work and inspiration of Fienberg and Goodman form a major basis for the new and important methodology that is only partially exposited in this study. Reviewer's comments on the first hastily written draft helped to trigger efforts that have led

to what I hope is a more interesting manuscript than would otherwise be the case. However, I am solely and personally responsible for the opinions and drafting errors that the text may contain, particularly since good advice may not always have been fully followed.

Leroy O. Stone,
Senior Advisor on Population Studies,
May 1977.

TABLE OF CONTENTS

TABLE OF CONTENTS – Concluded

LIST OF TABLES

LIST OF CHARTS

CHAPTER 1

PURPOSE AND MAIN FINDINGS

1.1. Purpose of Study

The social and economic conditions of a Canadian community depend partly upon the characteristics of the people who establish residence in it and the kinds of people who move out of it. People move into and out of almost all geographic areas that contain established communities. For a community that is confronted with opportunities and problems of economic and social change, the occupations of these migrants (the type of work they usually do) are among their most important characteristics.

If, for example, there is a surplus of teachers in a particular community, relative to the available teaching jobs, then the regional migration of teachers is a matter of more than passing interest to persons concerned with the problems of that community. In Northwestern Ontario, the migration of miners and loggers is a subject of considerable concern to persons dealing with the economic growth prospects of that region.

Citizens and public officials who are concerned with the role of population mobility in regional problems need to know what they can properly assume about the interrelation between the attributes of a region and the kinds of people who move to and from it. This brief analytical study attempts to contribute partially to the satisfaction of that need for knowledge: it uses 1971 Census data to measure the strength and pattern of association between the characteristics of the regions and the occupational composition of migration among those regions, and presents some of the results from the measurement exercise.

An example of the type of concern to which this study could be directed is the case of labour supply requirements in Northwestern Ontario. Sectors of the mining and forestry industries (including raw materials processing operations) are in substantial need of a stable labour supply with a particular mix of skills. Some company representatives in that region claim that industrial expansion in Northwestern Ontario is being threatened by difficulties in securing an adequately stable supply of labour with the necessary skills. A question that concerned citizens and officials are asking is: "How can Northwestern Ontario attract and retain workers with the types of skills that the region needs? ".

To answer this question, these persons must speculate about ways in which the conditions of work and social life in Northwestern Ontario communities help to determine the kinds of people who choose to reside in or leave the region. They must also consider the possibility that the factors which mould the occupational mix of the migration to and from Northwestern Ontario may have little to do with special features of the region.

The study adopts the working hypothesis that there is a systematic, though partial, connection between the occupational composition of a group of migrants and attributes of the group's areas of origin and destination. The study then develops and applies a method for measuring the extent to which the occupational composition of a group of migrants is systematically associated with measured characteristics of the areas of origin and destination for the migration, as well as with selected factors that are independent of such area characteristics. Both the directions and the relative weights of the factor's statistical contribution are considered. The main findings and interpretations from the statistical analysis are presented in Section 1.3.

For a number of reasons, some of which are offered in the following chapters, this analysis must be considered exploratory. The working hypothesis cannot be tested definitively with the available data. If the adopted methodology and its application seem reasonable in given circumstances, then the hypothesis may only be considered to be more or less plausible in the light of the patterns shown by the statistics. By presenting what are essentially preliminary results from an incomplete analysis, it is hoped that other researchers will be stimulated to re-examine more deeply the working hypothesis and the empirical patterns of association suggested by the census data.

1.2. Basic Concepts and Data Source

1.2.1. Concepts

Several concepts that will be used repeatedly require clarification.

"Migration" is generally considered to be the act of uprouting one's household and moving it from one location (an "origin") to another (a "destination"). Many studies and bodies of data limit the application of the term "migrant" to someone who actually crosses some defined boundary (e.g., a municipal boundary) in moving his household. (For further discussion of this concept, see Stone, 1969, pp. 6 - 8.) In the Canadian census statistics, for example, a move that fails to cross a municipal boundary is treated as an instance of geographic mobility, but it is not classified for statistical purposes as an instance of migration. This distinction is somewhat arbitrary. Generally, its usefulness rests on the assumption that moves that cross municipal boundaries are likely to be more consequential for the local communities as well as for the movers' households than those that do not cross such boundaries.

A "migration stream" refers to a group of migrants who resided in one specific area of origin at a particular time, but who resided in another specific area of destination at another given time. The timing of observation of place of residence, as well as the identities of the places of origin and destination, are needed in order for one to identify precisely a particular migration stream.

The "composition" of a migration stream with respect to a given attribute (e.g., occupation) means, in this study, the proportional distribution of the stream among categories of the pertinent variable. Thus by "occupational composition" of a migration stream is meant the proportional distribution of the migrants among categories of occupation.

In this text the term "region" simply means a defined geographic area. For example, it can signify a group of provinces (e.g., the Prairie region), a single province, a part of a province (e.g., metropolitan Toronto or the Gaspé), or a grouping of areas that may cut across provincial boundaries.

Typically, a region is considered to be an area whose parts are geographically contiguous. This study emphasizes another type of contiguity – economic contiguity (discussed in Chapter 2). In order to remind the reader that the regions of this study do not always have geographic contiguity among their parts, the term "region-group" will be used. A region-group is a collection of regions.

1.2.2. Data Source

The data base for this study is the 1971 Census of Canada, particularly the portion that pertains to geographic mobility. Two census questions dealt with geographic mobility. Both questions were put to a one-third systematic sample of the population aged 15 and older as of 1 June 1971. From the sample responses, estimates of the relevant totals for the whole population were developed (see Appendix A). The first of the two census questions asked about the respondent's place of residence on 1 June 1966 (commonly called the "five-year migration question"). The second question elicited information on the number of times respondents had changed their municipality of residence between 1 June 1966 and 1 June 1971.

The 1971 Census has also been used to provide data on a variety of characteristics of both the migrants and the non-migrants. Some of these characteristics were described using data drawn from the 100% count of the population at the 1971 Census date (e.g., sex, age, marital status, and mother tongue). Other characteristics were described using information gleaned from the one-third sample (e.g., educational attainment and occupation). The scope of this study does not permit a discussion of the quality of these data. (See Norland, *et al.,* 1977.)

However, some discussion of the quality of the migration data is given in Appendix A. In general, the amount of data quality evaluation that is feasible suggests that the data are reasonably reliable for the type of use to which they are put in this study. The statistics are faulty, but informed and judicious use of them permits the derivation of useful information about certain broad features of Canadian mobility. The text therefore emphasizes the general magnitudes of aggregates, and of differences, and the broad systematic patterns of variation among the numbers in any table. By emphasizing these aspects of the data rather than the exact values of the numbers shown, the reader will concentrate on information in which distortions due to data errors are minimized.

1.2.3. Universe of Observation

Only a portion of the population aged five and over on 1 June 1971 (referred to as the "reporting population" in this text) is relevant to the present concerns. As a study on internal migration (measured on the basis of the five-year migration question), it should and does exclude persons residing outside Canada on 1 June 1966. The study concentrates on the occupational distribution of migrants and places much emphasis on the mediating role of attained education. Thus, it is appropriate to exclude persons who had not worked in the calendar year preceding that of the 1971 Census, as well as those who attended school in the school year preceding the 1971 Census.

The universe has also been restricted with respect to sex and age. Family migration patterns might cause the data for wives and children to distort the statistical measures of association between area characteristics and the occupational pattern of migration. There are substantial difficulties with the quality of the mobility data with regard to the elderly. In the light of these considerations, the universe is further restricted to males aged 20 - 54 in 1971.

The group that remains after these exclusions is called the "sample population". The sample population for this study is thus comprised of males who resided in Canada on both 1 June 1966 and 1 June 1971, were out of school during the 1970 - 71 school year, had worked in 1970, and were aged 20 - 54. Actually, the analysis is focused on the inter-municipal migrants within this sample population. These migrants are classified into streams, groups of migrants going from one particular region-group to another. Each of the 42 selected streams is again simultaneously broken down into 120 categories of age, by schooling, and by occupation (see Appendix D).

1.3. Main Findings and Interpretations

The 1971 Census data indicate that the occupational distribution of a group of migrants in streams that involve metropolitan areas can be predicted with substantial accuracy through the use of information concerning the migrants' age composition and census data on characteristics of the regions of origin and destination of the migration.

In general, the pertinent regional characteristics include the occupational pattern of job opportunities in each region and community amenities that are attractive to particular educational groups. Since the census data do not provide direct measures of such opportunities and amenities, it is necessary to use proxy measures. For a particular region, the proxy measures used in this study are the educational and occupational composition of the region's non-migrants.

The industrial composition of economic activity and the technology of production in a region largely determine the kinds of occupation in which the region provides job opportunities. Differing growth rates among the industries cause job opportunities to expand rapidly in some occupations, to increase only modestly in others, and to decline in yet others. Such differences among occupations in the abundance of the job opportunities that they offer directly affect the occupational composition of groups of migrants going to and from that region. When there is a group of migrants moving from one specific area of origin to another region of destination, the differences between the two regions in the occupational compositions of their job opportunities (which are in turn affected by their differences in industry structure and pattern of industrial expansion) will help to determine the occupational composition of that group of migrants.

There is at least one other way in which the attributes of a region affect the occupational composition of the region's migrants. This way involves the educational composition of the migrants.

Regions differ in the amenities they offer to particular educational groups. For example, persons with university education will generally tend to find unattractive a community that fails to provide the diversity of recreational opportunities and intellectual stimulations that they expect. Also, persons seeking particular kinds of education will find some regions far more attractive than others, due to regional differences in the availability of certain types of educational opportunities. The amenities and educational opportunities offered by a region affect the educational composition of its in-migrants and out-migrants. Therefore, because of the marked association between the educational composition of a group and its occupational structure, the regional attributes again help to influence the occupational pattern of the migration.

The 1971 Census data indicate, however, that one would not achieve very high accuracy in predicting the occupational composition of a group of migrants only on the basis of information about the economies and amenities of the communities of origin and destination of those migrants. This situation arises because of another aspect of internal migration in Canada – a group of migrants tends, in a significant degree, to have certain distinctive features regardless of the peculiarities of the regions of origin and destination. For example, migrants tend as a group to be much younger and better educated than the non-migrant adult population (see Stone, 1978).

In Canada, the educational composition of a group of persons is markedly associated with its age composition (e.g., relatively high proportions with university training are at present most likely to be found among young adults). The occupational composition of the group is in turn related to its educational composition. Thus, the persistent age pattern of migration, which is relatively insensitive to the characteristics of the communities of origin and destination of the migrants, tends to generate certain occupational patterns that cannot be explained by these regional characteristics.

It is also likely that educational groups differ in their mobility propensity. The desire to enjoy new experiences and opportunities through a process of moving from one community to another is more heightened in some educational groups than in others. ("Footlooseness" is particularly noticeable among persons with higher education, for example.) As a result, one would find that, even after age is taken into account, some educational groups are more mobile than others. Such group differences in mobility tend to a significant degree to persist across a variety of origins and destinations of the migrants. Again, because of the association between the educational and occupational compositions of a group, these persistent educational differences in mobility would lead to migrant occupational patterns that are somewhat insensitive to characteristics of the regions of origin and destination of the migration. The net effect would be to lower the accuracy with which one could predict the occupational composition of a group of migrants using only information about characteristics of their regions of origin and destination.

Thus, while the economies and amenities of regions are important influences upon the occupational composition of migration to and from such regions, distinctive features of migration that tend to persist across a variety of origins and destinations decrease the accuracy of predicting migrant occupational composition from information about the regions. The census data do indicate, however, that when we combine the information about regions with data on the persistent features of migration (e.g., the age composition), we can achieve considerable accuracy in predicting the occupational composition of a group of migrants, provided that a metropolitan area is at the origin or destination of the migration. The materials presented below in this study show (in terms of specific mathematical formulas and statistical procedures) how to put such kinds of information together, within the limitations of census data. This report also indicates concrete empirical results of the application of such formulas and procedures to 1971 Census data for selected groupings of Canadian regions. Some of the concrete empirical results will now be summarized.

Lacking any knowledge about the economies and amenities of the regions to and from which migrants move, or about the persistent features in the composition of migrants, one might predict that the occupational structure of the migrants will be that of the universe of adults to which the migrants belong. (In this study, the universe of adults consisted of males aged 20 - 54 in 1971, who resided in Canada on 1 June 1966, were out of school and had worked at some time during 1970.) Such a prediction would be largely inaccurate. The 1971 Census data indicate that an approximate 70% improvement in the accuracy of prediction can be expected if one makes use of information about the occupational structures of the non-migrants in the regions of origin and destination, and about the educational structure of the migration, provided that metropolitan areas are involved. (The percentage improvement in the accuracy of predicting the occupational pattern of the migration varies from one area of origin and destination to another, and depends upon the specific ways in which education and occupation are broken down into categories.)

We also find that the accuracy of prediction of the educational composition of a group of migrants can be improved greatly by taking into account the age composition of the migrants and educational patterns of non-migrants in their regions of origin and destination. Thus, by taking into account the age composition of a group of migrants and the educational and occupational patterns of the non-migrants in the group's regions of origin and destination, we can substantially improve the accuracy of predicting the occupational composition of the migration.

Metropolitan areas in Canada tend to have significantly different economies than non-metropolitan areas. This difference affects the occupational pattern of migration from metropolitan to non-metropolitan areas, or vice versa. The 1971 Census data indicate that the metropolitan area characteristics tend to dominate the occupational pattern of the migration, relative to the non-metropolitan area attributes. Two underlying factors may be at work here. Firstly, many of the migrants from metropolitan region-groups may have continued to work in the former areas. Secondly, many of those who took up work in the non-metropolitan areas may have been attracted to opportunities similar to those in metropolitan areas (e.g., there has been a heavy flow of skilled engineers into the James Bay development area). In any event, whenever we are dealing with a group of migrants moving between metropolitan and non-metropolitan areas, it is the former that will be most pertinent to the explanation of the occupational composition of the flow.

Canadian metropolitan areas tend to resemble each other markedly in their industrial structures, at least in terms of broad groupings of industries. This resemblance produces significant similarities among the areas in the occupational compositions of their job opportunities. Therefore, when considering migration from one metropolitan area to another, it should be assumed that both the region of origin and that of destination will tend to exert a similar pattern of influences on the occupational composition of the migration. The 1971 Census data indicate that those regional influences work mostly through the impact of regions on the educational pattern of the inter-metropolitan migration. After this impact is taken into account, the features of the metropolitan areas contribute little to improving the accuracy of prediction of the occupational composition of migration among metropolitan areas.

The concentration of facilities for post-secondary education in metropolitan areas has special significance in explaining the pattern of educational attainment among migrants to and from such areas. As noted in an earlier census monograph (Stone, 1969, Chapter 4), migration streams going into or out of metropolitan areas show higher than average levels of educational attainment, even for age groups too old to have had much school attendance over the 1966 - 71 period. Migration streams flowing among regions with relatively low levels of urbanization show much lower than the average (among migrants) levels of educational attainment.

Further research is needed concerning the role of the metropolitan areas in moulding the educational compositions of Canadian migration streams. That these areas are the location of institutions of higher education, as well as of the businesses and industries that supply the most abundant opportunities to highly skilled or educated people, is clearly relevant. But these areas also seem to supply highly skilled or educated talent to non-metropolitan communities even for age groups beyond the ones that would contain many recent university graduates. Perhaps the opportunities that attract migrants to the non-metropolitan communities are such that they appeal to people who, as a whole, are significantly better educated than the non-migrants of these areas. Also, many of the out-migrants from Census Metropolitan Areas to other areas may actually live quite near to the CMA's and continue working in them.

It is difficult, nevertheless, to deny the relevance of regional differences in the occupational composition of labour demand to the educational composition of migration. Generally, the more sophisticated the industrial and distributive economy of a region, the greater is the relative share of jobs that require high levels of education. A region with a highly sophisticated economy also presents a mix of attractions that is especially appealing to relatively highly educated potential migrants. Thus, metropolitan areas, the regions with the most sophisticated economies, show migration flows with much higher than average overall levels of educational attainment.

The analytical models proved to be seriously deficient when we attempted to predict the schooling and occupational compositions of migrants moving from one non-metropolitan area to another. For this type of stream, the central working hypothesis of this paper may need significant modification. Unfortunately, defects in the census data make it difficult to decide readily if the problem is in the theoretical framework rather than in the data (and perhaps the methodology). The general situation is summarized in Sections 4.1.2 and 4.2.2. These summaries suggest the following hypotheses that may stimulate further research.

Regional "requirements" for certain levels of education in the working force pertain mainly to the higher levels of education. Sensitivity to community amenities is marked mainly among the more highly educated groups, and for these groups, the pertinent amenities are concentrated in and near metropolitan areas. Metropolitan areas are the major sources of domestic supply of highly educated persons, and they provide the most abundant employment and other opportunities to such persons. It is, therefore, mainly in the migration streams to and from metropolitan areas that a substantial impact of regional factors on the educational composition of a migration stream may be expected. Among other migration streams, the basic tendency for certain educational groups to be more mobile than others will tend to produce a typical educational profile of migration that does not vary much from one stream to another. We recommend that the foregoing set of hypotheses be subjected to theoretical scrutiny and empirical testing in other studies.

For the occupational composition of migration streams among non-metropolitan areas, no typical occupational pattern of migration can be cited. For example, the streams considered showed substantially higher than average (for all migrants) concentration in the professional groups of teaching and related occupations, and technological, social, artistic, religious and related occupations. The underlying data give no help in making sense out of these concentrations in terms of the overall occupational structures of the areas of origin and destination. We can only suggest that either the proxy measures of regional effects are unduly inappropriate (the issue of inappropriate proxy measures is discussed further in Chapter 4) or the central working hypothesis proves once again to be inapplicable to the data.

The critical hurdle here is the difficulty of bringing census data because of its confinement of the measurement of population characteristics to the end of the migration period for which migration is measured to bear effectively in resolving the dilemma. It could be, for example, that the proxy measures need to focus more on the occupational structure of job opportunities (as suggested in the theoretical rationale presented above) and less on the occupational structure of jobs held. But census data do not permit such proxy measures to be used.

The findings of this study are relevant to the concerns expressed by citizens and officials about the problem of achieving certain occupational mixes in the labour force that migrates to particular regions. We have noted, for example, the interest of persons in Northwestern Ontario in the role that migration can play in securing a labour supply with particular occupational skills for that region. By merging the findings of this study with those of others, we may draw some lessons. (For pertinent review articles that contain relevant bibliographies, the reader should see Stone, 1974a; Stone, 1976; Greenwood, 1975; and Shaw, 1975.) Selected lessons may be summarized as follows:

1. There are certain repetitive features in the age and educational patterns of geographic mobility that seem unrelated to regional characteristics, for reasons we do not yet understand very well (see Stone, 1978).

2. Prediction of the occupational composition of migrants into and out of a region can be improved considerably by taking into account the expected age distribution of the migrants and the educational and occupational compositions of the region's non-migrants if the streams have metropolitan areas as origins or destinations.

3. Whenever migration is taking place between a highly urbanized region and a largely rural one, the characteristics of the highly urbanized region will predominate in influencing the occupational composition of the migration, regardless of the direction of the migration.

It is extremely important to understand that none of these lessons can be considered proven by any one study, let alone one which is hampered by defective data. Only an accumulation of studies, each of which is not in itself definitive, can lend force to the lessons here suggested. The lessons are always subject to continued testing and possible reversal in future research.

1.4. Organization of Discussion

The reader who does not care to pursue the technical aspects of the development of the empirical analysis and its theoretical rationale should terminate detailed review of this report at this point.

The contents of the following chapters may be introduced as follows:

Chapter 2 prepares for the subsequent analysis by introducing the set of region-groups chosen for use. Systematic variations among the groups in terms of population size, population distribution, population composition, and economic structure are mentioned briefly. Some of the main features of the migration streams among the selected region-groups are presented, and prominent features of the pattern of differences among streams with respect to volume and to composition by education and occupation are indicated. The chapter is intended to provide the reader with an overall view of the variations which a later chapter analyses. Chapter 2 is not meant to be a treatise on the characteristics of the region-groups, as this is not a study on the regional geography of Canada.

Chapter 3 completes the preparation for the statistical analysis. A theoretical framework for subsequent interpretation of the data is presented. The theoretical discussion in this chapter naturally concentrates around the pivotal question: "Why is it reasonable to expect a relationship between the occupational composition of a migration stream and characteristics of its regions of origin and destination? ". In addition, to avoid discussions of a highly technical nature in Chapter 4, the general strategy of the explanatory analysis used there is outlined in Chapter 3. Readers who are especially keen to grasp the explanatory analysis strategy adequately should read Appendix C before proceeding to Chapter 4.

Chapter 4 presents the results of an analysis of the educational and occupational compositions of a migration stream. In this analysis, age and proxies for the characteristics of the places of origin and destination are the explanatory variables chosen to account for the educational composition of migration. The latter is viewed as a major intervening variable between regional characteristics and the occupational composition of a stream. Then, the chapter deals with the relative statistical contributions of schooling and characteristics of the regions of origin and destination to the shape of the occupational distribution of a migration astream.

The remainder of the text consists of brief concluding remarks and five appendices. To keep the size of this volume small, detailed technical discussion is avoided even in the appendices. In some cases, further relevant detailed discussion is contained in working papers prepared by the author.[1]

See footnote(s) on page 24.

FOOTNOTES

[1] These papers are cited and are available at photocopying cost (depending on the volume of the requests) by writing to the author. Also, special tabulations or background tables that are cited in this work, but not included here, are available (at cost of reproduction where necessary) by writing to the author. Costs will be payable to the Receiver General for Canada.

CHAPTER 2

PATTERNS OF VARIATION AMONG THE REGION-GROUPS AND THEIR MIGRATION STREAMS

2.1. Introducing the Region-groups

This chapter introduces the areal units used to define migration streams for this study. These areal units are called "region-groups". Selected characteristics of the region-groups, as well as of the migration streams among them, are discussed. The review of the characteristics of the chosen region-groups is brief because this study focuses on the occupational composition of geographically mobile Canadians and not on Canadian geography. The regions of origin and destination for migration streams are merely a subset of the collection of variables selected for use in analysing the composition of migration. Characteristics of the regions form only part of the battery of specified explanatory variables in an analysis that is not intended to focus on the natures of those regions.

What regional features are pertinent? To answer this question, it is helpful to consider what one means when designating a variable as being explanatory in an analysis. An explanatory variable is generally intended to reflect specific aspects of an object that are thought to be involved in influencing the behaviour of the subject of analysis. For example, if smoking is a subject of interest in an analysis of lung cancer, we would scarcely be interested in measuring the odour of smoke, because the odour of smoke is not thought to be involved in the influence that smoking has on lung cancer.

Similarly, given the chosen focus of this study, one is interested only in the aspects of the regional units that are thought to be involved in the influence that such units have upon the occupational composition of migration flows. The weight of evidence and testimony in the field of regional science suggest that the occupational distribution of the labour force of a region is closely related to the industrial structure of the region (see Denton, 1966, p. 13; Ostry, 1967, p. 1; Manpower and Immigration, 1975, pp. 23 and 24; and Meltz, 1969, p. 15). We hypothesize that variation in industrial structure is a major source of regional influence on the occupational composition of migration streams.

Thus, in highlighting aspects of the selected region-groups in this chapter, the focus is on industrial structure variation. This study hypothesizes that Canadian regions (using the word "region" in a broad sense to mean any defined geographic area) can be grouped into a relatively few classes based upon their industry structures (cf. Maxwell, 1965, p. 95; Ostry, 1968, Chapter 2; Brewis, 1969, p. 51; Camu, Weeks and Sametz, 1964, Chapter 10). These types range from highly diversified metropolitan areas, to medium-sized cities (outside metropolitan areas) that tend to specialize in limited types of manufacturing or service, to rural regions that are marked by heavy specialization in raw material extraction (mining, forestry, fishing, farming, etc.).

The above hypothesis is not tested in this study because we did not have the necessary resources. (The gathering and classification of industrial structure data for a large number of regions is an expensive and time-consuming process.) The hypothesis is therefore presented as a working assumption.

The region-groups are based upon a relatively new concept of the spatial structure of the Canadian economy (cf. Simmons, 1975; Ray, 1972). No single perspective on the spatial structure of the Canadian economy has the distinction of being the only correct one. Some perspectives emphasize the east-west and north-south dimensions with careful attention to the differing settlement histories of Lower Canada, Upper Canada, the Prairies, and the area in and beyond the Rockies. Another perspective focuses on the major political jurisdictions within Canada, especially the provinces. Yet another concentrates on areas that specialize in particular types of raw material extraction or processing.

A more recently articulated viewpoint, the one we emphasize in this study, recognizes a system dominated by a few large metropolitan areas. Around these areas, other major urban centres revolve like planets around their suns. The planets themselves may have one or more moons (small population centres) in their orbits of influence. This network of urban centres plays an important part in directing and co-ordinating the distribution of economic opportunities among the parts of a rural hinterland in which various kinds of raw material extraction take place (cf. Simmons, 1974, p. 6; Brewis, 1969, p. 58; Richardson, 1973, pp. 73 - 78; and Ray, 1972, p. 54).

Given this concept of an urban-centred system of economic control and activity, it seems reasonable for some purposes to define groups of regional units whose parts are not necessarily geographically contiguous. Instead, the criterion of grouping is economic contiguity.

2.2. Economic Contiguity and Delineation Criteria

What is meant by "economic contiguity"? To define this term precisely is difficult without a series of arbitrary stipulations, but its meaning is suggested by considering the following statement: the degree of economic contiguity between Montréal and Toronto far exceeds that between Montréal and Northern Saskatchewan. The web of daily economic interactions and the similarity of overall economic structure are much greater between the former pair of regions than between the latter. (Whether Montréal and Toronto are economically contiguous depends on the criteria and measures established for that type of contiguity; an elaborate discussion of such criteria and measures does not need to be done for the purposes of this study.)

Within each of the region-groups defined for use in this study is a collection of regions that enjoy a significant degree of economic contiguity. Measured in terms of regional economic interactions (see Simmons, 1975; Stone 1976; and

Ricour-Singh, 1977 for discussion of aspects of this concept in the context of relations among Canadian cities) and similarities in industry structures, the economic contiguity within the chosen region-groups is much greater than it is between them. The region-groups do not necessarily follow provincial boundaries.

In opting for this relatively unused concept of areal grouping, the author is not able to relate clearly the results of the analysis to more traditional regions. Those who consider it essential that the analysis go forward in terms of the traditional regions with which people are familiar will not be satisfied. We contend, however, that modern Canada cannot be adequately understood without the employment of a variety of regional perspectives (see Caves and Holton, 1961, p. 144; and Brewis, 1969, p. 45).

In delineating the region-groups for use in this study, great emphasis has been placed on the attainment of metropolitan or large city status. The region-groups range from areas with low levels of city development to those that are entirely comprised by Canadian Census Metropolitan Areas. The effect is to emphasize the urban dimension of regionalization rather than the political or the historical ones (Upper Canada, Lower Canada, the West, etc.). As a result, the reader gets a study on aspects of the composition of migration flows among region-groups delineated largely on urbanization criteria. It is assumed that by using such criteria, we will capture the main differences among region-groups (especially industry-structure differences) that help us to understand the occupational composition of migration streams.

In summary, this study uses an unusual concept in the grouping of migrants in terms of their areas of origin and destination. The territory that is considered as forming a region need not be comprised by a set of areas that touch each other (contiguous areas) in geographic space. More importance is given to the relative similarity of the areas in characteristics that are deemed, on the basis of existing theory and past research, to be especially relevant to the explanation of the occupational composition of geographic mobility. A number of the defined region-groups of origin and destination in this study are comprised by areal units that do not touch each other in geographic space – e.g., the group of metropolitan areas in Ontario and Quebec.

2.3. The Region-groups

Seven region-groups serve as the origins and destinations of the migration streams for this study.[1] Variations in urban settlement pattern and related economic structure were emphasized in defining these groups. The seven region-groups are defined as follows:

1. **Atlantic CMA's** – The Census Metropolitan Areas of Halifax, Saint John (N.B.), and St. John's (Nfld.).

See footnote(s) on page 47.

2. **Atlantic Non-CMA Area** – Parts of Newfoundland, Nova Scotia, and New Brunswick outside the Atlantic CMA's, and all of Prince Edward Island.
3. **Central CMA's** – The Census Metropolitan Areas of Chicoutimi-Jonquière, Hamilton, Kitchener, London, Montréal, Ottawa-Hull, Québec, St. Catharines-Niagara, Sudbury, Thunder Bay, Toronto, and Windsor.[2]
4. **Central CA's of 50,000 and Over** – The Census Agglomerations of Brantford, Guelph, Kingston, Oshawa, Peterborough, Sarnia, Sault Ste. Marie, Shawinigan, Sherbrooke, and Trois-Rivières.
5. **Central Predominantly Rural Area** – Parts of Ontario and Quebec outside of the Central CMA's and Central CA's of 50,000 and Over.
6. **Western CMA's** – The Census Metropolitan Areas of Calgary, Edmonton, Regina, Saskatoon, Vancouver, Victoria, and Winnipeg.
7. **Western Non-CMA Area** – Parts of Manitoba, Saskatchewan, Alberta, and British Columbia outside the Western CMA's.

Tables 2.1 - 2.3 provide summaries of population totals, urban population data, and industry composition data for the region-groups.

In 1971, the region-groups varied in population size from 461,000 (Atlantic CMA's) to over 8 million (Central CMA's). Four of the other five region-groups exceeded a million in 1971 population (Atlantic Non-CMA Area, Central Predominantly Rural Area, Western CMA's, and Western Non-CMA Area). The remaining region-group (Central CA's of 50,000 and Over) had slightly more than 800,000 persons in 1971.

Table 2.1 shows that among these populations, the fastest growth between 1966 and 1971 was attained by the three CMA region-groups. The Central and Western CMA region-groups each increased by an average annual growth rate of more than 2% between 1966 and 1971. Average annual growth rates of less than 1% were shown by the three regions with relatively low urbanization. The Central CA's of 50,000 and Over increased by an average annual rate of just above 1%. In short, there were substantial growth rate differences among the seven region-groups.

The level of urbanization varied strongly among the region-groups. As expected, very high percentages of the populations of the three CMA region-groups resided in urban centres with a minimum of 20,000 people. Table 2.1 shows that the percentage for Atlantic CMA's (82%) was far below that of the other two CMA region-groups. By reason of the census definition for Census Agglomeration, the Central CA's of 50,000 and Over had very large proportions of their population in urban centres of 20,000 and over. In both 1966 and 1971, over 92% of the population of the Central CA's of 50,000 and Over resided in one of the continuously built-up urban centres.

See footnote(s) on page 47.

TABLE 2.1. Population, Percentage of Population in Urban Centres of 20,000 and Over, and Population Size of Largest Urban Complex, for Seven Region-groups, Canada, 1966 and 1971

Region-group	Population			Percentage in urban centres of 20,000 and over		Size of largest urban complex	
	1966	1971	Growth rate[1] 1966-71	1966	1971	1966	1971
	thousands			per cent		thousands	
Atlantic CMA's[2]	432	461	6.8	84.2	82.3	184	190
Atlantic Non-CMA Area[3]	1,543	1,596	3.4	15.9	16.0	58	63
Central CMA's[4]	7,631	8,429	10.4	91.1	90.4	2,396	2,537
Central CA's of 50,000 and Over[5]	766	812	6.0	92.8	92.7	98	111
Central Predominantly Rural Area[6]	4,344	4,490	3.4	17.4	17.6	46	49
Western CMA's[7]	2,621	2,985	13.9	93.2	91.9	824	926
Western Non-CMA Area[8]	2,634	2,742	4.1	11.8	12.8	37	42

[1] The growth rate is defined as 100 x [(1971 population/1966 population) - 1].
[2] Atlantic CMA's – Halifax, Saint John (N.B.), and St. John's (Nfld.).
[3] Atlantic Non-CMA Area – Parts of Newfoundland, Nova Scotia, and New Brunswick outside the Atlantic CMA's and all of Prince Edward Island.
[4] Central CMA's – Chicoutimi - Jonquière, Hamilton, Kitchener, London, Montréal, Ottawa - Hull, Québec, St. Catharines - Niagara, Sudbury, Thunder Bay, Toronto, and Windsor.
[5] Central CA's of 50,000 and Over – Brantford, Guelph, Kingston, Oshawa, Peterborough, Sarnia, Sault Ste. Marie, Shawinigan, Sherbrooke, and Trois-Rivières.
[6] Central Predominantly Rural Area – Parts of Ontario and Quebec outside of Central CMA's and Central CA's of 50,000 and Over.
[7] Western CMA's – Calgary, Edmonton, Regina, Saskatoon, Vancouver, Victoria, and Winnipeg.
[8] Western Non-CMA Area – Parts of Manitoba, Saskatchewan, Alberta, and British Columbia outside the Western CMA's.

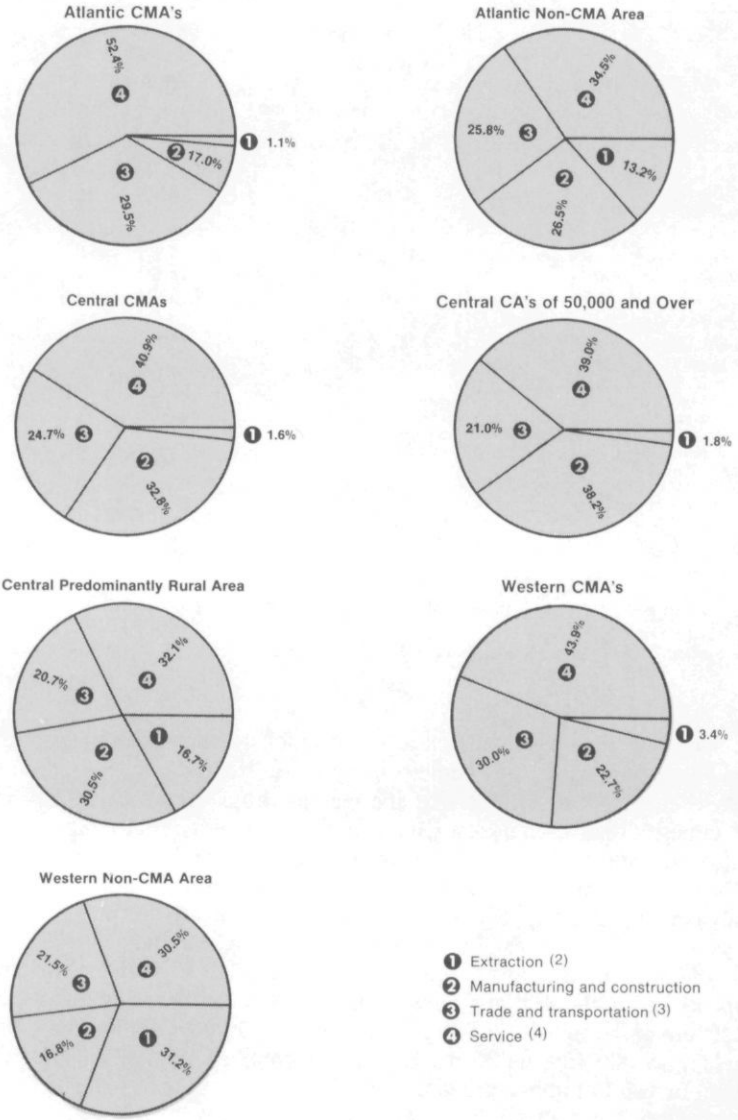
Source: 1971 Census of Canada, Bulletin 1.1 - 2, Table 1; and Bulletin 1.1 - 8, Tables 7, 8 and 9.

Defining an urban complex as a continuously built-up urbanized area which often extends beyond the municipal boundaries of the largest incorporated city within the area, it is notable that the region-groups vary strongly in the sizes of their largest urban complexes. The size of the largest urban complex is a crude index of an aspect of the character of the urban settlement pattern, especially the city size distribution, within a region-group. The urban settlement pattern is usually an important aspect of the economy of a region. As expected, the CMA groups showed the largest big city sizes. The largest urban complex in Central CA's of 50,000 and Over was substantial (111,000 for the Oshawa urban complex). Even the region-groups with relatively low levels of urbanization had largest urban complexes of significant size (63,000 for the Moncton urban complex, 49,000 for the North Bay urban complex, and 42,000 for the Prince George, British Columbia, urban complex).

Industry structure varied a great deal among the seven region-groups. For example, Chart 2.1 shows that the percentage of the labour force in manufacturing and construction industries ranged from a low of 16.8 for the Western Non-CMA Area to a high of 38.2 for the Central CA's of 50,000 and Over. One region-group showed more than 30% of the 1971 labour force in the extractive

Chart — 2.1

Industrial Distributions, Labour Force Aged 15 Years and Over in 1971 for Seven Region-groups (1), Canada, 1971



(1) See Table 2.1, footnotes 7 to 8, for precise definition of each region-group.
(2) Includes agriculture, forestry, fishing and trapping, mines (including milling), quarries and oil well industries.
(3) Includes transportation, communication and other utilities, and trade industries.
(4) Includes finance, insurance and real estate, community, business and personal services, public administration and defence.

Sources: 1971 Census of Canada, Bul. 3.4-4, Bul. 3.4-5, and Bul. 3.4-6.

industries (Western Non-CMA Area), whereas three other region-groups showed less than 2% of the labour force in extraction (Atlantic CMA's, Central CMA's and Central CA's of 50,000 and Over).

Large urban centres in North America tend to perform a large portion of business and other services for other regions (cf. Duncan, *et al.*, 1960, Part III; Simmons, 1974; and Stone, 1967, pp. 187 - 196). This tendency is often reflected in their markedly higher than average concentrations of work force in the industries from which such services are usually dispensed. The pertinent industry groups of Table 2.2 would be 1. finance, insurance and real estate; and 2. community, business and personal services. Table 2.2 shows the expected higher than average concentrations for each of the three CMA region-groups as well as for Central CA's of 50,000 and Over.

TABLE 2.2. Industrial Distributions, Labour Force Aged 15 Years and Over in 1971, for Seven Region-groups,[1] Canada, 1971

Industry[2]	Region-group						
	Atlantic CMA's	Atlantic Non-CMA Area	Central CMA's	Central CA's of 50,000 and Over	Central Predominantly Rural Area	Western CMA's	Western Non-CMA Area
	per cent						
Total labour force	100.0[3]	100.0	100.0	100.0	100.0	100.0	100.0
Agriculture	0.4	4.2	0.8	1.2	11.7	1.2	24.6
Forestry	0.1	2.6	0.1	0.2	1.9	0.4	2.7
Fishing and trapping	0.3	3.5	--	--	0.2	0.2	0.3
Mines (including milling), quarries and oil wells	0.2	2.9	0.7	0.3	2.9	1.7	3.6
Manufacturing	10.6	17.2	26.6	32.4	23.4	15.3	10.1
Construction	6.4	9.3	6.2	5.8	7.1	7.4	6.7
Transportation, communication and other utilities	10.8	10.0	8.2	5.9	7.0	10.7	8.0
Trade	18.7	15.8	16.5	15.1	13.6	19.3	13.6
Finance, insurance and real estate	4.8	2.1	6.0	3.5	2.5	5.8	2.4
Community, business and personal services	30.2	22.4	26.7	29.0	23.0	29.5	21.4
Public administration and defence	17.4	9.9	8.2	6.5	6.6	8.6	6.7

[1] See Table 2.1, footnotes 2 - 8, for a precise definition of each region-group.
[2] Excludes industry unspecified or undefined.
[3] Figures may not add to the total due to rounding error.
-- Means less than 0.05.

Source: 1971 Census of Canada, Bulletin 3.4 - 4, Table 3 A; Bulletin 3.4 - 5, Table 5; and Bulletin 3.4 - 6, Table 7.

Table 2.3 shows the scores of a special index that summarizes the degree of difference between the industry-group distributions of a pair of region-groups. This index is called a "dissimilarity index". Its values are computed from data on which Table 2.2 is based using the same industry groups as shown in that table (the computed value of the index in any given instance is sensitive to the choice of industry groups). The index score for a given pair of regions measures the percentage of one region's labour force that would have to be redistributed among the given industry groups in order that both regions might have the same industry-group distribution. A value of 10% in the dissimilarity index should be considered substantively significant.[3]

TABLE 2.3. Dissimilarity Indices[1] Among Industrial Distributions, Male Labour Force Aged 15 Years and Over in 1971, for Seven Region-groups,[2] Canada, 1971

Region-group	Region-group						
	Atlantic CMA's	Atlantic Non-CMA Area	Central CMA's	Central CA's of 50,000 and Over	Central Predominantly Rural Area	Western CMA's	Western Non-CMA Area
	per cent						
Atlantic CMA's	–	24.3	18.4	27.3	31.0	12.1	34.6
Atlantic Non-CMA Area	24.3	–	23.1	28.3	16.3	16.3	23.4
Central CMA's	18.4	23.1	–	10.5	18.5	11.7	34.1
Central CA's of 50,000 and Over	27.3	28.3	10.5	–	20.3	21.3	36.6
Central Predominantly Rural Area	31.0	16.3	18.5	20.3	–	22.5	17.3
Western CMA's	12.1	16.3	11.7	21.3	22.5	–	31.8
Western Non-CMA Area	34.6	23.4	34.1	36.6	17.3	31.8	–

[1] The dissimilarity index is a measure of the differences between two regions in the industrial distribution of their labour forces. The dissimilarity index is:

$$d\,(i, j, b) = \tfrac{1}{2} \sum_{b=1}^{11} \left| P_j(b) - P_i(b) \right|,$$ where $|x|$ is the absolute value of x, $P_j(b)$

is the per cent of the labour force of region j that is in industry b, and $i \neq j$. Roughly speaking, the index score is the percentage of one region's work force that would have to be redistributed among the selected industry groups in order for the two regions to have the same distribution. In this table the industry groups are: (1) Agriculture; (2) Forestry; (3) Fishing and trapping; (4) Mines (including milling), quarries and oil wells; (5) Manufacturing; (6) Construction; (7) Transportation, communication and other utilities; (8) Trade; (9) Finance, insurance and real estate; (10) Community, business and personal services; and (11) Public administration and defence.

[2] See Table 2.1, footnotes 2 - 8, for a precise definition of each region-group.

Source: 1971 Census of Canada, Bulletin 3.4 - 4, Table 3 A; Bulletin 3.4 - 5, Table 5; and Bulletin 3.4 - 6, Table 7.

On the average, over one fifth of the labour force in the region-groups would have to be redistributed among the chosen industry groups in order that the region-groups might have the same 1971 industry-group distribution of the labour force. Generally, there is a marked resemblance among the industry-group distributions of the three CMA region-groups (average index value is 14%). Even higher resemblance is shown between the Central CMA's and the group of Central CA's of 50,000 and Over. The industry structure of the latter region-group differs markedly from those of Atlantic CMA's and Western CMA's. The average degree of resemblance among the industry-group distributions of the three region-groups with relatively low urbanization was less (19% dissimilarity index score) than that among the CMA region-groups. As expected, the average difference between the

See footnote(s) on page 47.

industry-group distribution of a highly urbanized region-group and that of a region-group with relatively low urbanization was striking, the average dissimilarity score being 27%.

In sum, the delineation of region-groups for the analysis in later chapters has emphasized level of urbanization and associated industrial structure variation. The data show that major variation has been tapped in the regional delineation, especially when one considers the dominance of certain industry groups in the structure of the total Canadian economy. Much greater and more systematic variation is shown among these seven region-groups than would be shown among Canadian provinces. The region-groups tend to fall into two broad groups – areas of very high urbanization and areas of relatively low urbanization. Owing to the small size and the pattern of geographic concentration of the Canadian population, two or more suitable region-groups at middle levels of urbanization were not identified for the purposes of an analysis that treats the entire composition of a migration stream as an object of explanation.

2.4. Internal Migration Streams Among Region-groups

To complete the introduction to the region-groups chosen, this section comments on the variation among the 1966 - 71 migration streams that flowed between pairs of region-groups. The relative size of the streams and their compositions with regard to schooling and occupation are reviewed very briefly. The data pertain to the sample population.

Although this study does not attempt to explain the volume of migration from one region to another, a brief comment on the relative sizes of the migration streams among the seven region-groups is worthwhile. Table 2.4 shows a very wide variation in size among the migration streams, with a range from 200 - 189,600. The latter figure lies on the diagonal of the table and pertains to the inter-municipal migrants whose 1966 and 1971 places of residence were both within the Central CMA's region-group.

If we concentrate on the off-diagonal figures in Table 2.4, data that pertain to the migrants from one region-group to another region-group, the range of stream sizes is from 200 - 63,000. Generally, the smallest streams were those moving from Atlantic Canada to Western Canada or vice versa. The largest streams from one region-group to another generally involved Central Canada region-groups, either as origins or as destinations. Only nine of the 42 streams from one region-group to another exceeded 10,000. Fourteen of the streams were comprised of less than 1,000 persons.

Table 2.5 shows summary data on the amount of interstream variation of the two dependent attributes in the explanatory analyses presented in Chapter 4. The key figures in this table are the coefficients of variation shown in the third column. Each coefficient is expressed as a percentage and is the ratio of the standard deviation to the mean. To calculate the mean for a category of one of the given attributes, the pertinent percentages for the 49 streams shown in Table 2.4 are summed and divided by 49. The coefficient of variation is a measure of the degree of variability among the different streams relative to this mean.

TABLE 2.4. Sizes of the Streams of Migration Among the Seven Region-groups,[1] Canada, 1966 - 71

(Males Aged 20 - 54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970)

Region-group of origin	Region-group of destination						
	Atlantic CMA's	Atlantic Non-CMA Area	Central CMA's	Central CA's of 50,000 and Over	Central Predom-inantly Rural Area	Western CMA's	Western Non-CMA Area
	Inter-municipal migration streams (to nearest hundred)						
Atlantic CMA's	3,900	5,800	4,000	300	700	1,100	500
Atlantic Non-CMA Area	7,000	28,000	11,100	800	2,300	1,700	1,700
Central CMA's	2,600	6,900	189,600	9,700	54,500	14,300	4,900
Central CA's of 50,000 and Over	200	600	11,100	4,300	10,300	1,200	500
Central Predominantly Rural Area	500	2,000	63,600	9,400	88,800	4,100	4,000
Western CMA's	500	800	10,700	700	1,900	57,200	38,900
Western Non-CMA Area	200	700	4,000	300	2,200	45,400	82,900
	Each stream as a percentage of total migration						
Atlantic CMA's	0.48	0.73	0.51	0.04	0.09	0.14	0.06
Atlantic Non-CMA Area	0.88	3.51	1.38	0.10	0.29	0.22	0.21
Central CMA's	0.33	0.87	23.74	1.21	6.82	1.79	0.61
Central CA's of 50,000 and Over	0.02	0.07	1.39	0.53	1.30	0.15	0.06
Central Predominantly Rural Area	0.06	0.25	7.96	1.18	11.12	0.52	0.50
Western CMA's	0.07	0.09	1.34	0.09	0.24	7.17	4.87
Western Non-CMA Area	0.03	0.09	0.50	0.03	0.27	5.69	10.38

[1] See Table 2.1, footnotes 2 - 8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished tabulation. (The sources that are unpublished tabulations may be obtained from the author at the cost of photocopying where necessary – cost payable to the Receiver General for Canada.)

Substantial variation among streams is shown in the schooling distribution. The coefficients of variation for the proportions of streams within the four schooling categories range from 20% - 45%, the latter pertaining to the proportion with a university degree. Substantial interstream variability is also shown in occupational distribution. The range in the coefficients of variation for the proportions of streams in the 15 occupation groups is from 26% - 95%.

TABLE 2.5. Means, Standard Deviations and Coefficients of Variation for the Distributions of Regional Migrants[1] With Respect to Level of Schooling and Occupation, Canada, 1966-71

(Males Aged 20-54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970)

Item	Mean proportion	Standard deviation of proportions	Coefficient of variation of proportions
		per cent	
Schooling:			
Less than Grade 12	55.0	11.2	20.3
Grades 12 and 13 and non-university	21.7	5.4	25.1
Some university	8.5	2.9	33.9
University degree	14.8	6.7	45.2
Occupation:			
Managerial, administrative and related occupations	7.4	3.5	48.0
Teaching and related occupations	4.6	2.3	49.6
Occupations in medicine and health	2.5	1.1	43.1
Technological, social, religious, artistic and related occupations	11.6	4.0	34.1
Clerical and related occupations	7.1	2.5	35.2
Selected sales occupations[2]	3.1	2.0	63.7
Other sales occupations	8.8	2.9	32.8
Service occupations (excluding Armed Forces)	8.8	2.5	28.0
Farming, horticultural and animal husbandry occupations	2.1	1.9	92.3
Other primary occupations[3]	3.9	3.7	95.2
Processing occupations	4.6	2.5	55.2
Machining and product fabricating, assembling and repairing occupations	13.2	4.0	30.0
Construction trades occupations	10.1	2.9	28.6
Transport equipment operating occupations	6.7	1.8	26.0
Occupations not elsewhere specified	5.4	1.7	32.2

[1] In this table, the data include intraregional migrants, i.e., inter-municipal migrants who resided in the same region both in 1966 and 1971.
[2] Includes technical salesmen and related advisers, commercial travellers, street vendors and door-to-door salesmen, newsboys, insurance salesmen and agents, and driver-salesmen.
[3] Includes fishing, hunting, trapping and related occupations, forestry and logging occupations, mining and quarrying including oil and gas occupations.

Source: 1971 Census of Canada, unpublished tabulation.

More detailed information on the pattern of variation in educational and occupational composition of the migrants by region-group is given in Charts 2.2 and 2.3. Chart 2.2 deals with the educational composition of migrants. Each panel of the chart pertains to one region-group, and it provides information concerning those migrants who resided in that region-group on 1 June 1966. (Unfortunately, the educational attainment was measured as at 1 June 1971, and many of the migrants may have improved their educational attainment outside their 1966 region-group of residence.)

Chart 2.2 shows the pattern of differences between the educational composition of the migrants who originated in a specific region-group and that of all migrants, regardless of their place of origin. For each level of schooling there is a histogram. If the group of migrants from a specific region-group had a higher proportion than did all migrants for a particular level of schooling, the histogram is shown upright. If the migrants originating in the specified region-group had a lower proportion at that level of schooling than all migrants, the histogram is shown inverted. No histogram is shown when the two proportions are identical.

The height of the histogram indicates the degree of the difference between the two proportions. For example, migrants originating in the Atlantic CMA's had a slightly higher than average concentration at the level of less than Grade 12 education, and this is shown in Chart 2.2 by a low upright histogram. Those same migrants had a much lower than average concentration at the level of Grades 12 and 13 (without university) education, and this is shown by a histogram that is high but inverted.

Generally, the migrants who resided in Census Metropolitan Areas in 1966 tended to have the highest proportions with university education. For example, compared to the average proportion with a university degree, the proportion for out-migrants from Central CMA's was 25% higher. The corresponding figure for the Western CMA's was 30% higher.

Among the seven regional out-migration groups, the proportions with less than Grade 12 education were highest for the out-migrants from the regions that had no urban complex of at least 100,000 population. The proportion with less than Grade 12 education among out-migrants from the Atlantic Non-CMA Area, for example, was more than 15% above the corresponding proportion for the whole sample (see Chart 2.2). A similar figure is shown for the out-migrants from the Central Predominantly Rural Area.

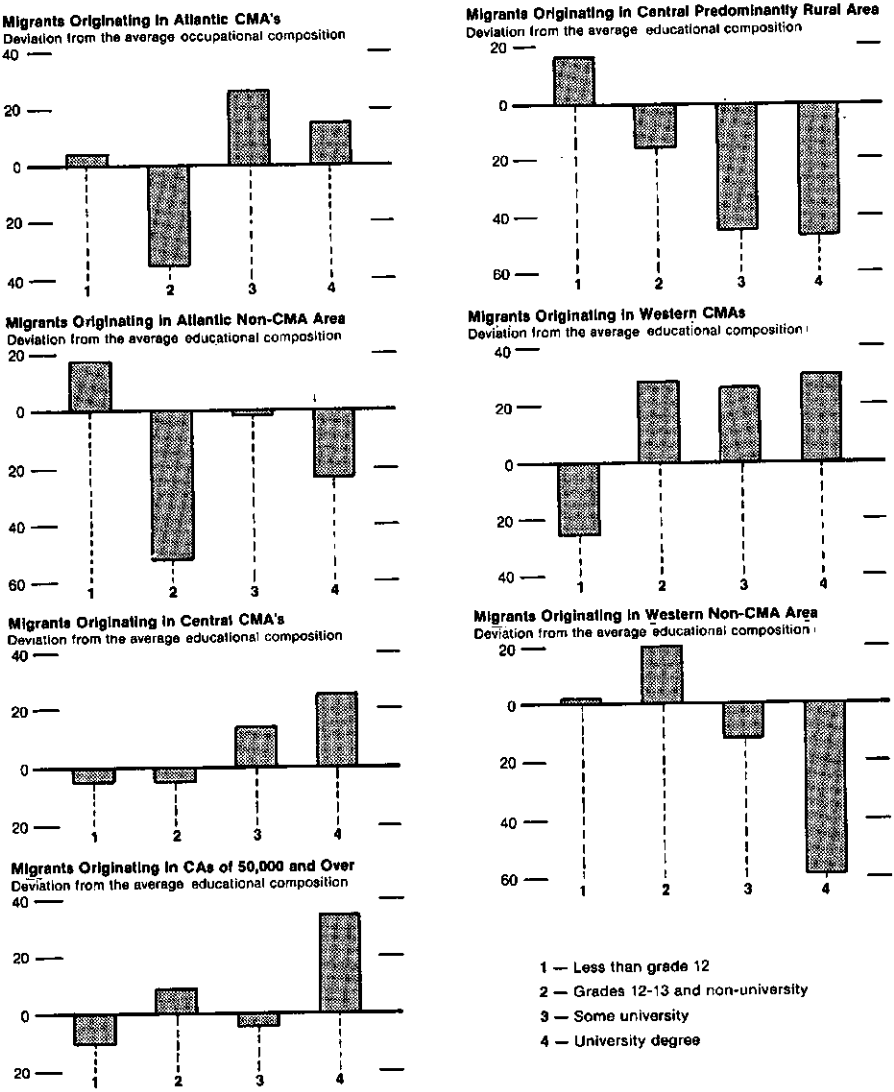
Table 2.6 shows data similar to those in Chart 2.2, but streams of migration are identified in the table. For example, the first panel of this table (the first seven rows) pertains only to the migrants whose destination was the Atlantic CMA's region-group. These seven rows break down the migrants to the Atlantic CMA's region-group according to their region-group of origin.

As in Chart 2.2, Table 2.6 shows ratios of proportions. The total out-migration from that single region of origin provides the average educational attainment distribution to which are compared schooling compositions of the seven individual streams leaving that region. For example, the figure of 1.09 in the top left-hand corner means that those migrants who went from one municipality to another but stayed within the Altantic CMA's were more heavily concentrated in the category of less than Grade 12 schooling than were the total out-migrants from Atlantic CMA's. Similarly, the figure of 1.93 in the seventh row shows that those who migrated from the Western Non-CMA Area to Atlantic CMA's had a much higher proportion with a university degree than did the total out-migrants from the Western Non-CMA Area.

Chart — 2.2

Relative Educational Compositions of Migration(1) by Region-Group(2) of Origin, Male Migrants, Canada, 1966-71

(Males aged 20-54 who resided in Canada on 1 June 1966, did not attend school in 1971, and worked in 1970)



(1) The relative educational compositions are ratios of proportions. Each ratio compares the proportion of migrants at a given level of schooling, for migrants originating in a particular region group, to the corresponding proportion for all migrants. The coefficients shown graphically on the chart are derived from the relative compositions. If a relative composition element (a ratio) is greater than or equal to one then the coefficient is 100 (value of the ratio -1). If the ratio is less than one then the coefficient is [100 (1 / value of the ratio -1)].

(2) See Table 2.1 footnotes 2 to 8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished data.

TABLE 2.6. Relative Educational Compositions[1] of Migration Streams, Among the Seven Region-groups,[2] Male Migrants, Canada, 1966 - 71

(Males Aged 20 - 54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970)

Region-group of origin	Level of schooling			
	Less than Grade 12	Grades 12 and 13 and non-university	Some university	University degree
	Region-group of destination, Atlantic CMA's			
Atlantic CMA's	1.09	0.92	0.92	0.76
Atlantic Non-CMA Area	0.89	1.13	1.37	1.27
Central CMA's	0.92	0.86	1.27	1.34
Central CA's of 50,000 and Over	0.92	0.84	1.11	1.50
Central Predominantly Rural Area	0.83	0.91	1.32	2.51
Western CMA's	0.78	0.88	1.07	1.87
Western Non-CMA Area	0.86	0.52	3.05	1.93
	Region-group of destination, Atlantic Non-CMA Area			
Atlantic CMA's	0.99	0.93	1.12	1.05
Atlantic Non-CMA Area	1.02	0.93	0.98	1.00
Central CMA's	1.27	0.51	0.81	0.79
Central CA's of 50,000 and Over	1.19	0.49	1.32	1.01
Central Predominantly Rural Area	1.01	0.81	1.15	1.30
Western CMA's	0.97	0.64	0.90	1.86
Western Non-CMA Area	1.06	0.72	1.07	1.50
	Region-group of destination, Central CMA's			
Atlantic CMA's	1.00	0.98	0.95	1.09
Atlantic Non-CMA Area	1.06	0.96	0.83	0.79
Central CMA's	1.00	0.99	1.01	1.01
Central CA's of 50,000 and Over	0.79	1.09	1.28	1.43
Central Predominantly Rural Area	0.89	1.14	1.33	1.35
Western CMA's	0.65	0.90	1.30	2.07
Western Non-CMA Area	0.73	1.02	1.28	2.80
	Region-group of destination, Central CA's of 50,000 and Over			
Atlantic CMA's	0.89	1.17	0.70	1.49
Atlantic Non-CMA Area	1.03	0.96	0.69	1.12
Central CMA's	0.75	1.28	0.98	1.54
Central CA's of 50,000 and Over	1.19	0.94	0.65	0.61
Central Predominantly Rural Area	0.88	1.34	1.02	1.18
Western CMA's	0.53	1.02	1.07	2.38
Western Non-CMA Area	0.77	0.98	0.86	3.10

See footnote(s) at end of table.

TABLE 2.6. Relative Educational Compositions[1] of Migration Streams, Among the Seven Region-groups,[2] Male Migrants, Canada, 1966-71 – Concluded

(Males Aged 20-54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970) – Concluded

Region-group of origin	Level of schooling			
	Less than Grade 12	Grades 12 and 13 and non-university	Some university	University degree
	Region-group of destination, Central Predominantly Rural Area			
Atlantic CMA's	1.07	1.22	0.46	0.78
Atlantic Non-CMA Area	1.00	0.97	0.82	1.18
Central CMA's	1.10	0.95	0.78	0.81
Central CA's of 50,000 and Over	1.18	0.92	0.74	0.65
Central Predominantly Rural Area	1.10	0.84	0.71	0.70
Western CMA's	1.12	0.79	0.74	1.20
Western Non-CMA Area	1.03	0.84	0.90	1.47
	Region-group of destination, Western CMA's			
Atlantic CMA's	0.80	1.42	1.17	1.27
Atlantic Non-CMA Area	0.82	1.53	1.33	1.21
Central CMA's	0.71	1.27	1.60	1.35
Central CA's of 50,000 and Over	0.69	1.30	1.63	1.29
Central Predominantly Rural Area	0.76	1.47	1.66	1.42
Western CMA's	0.98	1.09	1.01	0.88
Western Non-CMA Area	0.87	1.19	1.21	1.14
	Region-group of destination, Western Non-CMA Area			
Atlantic CMA's	0.89	1.34	1.12	0.98
Atlantic Non-CMA Area	0.97	1.33	0.91	0.74
Central CMA's	0.91	1.27	1.27	0.76
Central CA's of 50,000 and Over	0.88	1.06	1.35	1.14
Central Predominantly Rural Area	0.93	1.22	1.31	0.86
Western CMA's	1.13	0.92	0.91	0.82
Western Non-CMA Area	1.08	0.90	0.87	0.81

[1] The relative educational compositions are ratios of proportions. Each ratio compares the proportion of migrants at a given level of schooling for a particular migration stream to the corresponding proportion for all out-migrants from the region of origin of the given migration stream. Data in this table include migrants moving within a region.

[2] See Table 2.1, footnotes 2-8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished tabulation.

Table 2.6 shows some patterns of variation that are hidden in Chart 2.2. For example, Chart 2.2 shows that the out-migrants from the two sets of major urban areas in Central Canada (Central CMA's, and Central CA's of 50,000 and Over) had much higher than average proportions with a university degree. Table 2.6 shows that for each of these regions the proportion of out-migrants with a university degree varied considerably according to the destinations of the migrants.

Although all sample out-migrants from these Central Canada region-groups had much higher than average proportions with a university degree, the proportions were higher still for those streams moving towards Census Metropolitan Areas and away from Central Canada region-groups. In contrast, the concentration of the sample in university degree attainment was lower for those streams going toward regions that lacked major urban areas. For example, the proportion with a university degree among those moving to Atlantic CMA's from the Central Canada major urban areas was more than 34% higher than the corresponding proportion for all the sample out-migrants from the same region-groups. Similarly, the streams of migration from major urban areas of Central Canada to Western CMA's had proportions with university degree that are over 29% higher than the corresponding proportion for all sample out-migrants from the same major urban area. In sharp contrast, among those out-migrating from the Central CMA's and CA's of 50,000 and Over, the stream going to the part of Central Canada outside major urban areas had a proportion with university degree that was over 23% below the average. Table 2.7 shows that even after intraregion migrants are removed from the data, this pattern of variation just mentioned remains basically the same.

In sum, the data indicate that the regions with the most highly developed industrial economies and urban areas are the ones most likely to show migration streams with unusually high proportions of persons having university education. Regions lacking in urban complexes of at least 100,000 population tend to have migration streams with higher than average proportions of persons who did not graduate from high school. These generalizations apply both to the in-migration as well as the out-migration flows. Generally, in the sort of data generated from a census, the socio-economic composition of in-migration to and out-migration from a region tend to be markedly correlated (cf. Miller, 1967; Stone, 1971).

Regional variation in the occupational pattern of migration is shown in Chart 2.3 for the seven region-groups. This chart has the same sort of data as Chart 2.2, except that occupation is now the variable of interest, and the migrants are grouped according to their region-group of residence on 1 June 1971. The average occupational distribution against which the seven groups of in-migrants are being compared is that of the grand total of all inter-municipal migrants in the sample population. For example, the first histogram on the chart shows that in-migrants residing in the Atlantic CMA's on 1 June 1971 had a concentration in managerial, administrative and related occupations that was 21% higher than the average for all migrants. In contrast, the in-migrants to the Atlantic Non-CMA

TABLE 2.7. Relative Educational Compositions[1] of Migration Streams, Among the Seven Region-groups,[2] Male Interregional Migrants, Canada, 1966-71

(Males Aged 20-54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970)

Region-group of origin	Level of schooling: Less than Grade 12	Grades 12 and 13 and non-university	Some university	University degree
	Region-group of destination, Atlantic CMA's			
Atlantic CMA's	–	–	–	–
Atlantic Non-CMA Area	0.91	1.05	1.34	1.27
Central CMA's	0.92	0.85	1.31	1.37
Central CA's of 50,000 and Over	0.95	0.83	1.04	1.40
Central Predominantly Rural Area	0.93	0.78	1.00	1.90
Western CMA's	0.77	0.96	1.08	1.65
Western Non-CMA Area	0.99	0.45	2.53	1.49
	Region-group of destination, Atlantic Non-CMA Area			
Atlantic CMA's	1.02	0.90	1.10	0.98
Atlantic Non-CMA Area	–	–	–	–
Central CMA's	1.27	0.50	0.83	0.81
Central CA's of 50,000 and Over	1.23	0.49	1.24	0.95
Central Predominantly Rural Area	1.13	0.69	0.88	0.99
Western CMA's	0.95	0.71	0.91	1.64
Western Non-CMA Area	1.22	0.63	0.89	1.15
	Region-group of destination, Central CMA's			
Atlantic CMA's	1.02	0.95	0.93	1.01
Atlantic Non-CMA Area	1.08	0.89	0.81	0.79
Central CMA's	–	–	–	–
Central CA's of 50,000 and Over	0.82	1.08	1.20	1.34
Central Predominantly Rural Area	1.00	0.97	1.02	1.03
Western CMA's	0.64	0.99	1.32	1.83
Western Non-CMA Area	0.84	0.88	1.06	2.16
	Region-group of destination, Central CA's of 50,000 and Over			
Atlantic CMA's	0.92	1.14	0.69	1.39
Atlantic Non-CMA Area	1.05	0.89	0.68	1.12
Central CMA's	0.75	1.26	1.01	1.58
Central CA's of 50,000 and Over	–	–	–	–
Central Predominantly Rural Area	0.99	1.14	0.78	0.90
Western CMA's	0.52	1.12	1.08	2.10
Western Non-CMA Area	0.88	0.85	0.72	2.39

See footnote(s) at end of table.

TABLE 2.7. Relative Educational Compositions[1] of Migration Streams, Among the Seven Region-groups,[2] Male Interregional Migrants, Canada, 1966-71 – Concluded

(Males Aged 20-54 in 1971, Who Resided in Canada on 1 June 1966, Did Not Attend School in 1971, and Worked in 1970) – Concluded

Region-group of origin	Level of schooling			
	Less than Grade 12	Grades 12 and 13 and non-university	Some university	University degree
	Region-group of destination, Central Predominantly Rural Area			
Atlantic CMA's	1.10	1.18	0.45	0.72
Atlantic Non-CMA Area	1.02	0.90	0.80	1.18
Central CMA's	1.10	0.94	0.81	0.82
Central CA's of 50,000 and Over	1.22	0.91	0.69	0.60
Central Predominantly Rural Area	–	–	–	–
Western CMA's	1.10	0.87	0.75	1.06
Western Non-CMA Area	1.18	0.73	0.75	1.13
	Region-group of destination, Western CMA's			
Atlantic CMA's	0.82	1.39	1.15	1.18
Atlantic Non-CMA Area	0.83	1.42	1.30	1.21
Central CMA's	0.71	1.25	1.65	1.38
Central CA's of 50,000 and Over	0.71	1.29	1.53	1.21
Central Predominantly Rural Area	0.86	1.25	1.27	1.08
Western CMA's	–	–	–	–
Western Non-CMA Area	1.00	1.03	1.00	0.88
	Region-group of destination, Western Non-CMA Area			
Atlantic CMA's	0.91	1.31	1.10	0.91
Atlantic Non-CMA Area	0.99	1.23	0.89	0.74
Central CMA's	0.91	1.25	1.31	0.78
Central CA's of 50,000 and Over	0.91	1.05	1.27	1.06
Central Predominantly Rural Area	1.05	1.04	1.00	0.65
Western CMA's	1.11	1.01	0.92	0.72
Western Non-CMA Area	–	–	–	–

[1] See Table 2.6, footnote 1. Data in this table exclude migrants moving within a region.
[2] See Table 2.1, footnotes 2-8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished tabulation.

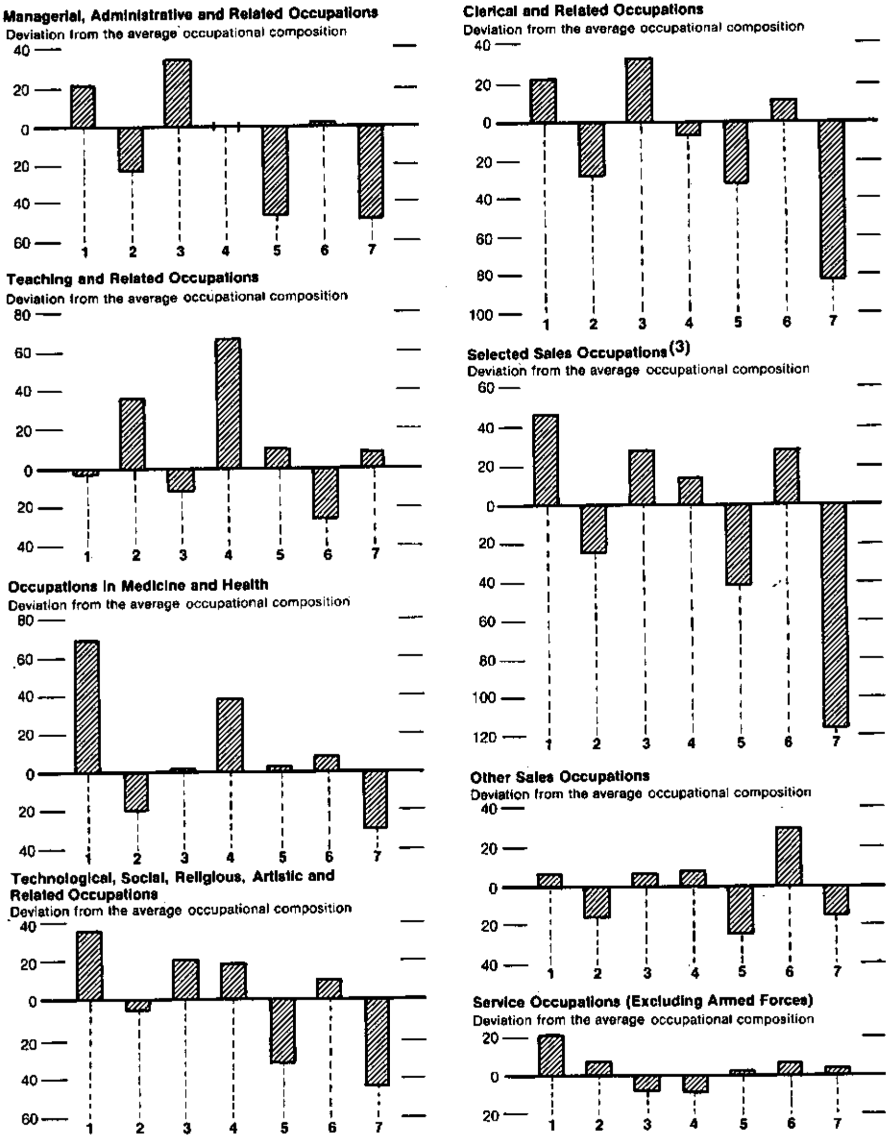
Area were 23% less concentrated in these occupations than was the aggregate of all migrants. (The occupation of a migrant was generally determined as at the date of the 1971 Census.[4] Many of these migrants may have been in a different occupation group in their regions of origin.)

See footnote(s) on page 47.

Chart — 2.3

Relative Occupational Compositions (1) of Migration by Region-Group (2) of Destination, Male Migrants, Canada, 1966-71

(Males aged 20-54 who did not attend school in 1971 and who worked in 1970)

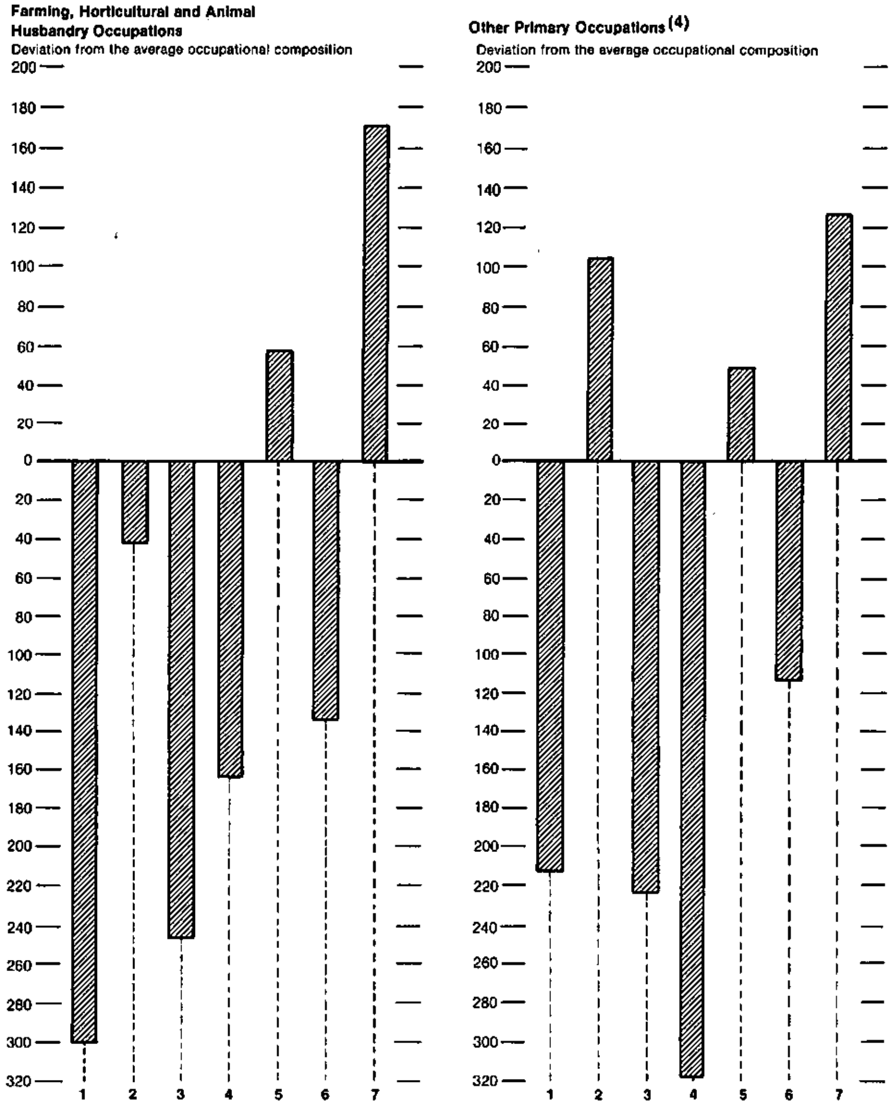


1 — Atlantic CMAs
2 — Atlantic Non-CMA Area
3 — Central CMAs
4 — Central CAs of 50,000 and Over
5 — Central Predominantly Rural Area
6 — Western CMAs
7 — Western Non-CMA Area

Chart — 2.3 (continued)

Relative Occupational Compositions(1) of Migration by Region-Group(2) of Destination, Male Migrants, Canada, 1966-71

(Males aged 20-54 who did not attend school in 1971 and who worked in 1970)

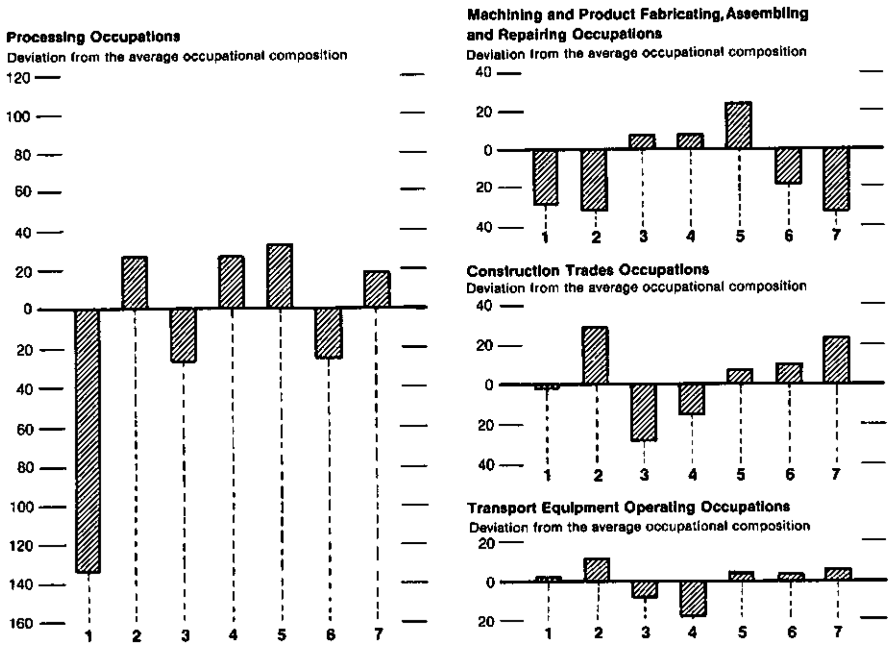


1 — Atlantic CMAs
2 — Atlantic Non-CMA Area
3 — Central CMAs
4 — Central CAs of 50,000 and Over
5 — Central Predominantly Rural Area
6 — Western CMAs
7 — Western Non-CMA Area

Chart — 2.3 (concluded)

Relative Occupational Compositions(1) of Migration by Region-Group(2) of Destination, Male Migrants, Canada, 1966-71

(Males aged 20-54 who did not attend school in 1971 and who worked in 1970)



1 — Atlantic CMAs
2 — Atlantic Non-CMA Area
3 — Central CMAs
4 — Central CAs of 50,000 and Over
5 — Central Predominantly Rural Area
6 — Western CMAs
7 — Western Non-CMA Area

(1) The relative occupational compositions are ratios of proportions. Each ratio compares the proportion of migrants in a given occupation for migrants going to a particular region-group of destination, to the corresponding proportion for all migrants. The coefficients shown graphically on this chart are derived from the relative compositions. If a relative composition element (a ratio) is greater than or equal to one, then the coefficient is [100 (value of rates -1)]. If the ratio is less than one, then the coefficient is [100 (1/value of the ratio -1).]

(2) See Table 2.1, footnotes 7 to 8, for a precise definition of each region-group.

(3) See Table 2.5, footnote 1.

(4) See Table 2.5, footnote 2.

Source: 1971 Census of Canada, unpublished tabulation.

The metropolitan area region-groups tended generally to have the strongest concentrations of in-migrants in managerial, technological, clerical, and certain sales occupations. The Central Canada region-groups had the highest concentrations in machining and product fabricating, assembling and repairing occupations. The regions with very few large cities had higher than average proportions in the primary and processing occupations. The highest concentrations of in-migrants in the transport occupations were in the Atlantic and Western region-groups, areas of the country where a large segment of industry is involved in the extraction of natural resources. The corresponding data (not shown here) that exclude intraregion migrants indicate a similar pattern.

2.5 Conclusion

It is a major working hypothesis of this study that certain characteristics of regional economies and populations are important factors in explaining the structural variation among regional migration streams. In order to apply this notion usefully in an analysis of that variation, the chosen areas of origin and destination of migration should differ substantially in economic structure. This chapter has demonstrated that the areas chosen for use in the analysis that follows meet this requirement in a significant, though by no means perfect, degree. This chapter has also shown that the migration streams among the chosen areas of origin and destination vary markedly in their educational and occupational distributions. Chapter 4 analyses the interstream variation of these distributions, and in Chapter 3, the relevance of the working hypothesis to the statistical analysis is developed from a theoretical point of view.

FOOTNOTES

[1] Some readers of this report may be disappointed to find only seven region-groups defined for use in the analysis. Indeed, a larger sample of area units is desirable. It is the modest size of the Canadian population that explains the use of such a small number of regions. The sample universe refers to persons in the prime working ages who resided in Canada in both 1966 and 1971, who were not attending school in the 1970 - 71 school year and who worked in 1970. The estimated total of 2.4 million migrants in the sample universe must then be simultaneously cross-classified (in order to conduct the analyses presented below) by sex, age, education, and occupation, as well as by area of origin and area of destination.

Now suppose there were 50 regions (a number we could quickly achieve by starting with the 22 1971 Census Metropolitan Areas as separate regions). Then we would have to distribute 2.4 million migrants among 600,000 cells (two sexes times two age groups times four educational groups times 15 occupation groups times 2,500 migration streams). The average of four persons per cell would be most unacceptable. In fact, the cells that pertain to migration streams from one region to another may contain an average of far less than four persons per cell, because a substantial proportion of the migrants may not have crossed the boundaries of those 50 regions (many of the migrants may be intraregion migrants). Only a cursory knowledge of the known defects in the quality of census data (a condition that applies to censuses in all countries) is sufficient to make one realize that such data would be dominated by unsystematic error variation. This remark does not imply that we merely wish to group data so as to maximize "explained variance". The issue is more basic than that – it is one of finding something to analyse, of making judicious use of the available statistics. It might be thought that the difficulties being outlined above could be avoided by using a sample of individual records. However, because the dependent variable for most of the pertinent analyses is the whole distribution of the migrants, from one specific region to another, with respect to a given attribute, we still need a large sample of migrants going in a specific direction between two particular regions, and this requirement leads us back to a relatively small number of regional units. (It should also be noted that the Public Use Sample Tape would have been useless, had it been available, for this study due to the highly aggregated level of the area codes attached to each record on this tape. A special sample of records constructed only for this study would have been prohibitively expensive.)

[2] Oshawa was designated as a census CMA for the first time in the 1976 Census of Canada. Therefore, in this study, data for Oshawa are not included in Central CMA's but in Central CA's of 50,000 and Over.

[3] A glance at the figures in Table 2.2 shows that Atlantic CMA's and the Western Non-CMA are two region-groups with very different industry structures. Yet the dissimilarity index score for this pair is only 35%. It is reasonable, on the basis of this observation, to suggest that a score of 50% on the dissimilarity index is a practical maximum, given the chosen industry groups and the dominance of certain industries in the total Canadian economy. In these terms, the lowest dissimilarity index score shown in Table 2.3, 10% for the difference between the two highly urbanized Central Canada region-groups, is one fifth of the maximum. The largest figure shown in that table is about three quarters of the approximate maximum.

[4] In the 1971 Census, occupation refers to the specific kind of work the person did on the job, as determined by the reporting of the kind of work, the description of the most important duties and the job title. Data relate to the respondent's job in the week prior to enumeration, if he or she had a job during that week, or to the job of longest duration since 1 January 1970, if he was not employed in that week. Persons with two or more jobs during the reference weeks were asked to give the information for the one at which they worked the most hours.

CHAPTER 3

THEORETICAL RATIONALE AND SOME CENTRAL METHODOLOGICAL CONCEPTS

3.1. Selecting the Strategy of Analysis

This chapter provides a loosely stated theoretical rationale for the central working hypothesis of the study and introduces the methodology that will be used in the analysis.

Many studies of geographic mobility set out to examine the causes of mobility, usually in terms either of factors that differentiate movers from non-movers, or of correlates of intergroup variations in levels of mobility, or of reasons persons give when they are asked why they move (see the review of literature in Stone, 1974a). Other studies, like this one, assume that persons have already become migrants, and concentrate on explaining other aspects of the phenomenon of mobility.[1]

The aspect of mobility that is of interest in this chapter is the occupational composition of a group of migrants who move from one particular area of origin to another specific area of destination. Can the occupational composition of that group of migrants be systematically related to characteristics of their areas of origin and destination?[2] If so, how strong is that relationship statistically, especially after holding constant other features of the migration stream (e.g., its educational composition) that may determine its occupational composition? These are the kinds of questions to be addressed.

In the next section, the theoretical framework for the analysis procedure is outlined. In effect, the theoretical rationale says that the occupational composition of a group of migrants results from two main sets of factors: 1. how the group is distributed according to other attributes that are causally related to occupation (e.g., education), and 2. the characteristics of the economies and populations of the places of origin and destination of the migration. Often there is interrelation between these two sets of factors.

The rest of this chapter is divided into two main parts. Sections 3.2 and 3.3 form the theoretical part. These sections attempt to explain how the occupational composition of migration is affected by the characteristics of the regions of origin and destination of the migration streams and other socio-economic characteristics of the migrant. Section 3.4 outlines the methodology for the analysis in Chapter 4.

See footnote(s) on page 58.

3.2. Regional Characteristics and Occupational Composition of Migration

Why do characteristics of the places of origin and destination of a group of migrants influence the occupational composition of that group? The comments that follow discuss this question from a theoretical viewpoint and sketch a partial answer. In formulating the discussion, a deliberate effort is made to avoid setting down a sequence of abstract principles in the type of jargon that is typical of a professional journal.

It is helpful in sketching this partial answer to assume that the group of migrants whose occupational composition is to be explained is going from the same place of origin to the same place of destination, i.e., a single migration stream. Further, for the sake of concreteness, it is useful to start by considering one migrant in that stream.

3.2.1. Place-of-origin Effects

In at least two ways, there can be a relation between the place of origin of a migrant and his or her occupation. Firstly, the migrant may have been prompted to leave the place of origin because of events that took place there and that particularly affected (or were otherwise related to) persons in certain occupations. For example, the closing of a major mine near a community in Northern Ontario may prompt miners, in particular, to leave that area. In short, one possible reason for a relation between characteristics of a place of origin and occupational composition of the out-migrants from that place is the occurrence at that place of events that tend to be related to persons who are unusually concentrated in particular occupations.

Secondly, the place of origin in question may have an unusually high proportion of persons in a particular education group. These persons may in turn (because of their education) tend to congregate in particular occupations. This tendency may influence the occupational composition of migration from that place of origin. For example, the Ottawa work force has an unusually high concentration of university-trained people. Whenever out-migrants from Ottawa comprise a representative cross-section of Ottawa's work force in education, the relation between education and occupation will tend to produce a distinctive occupational composition for those migrants.

3.2.2. Place-of-destination Effects

How can characteristics of the place of destination of a group of migrants in a particular stream influence the occupational composition of that stream? Consider the example of the miners who may have left a Northern Ontario mining community because of the closing of a major mine there. Suppose that, for the most part, they wished to remain in the occupation group of miners, and assume further that they were aware of job openings at mines near another Northern Ontario community in which they would have to reside in order to fill the openings. Under ordinary circumstances they would hardly be expected predominantly to choose to move to, let us say, Ottawa rather than to this other Northern Ontario community?

This example suggests the following principle. A place of destination may present a particular mix of attractions that appeals mostly to those potential in-migrants that have certain occupations. The more concentrated are the attractions of a given place, among the opportunities and amenities that appeal mostly to persons in certain occupations, the greater will be the tendency for the in-migrants to that place to have a particular occupational composition. Thus, the places of origin and destination of a stream of migration can influence the occupational composition of the stream either through the mixes of attractions that these places present to potential migrants or through peculiarities in the composition of the population at the place of origin.

3.2.3. Interrelation of Place-of-origin and Place-of-destination Effects

There is an interrelation between the forces exerted by the mix of attractions at the place of origin and those generated by the set of attractions at the place of destination. In our example of the migrant Northern Ontario miners, there is a correspondence between the occupational structure of job opportunities at the place of origin and that at the place of destination. The essence of this correspondence is that the particular types of skills that are overabundant at the place of origin are in short supply at the place of destination. This is the kind of correspondence that would tend to cause the out-migration from Montréal to Toronto to show a higher proportion of financial analysts than the out-migration from Montréal to Timmins. Without knowing a great deal about the circumstances surrounding each stream of migration, it is next to impossible to develop a cogent argument about the nature of the interrelation between origin and destination forces as they meet to help determine the composition of a particular migration stream, or even to effectively anticipate which of the two sets of forces are more important in determining the composition of the migration stream.

3.3. Occupational Composition of Migration and Other Socio-economic Characteristics of the Migrants

It is possible for the composition of a migration stream to be influenced by socio-economic factors that have nothing to do with the characteristics of the regions of origin or destination. For example, the effect of age on the educational composition of a migration stream may not vary much from one migration stream to another. In this study, it is hypothesized that the distribution of the migrant population with respect to education exerts an influence (that is partly independent of regional characteristics) on the occupational composition of migration. However, it may be extremely difficult to sort out statistically what portion of the influence of education is "independent" of regional characteristics, since the latter can affect the occupational composition of migration through education (acting as an intervening variable).

Thus the occupational composition of a migration stream could be the result of the distribution of the stream among the categories of another attribute to which occupation is related. This result may be partly "independent" of characteristics of the places of origin and destination of the stream, but it may also partly reflect such characteristics. (See Appendix B for additional discussion concerning the theoretical rationale for the analysis of the occupational composition of a migration stream.)

3.4. Method of Analysis

The foregoing sections contain a theoretical rationale for the working hypothesis that the occupational composition of a migration stream depends partly on characteristics of the stream's places of origin and destination and partly on the distribution of the stream with respect to related attributes such as education. The present section provides an outline of the selected method of analysis. Central to the type of analysis conducted in this study is a statistical model expressed concretely in one or more equations. This model is the formal expression of a series of hypotheses about the ways in which particular aspects of the socio-economic composition of the population and regional characteristics contribute to the shape of the overall occupational distribution of a migration stream. In the text, the model is called "the asserted explanatory model".

The kind of analysis that is used has some requirements that considerably limit its flexibility, but it is especially well adapted to problems where the "dependent variable" of the study is virtually the whole distribution of a population sample over the categories of a given attribute. In other words, it is especially well suited to the analysis of compositional differences among population groups, or compositional changes within a given population group.

3.4.1. Form of the Variables

Ideally, in an analysis of migration stream composition that uses the working hypothesis selected for this study, a specific place of origin or destination should be represented by the set of its scores on the variables that are chosen to reflect regional characteristics. For example, if we assume that regional industry structures are especially pertinent to the occupational compositions of inter-regional migration streams, we need a measure of the industry structure for each region in the analysis. But in order to conduct a traditional multivariate analysis (e.g., multiple-regression), we would need a substantial sample of associated migration streams for each industry structure. Due to the small size of the Canadian population (as explained in Chapter 2), the achievement of such a sample is not feasible using 1971 Census data.

Therefore, to exploit the available statistics in studying areal influences on the occupational composition of migration streams, we chose to represent the influence of a given region by a proxy variable which stands for the whole mix of unmeasured relevant characteristics of the region. The statistical effects of industry structure and other relevant aspects of the regions of origin and destination of a migration stream are deemed to be reflected in the occupational compositions of the non-migrants of those regions. This is an admittedly clumsy statistical procedure,[3] but we are trying to mine the 1971 Census data base to analyse an aspect of the composition of migration and, in the process, it is necessary to modify the traditional procedures of statistical analysis.

In order to put this procedure into use, the analysis is done for selected separate pairs of region-groups. In this way, we can interpret statistical coefficients for the effects of the regional dimension in terms of what we know about pertinent characteristics of each member of a given pair of region-groups. Thus, for example, we look specifically at the composition of the migration stream going from Central CMA's to the Central Predominantly Rural Area. Using data of the type shown in Table 2.3, we can look directly at the differences in industrial composition between these two region-groups and consider how these particular differences might influence the occupational composition of the migrants going from Central CMA's to the Central Predominantly Rural Area. We do not provide, in this report, a specific statistical coefficient that measures the relation of the industry structure difference to the occupational composition of the migration stream between Central CMA's and the Central Predominantly Rural Area in particular, but we do use information about that difference to interpret the coefficient that is associated with the proxy variable for regional factors used in the multivariate analysis.

The measurement of the statistical influence of education turns essentially upon the variations in occupational composition among the different categories of education. The influence of each education category is represented by a measure of its presence or absence.

We also consider measurements of the influence of regional characteristics on the educational composition of migration. The influence of education is viewed as an intervening variable between the regional factors and the occupational composition of migration streams (how this intervention can occur was illustrated in the preceding theoretical discussion). In conducting this analysis, age is introduced as an explanatory attribute. Each category of age is deemed as having a certain statistical influence on the educational composition of migration, and that influence is represented by the presence or absence of that category of age.

In sum, the attributes age, education and occupation are all viewed as collections of variables. Each category of each attribute is considered a variable (e.g., the age group 20-34). One of these variables (categories) is represented in the analysis by its presence or absence.

See footnote(s) on page 58.

The form of data that is consistent with these stipulations is the census cross-tabulation (or contingency table).[4] The statistical techniques to be used are a branch of the recently developed methods for the multivariate analysis of contingency tables. This study is an illustration (see Stone, 1977a for other examples) of the manner in which census data in their traditional format can be used to address questions that require multivariate analysis.

3.4.2. Technical Definitions and Notation

To introduce the type of model used in this study, a few technical definitions and some notations are necessary: the concept of an "nth-order conditional proportion" is followed by an exposition on the concepts of "zero-order effect" and "higher-order effects".

A "conditional proportion" is one whose denominator is the aggregate of a particular group in the population, and whose numerator is a subset of that group. The concept of a "condition" applies when a member of the population can be counted in the denominator only if he or she has a specified set of values of particular attributes. For example, the proportion that moved once among males aged 20 - 34 with university education is a third-order conditional proportion, because three attributes (sex, age and education) were cited in identifying the denominator of the proportion. Generally, the order of the condition is the number of attributes involved in specifying requirements for membership in the group that comprises the denominator of the proportion. The phrase "zero-order conditional proportion" means a proportion whose denominator is the total sample size for the analysis in question.

A zero-order effect is defined as the ratio of a first-order conditional proportion to a zero-order proportion. An "nth-order partial effect" (a higher-order effect) is the ratio of (n+1) th-order conditional proportion to an nth-order conditional proportion. The zero-order effect of a given explanatory attribute on the dependent one is an overall effect, a weighted average of higher-order interaction effects. The nth-order partial effect of two or more attributes, at a given order of interaction, upon the dependent one is a weighted sum of a subset of the interaction effects at the still higher-orders (see Stone, 1975a, Section 4). More precise definitions can be given in terms of the following notations:

Let "$Pr(Y_i)$"	mean the proportion of the entire sample that has the value i on attribute Y.
Let "$Pr(Y_i \mid X_j)$"	mean the conditional proportion who have the value i on attribute Y among those who have the value j on attribute X.
Let "$Pr(Y_i \mid X_j, U_k)$"	mean the conditional proportion who have the value i on attribute Y among those who have the value j on X **and** the value k on U.

See footnote(s) on page 58.

The measure chosen to reflect the zero-order effect of X_j on Y_i is $Pr(Y_i|X_j)/Pr(Y_i)$. In general, this measure indicates the extent to which our "prediction" (or the statistical probability) of Y_i is changed when we assume X_j compared with the case when we do not assume X_j. The more the knowledge of X_j alters this "prediction", the greater is the zero-order effect of X_j and the more the measure of the zero-order effect of X_j diverges from 1.0.

The "interaction effect" or the "nth-order partial effect" refers to a higher-order effect of an explanatory attribute upon the specified dependent one, given specific values of a particular set of other explanatory variables. The measure that reflects the first-order partial effect of U_k on Y_i, given X_j, is $Pr(Y_i|X_j,U_k)/Pr(Y_i|X_j)$. This measure tells us how much the **additional** knowledge U_k should alter our "prediction" (or the statistical probability) of Y_i given that we already know X_j. The alteration of our prediction is a result of the interaction effect of X_j and U_k upon Y_i. The more this ratio diverges from 1.0, the greater is the "partial effect" of U_k on Y_i, given X_j.

Whenever an effect is insignificant, its measure diverges from 1.0 by a negligible amount. If we hypothesize that a particular effect is insignificant, we may simply fail to specify it when the model is formulated. It should be noted that an unspecified effect is not, under this approach, assumed to be exactly equal to 1.0.

The zero-order effect of a given explanatory attribute on the dependent one should be specified in the model if the substantive theory, that is deemed pertinent, implies (by deductive argument) or strongly suggests (by plausible argument) that there is a substantial "direct" effect of the explanatory attribute upon the dependent one. A "direct" effect is generally one whose direction and strength are **not** greatly altered when we change the values assumed by the other explanatory attributes included in the model.

The higher-order effect of a given explanatory variable, X, with others, $(Z_1,Z_2, \cdots ,Z_n)$ upon the dependent attribute, Y, should be included in the model if the substantive theory strongly suggests that there is a substantial causal relation between X and Y whose strength and/or direction depend markedly on the values assumed by the other specified explanatory attributes $(Z_1,Z_2, \cdots, Z_n)$.

A statistical hypothesis about the **direction** of an effect is easily accommodated. We merely think in terms of whether the presence of absence, as the case may be, of a specific explanatory attribute value will increase or decrease the statistical probability of a given value of the dependent attribute. In dealing with the first-order term $Pr(Y_i|X_j, U_k)/Pr(Y_i|X_j)$, for example, we can formulate and test the hypothesis that the addition of condition U_k to X_j will increase the statistical probability of Y_i – a hypothesis that this term exceeds 1.0 significantly. By articulating a network of hypotheses of this sort, we can readily extend the explanatory model to include notions about the directions of co-variation between the dependent attribute and other variables in the model. We could go even a step further and specify a functional relation between an expected value of the dependent attribute and the given values of the explanatory ones.

3.4.3. The Multiplicative Power Model

An example of the kind of model used in subsequent chapters is now provided. This type of model is called "multiplicative power model" or "power model" because it always involves raising the terms in the model, the effect measures, by fractional exponents. Power models are a subclass of log-linear models (cf. Goodman, 1970, 1972, 1973a and 1973b; Fienberg, 1970a, 1972 and 1973; Bishop, Fienberg and Holland, 1974).

Using the notation introduced in the preceding section, an illustration of one type of power model may now be given:

$$\hat{Pr}(Y_i|X_j, U_k, Z_h) = Pr(Y_i) \cdot \left[\frac{Pr(Y_i|X_j)}{Pr(Y_i)} \cdot \frac{Pr(Y_i|U_k)}{Pr(Y_i)} \cdot \frac{Pr(Y_i|Z_h)}{Pr(Y_i)}\right]^{1/3}$$

$$\cdot \left[\frac{Pr(Y_i|X_j,U_k)}{Pr(Y_i|X_j)} \cdot \frac{Pr(Y_i|U_k,Z_h)}{Pr(Y_i|U_k)}\right]^{1/3} \cdot \lambda_{ijkh} \qquad (1)$$

The set of ratios in the first pair of square brackets on the right are the zero-order effects that are postulated as being significant. The ratios in the second pair of square brackets are those that measure the first-order effects postulated as being significant. The adjustment factor λ_{ijkh} is usually chosen so as to minimize the chi-square statistic generated by the model (see Appendix C), but in this study it is chosen so that the sum of the estimated values $\hat{Pr}(Y_i|X_j,U_k,Z_h)$ over i is equal to 1.0. The first-order effect $Pr(Y_i|X_j,Z_h)/Pr(Y_i|Z_h)$ is hypothesized as being insignificant and thus it is not specified in the model. The symbol "^" is a superscript denoting a value produced by a model.

Additional discussion of the properties of the power model illustrated by equation (1) and of aspects of the application of such models is provided in Appendix C.

3.4.4. Model Assessment Procedure

The analysis of the occupational composition of a migration stream is based on an assessment of the asserted explanatory model. The assessment of the performance of the model turns directly on a comparison of its "goodness of fit" with that of a "null" model[5] which, in contrast to the asserted one, hypothesizes no dependence between the occupational distribution being analysed and the chosen explanatory attributes.[6] The comparison is provided by a chi-square statistic. The null model yields a particular value of the chi-square statistic, and

See footnote(s) on page 58.

the asserted model (which claims a specific pattern of dependence) yields another value of the chi-square statistic. The performance of the asserted model can then be assessed by determining the extent to which its chi-square statistic is lower than that of the null model.

A chi-square statistic is also used to assess the overall contribution of each of the explanatory effects that are specified in the model. Given the measured reduction in chi-square of the null model that is achieved by the asserted model, we can then proceed to attribute a certain amount of that reduction to a particular explanatory effect that was specified in the asserted model. One of the major advantages of this method of analysis is that it permits a very simple procedure for accommodating in the statistical analysis substantive hypotheses about **any** order of interaction among the explanatory attributes. The procedure for the measurement of the contributions of individual specified effects is outlined in Appendix C.

The statistical procedure used to measure the separate contributions of zero-order and interaction effects to the overall performance of the model can also yield a value for the portion of the measured performance of the model that is **jointly shared** among the explanatory attributes. Two or more attributes will have a shared effect when they are mutually correlated. For example, since education and occupation are correlated, they will have a jointly shared contribution, in addition to their separate contributions, to the overall performance of the model. Generally, the stronger the correlation between two attributes, the greater is the portion of their total contribution that is jointly shared, and the less will we be able to measure the relative importance **(in statistical terms)** of these attributes to the performance of the model.

FOOTNOTES

[1] However, various assumptions about the explanation of mobility decisions are important in setting up explanatory theory for the study of these other aspects of mobility (see Appendix B).

[2] Underlying this question is the central working hypothesis of the analysis: characteristics of the regions of origin and destination of a migration stream help to determine its occupational composition. Some readers will wonder how this approach accommodates the generally accepted proposition that occupation is one of the determinants of geographic mobility. Although this proposition is deemed to be true by the writer, there remains considerable variation among migration streams with regard to their occupational composition. This variation is a legitimate problem for analysis because of its relevance to the concerns of officials and citizens who are attempting to deal with issues in the field of regional economic structure and development. Largely because of fundamental shortcomings in census data (shortcomings that are mostly tied to the fact that the occupation of a census respondent is determined only as of the end of the census migration period), we propose to deal with that problem of analysis without explicitly confronting the role of occupation as a determinant of mobility.

In an explicit confrontation with that role of the occupation variable, a procedure similar to the following might be followed. One might first measure statistically the effect of an occupation on the propensity to migrate (a measurement that is not feasible with existing census data without using a battery of untestable assumptions). Then, given the coefficients of "mobility propensity" for different occupation groups, and the occupational composition of the population at the start of the migration period or at the time of migration (neither of which are provided by census data), it would be a routine matter to compute a "predicted" (or expected) occupational composition of migration.

An important difficulty in implementing this strategy of analysis, aside from the decisive problems caused by the gaps in census data, would arise from the fact that occupation is not the only determinant of migration. (In other words, it is not the only factor which explains why some people are migrants and others are non-migrants.) Application of that strategy requires a method of analysis in which other determinants of migration propensity are considered while the coefficients associated with the occupation variable (i.e., a multivariate analysis) are being estimated. When this is done, the ever-present intercorrelation among the statistical measurements of the determinants of mobility propensity can produce a seriously distorted picture of the effect of each occupation upon mobility propensity. (In regression analysis this problem is called "multicolinearity".)

Even if multicolinearity is not a problem, the analysis procedure outlined would entail expensive wasted motion if all one is interested in understanding is the occupational composition of migrants. Some of the measured determinants of mobility propensity would be common to all occupation groups. Once we have estimated their coefficients (in the explanatory model of mobility propensity), we would promptly ignore their "unique" effects when focusing on the occupational composition of the migration, because we can compute occupational composition by merely focusing on the coefficients for the different occupation groups and the occupational composition of the population (migrants plus non-migrants) at the time of migration.

Another way of analysing the occupational composition of migrants is to focus the statistical analysis on the people who are already known to be migrants, and then to study occupation composition variation among different groups of migrants. This procedure has ample precedent in the design of experiments in the physical and biological sciences. The construction of a laboratory environment in which factors that are known to be important are physically "screened out" so as to permit a clearer view of other factors of immediate interest is quite common. For example, analyses of the potential effects of substances on the human body are often made by feeding exceptionally large doses of those materials to animals. That such a procedure can lead to highly misleading conclusions when generalizations made about animals are extended to human beings operating in their normal daily lives is perfectly obvious. Yet we are all anxious to opt for experimentation on animals, rather than on human beings. (We are happy to do what is feasible in the prevailing circumstances.)

It is fully recognized, however, that occupational differences in the propensity to migrate are important factors in a theoretical explanation of the occupational composition of migration. In principle, there is no neat separation of an analysis of the propensity to migrate from one of the composition of migration where that composition involves attributes that affect the propensity to migrate. Thus, suppositions about the effects of occupation on mobility propensity are involved in setting up the theoretical rationale for the central hypothesis of the study, as illustrated in Appendix B.

3 One of the notable problems with this procedure stems from the fact that out-migration over the 1966 - 71 period may affect the occupational composition of the non-migrants measured as at the 1971 date. It is assumed here that this factor does not substantially distort the extent to which the composition of non-migrants represents intrinsic characteristics (especially of the economy) of the region in question.

4 The meaning of "contingency table" may be clarified by the following illustration. Suppose the attribute age is categorized into two values (under 50, and 50 or older) and sex is categorized into two values (male and female): We can then stipulate four categories of the cross-classification of age by sex. A table that shows the numbers of people who fall into each of these four categories is a two-way contingency table. The order of a contingency table (two-way, three-way, etc.) is the same as the number of attributes that are cross-classified in defining the table. A contingency table is also called a "cross-tabulation".

5 Up to this point, several terms have been used with the double quotes ". . .". This practice signified that the terms are being used in a sense that does not quite conform to their conventional meanings. These terms include, "effect", "explanatory," "null" and "model". From this point onward, double quotes will be dropped from these terms; but the reader should always bear in mind that the terms are being used only advisedly. The context of the usage of these terms in the report should be sufficient to indicate reasonably well how they ought to be interpreted.

6 This strategy means that, in effect, we would not assess how well the model has performed by simply asking how well it "fits" the data, because we can, as in regression analysis, easily stack the model in such a way that its "fit" is perfect, or very nearly so.

CHAPTER 4

ANALYSIS OF THE OCCUPATIONAL AND EDUCATIONAL COMPOSITION OF MIGRATION STREAMS

This chapter presents the results of the statistical analysis of the occupational composition of migration streams flowing among the seven selected region-groups. Chapter 2 identified the seven region-groups and showed that there was substantial interstream variation in occupational composition. Chapter 3 rationalized the central working hypothesis of the analysis. This hypothesis states that the occupational composition of a migration stream depends partly on characteristics of the stream's areas of origin and destination, and partly on compositional features of the stream (such as education composition).

The industrial structures of the region-groups are thought to be especially pertinent to the effects of regional factors upon the occupational composition of migration. Chapter 2 showed that the seven region-groups varied substantially in the industrial structures of their labour forces. It is hypothesized that the most important compositional factor is the educational distribution of migration. Chapter 2 also showed that the migration streams among the seven region-groups varied considerably in their educational make-up.

It is also hypothesized that the interstream variation in educational composition partly reflects the influence of regional factors. Thus educational composition acts partly as an intervening variable between regional characteristics and the occupational composition of migration. The clarification, in terms of actual statistical data, of this intervening variable role for education requires an analysis of the interstream variation in educational composition. The analysis that follows therefore addresses two separate (though related) dependent variables – the educational composition of migration, and the occupational composition of migration.

To deal with these two dependent variables, a statistical model in two equations is developed and applied. The relative statistical strengths and the directions of the contributions of individual factors within each equation in the model are measured in a limited way and interpreted briefly. Through the chosen method of analysis, which was introduced briefly in Chapter 3, it is possible to evaluate the composite hypothesis that both equations substantially enhance our ability to "predict" the joint composition of a migration stream by occupation and education simultaneously. This composite hypothesis can be expressed as a third equation. However, shortage of time and other resources preclude testing the composite hypothesis. Appendix E formulates the three-equation model that includes the composite hypothesis, outlines the procedures for testing it, and indicates how its results may be used in projection or simulation work.

The essence of the two-equation model can be summarized in the following manner. The occupational compositions of migration streams vary systematically with change in the identities of the streams' areas of origin and destination because of industrial structures and related variations among the areas, as well as with shifts in the educational composition of the streams. The educational compositions of migration streams vary systematically with changes in the identities of the streams' areas of origin and destination as well as with shifts in the age composition of the streams.

The discussion is organized into two parts. The first section deals with the educational composition of migration streams. It reviews the issues addressed in the analysis of the educational composition of migration streams and introduces the analytical model and the data. Then, the main results of the application of the model are presented and interpreted briefly.

In the second section, occupational distribution of migration is treated similarly: the review of issues is followed by the presentation of the analytical model and the data, and then a short interpretation.

It should be noted here that the brevity of the interpretive text in this chapter is deliberate. More definitive discussion would involve a major synthesis of relevant literature relating the main points of that literature to the patterns shown below (a task that cannot be done quickly since almost all the relevant literature deals only indirectly and inferentially with the questions that are the focus of analysis here). Such discussion would also very likely involve experiments with a variety of alternative data sets and techniques. It is hoped that other researchers or university students will be stimulated by the present findings and commentary to embark upon the necessary synthesis of literature and supplementary statistical manipulations.

Several limitations and qualifications must be observed concerning the data and the analysis. Some of the pertinent qualifying remarks and cautions follow (others are presented in Chapter 3 and in Appendices C and D).

A specific sample from the 1971 Census population is designated for the analysis. It was desirable that the sample exclude, as much as feasible, a variety of groups such as persons who were still (in the year preceding the census) in the process of changing their educational attainment in a major way through enrolment in the formal educational system. The selected sample population consists of males who were in Canada on 1 June 1966, who were aged 20 - 54 on 1 June 1971, who were out of school during the school year 1970 - 71, and who worked during 1970. All generalizations offered below pertain to the universe represented by the sample population.

It was necessary to introduce somewhat arbitrary categorizations of the ranges of the selected attributes in order to proceed with the analysis. The general reasons for this situation are outlined in Appendix D. Two important points need to be made at this juncture. Firstly, the categories chosen are believed, on the basis of past experience with these sorts of data (and with some support from inspection of the relevant data distributions) to reflect meaningfully the shape of the related population distributions. Secondly, all numerical values of estimated parameters shown below are sensitive to the chosen categorizations of the ranges of these variables. It is expected that the broad patterns that are emphasized in the discussion will not be altered greatly with alternative meaningful categorizations. Unfortunately, the time cost of sensitivity testing for this problem proved prohibitive, and such evaluation work will have to be postponed for a future report (perhaps done by someone else).

4.1. Educational Composition of Migration Streams

Previous research and theory on migration suggest that a full explanation of the educational composition of a group of migrants in Canada leans markedly on the group's age distribution and on the characteristics of the pertinent regions of origin and destination (see Lee, 1966). This section pursues an explanatory multivariate analysis of the educational composition of Canadian 1966 - 71 internal migration in terms of these three classes of variables.

Two aims give direction to the analysis, though the analysis inadequately fulfils them. The aims are (a) to examine the extent to which selected explanatory attributes are simultaneously associated, through a specific model, with the educational composition of migration; and (b) to assess and interpret the relative statistical contributions of the attributes to the educational composition of Canadian internal migration in the 1966 - 71 period.

4.1.1. The Analytical Model

Section 3.4 has introduced the kind of model that will be used to conduct the main part of the analysis in this chapter. The procedure for developing a specific case of this general form involves the statement of a series of hypotheses concerning the substantively significant zero-order and higher-order effects of explanatory attributes upon the dependent one. Also, in order to choose among alternative ways of stating the measure of a given higher-order effect, some assumptions about the causal ordering of certain of the explanatory attributes may be needed. After stating the hypotheses and assumptions, which should be guided (ideally) by a formulated substantive theory, the specific form of the model is obtained by a simple procedure (see Appendix C). The following discussion commences with the statement of the substantive hypotheses and assumptions.

The age composition of a migration stream and the characteristics of its regions of origin and destination have substantial direct influences on the schooling distribution of a migration stream. Therefore, the analytical model includes the zero-order effects of selected measures of these explanatory attributes. The well-known tendency for younger age groups of persons who are out of school to have relatively higher levels of education than older age groups supports the decision to include the zero-order effect of age. Since the variation in educational attainment distribution among the regional population should tend to influence the variations in this distribution for the regional out-migration streams, a measure of the zero-order effect of region of origin is included. A similar measure for the region of destination is also included. The inclusion of a region-of-destination effect arises from two considerations. Firstly, a large proportion of young adult migration is related to the search for higher education, and the relevant educational facilities are heavily concentrated in particular regions. Secondly, regional variation in the occupational composition of labour demand affects the educational composition of migration through the interrelation of occupation with education.

In addition to the zero-order effects, this analysis considers higher-order effects involving age and the regional characteristics. It is hypothesized that there is an interaction of age with the regional attributes. In other words, the effect of the regional characteristics on the educational composition of migration varies somewhat depending on the age composition of the migration stream.

In the equations presented below, there is also a component intended to measure a direct interrelation of schooling and migration, independently of age and the measured factors. The direct interrelation of mobility and schooling would arise if there is a significant tendency for persons having differing schooling levels to differ in their mobility propensity, regardless of the values of other measured explanatory variables. If there is this direct interrelation, then it should be brought explicitly into the model so that it can be "held constant" statistically while the effects of age and of the regional characteristics are being estimated.

To introduce the model explicitly, additional notation is needed.

Let

"X_Y" mean the **Yth** category of schooling

"A_T" mean the **Tth** category of age

"M" mean migrant

"ϕ" mean region-group of origin

"D" mean region-group of destination.

Using these terms, the model can be expressed as follows:

$$\hat{Pr}(X_Y|A_T\phi DM) = Pr(X_Y)\cdot\left[\frac{Pr(X_Y|M)}{Pr(X_Y)}\cdot\frac{Pr(X_Y|A_T)}{Pr(X_Y)}\cdot\frac{Pr(X_Y|D)}{Pr(X_Y)}\cdot\frac{Pr(X_Y|\phi)}{Pr(X_y)}\right]^{1/4}$$

$$\cdot\left[\frac{Pr(X_Y|A_TD)}{Pr(X_Y|A_T)}\cdot\frac{Pr(X_Y|A_T\phi)}{Pr(X_Y|\phi)}\right]^{1/4}\cdot\lambda_{T\phi DM} \qquad (2)$$

The estimation procedures for this model illustrate one of the significant advantages of the multiplicative power model. The terms on the right-hand side of equation (2) are not estimated from the table that is used as observed data, to test the goodness of fit of each model. Instead, the regional effects are estimated from data for the "non-migrants" within each region ("non-migrants" here include intraregion migrants – i.e., persons who changed residence from one municipality to another but remained within the same region-group). The age effect and the direct migration-schooling interrelation are estimated from Canada-level data.

A concrete example illustrates the procedure. Consider the migration stream from the Central Predominantly Rural Area to the Western Non-CMA Area. It is proposed that the regional influences of each of these region-groups are reflected statistically in the educational compositions of their non-migrant sample populations. Thus, $Pr(X_Y|D)$ and $Pr(X_Y|\phi)$ were estimated from the non-migrant sample populations of areas ϕ and D. $Pr(X_Y|A_T\phi)$ was estimated from the data for age group T among the non-migrants of regions ϕ (the Central Predominantly Rural Area in this case) and D (the Western Non-CMA Area in this case), respectively.

We use Canada-level data to estimate $Pr(X_Y|A_T)$, $Pr(X_Y|M)$ and $Pr(X_Y)$. The first and third terms are estimated from the entire sample population in Canada (non-migrants as well as migrants). They represent, respectively, the average conditional distribution of education given age, and the average unconditional distribution of education in the entire sample population. Thus the ratio $Pr(X_Y|A_T)/Pr(X_Y)$ is a measure of the association of education with age in the entire population sample, and not just among migrants.

The second term, $Pr(X_Y|M)$, is estimated from the data for all intermunicipal migrants in the sample population of Canada regardless of the migration streams to which they belonged. It reflects the overall interrelation of schooling with migration independently of any specific migration stream.

The term $Pr(X_Y|M)/Pr(X_Y)$ is not included in the model as the measure of a relevant variable that was cited among the substantive hypotheses. It was introduced to prevent the measured statistical effects of age and regions of origin and destination from reflecting a component that arises from a direct interrelation of schooling with migration. This ratio represents a device for "statistical control" of that interrelation, while the effects of the other hypothesized explanatory attributes are measured. This procedure is analogous to an experimental design in which certain factors known to be relevant are arbitrarily "screened out" (usually held constant physically or randomized) in order to improve the chances that the "pure" effects of other variables can be seen more clearly.

In testing the fit of these models, the observed values for $Pr(X_Y|A_T\phi DM)$ are used. These are values for a specific migration stream, and are drawn from a much different table than any of those used to estimate the expected values $Pr(X_Y|A_T\phi DM)$. We thus have not estimated the parameters of the model from the identical data set used to test the fit of the model, and can refer to the values $Pr(X_Y | A_T \phi DM)$ as predicted distributions with some justification. However, the test of the model is not a comparison of the observed and predicted distributions. What is important is the comparison of the fit of the model with that of a null model which denies the dependencies asserted by the model of interest. The null model is:

$$\hat{Pr}(X_Y | A_T \phi DM) = Pr(X_Y) \qquad (3)$$

When we applied this model, we quickly ran into a condition of internal migration data that had previously been reported in the literature (see Miller, 1967; Stone 1971; Simmons, 1977). In their schooling composition, migration streams flowing into and out of metropolitan areas tend to be very similar. (Various possible reasons for this condition are outlined in Stone, 1971.) In consequence, when the highly atypical schooling distributions of non-metropolitan "non-migrant" populations were introduced into expression (2), in estimating certain of the effect measures, they reduced rather than enhanced the predictive accuracy of the model. In other words, they introduced a serious bias into the model and lead to the "anomaly" of a model that performs worse than one of its simplified modifications.

Section C.2 of Appendix C provides a detailed explanation of the bases of this "anomaly". It is sufficient in this chapter to point out that the effect measures in question refer to the characteristics of non-migrants in non-metropolitan areas, which are used as proxy measures for the characteristics (of such areas) that influence the educational pattern of migration to and from those areas. Appendix C, Section C.2, shows that a reasonable substantive interpretation can be placed upon the "misfiring" of these particular effect measures (as estimated).

However, the "anomaly" also has a simple substantive interpretation that could have been anticipated from already published literature (Miller, 1967; Stone, 1971). If the influence of the stream's origin (or destination) is dominant among regional factors, that dominance can be expected to turn up in both the stream and its counter-stream. This is precisely what the 1971 Census data show. The attributes of metropolitan areas tend to dominate educational compositions of the streams and counter-streams of migration of such areas. Thus, when we are dealing with migration from a metropolitan to a non-metropolitan area (or vice versa), we will bias the prediction model (if it has the mathematical structure of expression (2)) by including effect measures that are strongly influenced by the non-migrants in the non-metropolitan areas.

Our discussion also implies that as long as stream and counter-stream have very similar schooling distributions, virtually the same model may be used to predict their schooling distributions. The level of similarity that is involved in the present data series is shown in Table 4.1. This table shows conclusively that it would be a waste of time to look for substantially different models to predict the schooling compositions of the streams and counter-streams of migration between the two indicated pairs of metropolitan and non-metropolitan areas. Accordingly, we shall use essentially the same model for such pairs of migration streams. This model is expressed in two different formulas (expressions (4) and (5) below), depending on whether the metropolitan area is the origin or the destination of the migration stream. These truncated models, (4) and (5), may be seen as a testimony to the substantive finding that the schooling composition of migration between metropolitan and non-metropolitan areas is dominated in census data by attributes of the metropolitan areas (to the extent that regional characteristics are relevant), regardless of the direction of the migration flow.

TABLE 4.1. Schooling Distribution of Migrants in Six Selected Migration Streams,[1] Migrants in the Sample Population, 1966 - 71

Migration stream	Total	Level of schooling			
		Less than Grade 12	Grades 12 and 13 and non-university	Some university	University degree
		per cent			
Central CMA's to Central Predominantly Rural Area. . .	100.0[2]	61.1	20.5	6.8	11.6
Central Predominantly Rural Area to Central CMA's . . .	100.0	60.2	22.2	7.1	10.6
Western CMA's to Western Non-CMA Area	100.0	52.4	26.5	8.9	12.3
Western Non-CMA Area to Western CMA's	100.0	51.2	32.2	8.3	8.4
Central CMA's to Western CMA's	100.0	39.4	27.3	13.9	19.5
Western CMA's to Central CMA's	100.0	30.3	26.0	12.7	31.0

[1] See Table 2.1, footnotes 4 - 8, for a precise definition of each region-group.
[2] Figures may not add to the total due to rounding error.

Source: 1971 Census of Canada, unpublished tabulation.

In considering the migration from a non-CMA region-group to a CMA region-group, the model is specified as:

$$\hat{\Pr}(X_Y | A_T \phi DM) = \Pr(X_Y) \cdot \left[\frac{\Pr(X_Y | M)}{\Pr(X_Y)} \cdot \frac{\Pr(X_Y | A_T)}{\Pr(X_Y)} \cdot \frac{\Pr(X_Y | D)}{\Pr(X_Y)} \right]^{1/3}$$

$$\cdot \left[\frac{\Pr(X_Y | A_T D)}{\Pr(X_Y | A_T)} \right]^{1/3} \cdot \lambda_{T\phi DM} \qquad (4)$$

In considering the migration from a CMA region-group to a non-CMA region-group, the model specified is:

$$\hat{P}r(X_Y|A_T\phi DM) = Pr(X_Y) \cdot \left[\frac{Pr(X_Y|M)}{Pr(X_Y)} \cdot \frac{Pr(X_Y|A_T)}{Pr(X_Y)} \cdot \frac{Pr(X_Y|\phi)}{Pr(X_Y)} \right]^{1/3}$$

$$\cdot \left[\frac{Pr(X_Y|A_T D)}{Pr(X_Y|A_T)} \right]^{1/3} \cdot \lambda_{T\phi DM} \qquad (5)$$

As the preceding discussion suggests, the chosen analysis procedure is implemented by considering specific pairs of regions between which migration streams flow. It would be too time-consuming to consider all the possible pairs of the seven region-groups. Indeed, for several of these pairs, the sample of migrants is so small as to be clearly unusable (see Table 2.4).

Even after editing and imputing the raw data, the published 1971 Census data show an unusually large proportion of 1966 - 71 migrants who failed to state their 1966 places of residence. For Canada as a whole, the figure exceeded 12%, whereas in the 1961 Census, the figure was close to 5% (see Stone, 1969, Appendix B). This and other known sources of deficiency in the census data would indicate that very large samples ought to be used when analysing distributions. In Table 2.4, a figure of 1,000 implies a sample size of only 300. It would seem advisable to limit the analysis to cases that had much larger samples than 300 respondents. Table 2.4 shows that 14 of the 42 migration streams involved samples of less than 300 persons. An additional seven streams had samples of less than 600 persons.

Sixteen of the remaining 21 streams involved migration to or from a Census Metropolitan Area. Where migration between a CMA and a non-CMA area was involved, the results of the analysis were so repetitive that a sample of two pairs of such areas suffice to show the basic patterns arising in the statistical analysis. Data are presented for movement between the Western CMA's and the Central Predominantly Rural Area, and between the Western CMA and the Western Non-CMA Area. An analysis was also conducted of the migration from Central CA's of 50,000 and Over to Central CMA's. Data for this case repeat the pattern for the cases mentioned above.

Where migration from one CMA to another was at issue, the pattern shown in the statistical analyses was basically similar for each pair of areas. Data on migration from Western CMA's to Central CMA's is shown in this paper, and they are sufficient to reveal the basic patterns. Data on migration from Western CMA's to Central CMA's will be presented below.

With regard to the remaining five streams among non-metropolitan areas, three were chosen for study: 1. migrants from Central CA's of 50,000 and Over to Central Predominantly Rural Areas, 2. migrants from the Atlantic Non-CMA Area to Central Predominantly Rural Areas, and 3. migrants from the Central Predominantly Rural Area to the Western Non-CMA Area.

Thus, the model was applied to a sample of at least one migration stream among the three major types of region-group defined for this study: Census Metropolitan Areas, large Census Agglomerations, and Predominantly Rural Areas. As was noted in Chapter 2, these three classes of region tend to show markedly different industry structures.

In presenting data in tabular form, details are omitted for all streams that did not involve a metropolitan area at origin or at destination. The reason for this is that the specification error already noted (and discussed in more detail in Appendix C, Section C.2) caused serious biases in the model fits for those streams. Generally, whenever a migration stream went from one non-metropolitan area to another, the proxy measures of regional characteristics did very little to enhance the performance of the model. Those variables are technically inappropriate (see Section C.2 of Appendix C), but they were retained because of fundamental defects in the census data. Although no detailed tables are provided for the class of migration streams mentioned above, the following substantive interpretations take into account the patterns revealed in the analyses for these streams.

4.1.2. Results of Analysis

Table 4.2 summarizes the general results for selected migration streams. Generally, the asserted model is superior to the null model. There is a systematic multivariate association of the schooling distribution of a migration stream with its age distribution and with characteristics of the stream's areas of origin and/or destination. Each of these factors contribute to the strength of the accuracy of the models' predicted schooling distributions of migration. There is no uniform statistical predominance of the "age effect" over the "regional effect". The higher-order effects of regional characteristics are consistently below (statistically) the levels of the zero-order effects. Except for the streams between the two parts of Western Canada, the residual migration-schooling interrelation is also a significant contributor (data not shown on table).

The failure of the model to show a high coefficient of association in the migration from Western CMA's to Central CMA's invites comment. A detailed study of the underlying data reveals that some 30% of those migrants in the population sample had a university degree. The estimates of the model's effect measures show very similar patterns of statistical effects stemming from the schooling compositions of "non-migrants" in the two region-groups. These "regional effects" were clearly stronger (statistically) than the age effect (as Table 4.2 shows), but none of the specified variables in the model tended to "push" the predicted proportion with university education anywhere close to the target of

30%. Since this proportion is far above the average (11%) for all migrants in the population sample, it does not reflect a persistent feature of migration streams that is independent of regional characteristics. It more likely reflects peculiarities of the metropolitan areas that are not being tapped by the specified predictor variables – in other words, serious specification error in setting the model to deal with the inter-metropolitan migration.

Appendix C, Section C.2, points up such specification error in the proxy variables chosen to represent regional characteristics, and explains the need to accept the error because of a fundamental defect in the census data. It is noted that the more technically appropriate specification of $Pr(X_Y | \phi)$, for example, involves estimation of data for the group of all the out-migrants from area ϕ. The rejection of this approach, on the grounds that it would produce a situation in which the variable ceases to serve its substantive purpose (that of reflecting relevant conditions within region ϕ), may not be necessary when dealing with migration among metropolitan areas. This seems to be so because all migration from metropolitan areas seems to be dominated, in its schooling distribution, by the metropolitan area characteristics.

Since the metropolitan area "non-migrants" already show educational attainment distributions roughly similar to the average for all migrants, it is apparent that the proportion with university education would be significantly enhanced in data that are confined to the metropolitan area migrants. Thus a re-specification of the model's terms that deal with regional attributes, along the lines indicated in the "ideal solution" mentioned in Section C.2 of Appendix C, would almost certainly lead to a much more adequate predictor of the schooling composition of inter-metropolitan migration than is indicated in Table 4.1.

The chosen proxy measures for regional attributes may also be seriously deficient in the statistical analyses done for streams flowing among non-metropolitan areas. Among the streams examined (they are listed above), the proxy measures for regional characteristics consistently lowered the predictive accuracy of the model. This is because these measures were dominated by the non-migrants within non-metropolitan areas whose general level of education was much lower than that of the migrants among those areas. Thus, these migration streams flowing among non-metropolitan areas showed variations that can only be partially predicted from their age compositions among the predictor variables measured. The central working hypothesis of the study breaks down when we wish to relate the characteristics of non-metropolitan areas to the educational composition of their migration streams.

A number of interpretations of this situation may be offered. Firstly, as in the case of the inter-metropolitan migration, it could be that regional factors continue to be pertinent but they remain untapped on account of specification of the measures for regional characteristics. The data do not strongly suggest this interpretation, however. The migration streams flowing among non-metropolitan

TABLE 4.2. Multivariate Analyses[1] of the Schooling Distribution of Migration, Selected Region-groups[2] of Canada, 1966 - 71

Item	Model asserted	Zero-order effects			Higher-order effects	
		Age	Region-group of origin	Region-group of destination	Region-group of origin given age	Region-group of destination given age
	Migration from Western CMA's to Western Non-CMA Area					
Relative contribution to reduction in chi-square[3]	100.0	15.5	15.8	[5]	6.9	[5]
Coefficient of association[4]	0.76	0.17	0.17		0.07	
	Migration from Central Predominantly Rural Area to Central CMA's					
Relative contribution to reduction in chi-square[3]	100.0	22.4	[5]	13.1	[5]	3.2
Coefficient of association[4]	0.78	0.23		0.14		0.03
	Migration from Western CMA's to Central CMA's					
Relative contribution to reduction in chi-square[3]	100.0	2.4	31.2	12.4	[6]	12.4
Coefficient of association[4]	0.33	0.03	0.33	0.14		0.12
	Migration from Western CMA's to Western Non-CMA Area					
Relative contribution to reductiion in chi-square[3]	100.0	15.5	28.2	[5]	22.7	[5]
Coefficient of association[4]	0.78	0.20	0.36		0.23	

See footnote(s) at end of table.

TABLE 4.2. Multivariate Analyses[1] of the Schooling Distribution of Migration, Selected Region-groups[2] of Canada, 1966-71 – Concluded

Item	Model asserted	Zero-order effects			Higher-order effects	
		Age	Region-group of origin	Region-group of destination	Region-group of origin given age	Region-group of destination given age
	Migration from Western Non-CMA Area to Western CMA's					
Relative contribution to reduction in chi-square[3]	100.0	17.2	[5]	29.1	[5]	26.4
Coefficient of association[4]	0.76	0.23	[5]	0.40	[5]	0.26

[1] See Appendix D for a specification of the sample universe and the categorization of the variables.

[2] See Table 2.1, footnotes 4, 6, 7 and 8, for a precise definition of each region-group.

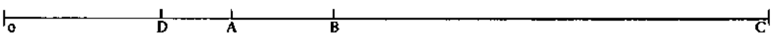
[3] The null model which denies any dependence between the dependent attribute and the explanatory ones generates the chi-square value with respect to which reduction is assessed. (This chi-square is roughly analogous to the variance to be explained in regression analysis (cf. Goodman, 1970, 1972).) The asserted model which does claim a particular pattern of dependence between the 'dependent' attribute and the explanatory ones normally generates a lower chi-square than the null model – thus the asserted model reduces or improves the chi-square of the null model. The reduction in chi-square attributable to the asserted model is the base on which we calculate the percentage contributions of the effects to the overall reduction in chi-square.

The 100% figure in the first column reminds the reader that the reduction in chi-square achieved by the asserted model is the total contribution. Each of the next five columns is the percentage of that total reduction in percentage that is attributed to a particular effect.

The percentage contributions of the effects are not necessarily additive. Additivity depends upon the precise definition of each effect and upon the existence of intercorrelations among the explanatory attributes.

[4] The coefficient of association is a measure of the strength of the relationship between a specified set of explanatory variables (effects in this case) and the designated dependent variable (the conditional distribution of the population over categories of the dependent attribute – schooling distribution of migrants in this case) within the context of a specified model. The first number in each row pertains to the overall relationship between all the specified effects, within the context of the asserted model, and the dependent variable. This measure is roughly analogous to R^2 in multiple regression analysis. The remaining numbers in each row are measures of partial association between a particular effect and the dependent variable, given that the other specified effects are "held constant" statistically. This measure is roughly analogous to partial r^2.

In assessing the partial coefficient of association for a single effect, we first generate the chi-square of a modified asserted model. This modification is achieved by deleting from the asserted model the effect in question plus all effects that are intrinsically related to the one in question (see Stone, 1978, Appendix C, Section C.4). We then compute the chi-square that is implied if we add to the modified asserted model only the effect in question. The difference between this latter chi-square and the one of the modified asserted model is the defined absolute reduction in chi-square attributable to the effect in question. This reduction is then divided by the difference between the chi-square of the null model and that of the modified asserted model (the latter measures the reduction in the chi-square of the null model that the modified asserted model achieves). The following diagram and symbols will illustrate the procedure:



Let A, B and C be chi-square values.

$A<B<C$

C is the chi-square of the null model.

B is the chi-square of the modified asserted model.

A is the chi-square that is obtained when the effect in question is added back into the modified asserted model.

$C - A$ is the reduction in chi-square achieved by the modified asserted model after it has been augmented by the effect in question.

$B - A$ is the measured contribution of the effect in question to $C - A$.

$(B - A) / (C - A) = \gamma$ is the defined coefficient of partial association for the effect in question.

The diagram above can also be used to illustrate the coefficient of association for the overall relationship between all the specified effects.

Let D be the chi-square of the asserted model.

$C - D$ is the reduction in chi-square achieved by the asserted model.

$(C - D) / C = \eta$ is the improvement in chi-square that the asserted model achieves, and is the defined coefficient of association for the whole asserted model.

[5] This effect is not applicable to the model for this migration stream.

[6] Formula number 4 for the model for the migration streams flowing between CMA's does not contain the first-order effect of region-group of origin given age. The pertinent higher-order effect is the first-order effect of age given origin. For this effect measure, the relative contribution to reduction in chi-square is 4.9 and the coefficient of association is 0.05.

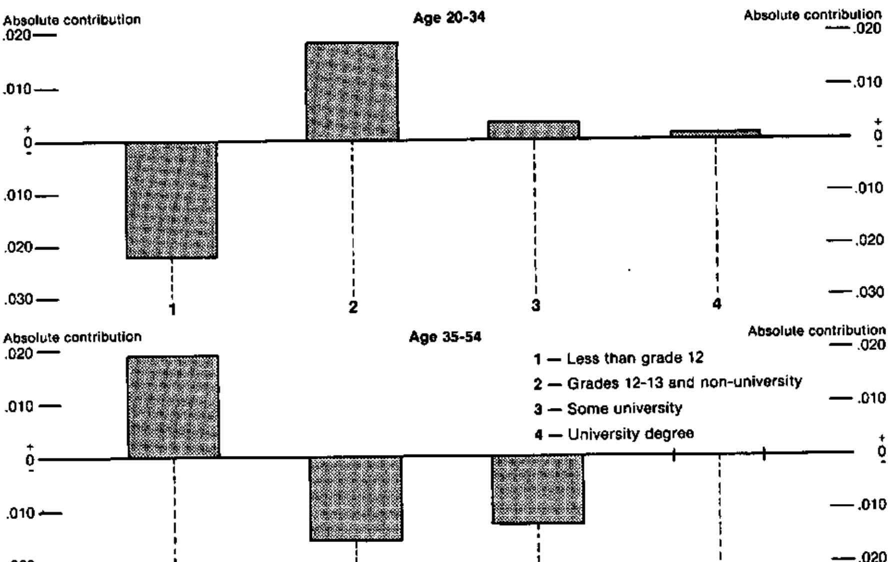
Source: 1971 Census of Canada, unpublished tabulation.

areas are somewhat less "highly educated" than the entire group of all migrants, but the divergence is neither sharp nor consistent over the various streams. For example, 11% of the migrants in the population sample had a university degree. The same figure holds true for the migrants from Atlantic Non-CMA Area to Central Predominantly Rural Area. Yet only 6% of the migrants from the latter areas to Western Non-CMA Area had a university degree. Since these percentages are much lower than those for the streams flowing among metropolitan areas, there is clearly still a hint of regional effects that the chosen proxy measure fails to reflect, but this hint must be considered weak.

The more plausible interpretation seems to be a basic inapplicability of the study's central working hypothesis to the migration streams flowing among non-metropolitan areas. These streams tend to bear the characteristic stamp of migrants (they are more highly educated than the rest of the population), and

Chart – 4.1

Absolute Contribution(1) of the Zero-order Effects of Age to the Schooling Distribution of the Migrants(2) from the Western Non-CMA Area to Western CMAs(3), 1966-71



(1) Let $\widehat{Pr}(X_Y | A_T ODM; H_I) \equiv Q_I$ be the proportion in category Y of X, given A_T,O,D,M, as **predicted** by a model that includes the effect in question. The corresponding **predicted** proportion when that effect is excluded is represented by $\widehat{Pr}(X_Y | A_T ODM; H_E) \equiv Q_E$. The absolute contribution of the effect to the level of the proportion for a specific category Y of X as predicted by the asserted model is measured by $Q_I - Q_E$. The model that produces Q_I may be the asserted model itself or a variant of it, depending on the order of the effect and the types of other effects specified in the model (see Appendix C, Section C.2). It is not necessary to use $Q_I - Q_E$ invariably. Sometimes $Q_E - Q_I$ is more appropriate, depending on the substantive hypothesis as to whether the effect in question ought to add to (or subtract from, as the case may be) the level of the proportion in category Y of X.

(2) See Appendix D for a specification of the sample universe and the categorization of the variables.

(3) See Table 2.1, footnotes 7 and 8, for a precise definition of each region-group.

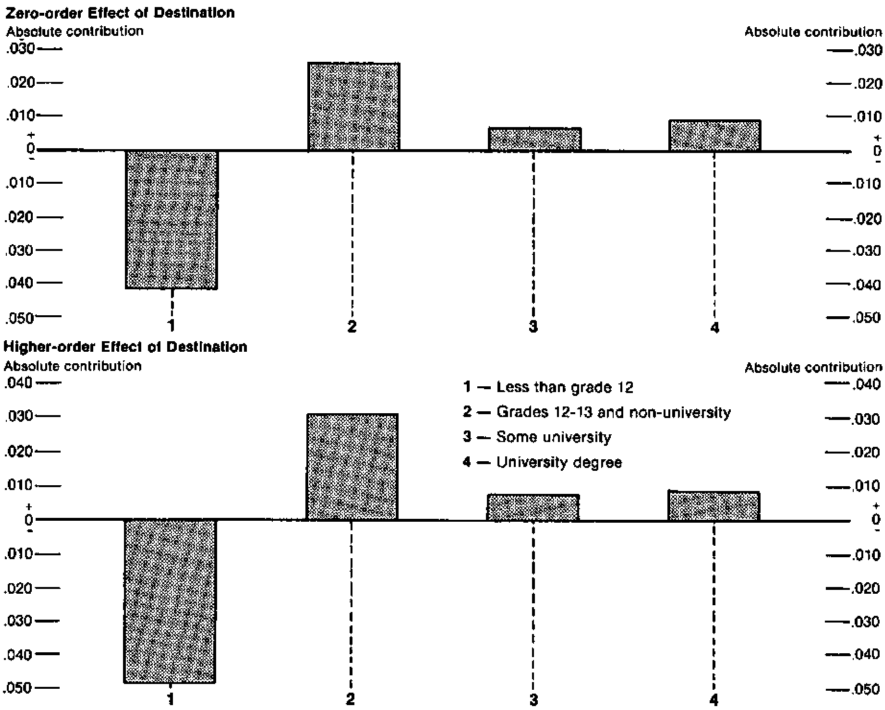
Source: 1971 Census of Canada, unpublished tabulation.

there is little in the regional attributes of non-metropolitan areas to produce much variation around the average educational pattern for migrants. In any event, the census data are fundamentally handicapped (because they measure population characteristics at the end of the migration period) in providing a basis for a more effective effort to identify the regional effects on migration stream composition. It is possible, however, that a different methodology might yield a more definitive outcome. Other researchers are invited to consider this prospect.

Charts 4.1 - 4.3 provide selected information concerning the directions of the contributions of the measured factors to the schooling distribution of the migration stream from the Western Non-CMA Area to Western CMA's. They show patterns that are generally repeated in the other streams that involve CMA's. Whether the metropolitan area region-group is the origin or the destination of a stream (for the migration stream involved in Chart 4.1, it is the destination), the

Chart — 4.2

Absolute Contribution(1) of the Effects of Region of Destination to the Schooling Distribution of the Migrants(2) from the Western Non-CMA Area to Western CMAs(3) 1966-71



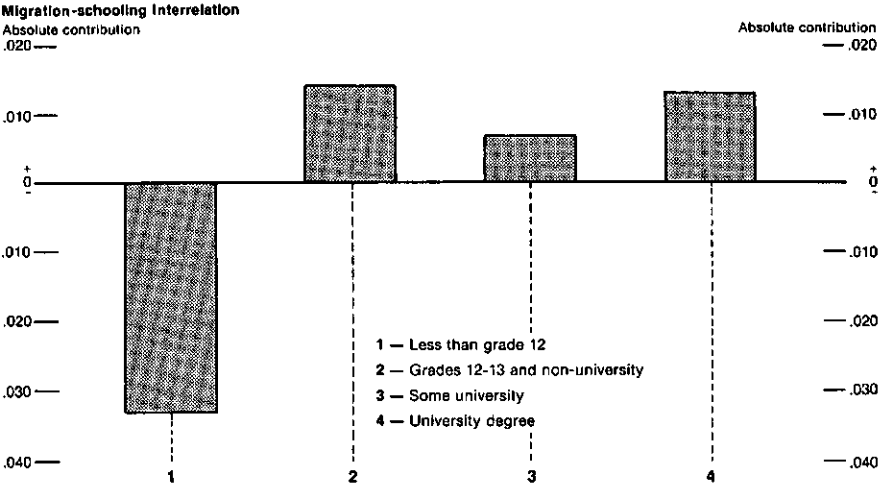
(1) See Chart 4.1, footnote 1.
(2) See Appendix D for a specification of the sample universe and the categorization of the variables.
(3) See Table 2.1, footnotes 7 and 8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished tabulation.

net effect of a metropolitan area's characteristics is to raise the level of education in migration streams going to and from CMA's. This observation is made even after age is allowed to "take out" its contribution.

Chart — 4.3

Absolute Contribution(1) of the Zero-order Effects of the Migration Schooling Interrelation to the Schooling Distribution of the Migrants(2) from the Western Non-CMA Area to Western CMAs(3), 1966-71



(1) See Chart 4.1, footnote 1.
(2) See Appendix D for a specification of the sample universe and the categorization of the variables.
(3) See Table 2.1, footnotes 7 and 8, for a precise definition of each region-group.

Source: 1971 Census of Canada, unpublished tabulation.

The question has been raised as to whether the measured "areal effects" on the educational composition of the migration streams are merely results of the concentration of higher-education institutions in metropolitan areas. Certainly much of the 1966 - 71 migration in the 20 - 34 age group involved movement to and from such institutions. To pursue this issue, the model was modified to eliminate the age attribute and applied separately to the data for the 20 - 34 and 35 - 54 age groups. Table 4.3 gives some pertinent results. Data for the migration from the Central Predominantly Rural Area to Central CMA's are shown in this table. It is in this particular migration stream that any effect of the spatial concentration of institutions for post-secondary education is most likely to be evident.

In the 20 - 34 age group, the measured area-of-destination effect actually detracts from the goodness of fit of the model (the possibility that an effect reduces the goodness of fit, when added to the model, arises because of the adopted estimation procedure) when the measure of the migration-schooling interrelation is also included. This means that after we take into account the average educational composition of all migration streams, the area-of-destination effect adds nothing to the goodness of fit. Since the great bulk of migrants were going to or from metropolitan areas, it is to be expected that the educational distribution of the migration stream to Central CMA's resembles the composition for all migrants. To reassess the area-of-destination effect, we computed its contribution before the measure of the migration-schooling interrelation was taken into account. When this was done, the area-of-destination effect reduced the chi-square of the null model by 50% (data not shown).

TABLE 4.3. Multivariate Analyses[1] of the Schooling Distribution of Migration from the Central Predominantly Rural Area to Central CMA's[2] for the Population in the Age Group 20 - 34 and for the Population in the Age Group 35 - 54, 1966 - 71

Item	Model asserted	Zero-order effects	
		Migration-schooling interrelation	Region-group of destination
	Age group 20 - 34		
Relative contribution to reduction in chi-square[3]	100.0	65.2	– 1.3
Coefficient of association[4]	0.95	0.65	0.01
	Age group 35 - 54		
Relative contribution to reduction in chi-square[3]	100.0	53.0	25.1
Coefficient of association[4]	0.83	0.53	0.25

[1] See Appendix D for a specification of the sample universe and the categorization of the variables.
[2] See Table 2.1, footnotes 4 and 5, for a precise definition of each region-group.
[3] See Table 4.1, footnote 2.
[4] See Table 4.2, footnote 3.

Source: 1971 Census of Canada, unpublished tabulation.

In the 35 - 54 age group, movement to and from institutions of post-secondary education could not have dominated the data. Table 4.2 shows that there is a considerable area-of-destination effect on the schooling composition of the migration even when the measure of migration-schooling interrelation is taken into account. In short, the influence of metropolitan area characteristics on the educational composition of the migration stream involves much more than the process of going to and from the post-secondary institutions located in metropolitan areas. This is not a surprising finding, since similar observations were made on the basis of 1961 Census data (Stone, 1969, Chapter 4).

4.1.3. Summary Remarks

This analysis started with a relatively simple question which we have addressed with less than ideal census data. We wanted to know whether it is possible to improve substantially the prediction of the educational composition of a migration stream by taking into account information about its age structure as well as about the educational compositions of the non-migrants living in the areas of origin and/or destination of the stream. Based on this analysis, the answer is affirmative when dealing with migrants to and from metropolitan areas. Taken together, the age distribution of a group of such migrants and the characteristics of the regions of origin and destination can to a great degree account for, in statistical terms at least, the educational composition of the migration stream. Each of the sets of explanatory factors has a substantively important contribution to the character of the educational attainment composition of a migration stream that is independent of the other factors within the context of the specified analytical model. None of these factors is of overwhelming statistical importance, as measured, relative to the others.

The age effect is well known. It operates consistently and independently of the stream's regions of origin and destination. The younger migrants tend to raise the overall educational level of the stream, and the older ones tend to lower it.

After the age effect is taken into account (or "held constant") statistically, numerous regional effects are evident in the educational composition of migration. The regional effect, measured through the educational composition of the regions' non-migrants, depends largely on whether a metropolitan area is involved in the migration flow. If the stream is flowing between a metropolitan area and a non-metropolitan one, then, regardless of the origin, the metropolitan area's characteristics will be predominant in moulding the educational composition of the migration flow.

The data also show a distinctive pattern of migration-schooling interrelation that cannot be accounted for by age or by the measured regional attributes. Migrants tend to be significantly better educated than the non-migrant population, independently of age and regional factors. The pattern of this interrelation is well known (see Stone, 1969, Chapter 3; McInnis, 1970). The present calculations show that the pattern does not arise merely because of the age distribution or because of the characteristics of the areas to and from which migrants go. Higher education may inherently tend to make people more mobile (see Stone, 1978) unless they are in certain occupations where success requires striking deep roots into a particular community. It appears that one must rely upon this interrelation in predicting the educational attainment pattern for migrants among non-metropolitan areas.

4.2. Occupational Composition of Migration Streams

In this section, the problem is to analyse variations in the occupational composition of a group of migrants. Fifteen broad occupation groups are used. A more detailed breakdown of occupation is highly desirable, but, for the reasons discussed in Chapter 2, a relatively crude breakdown is a practical necessity.

Special interest is attached to the influence of the characteristics of the regions of origin and destination on the occupational composition of a group of migrants. Very general reasons why the regional characteristics are deemed to be relevant are stated in Appendix B, and a more concrete discussion was given in Chapter 3. Essentially, these reasons emphasize the significance of the regional variations in economic structure, and particularly in the occupational composition of labour demand, in influencing the migration decisions of persons in different occupations. The educational composition of a group of migrants is also important in explaining its occupational structure. Part of the measured influence of education will arise because of the interrelations between regional characteristics and the educational composition of migration (as shown in the preceding section), but a part is also likely to be independent of regional factors.

The objectives that guided the analysis that follows are broadly similar to those outlined in Section 4.1, except here the dependent variable is the occupational distribution of migration. The aim is to measure the degree and pattern of a multivariate association between the chosen explanatory attributes (in this case schooling, characteristics of the region of origin and those of the region of destination) and the occupational composition of migration. The analysis also aims to assess and interpret the relative strengths and directions of the statistical contributions of the selected explanatory attributes that are included in the model that is put forward for measuring the multivariate association. The data base for the analysis is the same as that used for Section 4.1.

4.2.1. The Analytical Model

The general approach used to analyse the occupational composition of migration streams is broadly similar to that outlined for education (Section 4.1.1). The principal differences are these: Firstly, schooling replaces age as the specified explanatory compositional factor; secondly, effect measures for both place of origin and place of destination are always included in the model, since there is no a priori reason for such exclusion; thirdly, higher-order effects of schooling with destination and with origin are always included in the model.

Again, we follow the general estimation procedure outlined in Section 4.1.1. The data used to test the model are not the same as those used to estimate its parameters. Regional effects are estimated from tables dealing with the non-migrants in the region-groups. The "schooling effect" on the occupational composition of migration is estimated from a table for the entire population sample, non-migrants and migrants. This table provides the Canada-level conditional distribution of occupation given schooling. An average migration-occupation interrelation is estimated from data for all inter-municipal migrants taken together, regardless of their stream memberships.

Let "W_Y" mean the Yth category of occupation, and "X_T" mean the Tth category of schooling. Then the model can be expressed as:

$$\hat{Pr}(W_Y | X_T \phi DM) = Pr(W_Y) \cdot \left[\frac{Pr(W_Y | M)}{Pr(W_Y)} \cdot \frac{Pr(W_Y | X_T)}{Pr(W_Y)} \cdot \frac{Pr(W_Y | \phi)}{Pr(W_Y)} \cdot \frac{Pr(W_Y | D)}{Pr(W_Y)} \right]^{1/4}$$

$$\cdot \left[\frac{Pr(W_Y | X_T D)}{Pr(W_Y | X_T)} \cdot \frac{Pr(W_Y | X_T \phi)}{Pr(W_Y | \phi)} \right]^{1/4} \cdot \lambda_{T \phi DM} \qquad (6)$$

4.2.2. Results of Analysis

Table 4.4 is used to summarize the results of applying the model to the selected migration streams that involve metropolitan areas. Once again the asserted model is much superior to the null model. Knowing the educational composition of a migration stream and the economic structures of its areas of origin and destination considerably improves our ability to predict accurately the occupational composition of the stream. Seventy per cent or more of the chi-square of the null model is reduced (eliminated) by the asserted model, even though the observed occupational structure distribution of each stream was not used to estimate the model's parameters. The measure of the direct occupation-migration interrelation, $Pr(W_Y | M)/Pr(W_Y)$, does little to artificially enhance the strength of the model, in contrast with the schooling model (data not shown in table).

The compositional effect of schooling is consistently the dominant explanatory factor. The zero-order and first-order effects of this factor largely account for the performance of the model. As seen from the partial association coefficients associated with the first-order effect of schooling given region of origin, $Pr(W_Y | X_T \phi)/Pr(W_T | \phi)$, this attribute has an impact that varies with the nature of the region of origin (in addition to its separate zero-order contribution).

The patterns just mentioned hold true in all the observed migration streams – strong performance of the model and clear dominance of the schooling factor. When we turn to the other factors, substantial variation among migration streams is observed in the patterns of the coefficients.

TABLE 4.4. Multivariate Analyses[1] of the Occupation Distribution of Migration, Selected Region-groups[2] of Canada, 1966 - 71

Item	Model asserted	Zero-order effects			Higher-order effects	
		Schooling	Region-group of origin	Region-group of destination	Schooling given region-group of origin	Region-group of destination given schooling
	Migration from Central CMA's to Central Predominantly Rural Area					
Relative contribution to reduction in chi-square[3]	100.0	57.4	8.1	- 0.9	38.4	0.5
Coefficient of association[4]	0.72	0.93	0.13	- 0.01	0.38	0.00
	Migration from Central Predominantly Rural Area to Central CMA's					
Relative contribution to reduction in chi-square[3]	100.0	48.7	– 3.7	12.7	34.2	5.1
Coefficient of association[4]	0.79	0.81	0.06	0.13	0.34	0.05
	Migration from Western CMA's to Central CMA's					
Relative contribution to reduction in chi-square[3]	100.0	44.9	4.4	6.2	25.6	3.6
Coefficient of association[4]	0.74	0.64	0.06	0.06	0.26	0.04
	Migration from Western CMA's to Western Non-CMA Area					
Relative contribution to reduction in chi-square[3]	100.0	53.3	8.0	9.2	33.9	5.0
Coefficient of association[4]	0.71	0.89	0.12	0.10	0.34	0.05
	Migration from Western CMA's to Western Non-CMA Area					
Relative contribution to reduction in chi-square[3]	100.0	52.0	- 6.4	14.0	36.9	6.2
Coefficient of association[4]	0.77	0.91	- 0.10	0.15	0.37	0.06

[1] See Appendix D for a specification of the sample universe and the categorization of the variables.
[2] See Table 2.1, footnotes 4, 6, 7 and 8, for a precise definition of each region-group.
[3] See Table 4.1, footnote 2.
[4] See Table 4.2, footnote 3.

Source: 1971 Census of Canada, unpublished tabulation.

As was seen with the schooling distribution of migration, the occupational composition of the non-metropolitan areas exerted little influence on the structure of the migration streams flowing between non-metropolitan areas and metropolitan areas. This holds true whether or not the non-metropolitan areas are origins or destinations.

In the migration from Western CMA's to Central CMA's, the regional factors all have magnitudes far below the levels of the zero-order and first-order effect of schooling. This condition is a direct result of the fact that the occupational structures of the metropolitan areas' non-migrant populations are quite similar, and are not greatly different from that which is predictable on the basis of knowing the schooling composition of the migration only.

A brief comment is in order concerning the results for migration streams among non-metropolitan areas. Analysis was done for two of these streams: 1. migrants from Atlantic Non-CMA Area to Central Predominantly Rural Area, and 2. migrants from Central CA's of 50,000 and Over to Central Predominantly Rural Area. The schooling variable remained dominant in these analyses. The proxy measures for "regional effects" showed much poorer performance than they did in the streams covered in Table 4.3.

Table 4.5 shows that the results reported above are not merely a consequence of the mobility patterns of the young-adult group (aged 20 - 34) moving into and out of schools. This table presents the results of the multivariate analysis of the occupational composition of the migration stream aged 35 - 54 from the Central Predominantly Rural Area to Central CMA's. The pattern of variation shown in this table is quite similar to that for the same migration stream in Table 4.4. There is a strong zero-order effect of schooling, followed by a much weaker but notable zero-order region-of-destination effect.

TABLE 4.5. Multivariate Analyses[1] of the Occupation Distribution of Migration from the Central Predominantly Rural Area to Central CMA's[2] for the Population in the Age Group 35 - 54, 1966 - 71

Item	Model asserted	Zero-order effects			Higher-order effects	
		Schooling	Region-group of origin	Region-group of destination	Schooling given region-group of origin	Region-group of destination given schooling
Relative contribution to reduction in chi-square[3]	100.0	50.8	1.6	10.6	30.0	3.8
Coefficient of association[4]	0.72	0.77	0.02	0.11	0.30	0.04

[1] See Appendix D for a specification of the sample universe and the categorization of the variables.
[2] See Table 2.1, footnotes 4 and 5, for a precise definition of each region-group.
[3] See Table 4.1, footnote 2.
[4] See Table 4.2, footnote 3.

Source: 1971 Census of Canada, unpublished tabulation.

Charts 4.4 - 4.7 show data on the patterns of the directions of the schooling effects and regional factors. These are net effects, effects measured after the other factors in the model are held constant statistically. The patterns shown are just what one would expect from the literature on the occupational impact of schooling, and on regional differences in occupational structure. Generally, as the level of education rises so does the concentration of the migrants in predominantly professional, administrative and skilled occupations, even after the regional factors are taken into account. The occupational structure of the working force in a region tends to place its stamp on the composition of a migration flow to or from that region (the data show this pattern even for effects that make a weak contribution to the strength of the overall performance of the model). However, that stamp is in competition with the characteristics of the other region involved in the stream, and will likely be dominant only if the region in question is a metropolitan area (while the other is not).

4.2.3. Summary Remarks

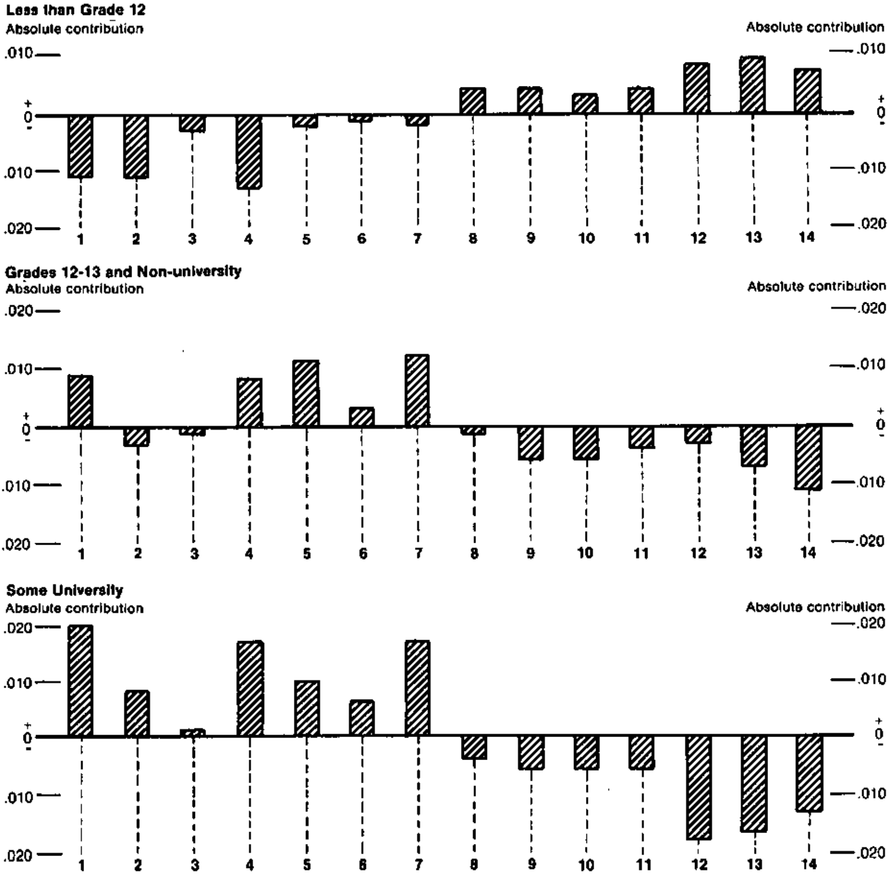
To speculate usefully about the likely occupational composition of a stream of migration from one region to another, we should look first for information about the distribution of the migrants by schooling. According to our analysis based on the 1971 Census data alone, we will go a long way toward adequate prediction of the migrants' occupational mix (given the selected occupation grouping), even if we know nothing about their regions of origin and destination. The educational distribution, however, is a function of the migrants' age composition and of characteristics of the areas of origin and/or destination.

Further improvements in the accuracy of predicting the occupational composition of a migration stream are achieved by taking into account the economic structures (and their associated occupational structures) of the areas of origin and destination. If one area is metropolitan and the other is not, primary emphasis should be placed on the economic structure of the metropolitan area. If both areas are metropolitan, the improvement in prediction accuracy afforded by the measurement of regional characteristics after the intervening variable of schooling is considered may be marginal. Regional characteristics tend to be marginal when (a) the regions resemble each other closely in economic structure, and (b) the occupational distributions of their non-migrant work forces are similar to those which are predicted for the migrants, based upon the schooling composition of the migration.

In general, migration streams with relatively high proportion of migrants with post-secondary educational attainment show higher than average concentrations of managers, professionals, and some related white collar occupations. Similar concentrations in the occupational distribution of migration are evident for the streams moving into and out of the regions with more sophisticated industrial economies (the metropolitan areas in particular).

Chart — 4.4

Absolute Contribution(1) of the Zero-order Effect of Schooling to the Occupational Distribution of the Migrants(2) from Western CMAs to the Western Non-CMA Area(3), 1966-71

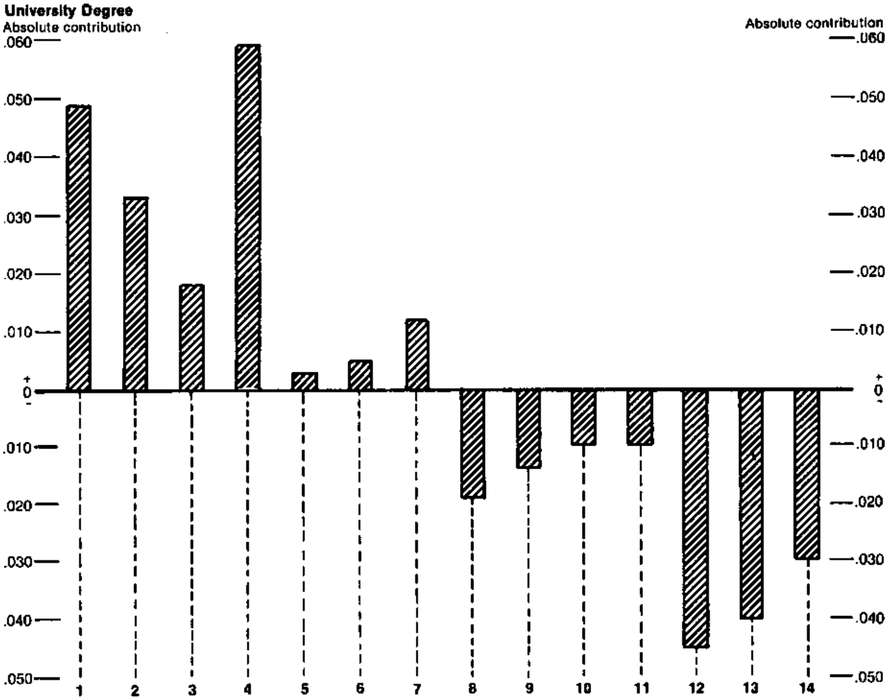


1 — Managerial, administrative and related occupations
2 — Teaching and related occupations
3 — Occupations in medicine and health
4 — Technological, social, religious, artistic and related occupations
5 — Clerical and related occupations
6 — Selected sales occupations(4)
7 — Other sales occupations
8 — Service occupations (excluding armed forces)
9 — Farming, horticultural and animal husbandry occupations
10 — Other primary occupations(5)
11 — Processing occupations
12 — Machining and product fabricating, assembling and repairing occupations
13 — Construction trades occupations
14 — Transport equipment operating occupations

Chart — 4.4 (concluded)

Absolute Contribution(1) of the Zero-order Effect of Schooling to the Occupational Distribution of the Migrants(2) from Western CMAs to the Western Non-CMA Area(3),1966-71

University Degree



1 — Managerial, administrative and related occupations
2 — Teaching and related occupations
3 — Occupations in medicine and health
4 — Technological, social, religious, artistic and related occupations
5 — Clerical and related occupations
6 — Selected sales occupations (4)
7 — Other sales occupations
8 — Service occupations (excluding armed forces)
9 — Farming, horticultural and animal husbandry occupations
10 — Other primary occupations (5)
11 — Processing occupations
12 — Machinery and product fabricating assembling and repairing occupations
13 — Construction trades occupations
14 — Transport equipment operating occupations

(1) Let $\widehat{Pr}(W_y | X_T ODM; H_I) \equiv Q_I$ be the proportion in category Y of W, given X_T,O,D,M, as "predicted" by a model that includes the effect in question. The corresponding **predicted** proportion when that effect is excluded is represented by $\widehat{Pr}(W_y | X_T ODM; H_E) \equiv Q_E$. The absolute contribution of the effect to the level of the proportion for a specific category Y of W as predicted by the asserted model is measured by $Q_I - Q_E$. The model that produces Q_I may be the asserted model itself or a variant of it, depending on the order of the effect and the types of other effects specified in the model (see appendix C, section C.2).
It is not necessary to use $Q_I - Q_E$ invariably. Sometimes $Q_E - Q_I$ is more appropriate, depending on the substantive hypothesis as to whether the effect in question ought to add to (or subtract from, as the case may be) the level of the proportion in category Y of W.

(2) See Appendix D for a specification of the sample universe and the categorization of the variables.

(3) See Table 2.1, footnotes 7 and 8, for a precise definition for each region-group.

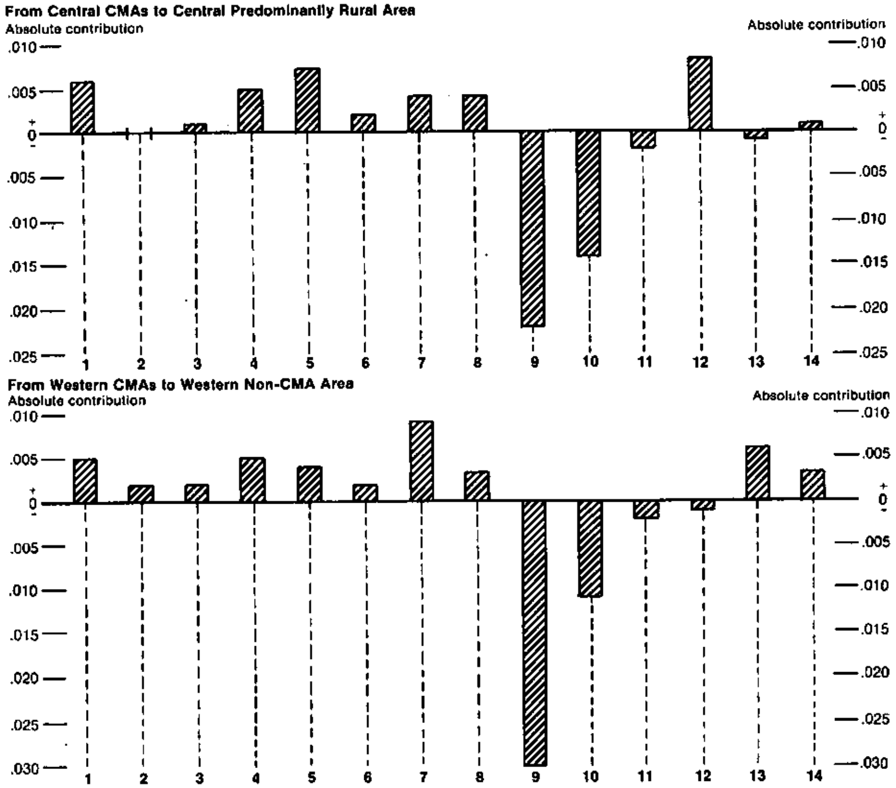
(4) See Table 2.5, footnote 1.

(5) See Table 2.5, footnote 2.

Source: 1971 Census of Canada, unpublished tabulation.

Chart — 4.5

Absolute Contribution(1) of the Zero-order Effect of Origin to the Occupational Distribution of the Migrants (2) from Central CMAs to Central Predominantly Rural Area and from Western CMAs to Western Non-CMA Area(3), 1966-71



1 — Managerial, administrative and related occupations
2 — Teaching and related occupations
3 — Occupations in medicine and health
4 — Technological, social, religious, artistic and related occupations
5 — Clerical and related occupations
6 — Selected sales occupations (4)
7 — Other sales occupations
8 — Service occupations (excluding armed forces)
9 — Farming, horticultural and animal husbandry occupations
10 — Other primary occupations (5)
11 — Processing occupations
12 — Machining and product fabricating, assembling and repairing occupations
13 — Construction trades occupations
14 — Transport equipment operating occupations

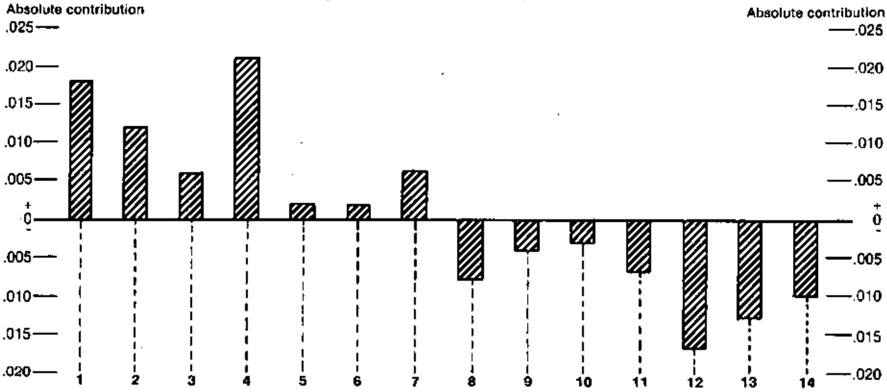
(1) See Chart 4.4, footnote 1.
(2) See Appendix D for a specification of the sample universe and the categorization of the variables.
(3) See Table 2.1, footnotes 4, 6, 7 and 8, for a precise definition of each region-group.
(4) See Table 2.5, footnote 1.
(5) See Table 2.5, footnote 2.

Source: 1971 Census of Canada, unpublished tabulation.

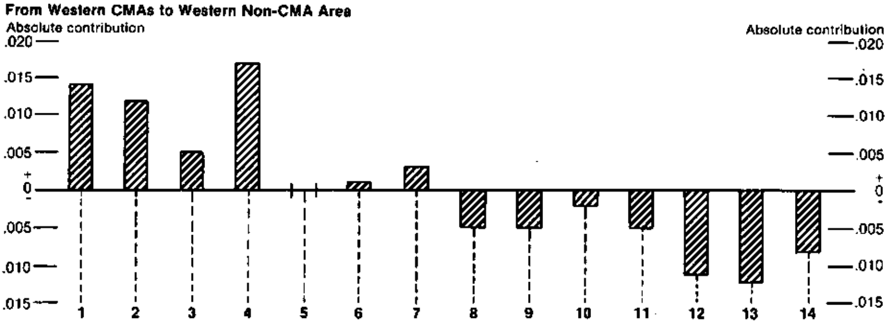
Chart — 4.6

Absolute Contribution[1] of the Higher-order Effects of Origin to the Occupational Distribution of the Migrants[2] from Central CMAs to Central Predominantly Rural Area and from Western CMAs to the Western Non-CMA Area[3], 1966-71

From Central CMAs to Central Predominantly Rural Area



From Western CMAs to Western Non-CMA Area



1 — Managerial, administrative and related occupations

2 — Teaching and related occupations

3 — Occupations in medicine and health

4 — Technological, social, religious, artistic and related occupations

5 — Clerical and related occupations

6 — Selected sales occupations[4]

7 — Other sales occupations

8 — Service occupations (excluding armed forces)

9 — Farming, horticultural and animal husbandry occupations

10 — Other primary occupations[5]

11 — Processing occupations

12 — Machining and product fabricating, assembling and repairing occupations

13 — Construction trades occupations

14 — Transport equipment operating occupations

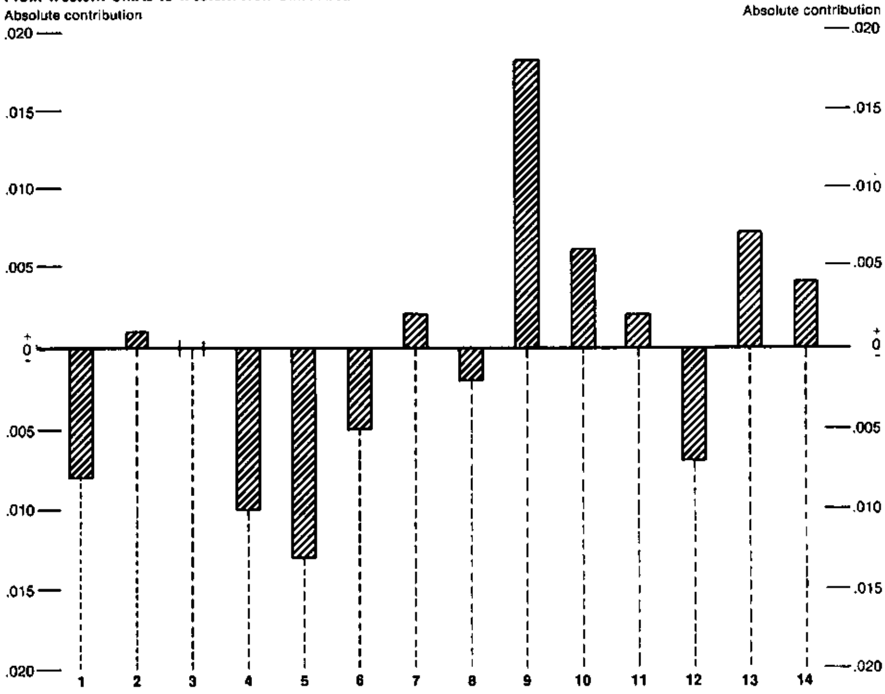
(1) See Chart 4.4, footnote 1.
(2) See Appendix D for a specification of the sample universe and the categorization of the variables.
(3) See Table 2.1, footnotes 7 and 8, for a precise definition of each region-group.
(4) See Table 2.5, footnote 1.
(5) See Table 2.5, footnote 2.

Source: *1971 Census of Canada, unpublished tabulation.*

Chart — 4.7

Absolute Contribution (1) of the Zero-order Effect of Destination to the Occupational Distribution of the Migrants (2) from Western CMAs to the Western Non-CMA Area (3), 1966-71

From Western CMAs to Western Non-CMA Area



1 — Managerial, administrative and related occupations
2 — Teaching and related occupations
3 — Occupations in medicine and health
4 — Technological, social, religious, artistic and related occupations
5 — Clerical and related occupations
6 — Selected sales occupations(4)
7 — Other sales occupations
8 — Service occupations (excluding armed forces)
9 — Farming, horticultural and animal husbandry occupations
10 — Other primary occupations(5)
11 — Processing occupations
12 — Machining and product fabricating, assembling and repairing occupations
13 — Construction trades occupations
14 — Transport equipment operating occupations

(1) See Chart 4.4, footnote 1.
(2) See Appendix D for a specification of the sample universe and the categorization of the variables.
(3) See Table 2.1, footnotes 4, 6, 7 and 8, for a precise definition of each region-group.
(4) See Table 2.5, footnote 1.
(5) See Table 2.5, footnote 2.

Source: 1971 Census of Canada, unpublished tabulation.

CONCLUDING COMMENTS

This has been a study in the demography of Canada. Applying to 1971 Census data some modifications and recently developed techniques for the multivariate analysis of cross-tabulations, the study indicates that information about the economies and population characteristics of the regions of origin and destination of migration can markedly improve our prediction of the migration stream's occupational distribution. The models that are articulated here and the related estimation procedures may be applicable in preparing intercensal estimates or in macro-simulation studies of the educational and occupational structures of migration flows. The full tables of the pertinent coefficients for effect measures are not printed here, but they are available by writing to the author.

Breaking ground into a relatively new area of migration research with techniques that accommodate the shortcomings of census cross-tabulations has been a major objective of this study. This work raises far more questions than it answers, as one should expect in ground-breaking exercises. It surely points up the severe limitations of data that indicate changeable characteristics of migrants only at the end of the migration interval.

The methodology used, multiplicative power models, is an elaborate application of a standardization strategy which has been used in demography for decades. Systematic patterns of coefficients that reflect statistical relationships among variables are shown effectively by this technique, if those relationships are strong in the underlying data; but these patterns only serve to raise questions as to what is going on behind them. This, of course, is a property shared by all forms of multivariate analysis.

APPENDIX A

1971 CENSUS MIGRATION DATA

A.1. Census Questions and Sample Estimation Procedure

The 1971 Census migration data used in this study are based mainly on two questions that were administered to a one-third systematic sample of private households in all self-enumeration areas, to all households in canvasser areas, and to all permanent residents of collective dwellings. In the 1971 Census, 97% of the population was enumerated by the self-enumeration method, while 3% was enumerated by the more traditional personal interview method. This 3% consisted mostly of residents of remote areas (northern areas of 10 provinces, Northwest Territories, Yukon, etc.) and residents of institutions.

Some general features of census migration data and their major sources of error were outlined in Appendix B of a 1961 Census monograph on migration (Stone, 1969). Readers who are interested in exploring data quality issues should consult that reference, since the general points made there will not be repeated below. For details of the sampling methods used for the 1971 Census, readers are referred to Dodds (1971).

The households of Canada were divided into two types for the 1971 Census. The first was a private household, which consisted of a person or a small group of persons occupying an ordinary dwelling. The second, the collective type of household, included hotels, large lodging-houses of 10 or more lodgers, institutions, hospitals, military camps, lumber camps, and other establishments of a similar nature. Persons living in collective households were subdivided into two groups, permanent and temporary. Permanent residents of collective dwellings had no usual place of residence elsewhere in Canada and were counted as part of the population of the collective dwelling. Temporary residents were enumerated at the collective dwellings, but included in the population count of their usual place of residence. In 1971, the number of Canadians residing in private households was approximately 97.5% of the total population.

The two questions which yielded the great bulk of the data used in this study were:

26. Where did you live 5 years ago, on June 1, 1966 ?

- ○ Same dwelling ——→ **SKIP TO QUESTION 28**
- ○ Same city, town, village or municipality (*not same dwelling*)
- ○ Outside of Canada
- ○ Different city, town, village or municipality in Canada, give its name ——↓

City, town, village, municipality, etc.

---------------------- ----------------------
County Province

IMPORTANT: *If outside city or town limit, specify name of suburban municipality and not of city or town.*

27. How many times have you **MOVED** from one Canadian city, town, village or municipality to another since June 1, 1966 ?
Count moving away and returning to the same place as 2 moves.

○ None	○ 2	○ 4
○ 1	○ 3	○ 5 or more

In addition, the study also made some use of mobility data generated from another question addressed to persons who were not born in Canada. This question is as follows:

12. If born **OUTSIDE** Canada, in what period did you first immigrate to Canada?

○ Before 1931	○ 1956–1960	○ 1967–1968
○ 1931–1945	○ 1961–1964	○ 1969
○ 1946–1950	○ 1965	○ 1970
○ 1951–1955	○ 1966	○ 1971

Unfortunately, year of immigration to Canada was not asked for persons who were born in Canada, but who had emigrated to another country and were returning as immigrants to Canada.

These questions were asked of all persons in the sample who were at least 15 years old on 1 June 1971 (i.e., born before 1 June 1956). Persons in the sample who were born since 1 June 1966 were excluded from the migration data. For family persons aged 5 - 14 on 1 June 1971, the migration status of the head of the family was assigned. For non-family members aged 5 - 14, the mobility status of the household head was assigned. In addition to the mobility status assignments for persons aged 5 - 14, there was also an elaborate procedure of data editing and mobility status assignments for persons who gave certain kinds of incomplete answers to the mobility questions (see Section A.2).

Estimates of totals were derived from the sample responses by a ratio estimation procedure that was a sophisticated version of the one used for the 1961 Census (see Stone, 1969, p. 329). The ratio estimation technique made use of population figures from the 100% count to inflate figures drawn from the sample counts. The estimate for the total population having a characteristic "x" is given by a formula that has the following general form:

$$N_X = C_X \cdot \left(\frac{N}{C}\right) \text{where}$$

C_X is the sample count of persons with characteristic x; N/C is a weight; C is a function of the size of the sample; and N is a function of the size of the total population.

In principle, the functions N and C are defined for a particular subgroup of the population – e.g., males aged 20 - 34, and residing in Ottawa. The attributes in terms of which the subgroup is defined were all covered in the 100% count of the 1971 Census. Thus, the functions could be evaluated by simply letting N be the total number of enumerated members of the designated subgroup, while C is the portion that actually fell in the sample. However, when we sum N_X over all values of x and over all categories of the attributes that define the above-mentioned subgroup, we would not necessarily obtain the same total as that enumerated in the 100% count. To resolve this problem, a multistep calculation procedure was used to arrive at the chosen value of N/C.

The first step in the calculation of the weight, N/C, was the determination of a geographical level where agreement between sample estimates and comparable population counts was to be ensured. The next step was to specify the subgroups to which the estimator would be applied. The subgroups were defined in terms of individual cells of a cross-classification of the variables: language (English, French or other), age, sex, marital status, whether or not a person's residence is a farm, the person's status within his/her family, and his/her family's composition. Weights were then calculated for each cell (defining a specific subgroup) of the cross-classification in such a way that selected sums (over cells) of the estimates equalled predetermined control totals drawn from the 100% census count. In order to achieve this result, an iterative calculation algorithm was used (Nargundkar, 1971; and Brackstone, 1971). The final step in the calculation of the weights was the conversion of the weights to integers. This was an innovation from the 1961 Census where fractional weights were accepted.

A.2. Data Editing and Imputation

The scale of data editing and imputation for the migration questions was much greater with the 1971 Census returns than with those of the 1961 Census. There was considerable imputation of responses for many persons who failed to provide complete answers to the questions on migration. Generally, where the respondent failed to answer a portion (or the whole) of the migration questions an artificial answer was coded on his record (using a computer programme).

A definite priority order was assigned to the source of the auxiliary information used to estimate the missing response. Firstly, related, presumably correct information provided by the respondent was used. For example, if a person recorded that he/she immigrated to Canada in a year after 1966 (Question 12), then a portion of his/her response to the five-year migration question could be imputed.

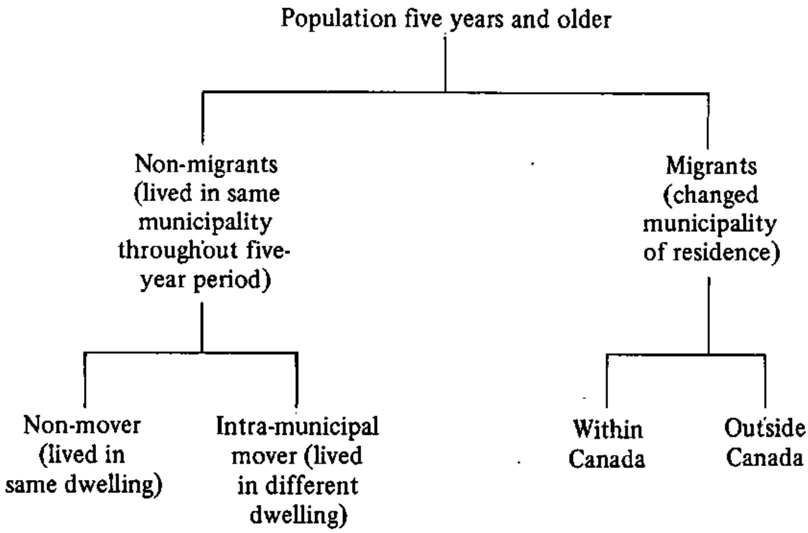
Secondly, presumably correct information provided by members of the pertinent respondent's family was used. Priority was given to close relatives where a match existed between certain information on a close relative's record and that on the respondent's record. For example, if a spouse was found, and if it was determined that the spouse and the respondent had the same date of first marriage, then the migration information recorded by the spouse was coded on the respondent's record, assuming that the latter was incomplete.

Thirdly, an artificial response was imputed from information given by the last respondent examined who had the same values on a specific set of variables (e.g., sex, age, and education) as those of the pertinent respondent.

A.3. Mobility Concepts

Question 26 on the 1971 Census questionnaire is the basis of several of the mobility concepts used in this report and in census bulletins. This question yields what are known as "five-year migration data", whose main features and limitations have already been outlined in Stone (1969, pp. 6 - 8, 329 and 330). In this study, the following network of migration concepts was used:

Migration Status



Migrants are persons who changed municipality of residence during the five-year period, 1 June 1966 to 1 June 1971;

non-migrants are persons who were living in the same municipality throughout the five-year period; and

intra-municipal movers are persons who were living in the same municipality throughout the five-year period but who were living in a different dwelling on 1 June 1966 and 1 June 1971.

It is important to note that unlike the 1971 Census bulletins or the 1961 monograph on migration, this study includes, in the category of migrants, people who had changed municipality of residence at least once between 1 June 1966 and 1 June 1971, but who reported themselves as residing in the same municipality at both dates. Except for this difference, however, the concepts are the same as those used in 1971.

A.4. Evaluation of Data

As a result of the edit and imputation procedure described in Section A.2, no totals of non-respondents from the census main file gives, with assurance, an adequate picture of the non-respondents' attributes. To obtain that picture, we would need a representative sample of the census records prior to the initiation of the imputation steps. Five per cent of the records were preserved in the form they had before the initiation of the computer edit procedure. However, difficulties in accessing the file of unimputed data and peculiarities in the format of the data on this file preclude the analysis of these records within the time and other resources available for the production of this monograph. The records could not be truly described as unedited because they have been subjected to some clerical editing procedures. In consequence, no 1971 Census tables similar to the key tables shown in Appendix B of the author's 1961 Census monograph are available. We are simply unable at this time to comment extensively on the possible biases in the census migration data because of incomplete response to the migration questions.

As was the case at the time of writing the 1961 Census migration study, the data quality evaluation studies for the 1971 Census migration data are not available (cf. Norland, *et al.*, 1977). The possible types of evaluations and the relevant sources of error have been previously outlined in Stone (1969, pp. 330 and 331).

It is hoped that a 1971 Census Evaluation Study that includes the quality of responses to the migration questions may be undertaken in due course, perhaps based on a sample of unedited records obtained from the original questionnaires. Secondly, this influence depends on flows of information and on the access that the potential migrant has to such information flows. Thirdly, there are time changes in the volume of information concerning opportunities. Fourthly, regional factors exert an influence on the interaction among different types of mobility. For example, if employment opportunities are concentrated in specific occupations while these occupations are themselves concentrated in specific regions, one may expect that many of the people who wish to enter such occupations (i.e., to be occupationally mobile) will have to migrate to those regions. (For more details, see Stone, 1973a and Stone, 1975.)

One aspect of the influence of regional factors may be outlined in terms of the following illustrative example concerning the occupational composition of a migration stream. The composition of a migration stream for a given socio-economic variable, e.g., occupation, is indicated by a collection of proportions (the set of proportions of migrants in each occupation group). Thus, when we talk about relating stream composition to regional characteristics, we are implicitly referring to relations among sets of numbers. To help keep this discussion concrete, let us consider just one number from the set that describes stream composition – e.g., the proportion of migrants who are financial analysts. Now consider two migration streams, say one flowing from Montréal CMA to Toronto CMA and the other flowing from Montréal CMA to Timmins. Suppose the proportion of financial analysts is greater in the former than in the latter: This differential would be one element of the variation in composition between the two streams.

APPENDIX B

REGIONAL CHARACTERISTICS AND THE SOCIO-ECONOMIC COMPOSITION OF MIGRATION STREAMS

Chapter 3 has outlined some ways in which the characteristics of the regions of origin and destination of a migration stream may influence its occupational composition. That discussion does not make explicit some of the pertinent theoretical assumptions that need to be mentioned. Appendix B provides some of the omitted assumptions. (The following comments are a summary of more extended discussion contained in Stone, 1973a.)

The theoretical rationale provided in Chapter 3 addressed itself basically to the question of why one would expect a relation between the socio-economic composition of a migration stream and the characteristics of the stream's regions of origin and destination. In approaching this question, it is desirable to begin with some thoughts on the explanation of migration decisions, and on the integration of factors operating at the regional level with processes of decision-making at the individual or household level.

A migration stream, a macro-level phenomenon, is basically the collective result of processes at the micro-level. Thus, the explanatory theory should include some postulates on the subject of the links between factors operating at the regional level and the migration decision-making processes. One of the basic postulates is that in the decision-making of potential migrants who occupy a specific socio-economic status and reside in the same region, a given regional factor tends to exert a relatively similar influence.

Additional postulates concern the probability of an individual migrating from one region to another. This probability is a function of attributes of the individual, relevant regional characteristics or "opportunities" whose "values" vary from one region to another, information flows about such characteristics and the perception of such information by potential migrants. In general we would expect that, for any potential migrant from a given region, the probability of moving to another region will increase as the values of the potential destination on relevant economic opportunity dimensions exceed those of the region of origin. However, the influence of these regional-level factors is mediated by the information flows about relevant economic and social conditions among the regions. These flows affect the potential migrant's perception of that region's attractive conditions and amenities, which in turn enhances the probability of moving to that region. This probability is, however, also a function of household-level factors like life-cycle changes, household size changes and social ties in the region of current residence. Thus, there is a whole network of economic, social and demographic forces working upon the mobility probability – some tending to increase it, others to decrease it, and many of the forces interact with each other.

Macro-level factors exert at least four types of influence on the process of migration decision-making. Firstly, there is the actual spatial distribution of various opportunities which influence the migration decision.

For potential out-migrants from Montréal who are financial analysts, Toronto and Timmins have networks of characteristics that may be said to comprise attraction forces. From the vantage point of potential out-migrants (financial analysts) from Montréal, the Toronto attraction forces are far more weighty than those of Timmins. The actual migrating financial analysts are thus much more likely to choose Toronto over Timmins. Then, depending on the relative volumes of other kinds of migrants from Montréal to Timmins and to Toronto, we can expect the flow to Toronto to show a higher proportion of financial analysts. In general, it is hypothesized that where two regions are receiving migrants from a common origin, variation in their structures of migration attractions will be related to differences between the two stream compositions. Of course, the structure of migration attractions is essentially unobservable. We shall assume that various aspects of this structure are indicated in different ways by several regional measures with respect to (a) characteristics of economic activities, (b) population and working force compositions, and (c) other factors.

We can briefly indicate the nature of some of the forces underlying the linkage of the structure of regional attractions to migration stream composition for occupation. Both the demand and the supply of labour in a given region have certain skill structures that reflect the character of the region's economy. The processes of replacing lost workers in given occupations, of recruiting new workers to expanding occupations, and of achieving working force readjustments necessitated by economic structure changes are not normally achieved adequately (in the short run at least) by retraining, labour force entries and exits, and occupational mobility within the same region. Occupation-specific labour demand helps trigger interregional mobility. The structure of such demand is an important aspect of the structure of migration attractions. To quote Blau and Duncan (1967, p. 243) " . . . the division of labor varies from one type of locality to another, and with it the opportunity structure that affects men's careers. . . . Migration provides a social mechanism for adjusting the geographical distribution of manpower to the geographical distribution of occupational opportunities".

APPENDIX C

MULTIPLICATIVE POWER MODELS FOR THE MULTIVARIATE ANALYSIS OF CONTINGENCY TABLES

The purpose of this appendix is to supplement the methodological discussion provided in Chapter 3 in order to offer further information about the type of analysis being used. A variety of features of the chosen technique of analysis are outlined briefly. A more complete discussion is outside the limits of this study, but several references will be made to Appendix C of Stone (1978). Still further details are provided in Stone (1975a); Goodman (1972, 1973a, 1973b); and Bishop, Fienberg and Holland (1974).

The method whose features are outlined here is designed for use when a problem of explanatory analysis arises in the field of population composition with data that are in the form of cross-tabulations (contingency tables). When such a problem arises, there should be, where feasible, a substantive theory that serves to rationalize either by strict deduction or plausible argument a specific explanatory model.[1]

Typically, the model is then applied empirically in the conduct of one type of multivariate analysis of a contingency table, where it postulates a definite set of dependent attributes and some specific causal ordering among these and a set of explanatory attributes.[2] (The kind of model being used here is called "multiplicative power model".) A variable may be explanatory in one equation but dependent in another. Use of the model may serve a variety of objectives. One objective may be to see how well the distribution of a population over categories of the dependent attributes can be predicted from defined and estimated statistical "effects" of the explanatory attributes. Another possible objective may be to compare the fit or predictive performance of a model that assumes a specific pattern of relationship between the dependent distributions and values of selected explanatory attributes with that of a null model that denies any such dependence. Yet another kind of aim would be to use variants of the model to examine the strength and pattern of the contributions of individual explanatory attributes to the performance of the model. A combination of one or more of these objectives may be pursued in a given empirical analysis. The results of such analysis ideally involve revision and/or elaboration of the theory.

Multiplicative power models are developed in terms of the concepts of "dependent distribution", "nth-order conditional proportions", and "effect measures". These concepts, and related notation, were introduced in Sections 3.4.2 and 3.4.3 of the main text. The reader who wishes to have them discussed in greater detail should see Stone (1978, Appendix C).

See footnote(s) on page 106.

A model is normally expressed as a function of the product of selected effect measures. What happens when the model is applied to data may be summarized by a rough analogy with analysis of variance. In this analysis, variance is partitioned into components according to a network of defined "between and within group" variations. In multiplicative power models we first develop the concept of the conditional expected frequency distribution over the categories of a designated dependent attribute, given the joint distribution over the cells of the simultaneous cross-classification of a set of explanatory attributes. The model that is specified, usually on the basis of substantive theory, partitions that conditional expected frequency distribution into components that are viewed as measures of specified effects of the explanatory attributes. A concrete application of the model involves estimation of the effect measures and "prediction", in terms of the model, of the conditional expected frequency distribution over categories of the dependent attribute. On the basis of various comparisons of observed and expected frequency distributions, several tests are then conducted.

To set up a model, we specify the explanatory attributes, their significant effects (using an appropriate set of substantive hypotheses), and the form of the model in which these effects appear. Specifying the form of the model involves the strategy of partitioning a conditional proportion (see Stone, 1978, Appendix C, Section C.2). We apply this strategy to find the value of the exponent. The value of this parameter is given by the minimum number of different partitionings needed to include all the effects specified as being significant. With the specification completed we move to estimation and hypothesis testing.

C.1. Estimation and Hypothesis Testing

Except for computational difficulties that arise with the more complex models that may be devised, estimation of effect measures and the "predicted" dependent distributions is fairly routine using an electronic computer. Alternative procedures are outlined in Stone (1978, Section C.3).

Goodman has provided several useful discussions of the procedures for hypothesis testing by partitioning chi-square (cf. Goodman, 1970, pp. 247 - 249; Goodman, 1972, pp. 1049 - 1056; Goodman, 1973b, pp. 181 - 183). This paper offers an introductory commentary on the procedures only because the nature of the appropriate tests for the kind of model discussed in this appendix is not immediately obvious from a study of the cases covered in Goodman's discussions.

All of the tests envisaged in this discussion involve comparisons of estimated expected cell frequencies with observed cell frequencies for the contingency table in question. The initial test may be of the goodness of fit of the model – a test of how closely the estimated expected and the observed frequencies match. Typically, this is a routine test that involves merely the calculation of an appropriate chi-square statistic. However, this is not the critical test.

To envisage the critical test clearly, the more familiar procedures in regression analysis may be cited. In this analysis, the critical test of the overall performance of the model involves the question of the proportion of the variance of the dependent variable that is accounted for by the explanatory variables. A relevant measure is R^2, and in regression analysis we ask whether R^2 is significantly greater than zero. Now, zero R^2 is observed when there is no systematic relation between the dependent and the explanatory variables. Thus, we do not ask how significant is the divergence of the "predicted" from the observed values of the dependent variable (which is the question asked in the initial goodness of fit test mentioned above). Instead, we ask how much of an improvement in "prediction" we have when we use the specified model as compared with the assumption that there is no relation (zero systematic covariation) between the dependent and the explanatory attributes.

What we need then is a specified null hypothesis, and a value of the chi-square statistic associated with it. Since the model essentially postulates that Y_i is dependent on (X_j,U_k,Z_h) in a certain way, a suitable null hypothesis is that there is no such dependence. The appropriate test of the performance of the model examines whether the chi-square statistic associated with the model is substantially lower than the chi-square statistic associated with the null hypothesis. In order to conduct this test, we will have to derive the model that is implied by the null hypothesis, compute the estimated expected cell frequencies implied by this model, and then compute the associated chi-square statistic. In the case of the model which (1) expresses, the corresponding null hypothesis would be given by:

$$\hat{Pr}(Y_i|X_j,U_k,Z_h) = \hat{Pr}(Y_i) \qquad (7)$$

The estimated expected cell frequencies under the null hypothesis are obtained by formulas shown in Stone (1978, Section C.4).

Let "χ^2 (m)" mean the chi-square statistic associated with the model, and "χ^2 ($\bar{m}$)" be the chi-square statistic associated with the null hypothesis. Then the statistic that tests the performance of the model is $\chi^2(\bar{m}) - \chi^2(m)$ and its number of degrees of freedom is equal to the difference between the numbers of degrees of freedom associated with $\chi^2(\bar{m})$ and $\chi^2(m)$, respectively.

We can get into some minor complications when we wish to test the significance of individual effects. In principle, to test the significance of a given effect, we can start with the original model and then derive the modified model that is implied by assuming the effect to be insignificant. Then if the chi-square statistic associated with the modified model is significantly larger than that associated with the original model, the effect in question would be judged to be significant. In some cases, it is quite a routine matter to set a particular effect measure in the original model equal to 1.0, and immediately derive the modified model.

However, often one particular effect measure is necessarily interrelated with others and it cannot be set equal to 1.0 without altering the values of the others. In these cases, the proper test of significance for a single effect may not be immediately obvious (see Stone, 1978, Section C.4, for a discussion of the pertinent procedures).

C.2. Coefficients of Association

Measures of degree of goodness of fit of the model in terms analogous to multiple correlation, and of the overall contribution of individual effects in terms analogous to partial correlation, have been presented by Goodman (1972, pp. 1056 - 1058). They are applicable in the approach outlined in this paper.

The coefficient of association for the model given in (1), the degree of goodness of fit of the model, pursuing the illustration explored above, is measured by

$$0 \leqslant \frac{\chi^2(\bar{m}) - \chi^2(m)}{\chi^2(\bar{m})} \leqslant 1 \qquad (8)$$

The formulas for the coefficients of partial association are indicated in Stone (1978, Section C.5).

The more closely each of the above-mentioned coefficients approximates the value of 1.0, the better is the fit of the model or the stronger is the contribution of a given effect (as the case may be).

In Section 4.1.1, the text alludes to the peculiar situation in which the "addition" of an effect measure into an asserted model causes its coefficient of association to be lowered (or, as the text says, the "model performance deteriorates"). How this situation could arise mathematically is evident from the structure of expression (2) and the definition applied to the lambda term "λ". Expression (2) essentially asserts that the conditional distribution of X (given A_T,ϕ,D,M) is a geometric mean of some lower-order conditional distributions, subject to an adjustment factor $\lambda_{T\phi DM}$. Clearly, if any one of those lower-order distributions diverges far from the one being predicted and if there is no correction of the divergence built into either another lower-order distribution specified in the model, or $\lambda_{T\phi DM}$, or a special parameter provided to effect such adjustment (perhaps under a least-square or chi-square minimization algorithm), then the inclusion of the "offending" lower-order distribution will certainly produce a bias in the model's predictions. Removing the bias will improve the predictions (or the model's goodness of fit). In the same manner, the null model could conceivably provide more accurate predictions than a badly biassed asserted model.

Let us now consider the three possibilities of correcting the "anomaly" of a model that performs worse than an altered model that uses only a portion of all its components (a common experience in population forecasting, incidentally, where greater complexity of a model's structure and its parameters may cause the model to predict certain key aggregates less well than at least one simpler model).

In the tradition of multivariate analysis, the first thought is to associate each effect measure with a unique parameter and estimate this new set of parameters by some "error minimization" algorithm. This "solution" could easily raise as many problems as it solves, as any textbook on econometrics will show. It is sufficient to observe that it requires a large number of observations for each effect measure. In the present dataset there is a maximum of 10 independent observations on each effect measure in expression (2) (see the discussion on the choice of migration streams in Section 4.1.1). From these 10 sample points, expression (2) would require us to estimate six parameters through some partial differentiation process (involving such criteria as least squares, minimum chi-square or maximum likelihood). This approach is not acceptable for this study.

The second "solution" that quickly comes to mind is the resort to the usual estimation strategy in contingency table analysis. This is to use an identical set of numbers to estimate the parameters as well as to test the goodness of fit of the model. This procedure is usually involved in the iterative fitting algorithms. Not only is this an impractical procedure for terms like ($Pr(X_Y \mid D)$ and $Pr(X_Y \mid \phi)$ in expression (2), but such terms would also measure properties of the migration stream whose schooling composition is being predicted, and not attributes of the areas of origin and destination of the stream. This procedure is also not acceptable.

The third "solution" is a slight modification of the second. For example, instead of estimating $Pr(X_Y \mid \phi)$ from data on the non-migrants of region ϕ, we should use data on all the **out**-migrants from ϕ (including those going to regions other than D). This does appear to be the theoretically correct solution. Indeed, it is quite apparent that this approach would correct the "anomaly", since the schooling distribution of all out-migrants from a non-metropolitan area differs very sharply from that of the non-migrants within such an area.

Unfortunately, we must avoid this ideal solution because of a major problem in the census data. Educational distributions are measured at the end of the migration interval. If we allowed the out-migrants from ϕ to dominate the estimate of $Pr(X_Y \mid \phi)$, we would be reflecting conditions in regions outside ϕ to a large degree. Thus we can only conclude that the "anomaly" under discussion here arises because we have specified inappropriate proxy measures for the regional characteristics. This is done because census data do not provide measures of schooling at the start of the migration period.

C.3. Some Limitations of the Method

Like most techniques of multivariate analysis, the type discussed in this paper has significant disadvantages or limitations (Goodman, 1972 and 1973a, has commented upon the advantages). Another paper is needed to outline the limitations and comment on the ways and degrees to which they can be overcome. For the present, we shall merely list a number of the more important limitations:

1. In most applications involving demographic data, arbitrary categorization of some of the variables is unavoidable, and the precise values of coefficients are sensitive to the categories chosen (see Appendix D).
2. Like most multivariate analysis techniques, the coefficients derived for the effects of a given variable are meaningful only within the context of the specified model and variables. Some potentially serious degrees of "specification bias" can result from a variety of sources such as serious omission of relevant variables. This problem has received almost no attention so far in the literature on multivariate contingency table analysis.
3. Unless the explanatory variables are mutually independent, statistical indications of the relative importance of the different specified effects can be seriously misleading.
4. Serious practical difficulties can arise when the tables become very large (in terms of their number of cells). The tendency in dealing with this problem is to collapse categories (as is done in this study). The reader should consult the related discussion by Bishop, Fienberg and Holland (1974, Chapter 5).
5. Special procedures are often required for handling tables that contain cells that are necessarily zero or relevant zero marginals (cf. Bishop, Fienberg and Holland, Chapter 5).
6. An important problem that deserves further attention concerns the applicability of the formal theory and procedures of statistical inference with certain types of data. The question of the appropriate statistical theory rationale arises naturally because the human population that is distributed over the relevant attribute space is usually a sample from a larger human population. Even when the entire relevant human population is involved, the analyst often wants to rationalize the use of statistical inference in tests of goodness of fit of the explanatory model or of individual effects of selected attributes. In this connection, the analyst must deal explicitly with the question of error (sampling) distributions. Also, there is the problem of estimating theoretical expected values from sample observations in a contingency table. Here the analyst must be concerned with the properties of the chosen estimates in the light of the nature of the sampling process that generated the observations in the table.

The statistical inference aspects of contingency table analysis have been discussed at length by several mathematical statisticians (cf. Fienberg, 1970; Goodman, 1970, 1972, 1973a; and the references cited there). It would appear that the practicing analyst already has a battery of published results for conducting a wide variety of tests of significance in connection with his statistical model. Typically, these tests involve one or two forms of the chi-square statistic, and the assumption that for the problem and data in hand this statistic has the chi-square sampling distribution. The validity of this assumption typically requires that the observations within any cell of the contingency table are independent and were generated by some simple sampling scheme, such as simple random sampling. An identification of pertinent sampling schemes is given by Fienberg (1973, p. 6) and Goodman (1971, pp. 37 and 38).

Difficulties in satisfying the conditions for valid application of the chi-square sampling distribution for the tests mentioned above are likely to arise when the data are drawn from administrative records or from a full-count census. In both types of data source, it is often difficult to specify the sampling process that generated the observations in the contingency table. However, in some problems, the nature of the data may be such as partially to warrant the assumption that the observations within any particular table cell are largely independent and that the total number of observations is fixed.[3] In any event, with the two kinds of data source in question, the application of statistical inference procedures should probably be justified principally on the grounds that the chosen procedures embody a set of ground rules for decision-making which are sufficiently objective that any two investigators using the same rules ought to reach the same conclusion, that is, take the same acceptance or rejection decision (cf. Fienberg, 1970, pp. 424 - 427; Bishop, Fienberg and Holland, 1974, Chapter 3).

Another kind of problem arises in regard to the estimation of the theoretical expected cell counts from the observations in a contingency table. Fienberg's (1970a, pp. 421 - 424) discussion suggests that we have to make rather strong assumptions about the sampling process generating the observations in the table in order to derive sharp information about properties of certain estimators being used, especially iterative fitting of cell counts to fixed marginals. The reasonableness of such assumptions may be quite clear for certain kinds of sample survey data, but there are large blocks of demographic data, presentable in a contingency table format, in which such assumptions will instantly be seen to be rather arbitrary. In this situation, it would be wise to attempt to rationalize the results of a particular estimation procedure on informal substantive grounds, if we are unable to be reasonably sure that a suitable sampling process generated the observations in our contingency tables.

See footnote(s) on page 106.

FOOTNOTES

[1] The term "model" is subject to a variety of meanings in the social science literature. In one usage it means the same as "theory". In the present work we have a different usage that may be illustrated as follows. By an "explanatory theory" of the relations between y a dependent variable, and x and z explanatory variables, we mean a network of statements that purport to (a) give reasons why such relations exist and/or (b) characterize generally the nature of such relations under specified conditions. A specific model that may seem plausible in the light of the theory may then be $y = a+bx+cz$ (where a,b and c are constants). Another specific model that may seem plausible might be $y = a \cdot x \cdot z$. The model may thus be seen as being one kind of hypothesis (one about the functional form of the relation being studied) that may be said to be plausible in the light of the theory.

In referring to a hypothesis that seems plausible in the light of the theory, it is assumed here that logical implication of the hypothesis by the theory is not required in order to claim that the latter rationalizes the former. The author adheres to the philosophy of science school which holds that the probabilistic rationalization (signified by a statement to the effect that the hypothesis is plausible in the light of the theory) of hypotheses by theory is a legitimate and common practice in scientific work (see Hempel, 1965, pp. 381 - 412 and Chapter 12).

[2] Two different, though related, kinds of dependent variable are mentioned in this paper. The first refers to the dependent attribute (e.g., educational attainment) that is considered in the theoretical discussion to be related to certain explanatory attributes (e.g., age). In some explanatory analysis problems, it is considered that what the formal explanatory model will treat as the dependent variable is not the dependent attribute *per se,* but the distribution of a population over designated categories of this attribute. For example, explanatory analyses of migration often are developed in such a way that although the theory deals with the factors and mechanism by which a person becomes a migrant (the attribute here is migration status), the dependent variable of the formal model is the proportion who are migrants out of a given population (the distribution of population by migration status).

In this report, the phrase "dependent variable" usually refers to the distributional variable, while the phrase "dependent attribute" usually refers to the variable that is initially considered in the relevant theory. The distinction between the two terms is rather arbitrary, and it is made only to reflect the fact that we are dealing with two levels of dependent variables. Usually, the context of a particular remark should clear which level is in question.

[3] For a discussion of the significance of this point, see Goodman, 1970, p. 232.

APPENDIX D

CATEGORIZATION OF ATTRIBUTES

The specification of variables has been guided by the requirement that all variables be represented as attributes with more than two categories, and by concern for economy in running the various tests. Accordingly, relatively crude categorization is used for certain of the attributes. Based on past experience in research using these attributes, however, it is believed that the categories chosen will capture most of the crucial aspects of the pattern of population distribution for each attribute chosen. Ideally, the general pattern of the results of statistical analysis ought to be tested for sensitivity to reasonable alterations of attribute categories.

Two factors helped to dictate the chosen detail of categorization. Firstly, the computer core storage area available for running the analysis programme limited the input matrix size. Secondly, detailed categorizations of such variables as occupation and schooling would have provided unduly large matrices of "effect" measures, given the relatively small number of observations. It was assumed that although the categorization of a variable like occupation is quite inadequate by itself, when we consider the simultaneous cross-classification of occupation and other variables, we should be able to pick up a good deal of "variance explaining power" that a more detailed breakdown of occupation alone would provide.

The need for relatively crude categorizations of variables in any problem where a large number of variables is involved, and the sensitivity of the results of the analysis to the specific categorization chosen, are important limitations of the analysis of contingency tables. No multivariate analysis procedure is without limitations, however, and the one chosen here does have significant advantages that should be balanced against its shortcomings (cf. Goodman, 1972).

The following is a list of the sample universe and the attributes used in Chapter 4, showing the relevant categorizations.

Sample Universe

Males aged 20 - 54 in 1971, not attending school in the 1970 - 71 school year, who worked in 1970, who moved at least once inter-municipally between 1 June 1966 and 1 June 1971, and who resided in Canada on 1 June 1966.

Attributes

Age (2)

20 - 34 years

35 - 54 "

Level of Schooling (4)

Less than Grade 12

Grades 12 and 13 and non-university

Some university

University degree

Occupation (15)

Managerial, administrative and related occupations

Teaching and related occupations

Occupations in medicine and health

Technological, social, religious, artistic and related occupations

Clerical and related occupations

Selected sales occupations (include technical salesmen and related advisers, commercial travellers, street vendors and door-to-door salesmen, newsboys, insurance salesmen and agents, and driver-salesmen)

Other sales occupations

Service occupations (excluding Armed Forces)

Farming, horticultural and animal husbandry occupations

Other primary occupations (include fishing, hunting, trapping and related occupations, forestry and logging occupations, mining and quarrying including oil and gas field occupations)

Processing occupations

Machining and product fabricating, assembling and repairing occupations

Construction trades occupations

Transport equipment operating occupations

Occupations not elsewhere specified (excluding not stated)

Region-group of Origin (7)

Atlantic CMA's

Atlantic Non-CMA Area

Central CMA's

Central CA's of 50,000 and Over

Central Predominantly Rural Area

Western CMA's

Western Non-CMA Area

Region-group of Destination (7)

Atlantic CMA's
Atlantic Non-CMA Area
Central CMA's
Central CA's of 50,000 and Over
Central Predominantly Rural Area
Western CMA's
Western Non-CMA Area

APPENDIX E

A MODEL FOR THE JOINT DISTRIBUTION OF EDUCATION BY OCCUPATION

This appendix states the full three-equation model that was set up for analysis of the joint education-by-occupation composition of a migration stream. The full model was not tested in the statistical work for this study because of the shortage of time and the assumption that most readers would be interested mainly in the separate analyses of the educational and occupational patterns of migration. It was thought worthwhile, however, to state the full model and indicate how it would be fitted and evaluated. This step will indicate how one can take the work (presented in the main text) additional steps into further analysis. It also permits the author to indicate how the results of this work can be used in two common types of practical simulation problems.

Recall two of the general models used in Chapter 4. (The notation will not be reintroduced below, as it is given in Chapter 4 and in Appendix C.) The first equation, used for the schooling composition of migration, will be called H_1 (meaning hypothesis number one).

$$\hat{\Pr}(X_Y | A_T \phi D M) = \Pr(X_Y) \cdot \left[\frac{\Pr(X_Y | M)}{\Pr(X_Y)} \cdot \frac{\Pr(X_Y | A_T)}{\Pr(X_Y)} \cdot \frac{\Pr(X_Y | \phi)}{\Pr(X_Y)} \cdot \frac{\Pr(X_Y | D)}{\Pr(X_Y)} \right]^{1/4}$$

$$\cdot \left[\frac{\Pr(X_Y | A_T D)}{\Pr(X_Y | A_T)} \cdot \frac{\Pr(X_Y | A_T \phi)}{\Pr(X_Y | \phi)} \right]^{1/4} \cdot \lambda_{T \phi D M} \qquad (9)$$

The second, used for the occupational composition of migration, is called H_2. In this equation, W_U is the Uth category of occupation.

$$\hat{\Pr}(W_U | X_Y \phi D M) = \Pr(W_U) \cdot \left[\frac{\Pr(W_U | M)}{\Pr(W_U)} \cdot \frac{\Pr(W_U | X_Y)}{\Pr(W_U)} \cdot \frac{\Pr(W_U | \phi)}{\Pr(W_U)} \cdot \frac{\Pr(W_U | D)}{\Pr(W_U)} \right]^{1/4}$$

$$\cdot \left[\frac{\Pr(W_U | X_Y D)}{\Pr(W_U | X_Y)} \cdot \frac{\Pr(W_U | X_Y \phi)}{\Pr(W_U | \phi)} \right]^{1/4} \cdot \lambda_{Y \phi D M} \qquad (10)$$

The third hypothesis, H_3, relies upon the first two,

$$\hat{P}r(X_Y W_U | A_T \phi DM) = \hat{P}r(X_Y | A_T \phi DM) \cdot \hat{P}r(W_U | X_Y \phi DM) \cdot \alpha_{T|YU\phi DM} \quad (11)$$

$$\alpha_{T|YU\phi DM} \equiv \frac{Pr(A_T | X_Y W_U \phi DM)}{Pr(A_T | X_Y \phi DM)} \quad (12)$$

which can be readily verified by a few algebraic manipulations of the terms in (11).

Sections 4.1.1 and 4.2.1, in the text, have discussed the fitting procedure for H_1 and H_2. Thus, to fit H_3 we have only to deal with the α term. At least three procedures are possible. Each has its own requirements and advantages, and the use to which H_3 will be put should determine which procedure is followed.

The first is to estimate the α term directly from the set of observed frequencies that will be used to test the goodness of fit of the model. The second is to use a chi-square minimization algorithm to estimate α (this comment is elaborated below). The third is to estimate α by means of:

$$\alpha_{T\phi DM} = 1 \Big/ \left[\sum_{Y;U} \hat{P}r(X_Y | A_T \phi DM) \cdot \hat{P}r(W_U | X_Y \phi DM) \right] \quad (13)$$

To adopt the chi-square minimization procedure, consider that Y,U,T,ϕ, D,M is the observed frequency for persons with the combination of attributes Y,U,T,ϕ,D,M. Let us define the corresponding expected frequency, if H_3 is true, as

$$F(H_3)_{YUT\phi DM} = \hat{P}r(X_Y | A_T \phi DM) \cdot \hat{P}r(W_U | X_Y \phi DM)$$

$$\cdot \alpha_{T|YU\phi DM} \cdot \eta_{T\phi DM} \quad (14)$$

where $$\eta_{T\phi DM} = \sum_{Y,U} \eta_{YUT\phi DM} \quad (15)$$

Defining the chi-square value as

$$\chi^2(H_3) = 2 \sum_{Y,U,T} \eta_{YUT\phi DM} \ln \left[\eta_{YUT\phi DM} / F(H_3)_{YUT\phi DM} \right] \quad (16)$$

it is possible to construct an iterative fitting algorithm that finds the values of $\alpha_{T|YU\phi DM}$ which will minimize the chi-square value. The initial values of $F(H_3)_{YUT\phi DM}$ are given by (14) after setting the values $\alpha_{T|YU\phi DM} = 1$. The minimization is done for given marginal distributions $\eta_{YT\phi DM}$ and $\eta_{YU\phi DM}$, which are computed from H_1 and H_2.

It should be noted that there is a simple formula that links the chi-square of H_3 with the expected frequencies associated with H_1 and H_2, i.e., $F(H_1)_{YT\phi DM}$ and $F(H_2)_{YU\phi DM}$. Thus it is unnecessary to compute $F(H_3)$ if $F(H_1)$ and $F(H_2)$ are already available. In some conditions (unfortunately not satisfied by the models in this paper), it is possible to compute $\chi^2(H_3)$ directly from $\chi^2(H_1)$ and $\chi^2(H_2)$. (See Goodman, 1973.)

Once we have estimated the parameters of H_1 and H_2, expression (11) can be used to do two different kinds of simulation. If the joint frequency distribution $F(H_3)_{YUT\phi DM}$ is not available, then expression (11) can be used to estimate it. This application is useful in the situation where one wants the detailed education-by-occupation breakdown of migration but cannot get it. Another class of practical problems arises when joint distributions $Pr(X_Y W_U | A_T \phi DM)$ are needed to generate events in a micro-simulation model or to do macro-level projection. Expression (11) can be applied in this circumstance. In both of these kinds of application, it will normally be necessary to use the third of the three procedures outlined above to estimate the α term, unless previous estimates from a relevant data set are simply adopted for use in the application.

This specific example, dealing with the joint composition of a subpopulation by education and occupation simultaneously, may not be of much practical interest. However, the general strategy can be applied in other problems, e.g., projecting the future pattern of shifts from one mother tongue to another language of current use in a population.

SELECTED BIBLIOGRAPHY

Bishop, Yvonne M. 1969. Full Contingency Tables, Logits, and Split Contingency Tables. *Biometrics 25*: pp. 383 - 400.

Bishop, Y.M., Fienberg, S.E., and Holland, P.W. 1974. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.

Blau, Peter M., and Duncan, Otis Dudley. 1967. *The American Occupational Structure*. New York: John Wiley and Sons, Inc.

Brackstone, G.J. 1971. The 1971 Census Weighting Procedures, *Population and Housing Research Memorandum*, No. PH - Gen - 9, Statistics Canada, Ottawa.

Brackstone, G.J. 1976. Raking Ratio Estimators. *Survey Methodology 2:* pp. 63 - 70.

Brennan, Michael J. 1965. A More General Theory of Resource Migration, in Michael J. Brennan, ed., *Patterns of Market Behaviour*. Providence: Brown University Press: pp. 45 - 64.

Brewis, T.N. 1969. *Regional Economic Policies in Canada*. Toronto: MacMillan Company of Canada Limited.

Camu, Pierre, Weeks, E.P., and Sametz, Z.M. 1964. *Economic Geography of Canada*. Toronto: MacMillan of Canada.

Canada. Manpower and Immigration. 1975a. *Canadian Occupational Forecasting Program*. Ottawa: Information Canada.

Canada. Manpower and Immigration. 1975b. *Internal Migration and Immigrant Settlement*. Ottawa: Information Canada.

Canada. Statistics Canada. 1973a. *1971 Census of Canada, Population, Cities, Towns, Villages, Census Metropolitan Areas and Census Agglomerations*. Bulletin 1.1 - 8, Catalogue 92-708. Ottawa: Information Canada.

Canada. Statistics Canada. 1973b. *1971 Census of Canada, Population, Census Subdivisions (Historical)*. Bulletin 1.1 - 2, Catalogue 92-702. Ottawa: Information Canada.

Canada. Statistics Canada. 1974. *1971 Census of Canada, Population, Internal Migration*. Bulletin 1.2 - 7, Catalogue 92-719. Ottawa: Information Canada.

Canada. Statistics Canada. 1975a. *1971 Census of Canada, Industries, Industry Divisions by Sex, for Canada, Provinces and Census Divisions*. Bulletin 3.4 - 4, Catalogue 94-741. Ottawa: Information Canada.

Canada. Statistics Canada. 1975b. *1971 Census of Canada, Industries, Industries by Sex, for Census Metropolitan Areas, Place of Residence and Place of Work.* Bulletin 3.4 - 5, Catalogue 94-742. Ottawa: Information Canada.

Canada. Statistics Canada. 1975c. *1971 Census of Canada, Industries, Industries by Sex, for Census Agglomerations of 25,000 and Over (Place of Residence) and Census Agglomerations of 50,000 and Over (Place of Work).* Bulletin 3.4 - 6, Catalogue 94-743. Ottawa: Information Canada.

Caves, R.E. and Holton, R.H. 1961. *The Canadian Economy: Prospects and Retrospect.* Cambridge: Harvard University Press.

Courchene, Thomas J. 1970. Interprovincial Migration and Economic Adjustment. *Canadian Journal of Economics III*: pp. 551 - 575.

Courchene, T. 1974. *Migration, Earnings, and Employment in Canada: 1965 - 68.* Montréal: C.D. Howe Research Institute.

Davis, James A. 1974. Hierarchical Models for Significance Tests in Multivariate Contingency Tables, in Herbert L. Costner, ed., *Sociological Methodology, 1973 - 1974.* San Francisco: Jossey-Bass, Inc.

Denton, Frank T. 1966. *An Analysis of Interregional Differences in Manpower Utilization and Earnings.* Staff Study No. 15, Economic Council of Canada, Ottawa: Queen's Printer.

Dodds, D.J. 1971. Sampling in the Self-Enumeration Areas of the 1971 Census. *Population and Housing Research Memorandum,* No. PH - Gen - 8. Statistics Canada, Census Field, Ottawa.

Duncan, O.D., *et al.,* 1961. *Metropolis and Region.* Baltimore Md.: Johns Hopkins Press.

Eldridge, Hope T. and Thomas, Dorothy Swaine. 1964. *Population Redistribution and Economic Growth, United States, 1870 - 1950, Vol. III: Demographic Analyses and Interrelations.* Philadelphia: American Philosophical Society.

Fienberg, Stephen E. 1970. The Analysis of Multidimensional Contingency Tables. *Ecology 51*: pp. 419 - 433.

Fienberg, Stephen E. and Holland, P.W. 1970. Methods of Eliminating Zero Counts in Contingency Tables. in G.P. Patil, ed., *Random Counts in Models and Structures.* University Park: the Pennsylvania State University Press.

Fienberg, Stephen E. 1972. The Analysis of Incomplete Multi-Way Contingency Tables. *Biometrics 28*: pp. 177 - 202.

Fienberg, Stephen E. 1973. *The Analysis of Cross-Classified Data*. St. Paul: Dept. of Applied Statistics, University of Minnesota.

Gallaway, Lowell E. 1967. Industry Variations in Geographic Labour Mobility Patterns. *Journal of Human Resources 2*: pp. 461 - 474.

George, M.V. 1970. *Internal Migration in Canada*. Ottawa: Queen's Printer.

Goldstein, Sidnes. 1964. The Extent of Repeated Migration: An Analysis Based on the Danish Population Register. *Journal of the American Statistical Association LIX*: pp. 1121 - 1132.

Goodman, Leo. A. 1970. The Multivariate Analysis of Qualitative Data. *Journal of the American Statistical Association 65*: pp. 226 - 256.

Goodman, Leo. A. 1971. The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods. *Technometrics 13*: pp. 33 - 61.

Goodman, Leo. A. 1972. A General Model for the Analysis of Surveys. *American Journal of Sociology 77:* pp. 1035 - 1086.

Goodman, Leo. A. 1973a. Causal Analysis of Data from Panel Studies and Other Kinds of Surveys. *American Journal of Sociology 78:* pp. 1135 - 1191.

Goodman, Leo. A. 1973b. The Analysis of Contingency Tables When Some Variables are Posterior to Others. *Biometrika 60*: pp. 179 - 192.

Grant, E. Kenneth and Vanderkamp, John. 1976. *The Economic Causes and Effects of Migration*. Ottawa: Minister of Supply and Services.

Greenwood, M.J. 1975. Research in Internal Migration in the United States: A Survey. *Journal of Economic Literature 13:* pp. 397 - 433.

Hamilton, C. Horace. 1959. Educational Selectivity of Net Migration from the South. *Social Forces 38*: pp. 33 - 42.

Hempel, Col. G. 1965. *Aspects of Scientific Explanations*. New York: The Free Press.

Hunter, Lawrence and Reid, Graham L. 1968. *Urban–Worker Mobility*. Paris: Organization for Economic Co-operation and Development.

Illsley, Raymond, Finlayson, Angela, and Thompson, Barbara. 1963. The Motivation and Characteristics of Internal Migrants. *The Milbank Memorial Fund Quarterly 41*: pp. 115 - 143 and 217 - 247.

Irving, R.M., ed., 1972. *Readings in Canadian Geography.* Toronto: Holt, Rinehart and Winston of Canada, Ltd.

Kasahara, Yoshiko. 1965. Internal Migration and the Family Life Cycle, Canadian Experience, 1956 - 61. Paper presented to the 1965 World Population Conference in Belgrade.

Kosinski, Leszek A. 1976. Internal Migration in Canada. *Geographical Journal.*

Kuznets, Simon and Thomas, Dorothy Swaine. 1957. Introduction, in Everett S. Lee, *et al., Population Redistribution and Economic Growth, United States, 1870 - 1950.* Philadelphia: American Philosophical Society.

Labor, Gene and Chase, Richard S. 1971. Interprovincial Migration in Canada as a Human Capital Decision. *Journal of Political Economy:* pp. 795 - 804.

Ladinsky, Jack. 1967. The Geographic Mobility of Professional and Technical Manpower. *The Journal of Human Resources 2*: pp. 475 - 494.

Lansing, John and Mueller, Eva. 1967. *The Geographic Mobility of Labour.* Ann Arbor: Institute for Social Research, University of Michigan.

Lee, Everett S. 1966. A Theory of Migration. *Demography 3:* pp. 47 - 57.

Leslie, Gerald R. and Richardson, Arthur H. 1961. Life-Cycle, Career Pattern and the Decision to Move. *American Sociological Review 26*: pp. 894 - 902.

Long, Larry H. 1970. On Measuring Geographic Mobility. *Journal of the American Statistical Association 65.*

Lowry, Ira S. 1966. *Migration and Metropolitan Growth.* San Francisco: Chandler Publishing Company.

Lycan, Richard. 1969. Interprovincial Migration in Canada: The Role of Spatial and Economic Factors. *Canadian Geographer 13*: pp. 237 - 254.

Maxwell, J.W. 1965. The Functional Structure of Canadian Cities: A Classification of Cities. *Geographic Bulletin 7*: pp. 79 - 104.

McInnis, Marvin. 1971. Age, Education and Occupation Differentials in Interregional Migration: Some Evidence for Canada. *Demography 8*: pp. 195 - 204.

Meltz, Noah M. 1969. *Manpower in Canada 1931 - 61.* Ottawa: Queen's Printer.

Miller, Ann R. 1967. The Migration of Employed Persons to and from Metropolitan Areas of the United States. *Journal of the American Statistical Association 62*: pp. 1418 - 1832.

Nargundkar, M.S. and Arora, H. 1971. The Raking-Ratio Estimation Procedure for the 1971 Census, Unpublished paper, Statistics Canada, Ottawa.

Nelson, Philip. 1959. Migration, Real Income and Information. *Journal of Regional Science 1*: pp. 43 - 75.

Nickson, May. 1967. *Geographic Mobility of Labour in Canada October 1964 - October 1965*. Special Labour Force Studies No. 4. Ottawa: Queen's Printer.

Norland, *et al.* 1975. Evaluation in the 1971 Census of Canada: Overview and Selected Findings.*Canadian Studies in Population 2*: pp. 65 - 90.

Ostry, Sylvia. 1967. *The Occupational Composition of the Canadian Labour Force*. Ottawa: Queen's Printer.

Ostry, Sylvia. 1968. *Geographic Composition of the Canadian Labour Force*. Ottawa: Queen's Printer.

Ray, D.M. 1972. The Economy in L. Gentilcore, ed., *Studies in Canadian Geography: Ontario*. Toronto: University of Toronto Press.

Richardson, H.W. 1973. *Regional Growth Theory*. Toronto: Halsted Press, John Wiley and Sons, Inc.

Ricour-Singh, Françoise. 1979. *Les pôles d'attraction et les zones d'influence du Canada*. Census Analytical Study. Ottawa: Minister of Supply and Services (forthcoming).

Robert, Bernard. 1971. *Profils Migratoires, Comtés et Régions, Province de Québec, 1961 - 1966*. Québec: Bureau de la Statistique du Québec.

Robert, Bernard. 1972. *Évolutions démographiques régionales et migrations intérieures de la population, Province de Québec, 1941 - 1966*. Québec: Bureau de la statistique du Québec.

Rossi, Peter H. 1955. *Why Families Move*. Glencoe: Free Press.

Rottenberg, Simon. 1956. On Choice in Labour Markets. *Industrial and Labour Relations Review*: pp. 183 - 199.

Shaw, R.P. 1975. *Migration Theory and Fact, a Review and Bibliography of Current Literature.* Philadelphia: Regional Science Research Institute.

Shryock, Henry S. 1964. *Population Mobility Within the United States.* Chicago, Community and Family Study Centre. University of Chicago.

Shryock, Henry S. and Nam, Charles B. 1965. Educational Selectivity of Interregional Migration. *Social Forces XL*: pp. 299 - 510.

Simmons, James W. 1968. Changing Residence in the City. *Geographical Review*. pp. 622 - 651.

Simmons, James W. 1974a. *Patterns of Residential Movement in Metropolitan Toronto*. Toronto: University of Toronto, Dept. of Geography.

Simmons, James W. 1974b. Canada as an Urban System: A Conceptual Framework, Research Paper No. 62, Centre for Urban and Community Studies, University of Toronto.

Simmons, James W. 1977. Migration and the Canadian Urban System: Part I, Spatial Patterns, Research Paper No. 85, University of Toronto, Centre for Urban and Community Studies.

Sjaastad, Larry A. 1960. The Relationships Between Migration and Income in the United States. *Papers Proceedings of the Regional Science Association 6*: pp. 37 - 64.

Sjaastad, Larry A. 1962. The Costs and Returns of Human Migration. *Journal of Political Economy Vol. LXX*: pp. 80 - 93.

Speare, Alden. 1970. Home Ownership, Life Cycle Stage and Residential Mobility. *Demography 7*: pp. 449 - 458.

Speare, Alden. 1971a. A Cost-Benefit Model of Rural to Urban Migration in Taiwan. *Population Studies 25*: pp. 117 - 130.

Stone, Leroy O. 1969. *Migration in Canada: Some Regional Aspects*. Ottawa: Queen's Printer.

Stone, Leroy O. 1971a. *On the Analysis of the Structure of Metropolitan Areas Migration Streams: A Theoretical Framework with Empirical Glimpses from Canadian and American Census Data*. Educational Planning Occasional Paper No. 1/71, Toronto: Ontario Institute for Studies in Education.

Stone, Leroy O. 1971b. On the Correlation Between In- and Out-Migration by Occupation. *Journal of the American Statistical Association*.

Stone, Leroy O. 1972. Some Theoretical Aspects of the Study of Migration Decisions: A Working Paper. Unpublished paper.

Stone, Leroy O. 1973a. Integration of Behavioural and Macro-Level Perspectives Toward Analysis of Compositional Variations Among Interregional Migration Streams. Unpublished paper.

Stone, Leroy O. 1973b. Latent Parameter for the Mobility Matrices of Real and Synthetic Cohorts, with Illustrations from Canadian Data. Unpublished paper.

Stone, Leroy O. 1974. What We Know about Migration Within Canada: A Selective Review and Agenda for Future Research. *International Migration Review.*

Stone, Leroy O. 1975. On the Interaction of Mobility Dimensions in Theory on Migration Decisions. *Canadian Review of Sociology and Anthropology 12*: pp. 95 - 100.

Stone, Leroy O. and Fletcher, Susan. 1976. *Migration in Canada: 1971 Census Profile Study.* Ottawa.

Stone, Leroy O. 1976a. Some Principles of Demometric Analysis. Unpublished paper, Statistics Canada, Ottawa.

Stone, Leroy O. 1976b. Quelques nouveaux développements dans l'analyse de la migration interne au Canada. Paper presented to L'Association des Démographes du Québec, in Sherbrooke.

Stone, Leroy O. 1978. *The Frequency of Geographic Mobility in the Population of Canada*. Census Analytical Study. Ottawa.

Suval, Elizabeth S. and Hamilton, Horace C. 1965. Some New Evidence on Educational Selectivity in Migration to and from the South. *Social Forces 43*: pp. 536 - 547.

Tarver, James D. 1964. Occupational Migration Differentials. *Social Forces 43*: pp. 231 - 241.

U.S. Bureau of the Census. 1973. *1970 Census of Population, Mobility for States and the Nation*. Washington, D.C.: U.S. Government Printing Office.

Vanderkamp, John. 1968. Interregional Mobility in Canada. *Canadian Journal of Economics*: pp. 596 - 608.

Vanderkamp, John. 1970. The Effect of Out-Migration on Regional Employment. *The Canadian Journal of Economics III*: pp. 541 - 549.

Vanderkamp, John. 1972. Migration Flows and their Determinants, and the Effects of Return Migration. *Journal of Political Economy 79*: pp. 1012 - 1031.

Vanderkamp, John. 1973. *Mobility Patterns in the Canadian Labour Force.* Economic Council of Canada. Special Study No. 16. Ottawa: Queen's Printer.

DATE DUE
