REPEATED MIGRATION
A SIMULATION EXPERIMENT BASED ON SCOTTISH MATERIALS

by

M.G. EL-ROUBY
(B.Com., M.Sc.)

THESIS

Submitted in Fulfilment of the Conditions for the Degree

of Doctor of Philosophy

to

THE UNIVERSITY OF GLASGOW

Research Conducted in

the

Department of Social and Economic Research

Faculty of Arts

September 1976

To my wife, my daughter, and my new-born son

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D.)

TABLE OF CONTENTS (CONT'D.)

LIST OF TABLES

LIST OF TABLES (CONT'D.)

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my profound indebtedness to my supervisor, Dr. T.H. Hollingsworth, Senior Lecturer in the Department of Social and Economic Research at Glasgow University, who first stimulated my interest in the subject of migration. His insights, guidance and inspiration have provided me with vital support throughout the work for this thesis. I also wish to express my gratitude to Professor L.C. Hunter, Head of the Department, for his continual encouragement and readiness to help in overcoming any academic or non-academic problems. Thanks are also due to Mr. B. Torsney, a research colleague in Statistics Department, Glasgow University, for the many valuable discussions with him on statistical methodology problems.

Much of the computing work in this thesis has been carried out on Edinburgh Regional Computing Centre - IBM 370/155 and 158 computers via Glasgow-Edinburgh Data Link and I wish to acknowledge the assistance of Glasgow Computing Service and its consulting staff. A special debt of gratitude is owed to Dr. P. Rosenberg for his valuable suggestions regarding the solution of various computational problems.

The assistance of the Ford Foundation, The Institute of Statistical Studies and Research, Cairo University, and the Egyptian Government in providing the necessary financial support for the undertaking of this study is gratefully acknowledged.

Though my graduate education has been both enjoyable and rewarding for me, it has been a strenuous experience for my wife and daughter. I feel deeply grateful to them for their patience and moral support without which the task of the research would not have been completed so smoothly.

My thanks are extended to Miss Maureen McCrorie and Mrs Christine MacCallum who took on the tedious work of typing the final draft of the thesis.

ABSTRACT

A random sample of persons drawn from the Scottish Central Register was used to study the phenomenon of repeated migration in Scotland for the years 1939-1964. The existence of repeated migration flows in Scotland is established by the fact that a migrant, on the average, made approximately 2.315 moves compared with an overall average of 1.038 moves during that period.

A study of the frequency of movement of the sample revealed that repeated migration is a characteristic of the single and young persons. Sex differential is of less importance than the age or marital status differentials. In general, changes in the life cycle from family formation to family dissolution are major stimuli for migration.

The influence of past mobility experience in determining levels of future mobility, whether this past experience is expressed in terms of the frequency of previous movement or in terms of previous duration of residence, is investigated. The hypothesis that people who were more mobile in the past are more likely to move again in the future is tested and validated.

These findings have considerable implications for the development of a general stochastic model of repeated migration. The model assumes that the propensity to migrate is not constant over time, but rather decreases as duration in the same place of residence increases, implying a linkage between previous mobility experience and future mobility prospects. The model also assumes that this propensity to migrate does not build up immediately after a move has occurred. There is probably some delay time during which the likelihood of a further move is negligible. The model is formulated in a simple mathematical form. It includes only three parameters defining respectively the general

probability of moving, the delay time between a move and a subsequent move, and a constant term used to identify the mode of decline in the probability of moving in relation to the duration of residence since last move.

The model is tested by means of a simulation technique. Hypothetical samples are simulated and a distribution of movement for each sample is generated. Such a distribution is compared with the actual distribution of movement and the Kolmogorov-Smirnov two sample test is used as a measure of goodness of fit of the model to the observed data. Simulation experimentation is a necessary media in order to minimize the value of Kolmogorov-Smirnov statistic to imply non-significant differences between the simulated and actual distributions of movement.

Methods of locating optimum conditions operating upon a system which maximize or minimize features of it are utilized. These methods were developed by Box and associates for the study of chemical processes.

According to these methods, the response surface is first represented by a plane whose coefficients are estimated using a suitably arranged experiment in the experimental region. The signs and magnitudes of the coefficients determine the direction of greatest gain in the response towards a near-stationary region. The response surface in this region is represented by a polynomial of higher degree than the first. The signs and magnitudes of its coefficients determine the nature of the surface and consequently the minimum or mximum can be located using a few more extra points.

These methods are adopted to determine the levels of model parameters which produce the best fit of the model to the observed data.

Kolmogorov-Smirnov statistic is treated as the response to be minimized and the model parameters as the independent variables that yield this response. Four consecutive experiments comprising a total of 77 simulations are performed along the above lines. The final outcome of these experiments is a fall in the value of response to a level equal to 0.017 implying that the differences between the simulation and actual samples, with a high level of probability, are random variations and that the two samples are drawn from the same population. This indicates that the model efficiently describes the actual patterns of repeated migration in Scotland and may eventually be used as a general explanation of the repeated migration phenomenon.

The effect of disaggregating the data by age and sex is assessed by fitting the model to selected age and sex subgroups. But certain constraints prevent such analysis from being thoroughly completed. However, the analysis suggests that sex and age variables as well as other relevant variables are important factors in formalizing a significant relationship between the probability of moving and the past history of movement.

Chapter (1)

INTRODUCTION

As one of the major components of population change, migration has drawn growing attention from various research disciplines in recent decades. Not only does internal migration contribute to the welfare of the society as a whole, but also it has great implications for the well-being of individuals within the society.

So many problems have been solved or hoped to be solved, both on aggregate and individual scales, by the redistribution of population through migration. The creation of a state of equilibrium and the restoration of social and economic balances between different areas have been major contributions of the migration mechanism. People are shifted from declining localities to other progressive localities where they can make more profitable use of their abilities and qualifications.

However, the outcome of such migration flows has not always been a success. For problems of various types have arisen as a result of migration, both in the sending and the receiving areas, and on both the national and individual levels. The migration of highly qualified members of the population in depressed areas, seeking greater opportunity elsewhere, has aggravated the social and economic problems of these areas. On the other hand, serious pressures have been imposed on the prosperous areas where these flows of migrants are accommodated. For example, the problem of congestion, the excessive demand on public services, the spread of diseases and health hazards, and the cases of social disorganization are the sorts of problems that may be created by migration in specific areas (van Arsdol, 1966; Willis, 1974).

Internal migration is, therefore, a vital demographic phenomenon which has had and continues to have a great appeal to scientific inquiry. Since 1885, when Ravenstein formulated his first 'Laws of Migration', a massive collection of migration studies has been produced. Diverse interdisciplinary efforts have been directed towards the verification of existing migration hypotheses as well as the development of new hypotheses.

However, the attention given to internal migration is far less than that given to other components of population change (fertility and mortality), both in quality and magnitude. Many reasons lie behind such a relative backward state of migration research. The most commonly acknowledged reasons are: Firstly, the data on migration are usually inadequate and unreliable. This has led to difficulty in measuring and analysing migration phenomena. Secondly, studies on migration have always been treated individually, each as a unique case and seldom with linkage to other studies. Each group of researchers, with varying research inclinations, have confined themselves to specific problems with little effort to develop a comprehensive, general model of migration. Finally, there has been a lack of conceptual framework that incorporates the enormous mass of empirical findings into a well-defined theory which enables the study of internal migration to be carried out on a solid and meaningful basis.

Repeated migration (moving more than once) is one of the concomitant features of the migration process. It has been noted by Goldstein that the observed high mobility rates which seem to characterize certain communities do not necessarily mean the instability of the majority of populations in these communities. Rather, they represent movements of the same migrant elements of population who change residence more frequently than the others, thus indicating a basic stability of the population as a whole but with a marginal fringe of highly mobile persons 'chronic movers' (Goldstein, 1954 and 1964). Goldstein has also emphasized that the high degree of residential stability for most of the population is not contradictory with high general rates of mobility. "The repeated moves of a small segment of the population inflate the overall mobility rates well beyond what they would be if mobility were based on the number of migrants instead of on the number of moves." (Goldstein, 1964; p. 1121).

The existence of repeated migration phenomenon is well known but its incidence has been difficult to record. The same reasons behind the existing lag in migration knowledge, in general, equally apply in the case of repeated migration.

There has been some attempts to develop and utilize new sources of data which allow a longitudinal approach to the analysis of migration that provides a full history of migration throughout the whole life span of individuals or part of it. The work by Hollingsworth (Internal Migration Statistics from the Central Register for Scotland of the National Health Service) was one such attempt (Hollingsworth, 1970). However, the effects of repeated and return migration phenomena on migratory mobility were not fully explored in this earlier work, though their existence were acknowledged.

> "The thesis that return migration is very common is thus not proved but does seem likely .... Further research on the Central Register Records could establish this beyond doubt .... We can infer from Table 4.1 that repeated migration is also very common ...... Although about half the sample population did not move even once, therefore, considerably more than half of those who moved once moved twice."
>
> (Hollingsworth, 1970; p. 70)

The work of this thesis seeks to at least partially fill this gap in our knowledge of Scottish migration, using the same materials, and at the same time to contribute to the existing attempts to develop a more refined theory of the repeated migration phenomenon. The main purpose is to attempt some explanation of the actual patterns of repeated migration, as evident in the data on Scottish migration with a view to developing a theoretical framework that may serve as a general explanation of repeated migration.

The test of success of such theoretical construct is whether the actual patterns could have been produced by the suggested theory. Failure to do this will indicate that either the phenomenon under investigation is a result of complex and interrelated forces that can not be accounted for by any model or the theory itself is incomplete or inadequate.

The study is mainly based on materials extracted from the Scottish Central Register (henceforth referred to as S.C.R.) between 1939-1964. The data set used consists of a systematic random sample of the registered population. The sample was originally drawn by Hollingsworth for his earlier studies on migration (Hollingsworth, 1968 and 1970).

The thesis consists of eight chapters the first of which is the introduction. Chapter two reviews the major aspects of internal migration as described in the migration literature. The chapter begins with a discussion of how the term 'migration' is defined and why migration is vital to the welfare of society and individuals. It proceeds to investigate the determinants of migration by providing answers to the questions: Who migrates? and Why? Migration theories and models of measurement are described in the same chapter with emphasis on the reasons behind the observed lag in the development of a general theory of migration. The chapter ends with a description of the major sources of migration data and the advantages and disadvantages of each of these sources.

In chapter three S.C.R., the basic source of information, is described. Some of the deficiencies of the British registration system are mentioned.

The method of selecting the sample is outlined and the characteristics of the sample revealed by first tabulation are illustrated.

Chapter four is devoted to the study of the frequency of movement of the Scottish population, as represented by the S.C.R. data, during 1939-1964. The distribution of individuals according to their recorded number of moves is analysed with reference to variables related to the individuals' characteristics or changes in their life histories. The basic variables studied are: sex, age, marital status, and change in marital status. Chapter four also investigates the effect of past mobility experience on mobility behaviour in the near future with the investigation of sex and age variables on the relationship between both sequences of mobility.

A simple, practical model for repeated migration is developed in chapter five. Testing the validity of such a model and deriving the properties of interest through ordinary analytic techniques appear to be impractical and require a great deal of mathematics.

However, experimentation using a computer simulation model is thought to be a valuable technique for analysing the patterns evident in the S.C.R. data. A discussion is therefore given in chapter five, of simulation; what is meant by simulation? Why we resort to it? and What are the methodological considerations in a simulation experiment, in general?

For simulating the proposed model, a specific simulation experiment has to be designed, It describes how the model can be operationally adjusted to meet computer simulation requirements, how the initial values of the model parameters and other simulation parameters are determined, and how the logical steps for simulating individuals are set out. Such specific computer simulation experiment is explained in the appendix.

Chapter five also describes the criteria employed for testing the validity of the suggested model. It includes a preliminary assessment of the model and the parameters involved through simulation of small fractions of the whole sample at risk, and using 'Crude' search procedures in which only one parameter at a time is changed.

A full testing of the model requires experimentation using the whole sample at risk. But since this is almost impossible with such 'Crude' search procedures, more efficient techniques have to be utilized. Methods for locating optimum conditions that yield optimum result of a process within a system of any kind were developed by Box et. al. for the study of chemical processes (Box and Wilson, 1951; Box et. al., 1953; Box, 1954). Such methods are adopted here for the determination of parameter levels that produce the best fit of the model to the actual data. An outline of the methods, theoretical background as well as practical considerations is presented in chapter six. This is followed by an application of the methods to the problem under investigation from which a final assessment of the model and its relevant characteristics can be made.

Chapter seven investigates the effects of sex and age variables on the construction of repeated migration models. This is carried out through the application of the model to some selected sex and age subgroups. The aim is to show how far a further disaggregation of the data by sex and age could affect the process of fitting the model to the observed data.

Finally, a summary of the thesis along with general conclusions and implications for futher research are presented in chapter eight.

Chapter (2)

A REVIEW OF MIGRATION LITERATURE

2.1: Introduction:

Although the literature on migration is diverse and voluminous, migration has received less attention than other components of population change have received. Among the reasons that contributed to this are the paucity of reliable data, the difficulty of measuring and analysing the migration incident, and the lack of a conceptual framework that enables the study of migration and other related phenomena to be done effectively.

In this chapter, the major aspects of migration are reviewed as described in the migration literature. The chapter consists of five sections, the first of which is this introduction. A discussion is given in section 2.2. on how the term 'migration' is defined and why migration is vital to the welfare of society and individuals. Section 2.3 contains a description of the determinants of migration. Such a description is given in two subsections. The first deals with migration motivation by providing an answer to the question: why people migrate? The second is concerned with differential migration and the hypothesis of migration selectivity by answering the question who migrates. Migration theories and models of measurement are presented in section 2.4. In this section, the reasons behind the existing lag in the development of a general theory of migration are illustrated. Finally, section 2.5 describes the major sources of migration data and the merits and deficiencies of each of these sources. In general, the discussion of these aspects of migration will be descriptive in nature rather than analytic.

2.2: What is Migration and Why It Is Important?

There has been no unanimity among the students of population movements over what is meant by the term 'migration'. The major factor behind such dissension is that most studies on migration, with varying research inclinations, have been concerned with different and individual aspects of the phenomenon. Three types of population movement have been identified, however. They include local movement or change of residence within the community, internal migration or change of residence from one community to another within the national borders, and international migration or change of residence across the national borders, i.e. from one country to another. A distinction between international migration and other types of movement can be easily maintained but the difficulty remains in separating local movement and migration (Bogue, 1959; p.489). Local mobility can be differentiated from migration on both spatial and temporal grounds. The spatial dimension relates to distance of movement whether conceived in physical terms (mileage), in economic terms as in Stouffer's concepts of intervening opportunities,[1] or as within administrative units/areas (Gould,and Prothero, 1975; p.39). The last of these concepts, differentiating distance within administrative units/areas, is more common and easier to apply. Still the difficulty of defining what size of administrative unit/area should be considered to distinguish local mobility from migration. Should these units be defined within large provincial boundaries, intermediate commune or county boundaries or minor civil boundaries like cities or small towns? Bogue suggests, for practical purposes, the commune or county boundaries

---

1. See section 2.4 below on migration theories and models of measurement.

as medium choice between two extremes (Bogue, 1959; p. 489). The spatial dimension of movement may also relate to the direction of the move. It is sometimes preferred to physical distance for its easier applicability. It also makes possible the differentiation between certain types of migration, rural-urban for example, from the rest of movement types.

The temporal dimension in defining migration is not of less importance than the spatial dimension. Population movement entails a continuum from the often frequent movements that last only a few hours within a limited area to a permanent change from one place to another over greater distances (Gould and Prothero, 1975; p.40).

The following are some examples of the diversity of migration definitions within time. On one extreme, migrants have been defined by Barclay as the persons who travel (in Willis, 1974; p.3). But if such a definition has to be considered, certain categories of people who are not definitely migrants would have to be included. These are the daily travellers to schools or to work and the tourists. When the travel is tied with a change of residence, such categories may be excluded. Lee's definition of migration as the 'permanent or semi-permanent change of residence' is bound at that direction (Lee, 1969; p.285). But as Lee himself explained, no restriction is placed upon the distance of move, the voluntary or involuntary nature of the act, and no distinction is made between internal and external migration. The presence or absence of permanence in the change of residence has been used by Zelinsky as a criterion to differentiate between migration and circulation. The latter includes all types of moves of 'short-term repetitive or cyclical' character but all sharing 'the lack of any declared intention of a permanent or longlasting change of residence' (Zelinsly, 1971; p. 226). The major shortcoming of such a criterion is that intentions and actual performances of migration do not highly correlate. In addition, the time implied by the term 'permanent' is by no means standard. The most common period of time used in migration analysis is one year according to the U.N. definition of migrations as those movements having a duration of more than one year (U.N., 1970).

The above analysis shows how greatly the definition of migration varies with regards to space and time. Such a variation explains the observed discrepancies in the estimates of migration and the lack of comparability among results. However, the underlying character of migration does not vary a great deal whatever definition is adopted.

The Importance of Migration:

Internal migration as a demographic phenomenon is a vital subject for scientific inquiry for it affects the society on both individual and aggregate levels. The following is a discussion of the reasons why the study of migration has such an importance.

Migration has been claimed to be a necessary element for population adjustment and equilibrium. The incidence of migration stems from the disparity of opportunities and economic imbalances among localities. By shifting population from declining or stagnant areas to areas of greater prosperity, migration serves as a tool for achieving equilibrium and restoring social and economic balance (Bogue, 1959; p. 487). At the individual level, migration helps to make maximum use of human abilities and qualifications by moving away persons of special skills to places where these skills are needed and can be most effectively used. Migration is also a vehicle for cultural diffusion and social integration. Migrants as bearers of their own cultures and attitudes move to areas where they affect and are affected by the cultures and attitudes of the native population. As a result, intersectional differences are greatly lessened (Bogue, 1959; p.487).

The increasing importance of migration is reflected not only by serving as a useful instrument for solving many social and economic problems, but also by creating problems in some specific areas. Migration motivated by economic distress poses heavy burdens both on the sending and receiving areas. It accelerates economic distress in the sending areas by drawing away the more qualified persons leaving behind persons of less qualifications. This reduces the productivity in these areas and lessens their attractiveness to new industry and capital investment. (Morrison, 1975; p.233) For the receiving areas, the flow of migrants may exceed the requirements of these areas or may bring persons with qualifications different from those needed. On the other hand, it has been maintained that the characteristics of migrants tend to be intermediate between the characteristics of populations at origin and destination. Education of migrants is higher than education of non-migrants at origin, but less than that of population at destination. This has led Lee to believe that a migration 'paradox' is that the movement of people tends to lower the quality of population at both origin and destination with respect to some particular characteristic (Lee, 1969; p.296).

Another aspect of the problems created by migration to the migrants themselves and to the areas where they go is that migrants or at least some of them find it difficult to adjust to new ways of life and cultures that are different from what they are used to. This results in personal disorganization which in turn increases the number of crimes, delinquencies, broken family ties, and other aspects of social disorganization (Bogue, 1959; p.488). Finally, the influx of migrants and population concentration have created problems to the environment. There have been a mounting pressure on land resources, an excessive demand on services, a probagation of air pollution, and a spread of diseases and health hazards in areas of settlement (Willis, 1974; p.2; van Ardsol, 1966; p.50).

Implications for Public Policies

The foregoing discussion on migration importance to the community both on the individual and national scales indicates the bearing of migration dynamics on the adoption of policies that achieve social, economic and even political goals. Migration can be considered as a means of solving the problem of unemployment by adopting policy measures which induce movement of people from the depressed areas to areas where new industries are established. Such measures may take the form of financial assistance to migrants to help them and their dependants to overcome the large expenses of moving. Assistance is not only provided for migrants from the depressed areas or for unemployed persons, it may also be granted to members of the labour force and key personnel to achieve mobility within and between regions. In Britain for example, the Employment Transfer Scheme was introduced to encourage outmigration, and the Key Workers and Nucleous Labour Force Schemes were introduced to encourage inmigration. Government policies may be directed as well to the development and enlargement of the economic and social structure of the declining areas by encouraging new industries to build up and capital investment to develop in these areas. This takes the form of loans and tax incentives for new entreprises. Not less important are policies directed towards the limiting of development in the prosperous areas by restricting the establishment of new enterprises in congested areas (Willis, 1974; pp.40-44).

However, policy measures adopted for restoring the economic and social balance between the depressed and the prosperous areas seem to have not achieved their objectives with the anticipated success in some countries. In Britain, Beaumont's study of the problem of return migration under a policy of assisted labour mobility concluded that a general deficiency of manpower policies in Britain is the tendency to concentrate on the recruitment aspect of the labour supply problem and neglect the issue of retention (Beaumont, 1976). This was evidenced by the sizeable wastage rates under the E.T.S. in a number of Scottish user exchange areas. For example, there were 63.8% unsuccessful relocatees in Greenock. (Beaumont, 1976; table 1, p.83).

2.3: The Determinants of Migration:

The study of the determinants of migration is mainly concerned with the analysis of the factors that influence the decision to migrate. There is a wide variety of such factors; whether economic, ecological factors, or factors related to the personal characteristics of individuals which cause people to migrate and thus differentiate between migrants and non-migrants. Recognition of these factors provides answers to two, though indistinguishable, queries on migration: who migrates? and why? The following section deals with the second of these queries.

2.3.1: Why People Migrate:

The major causes of migration are generally stated as:

(1) Changes in the Life Cycle:

There are certain changes or turnpoints in the life cycle of the individual that accentuate migration. Migration rates vary considerably with age. Young adults are more migratory than people in other age groups. Large proportions of adults are just completing their formal education, entering the labour force for the first time, and forming their families through marriage. An economic interpretation to this phenomenon of high mobility of young persons is that rational migrants maximise their expected earnings by moving early in their productive ages (Sjaastad, 1962; p.88). Changes in marital status are also events that induce migration. Fámily formation through marriage and family dissolution through death of spouse or divorce are usually periods marked of greater mobility. Another aspect of the relationship of life cycle changes to residential mobility is the effect of the structure of the household and the need to adjust housing to the housing requirements indicated by the change in family composition (Rossi, 1955). However, the degree of accentuation of life cycle changes to migration depends on social-psychological factors such as the way in which these changes are evaluated and related to housing requirements and the availability of housing units within the range of family financial resources (Willis, 1974; p.60).

(2) Occupation and Employment Status:

Most migration studies have shown that occupation is one of the important factors that affect the migration incidence. Persons in certain occupations have been shown to have more inclination to migrate than people in other occupations. Within the labour force, white-collar workers migrate more frequently than blue-collar workers (mannual and service workers). Members of the latter group move to avoid unemployment, find work, or to obtain a steadier work, whereas members of the former group move to increase their renumeration by stepping up the occupational ladder (Lansing and Mueller in Morrison, 1975; p.226). Within the white-collar group professional workers, especially salaried rather than self employed workers, have the greatest inclination to migrate. A self employed worker is tied down by the constraints of capital investment and a clientele built up over so many years while a salaried worker is free of such constraints (Ladinsky, 1967 (a); p. 257). The relationship between residential mobility and occupation is influenced by the effects of other variables such as age, marital status, income, etc. Ladinsky indicated that young married professionals are the most migratory groups and that low income and high education stimulate migration while increases in family size and age discourage it (Ladinsky, 1976 (b); pp. 298-302). Leslie and Richardson (1961) found that career-pattern variables had higher correlation with residential mobility than had life-cycle variables. This was attributed, however, to the homogeneity of the data utilized.[2] Jansen showed that the correlation between a combination of life-cycle and career-pattern variables and future mobility intentions was very highly significant. When separating the two groups of variables, both life-cycle and career-pattern variables have had similar degrees of significance (Jansen, 1968; p. 72).

Unemployment also is an influential factor in determining the decision to migrate. Lansing and Mueller proved that unemployed workers are more inclined to migrate than employed workers. The migratory response to unemployment was shown to be 'weak' and 'uneven'. Weak because people with unemployment experience continue not to move and uneven because people who are more prone to unemployment are usually more immobile than others (Lansing and Mueller <u>in</u> Morrison, 1975; p. 226).

---

2. There were only seven out of 201 heads of household over 50 years old, no one-person households, and only 18 of 201 households were renting their homes (Leslie and Richardson, 1961, pp 899-900).

The influence of unemployment rates in different areas on the volume of migration among these areas has been investigated. Areas with higher unemployment rates are expected to lose population through outmigration to areas with lower unemployment rates. Several studies, however, produced either insignificant coefficient or a coefficient with unexpected sign for the unemployment variable (Nelson, 1959; Lowry, 1966). The failure to show the influence of unemployment rates on migration in the right direction has been attributed to the fact that migration, in these studies, has been measured by a single equation type of model which is characteristic of an inherited simultaneous equation bias. This is particularly noticed in cases where the explanatory variables are defined for the end of the period to analyse migration during the period (Greenwood, 1975).

(3) Education:

Migration inclinations vary with the educational attainment of individuals. The better educated persons are known to be considerably more migratory than the lower educated persons. The following explanations are often given in the literature. First, information on employment and job opportunities are likely to increase with increased education. Second, migration risk and uncertainty are lower for the educated people since they are more likely to have a job prior to migration. Finally, education causes the individual to be less attached to the local traditions and family surroundings and more aware of the chances available in other localities and hence the educated person is more receptible to migration than the uneducated. The importance of education stems from the fact that increasing education diminishes the effect of distance as the major deterrent to migration (Folger, 1958), (Schwartz, 1973). As information on job opportunities are more abundant to the better-educated than to the lower-educated, individuals in the former group are likely to travel longer distances than do individuals in the latter group.

(4) Social and Physical Environment:

It has been recognized that there is a strong relationship between migration behaviour and changes in the social and physical structure of the environment and the neighbourhood. Changes in residential environment of invidivuals and families increase their propensity to migrate. On the other hand, changes in the demographic and social composition in the neighbourhoods may stimulate changes in residence (Willis, 1974, p.61). The role of environmental hazards in activating population movement away from the affected areas has been illustrated by van Arsdol (van Arsdol, 1966). Using data from Los Angeles in 1960, he suggested that urban settlement intensifies certain natural hazards which accompanied the conversion of land to urban use. Examples of these natural hazards are: bush fires, earth slides and floods. Increasing densities of settlement added artificial types of hazards such as air pollution and aeroplane noise. Subsequent population changes then favoured hazard free areas (van Arsdol, 1966; p.50).

An ecological model was suggested by Wolpert to structure the relationship between individuals and their social and physical environment and other conspicuous places (Wolpert, 1966). Wolpert indicated that at critical points of the life cycle, certain stressors will face the individuals and adjustment processes will be initiated with respect to possible places in the space action. Sources of stress are noise, air and water pollution, congestion and lack of open places, and lawlessness. Individuals vary in their ability to remove or lessen the pressure of these sources of stress (Wolpert, 1966; pp.95-99).

In a survey of mobility and environmental attitude among blue-collar workers in Jarrow and Wickham on Tyneside 1970, Willis noted that environmental reasons for moving formed a small proportion of total reasons. In some cases, however, environmental disamenities have positive impact on the desire to move (Willis, 1974; pp. 61-62).

(5) Past Mobility Experience:

The importance of residential mobility in the past as a factor in determining future mobility behaviour has come under scrutiny by many researchers in recent decades. Though there is a growing support given to the existence of high correlation between future mobility and mobility past experience, the precise mode in which such experience should be expressed has been subject to disagreement among researchers. Most studies have concentrated on the duration of residence since the individual's last move as the key variable for determining future mobility. In a study of the determinants of family migration in 1943, Rider and Badger have shown that the probability of moving decreases as the length of residence increases. Persons who have moved recently are more likely to move again in the future than those who have not moved recently (Rider and Badger, 1943; p.126). Some other works have demonstrated that the frequency of past movement is a powerful expression of such mobility experience. Goldstein in his studies on repeated migration in Norristown concluded that the observed high mobility rates are the outcome of repeated movement of the same individuals rather than single moves of a large number of individuals from one place to another within the community (Goldstein, 1954 and 1964). Rowntree, in the study of the frequency of movement of migrants in England and Wales during 1948-1950, indicated the rapidity of movement among those who moved more than once (Rowntree, 1957). Morrison (1971) also noted that migration is a 'repetitive episode'. He substantiated Goldstein's proposition that the observed mobility rates reflect repeated and frequent moves by the same people 'chronic movers' rather than single moves by others. As a result, people with a past history of movement are more disposed to move again than people without any such movement history. The following explanations for such a hypermobility were given by Morrison in 1975. Some of the moves are regarded as failures and the people who made these moves either return back to their homes or try again somewhere else. Successful moves lead people to attempt achieving new success by moving again. On the other hand, some people move with the intention that it will be just a temporary move. A typical example is the transfer of labour with the knowledge that the transferred will move again either back or to a new locality (Morrison, 1975; p.225).

(6) Availability of Information:

The volume and quality of information available on different localities play an important role in determining the decision of a potential migrant as to which destination to choose. People are more likely to migrate to places which they know better than to places about which they have little information. One explanation to the common knowledge that distance is a major deterrent to migration is that information is likely to decrease and uncertainty increases as the distance of move increases. In interpreting the effect of distance on migration, Schwartz concluded that the acceptance of the hypothesis that aging does not affect the elasticity of distance whereas increasing education diminishes the absolute value of this elasticity implies that the deterrence effect of distance to migration is a 'diminishing-information phenomenon' (Schwartz, 1973). A major source of information to the migrant is the presence of relatives and friends in the area of destination (Nelson, 1959). In addition to the information that relatives and friends can provide to persons in the area of origin, they can provide the recent migrants with various kinds of help including food, shelter, and means of communication. The psychological and financial barriers associated with long distances and known to discourage migration are likely to be reduced. As a result, chain migration - joining of family members who have stayed behind - is greatly encouraged. Fabricant (1970) and Greenwood (1970) have tested the hypothesis that the current allocation of migrants is a function of the allocation of past migrants. Greenwood furtherly suggested the introduction of the 'migrant stock' variable into the estimated relationship between migration and the explanatory variables (Greenwood, 1970, 1972, 1975). Greenwood defined the 'migrant stock' variable by the number of persons born in area i (origin) and living in area j (destination) at the beginning of the period over which the migration from i to j occurs (Greenwood, 1975; p.406). Empirical analysis proved that such a variable is highly significant.

2.3.2: Differential Migration and the Hypothesis of Selectivity:

The above discussion on the determinants of migration has indicated the reasons why people migrate. It also presented a general framework to the study of who migrates which is the main concern of this section. The study of migration differentials has been the focus of a large volume of migration researches. The central theme of these researches was to show the attributes which differentiate the migrant from non-migrant sectors of population and whether migrants really expose superior qualities to those who remain in their areas of origin. In 1938 and after an extensive study of migration, Thomas concluded that the only differential that persisted in most studies and over a long period of time is that persons in their late teens, twenties and early thirties are more migratory than others (in Jansen, 1968). Bogue reached the same conclusion by emphasizing that differentials other than age do not exist and should not be expected to exist. He suggested, alternatively, to test certain hypotheses related to migration selectivity with regard to population and environment conditions (<u>in</u> Jansen, 1968). Among the hypotheses that Bogue suggested for testing were:

1. In the early stages of migration, males outnumber females while in the settlement stage, this differential disappears or even favours females over males.
2. When migration is stimulated by economic growth, stagnant areas lose the better educated and higher skilled of their populations to the economically progressive areas.
3. In the migration streams where the flow in one direction is greater than in the other, there is a great selectivity among migrants. The receiving areas attract more young adults, single males and the sending areas have a high proportion of failure migration (returnee).
4. Selectivity is at a minimum when migration is stimulated by 'push' factors while it is at a maximum when the stimuli is 'pull' factors (Jansen, 1968; pp.63-64).

In a study of migration to Bristol, Jansen indicated the following differences between migrants and non-migrants. Migrants come from a higher educational background, a higher social and occupational status and are more aware of their high status than are residents of the same characteristics (Jansen, 1968).

From the mental hygiene point of view, Thomas presented a summary of the empirical foundation of knowledge as regards selective internal migration, particularly cityward migration (Thomas, 1958). She noted that the studies on the selection of physically fit demonstrated that the losses and gains from migration had been age-selective of adolescents and sex selective of females. On the other hand, the studies on the possible selective action of migration in drawing off the more intelligent elements of the rural population to cities have been speculative and produced contradictory results, some suggesting random selection and others suggesting the selection of higher grades of intelligence. As far as the mental fitness is concerned, disorganization of migrants has been claimed but no satisfactory verification has been given. Finally, the relative success or failure of young migrants in establishing new family ties has been investigated. A favourable differential to migrants has been indicated for both sexes and for all ages. Although the evidence is not conclusive, there appears that selection is at least one of the factors involved in the success of migrants in achieving family adjustment (Thomas, 1958).

Several modern concepts have been developed in an attempt to establish a general theory of differential migration. The following are some examples. In a study of the patterns of influence in a community, Merton introduced the concepts of 'Cosmopolitanism' and 'Localism'. The distinction between cosmopolitans and locals among the influential people is based upon the orientation of the individual to groups within or outside the community. Following Merton's classification, the 'Local' confines his interests to his community and is preoccupied with local problems while the 'Cosmopolitan' has some interests in his local community but he is "... also oriented significantly to the world outside..., and regards himself as an integral part of that world" (Merton, 1957; p.393). Brown and Belcher used such concepts to study residential mobility. They concluded that cosmopolitanism or localism played more important roles in determining the propensity of physicians in Georgia to move than traditionally discussed variables in migration. 'Cosmopolitans' are those committed to professional skills, having little loyalty to the community in which they lived (Brown and Belcher, 1966). In the Bristol study, Jansen showed that migrants had a significantly higher cosmopolitan score than all residents. This approach, however, has not been tested in similar studies in the developing countries.

In 1961, Beshers and Nishiura attempted a general conceptual scheme for the prediction of differentials in the various streams of migration (Beshers and Nishiura, 1961). They classified persons who are faced with a decision to migrate or not into several types according to the mode of orientation toward the occupational goals. Persons who have long range occupational goals are said to be of a 'purposive-rational' mode of orientation. Within this group, if the future plans are made in relation to a particular territory, the mode is of a 'localized' nature, otherwise it is of an 'extended' nature. On the other hand, persons without long-term future plans are said to be of a 'short-run hedonistic' mode of orientation. Only situational factors will determine their decisions to move. Given such a conceptual framework, Beshers and Nishiura concluded the following general statements. People having a 'purposive-rational' mode of orientation will decide to migrate or not depending on whether the move will encourage or discourage the attainment of their future goals. Situational factors, on the other hand, are the stimuli for migration for people with a 'short-run hedonistic' mode of orientation. In this case, prediction of differentials will depend upon the state of the environment at any particular time. Territorial constraints will discourage migration among persons whose future orientation is of a 'localized nature'. The above considerations lead to some hypotheses related to such characteristics of migrants as occupation, age, education etc. Beshers and Nishiura have tested and confirmed these hypotheses using data on migration to Indiana (Beshers and Nishiura, 1961).

The classification of migrants into 'Resultants' and 'Aspiring' groups, postulated by Taylor in a study of migration motivation and types, resembles Beshers and Nishiura's classification into 'purposive-rational' and 'short-term hedonistic' orientated migrants (Taylor, 1969). 'Resultant' migrants are those leaving in respond to socio-economic conditions while 'Aspiring' migrants are those people who leave because of their dissatisfaction with the actual conditions and the wish to better their life further.

Within the frame of 'push-and-pull' factors that affect the decision to migrate, Lee stressed the fact of selective migration (Lee, 1969). Since persons respond differently to the 'plus-and-minus' factors at areas of origin and destination according to their personal characteristics, migrants tend to be a non-random sample of the population at origin. If migrants were of high quality, the selection is positive. If they were of low quality, the selection is negative, Those who respond to 'pull' factors at destination are positively selected whereas those responding to 'push' factors at origin are negatively selected (Lee, 1969; pp.294, 295).

2.4: Migration Theories and Models of Measurement:

The need for a general theory that deals with migration and related phenomena has been expressed in many ways by contemporary scholars. Mangalam and Schwarzweller manifested that, the study of migration has not been an object of concern for leading theorists nor has the enormous mass of findings produced in the field been incorporated into a general well-defined theory (Mangalam and Schwarzweller, 1968; p.17). They referred to some misconceptions which have led to the observed lag in the development of a general migration theory. First, migration has been considered as a random phenomenon and the decision to migrate as a unique individual response to some situational accident. Such notions, as Mangalam and Schwarzweller argue, no longer exist and the existence of a basis for migration selectivity has been empirically proved. Second, 'reductionism' is a more serious misconception that marred the building up of a migration theory. 'Reductionism' in Mangalam and Schwarzwellers' explanation is the belief that migration phenomena must be reduced to and conceptualized in 'physical' and 'biological' terms. The use of distance and age variables is an example of physical reductionism while the use of sex variable is an example of biological redctionism (Mangalam and Schwarzweller, 1968; p.12). Third, there has been a concentration on individual characteristics and a neglect of human interactional elements in the study of migration. Fourth, migration studies have been treated individually, each as a unique case. Little attention is paid to other relevant studies. Fifth, the existing materials on migration are weak and non-reliable. The need is urgent for the improvement of the existing knowledge and the development of new sources of knowledge on migration. Finally, a 'sociological avoidance behaviour' has been observed. By this, Mangalam and Schwarzweller mean the confinement of each group of researchers with their own orientations and attitudes to specific areas of research.

It was Ravenstein's famous leading articles in the 1880's, in which he postulated his 'laws of migration', that formed the first conceptual analysis of migration and marked the beginning of its theory (Ravenstein, 1885 and 1889). Although Ravenstein's laws of migration have been criticised for not being formulated in a categorical order, they remained the starting point for any further work on migration theory. The first half of the twentieth century showed no comparable work to Ravenstein's for the same reasons identified above by Mangalam and Schwarzweller. Lee followed the same footsteps of Ravenstein and developed a general scheme into which a variety of spatial movements can be placed and a number of conclusions concerning the volume and streams of migration and the characteristics of migrants can be deduced (Lee, 1969). In such a scheme, Lee recognized four types of factors which affect the decision to migrate. Some factors relate to areas of origin and destination ('push-and-pull' or 'plus-and-minus' factors), some relate to the 'intervening obstacles' between these areas, and others relate to the personal characteristics of invividuals. A person faced with a decision to migrate compares the 'plus' elements which attract him and 'minus' elements which repel him. In this comparison, the person is affected by the set of intervening obstacles which exist between the points of origin and destination and his personal characteristics which shape his response to these factors. Recently in 1971, Zelinsky attempted the application of the principle of industrial innovation to the 'laws' of migration in the aim of producing a general framework for the study of the process of migration. The attempt resulted in what Zelinsky termed 'the hypothesis of mobility transition' which is parallel to and acting interdependently with 'the hypothesis of demographic transition'. The hypothesis says, "there are definite, patterned regularities in the growth of personal mobility through space-time during recent history, and these regularities comprise an essential component of the modernization process" (Zelinsky, 1971; pp.221-222). Five distinctive phases of the mobility transition parallel to the five phases of demographic transition are considered to describe the sequential spatiotemporal processes among modernized populations. They are shown in the following paragraphs.

Phase 1: The pre-modern traditional society, both fertility and mortality are at high levels and the population size is relatively stable. There is little genuine residential mobility and only limited circulation[3]

---

3. See definition, page 9.

sanctioned by customary practice in land utilization, social visits, commerce, warfare or religious observances.

<u>Phase II: The early transitional society</u>, fertility retains its high level, mortality declines rapidly and thus population size is growing. In the meantime, a massive movement from countryside to cities and to colonization frontiers, emigration to attractive foreign countries, small immigration of skilled and professionals, and growth of circulation are observed.

<u>Phase III: The late transitional society</u>, fertility declines, mortality decline slackens, and population rate of growth slows down. On the mobility sphere, movement to cities and colonization frontiers is lessened, emigration decreases, and circulation increases with growing structural complexity.

<u>Phase IV: The advanced society</u>, fertility decline terminates at a low level, mortality is stabilized at a level below that of fertility and a very slight natural increase occurs. On the other hand, residential mobility levels off and oscillates at a high level, movement to cities is further reduced, inter and intra-urban migration predominates, immigration of unskilled or semi-skilled persons is observed, and economic orientated circulation accelerates.

<u>Phase V: The future super advanced society</u>, more carefully controlled fertility is likely and a stable mortality pattern below present level is predictable. In this final phase of mobility transition, a decline in residential movement and some forms of circulation will occur, migration will be mostly of inter and intra-urban type, further immigration of unskilled workers from underdeveloped areas is possible, and strict political control of internal as well as international migration may be imposed.

Apart from the above few generalizations on migration theory, the huge and ever-increasing number of migration studies launched in recent decades failed to provide a comprehensive theory for such a complex phenomenon.

Migration Models:

Migration is one of the few fields of sociology that has been characterized by the invention of a wide range of mathematical formulas for prediction purposes. The most striking fact observed in migration was the relationship between distance and the number of moves. It was Ravenstein who first noticed that migration decreases with increasing distance between areas of origin and destination (Ravenstein, 1885). Such a fact has been mathematically formulated in a number of ways most of which are based on what is known as gravity-type or human interaction models. The most famous of these models, from which subsequent models were originated, is the well-known Zipf's $P_1 P_2 / D$ hypothesis. Such a hypothesis implies that migration and other types of interregional exchange is directly related to the sizes of populations of the two regions involved and indirectly related to the distance between them (Zipf, 1946). In mathematical terms the model takes the form:

$$M = P_1 P_2 / D$$

or as some researchers put it in a more general form as

$$M_{ij} = k P_i P_j D_{ij}^{-\alpha}$$

where, M_{ij} denotes gross migration between regions i and j,

P_i P_j are the total populations of the regions,

D_{ij} is the distance between them, and

K, α are positive constants to be determined, (ter Heide, 1963).

Various explanations have been given by many authors to the inverse relationship between distance and migration. The most important of these explanations are that cost and difficulty of moving increase over long distances, and that information on economic opportunities available at different places is more abundant at shorter distances. Zipf's formula has been tested and shown to yield satisfactory results (Folger, 1953; Anderson, 1955). The major deficiency of the model, however, is the tendency to overestimate short-distance migration. This is a typical characteristic of all pareto-type models. For that reason, some authors have attempted other formulas such as Lognormal, Exponential, and Gamma distributions for the distance variable (Somermeijer, in ter Heide, 1963; Morril and Pitts, 1967). There has also been disagreement about the power to which the distance variable should be raised and about whether or not the population variables should be given different weights (Anderson, 1955; ter Heide, 1963). Moreover, as there has been a differentiation between technical distance (means of transportation) and social distance (differences in culture or social status among regions), attempts have been made to introduce to Zipf's formula some measures of the relative attractiveness of regions which account for differences in such physical and social distances.

One of these attempts was carried out by Somermeijer, 1961 (<u>in</u> ter Heide, 1963). Somermeijer calculated R_{ij}, a measure for the differences in religous composition in areas, i, j. The modified formula became:

$$M_{ij} = K P_i P_j D_{ij}^{-\alpha} (1 + \beta R_{ij})^{-1}$$

The measure of relative attractiveness was furtherly developed to include other aspects than religous differences.[4]

A different version of the gravity type models was introduced by Stouffer in 1940. Stouffer completely denied any relationship between migration and distance and rather stressed the notion of 'opportunities'. The original hypothesis of Stouffer stated that "the number of persons going a given distance is directly proportional to the number of opportunities at that

---

4. A more detailed account of Somermeijer's attempts is found in ter Heide's article (1963).

distance and inversely related to the number of intervening opportunities" (Stouffer, 1940; p.846). The intervening hypothesis was later extended by Stouffer to include the 'competing migrants', i.e. the potential migrants to area of destination from other areas (Stouffer, 1960). Stouffer's hypotheses have also been tested and shown to yield similar (in some cases better satisfactory results to Zipf's hypothesis (Strodtbeck, 1950; Folger, 1953; Galle and Taeuber, 1966; Isbell, 1966). However, the methodological weakness of the model as far as the operational definitions of the variables involved are concerned has been criticised by Stouffer himself and by others (Galle and Taeuber, 1966). The measurement of opportunities depends on the knowledge of the numbers of in and out migrants which may be unknown and hence the model will fail in predicting future migration streams. In addition, a common deficiency between Zipf and Stouffer's hypotheses is the neglect of the psychological behaviour or motives of migrants which may be necessary or at least relevant to the analysis of migration. A model which takes into account both the characteristics of areas and those of individuals was suggested by Price (Price, 1959). Price presented his model by considering five individuals each with four characteristics and three areas each with three characteristics. Denoting individuals by the sumbol I_i, $i = 1,2,\ldots 5$, and their characteristics by the symbols A_i, B_i, C_i, and D_i, and denoting the areas by A_j, $j = 1,2,3$, and their characteristics by X_j, Y_j, and Z_j, the probability of individual I_i moving from area A_j to area A_k will be:

$$P\ (I_i, A_j, A_k) \quad = \quad F\ (A_i, B_i, C_i, D_i;\ X_j, Y_j, Z_j;\ X_k, Y_k, Z_k).$$

Developing the model further gives:

$$\begin{aligned} P\ (I_i, A_j, A_k) \quad = \quad & e_1\ (A_i) + e_2\ (B_i) + e_3\ (C_i) + e_4\ (D_i) \\ + \quad & f_1\ (X_j) + f_2\ (Y_j) + f_3\ (Z_j) + g_1\ (X_k) + g_2\ (Y_k) + g_3\ (Z_k) \\ + \quad & h_1\ (X_j - X_k) + h_2\ (Y_j - Y_k) + h_3\ (Z_j - Z_k) \end{aligned}$$

The model was developed on the basis that it will be simulated on electronic computers. First, the probability of moving based on individual's characteristics such as age, sex, marital status etc. is calculated. Second, this probability is partitioned into probabilities of moving to different possible destinations according to certain criteria such as population, contiguity, urbanization and others. Population counts from year to year can be computed and changes in the characteristics of areas or individuals can be introduced. Although Price's model considers the behavioural aspects of migration and the relative attractiveness of areas by incorporating the relevant characteristics of individuals and areas, it suffers from the following deficiencies. There is the neglect of distance as an important factor per se and as a proxy for other factors which do not appear in the model. There is also the use of additive rather than multiplicative formulas. The latter have proved to be far more superior than the former.

Migration as a Human Capital Investment:

Economists have regarded migration decisions as part of general investment theory, and in particular as part of human capital investment theory (Laber and Chase, 1971 ; Sjaastad, 1962). Sjaastad placed migration in a human investment context by considering the returns and costs of moving. He identified two types of costs. Monetary costs are represented in the increase in individual's expenditure on food, lodging and transportation in the course of moving. Non-monetary costs include opportunity costs such as the earnings foregone while travelling, searching for and learning a new job. They also include 'psychic costs', the costs of leaving familiar surroundings, family, relatives and friends. Sjaastad argues that 'psychic costs' would be difficult to quantify and if they were quantified, they should be treated differently from other types of costs (Sjaastad, 1962; p. 85). As for the returns to migration, they include the positive or negative increment to the real earnings that the migrant obtains by leaving his place of origin to a new place. There is also a non-monetary component in migration returns. This includes locational preferences on the part of the migrant and the satisfaction or dissatisfaction that he may receive in his actual move. Again, this type of returns is difficult to quantify.

The framework of the economic models which treat migration as a human capital investment is generally set as follows.

Let the present value of the differences between earnings in area j and area i in the migration stream between i and j be denoted by $\sum_{t=1}^{n} (E_{jt} - E_{it})/(1 + r)^t$, where r is the rate of discount. Similarly, let the present value of net costs associated with residence in these areas be denoted by $\sum_{t=1}^{n} (C_{jt} - C_{it})/(1 + r)^t$.

It follows that the present value of investment in migration from i to j [$(P V)_{ij}$] will be: $(P V)_{ij} = \sum_{t=1}^{n} (D_{jt} - E_{it})/(1 + r)^t - \sum_{t=1}^{n} (C_{jt} - C_{it})/(1 + r)^t$

A potential migrant will decide to migrate if $(PV)_{ij}$ is positive and will choose the destination that makes $(PV)_{ij}$ at maximum (Greenwood, 1975; p. 399). A collection of models have been developed following the same lines of reasoning. They are all set in the frame of cost-return or loss-benefit of migration (Vanderkamp, 1969 ; Speare, 1971; Laber and Chase, 1971).

Cost-benefit analysis of migration is also concerned with the economy as a whole as well as parts of it (individuals or regions). The main concern in the aggregate case is whether the society as a whole will be better off by allowing migration from lagging areas to areas of prosperity or following the alternative policy of developing the lagging areas through the encouragement of industry and investment to move to these areas (Willis, 1974; p.47). The analysis on the aggregate scale will however be more difficult than that on the invidual scale. In the former, it is not easy to define what is considered as cost or return element.

Stochastic Process Models of Migration:

A special class of models for migration analysis is the one in which migration is treated as a stochastic process, a process that develops with time and can not be predicted in the future with great certainty. Specifically, the well-established theory of Markov chains has been used to present changes with time in the population as a whole or specific segments of it, the labour force for example (Tarver, 1965; Blumen, Kogan, and McCarthy, 1955). The probabilities of moving from one place to other possible places at a given point of time are treated as a matrix of transition probabilities. Using such a matrix and given a vector of probabilities representing the proportion of population at various places at an initial time point, the structure of population at any time point in the future can be determined. The procedure is as follows:

If $[P = P_{ij}]$ is the matrix of transition probabilities, with non-negative elements and unit column totals,

and $p_i^{(0)}$ is the population structure at the initial time point $(n = 0)$,

then $p_i^{(1)} = P\, p_i^{(0)}$,

and $p_i^{(n)} = P^n\, p_i^{(0)}$

The use of simple Markov chains in migration has been criticized by not satisfying the necessary conditions of a Markov chain. These are the constancy of probabilities over the period of estimation and the irrelevance of information prior to the observed period. In migration, such conditions are unrealistic. Attempts have been made to relax these conditions by the use of non-stationary probabilities and the permission of a change in the transition matrix to occur according to some factors related to the knowledge of individual's previous migration such as the duration of residence prior to the move (Myers, McGinnis, and Maznick, 1967). However, the dynamic models which are based on more elaborated Markov chains require stationary probabilities for each category of population (of specific duration of residence) which is still an unrealistic assumption.

2.5: Migration Data:

While migration is at least of equal importance to other components of population change (fertility and mortality) it is the least satisfactorily studied of these phenomena. The inadequacy and non-reliability of migration data are the major factors, among others, in this respect. Three possible sources of data on migration are generally recognized. The following is a description of these sources along with an illustration of the merits and deficiencies of each of them.

(1) Censuses and Surveys:

The most-widely used source of information on internal migration is population censuses. In addition, special field surveys can be carried out by an official or private agency using a small sample of population. Migration data from these sources can be derived in two ways. The first is to ask the interviewed direct questions on place of residence, place of residence at any specified time, duration of residence, number of moves, ... etc. The second way is to compare population counts at two fixed dates in order to compute population total increase. When natural increase (births minus deaths as obtained from birth and death registers) is considered, an estimate of migration can be derived. This is what is known as the 'residual method' of estimation. Population censuses favour sample surveys in their coverage of the entire population. Sample surveys, on the other hand, have the advantage that valuable statistics can be obtained without the need to a large geographic detail. The increased costs of coding and enumeration for collecting census data must be weighted against the reliability and sampling limitations in sample surveys. Sample surveys have the additional advantage of the possibility of experimenting with novel questions on migration which helps in exploring the subject in greater depth (Shryock et.al.,1973). Both sources of data, however, share the common disadvantage of relying on retrospective recording of events. There is always a tendancy with human memories to forget the exact date or the original motivation for a move (Hollingsworth, 1970; p.2).

(2) Population Registers:

The second major source of migration data is a system of continuous population registers in which a complete and updated history of changes in residence is recorded. From such registers, valuable migration statistics can be compiled and analysed. Continuous population registers are not restricted only to migration data but are extended for political, sanitary or educational usages. Many of the Scandinavian and western European countries adopt the system of continuous population registration. In Britain, the National Registration system was set in September 1939. Each individual having an identity card and a ration book has to report to his local office all changes of residence except the temporary ones. The National Health Service registers replaced the National Register in February 1952. The N.H.S. records contain all changes in physician and changes of address.[4]

On theoretical grounds, continuous population registers can provide valuable data on migration. For up to date information are collected and the interaction of different types of data from various sources can easily

---

4. The N.H.S. records are the source of migration information for this research. A more detailed discussion on these records and their shortcomings is in chapter III.

be maintained (Willis, 1974). However, up to now, little use has been made of such registration systems, especially in the field of migration.

(3) <u>Auxiliary Sources</u>:

These are the least accurate sources of information on migration. They include electoral rolls, electricity board records, trade union records, and other miscellaneous sources. The use of such records for migration analysis is more tedious than the use of other sources. In the former, additional elaboration of the records is needed since they are not originally designed for migration.

<u>Longitudinal Approach to Migration Data</u>:

The collection of complete residence histories has been recently suggested by Taeuber as the most suitable approach to obtaining migration data (Taeuber, 1966). A longitudinal study of migration is the one which follows up the life span or part of the life span of individuals under investigation and record all their migratory acts in that span. In his article 'cohort migration', Taeuber presented some of the advantages of longitudinal analysis of migration. Such an approach implies the study of individual behaviour through time since migration is considered not only a function of current characteristics of individuals but also a function of their past behaviour and changes in their characteristics. Taeuber adds that with residences history approach where each move is assigned to a specific cohort and a specific age period during which the move occurred, several inter-and intra-cohort comparisons can be facilitated. Eldridge maintained that residence histories approach allows the study of migration by the type of move where the extent of return migration, primary and progressive (secondary) migrations can be measured and analysed (Eldridge, 1965). Residence histories approach also permits the study of the effect of past migration experience on future migration whether such an experience was expressed in terms of last duration of residence (Rider and Badger, 1943) or in terms of the frequency of past movement (Rowntree, 1957; Morrison, 1971). The concept of 'exposure residences' was adopted by Taeuber and associates to summarize residence histories. "An individual is assigned an exposure residence in a size-of-place interval if he has lived for ten or more years in places within that interval. The residence does not have to be continuous nor does it have to be in only one place". (Taeuber et.al., 1961; p.826). The size of place and the length of residence required to define 'exposure residences' may vary according to the purpose of study. The concept of exposure residences as a measure of migration is used as an indication of the individual's exposure to social and physical environmental conditions of specific areas.

Conclusion:

The above account of previous studies of the determinants, as opposed to the consequences, of migration suggests that certain characteristics differentiate migrants from non-migratns. The most important differences are those related to changes in the individual's life cycle (age, marital status, etc.), occupation, educational attainment, and past mobility experience. However, the first three of these factors have received a disproportionate amount of attention, relative to past mobility experience in previous migration researches; both on theoretical and empirical levels.

The intention of the present study is to take a first step in the direction of remedying such a limited attention paid to past mobility experience as an important factor in the decision to migrate. Recognition of the importance of this factor, in general, establishes the presumption that the phenomenon of repeated migration will be an important area for research. This is the central theme developed in the work of this thesis.

Accordingly, previous mobility experience is not studied in its own right, but rather as an important explanatory power in our attempt to develop a general model designed to explain and predict the patterns of movement experienced by repeared migrants in Scotland during 1939-64.

Before undertaking the task of developing such a model it is essential that we consider in some detail the actual patterns of movement exhibited by the Scottish Central Register statistics for the years 1939-64 with regards to repeated migration. The account of these patterns is given in chapter four. Prior to this, we present in chapter three a brief discussion of the Central Register for Scotland as the basic source of information from which our sample is drawn. The drawing of the sample, some possible deficiencies of the data together with an outline of the sample characteristics as revealed by first tabulations are presented there. It is to this discussion that we now turn.

Chapter (3)

DESCRIPTION OF RESEARCH MATERIALS

In this chapter, the Central Register for Scotland, which is the basic source of information utilized in the present study, is described. Some of the deficiencies of the registration system in Britain are mentioned, the method of selecting the sample is illustrated, and the characteristics of the sample revealed by the first tabulations (after preliminary editing) are presented. Finally, a procedure for filling in the missing dates of death and the missing dates of subsequent moves is discussed as a partial correction of some of the data deficiencies.

3.1 Data Sources:

The study reported here is mainly based on materials extracted from the Scottish Central Register between 1939 and 1964. The S.C.R. constitutes a part of the National Register in Britain from 29th September 1939 to 21st February 1952 and a part of the National Health Service Register from the 21st February 1952. The National Registration system was first introduced in 1939 for purposes of national security during the war and rationing. It was replaced by the National Health Service Registration on 21st February 1952. Changes of address were recorded under the National Registration system while only changes of physician were notified under the N.H.S. registration system.

Four earlier studies have used the N.R. statistics in analysing internal migration. The first two by Newton and Jeffrey (1951) and by Rowntree (1957) related to migration in England and Wales over a three year period 1948-1950. The other two studies by Hollingsworth (1968 and 1970) related to migration in Scotland and covered the period 29th September 1939 to 31st August 1964. The data set used in the present study is exactly the same data set originally used by Hollingsworth for his migration studies. Although Hollingsworth's studies have covered a wide variety of migration patterns in Scotland, the effect of repeated movement on migratory mobility was not fully explored. A re-examination of the S.C.R. records is intended here to fill such a gap.

The statistics based on the S.C.R. are far from complete and may need correction before analysis. An account of the deficiencies of these data and possible ways of correcting them has been given by Hollingsworth (1968). The major shortcomings of the data under the N.R. system (up to 21st February 1952) were the under-recording of births and deaths, the treatment of a large number of people demobilized from the armed forces as inmigrants and the inevitable neglect of persons moving to, from, or within Scotland who had first registered in some other part of the U.K. Moreover, new defects have arisen when the N.R. was replaced by the N.H.S.R. in 1952, since only changes of physician and not changes of address were the basis of registration. Because many people moved without bothering to change their doctor, the number of moves is likely to have been under-stated. In addition, a considerable time could elapse between a move and its recording. Only a partial correction of some of the entries of these records has been attempted here. This will be shown in later sections.

The data set, however, is still regarded as a worthwhile source of information on internal migration in general and on repeated migration in particular.

3.2: The Sample:

The data set employed here consists of a systematic random sample of 1 in 440 of the registered population stratified by the place of first registration. One record was taken randomly from every tenth page of each register book, constituting approximately 20,000 cases. The basic information included in these records are.

1. Type of registration. There are six different types of registration: Original registration on September 1939, new entrance on the Scottish Register since 1952, and new registration on loss of identity cards and/or ration books.
2. Date of registration.
3. Area of registration. It may be one of the 25 local executive areas contained in Scotland. There have been some changes in area boundaries since 1939. Most of these changes however have been by merging areas together so that local movements within areas can be easily eliminated.
4. Sex.
5. Marital status (Single, married, widowed and divorced).
6. Change in the marital status. Any of four alternatives might have occurred: no change in the initial status at registration, change from single to married, change of name only, and change of name but from adoption.
7. Year of birth.
8. Year of death if the person is reported dead.
9. Occupation.
10. Total number of moves. Any change from one local executive area to another is counted as a move. Any change of address or physician within one local executive area is not included in the Central Register and is not thus counted as a move.
11. Area and date of all subsequent moves. Up to only thirteen moves are recorded.

3.3: First Tabulations of the Data:

In the first attempt to process the sample data, some unexpected codes were encountered on the cards during the punching stage. Such odd codes were checked up and amended. In addition, an overall transformation of the existing alphabetic coding system to a numeric coding system was performed to render future computation easier. A one-way tabulation of the data is shown in table 3.1. Inspection of the entries of this table reveals the following characteristics of the sample.

(1) The total number of observations is 19886, fewer by less than one per cent than 19993, the original number drawn by Hollingsworth (1968).

(2) Type, year, and area of registration. Nearly fifty-six per cent of the total sample (11060 cases) were originally entered on the National Register in September 1939. Twenty-seven per cent (5264 cases) were counted as births during the whole period 1939-1964. The registered on demobilization from the armed forces constituted almost 6.5 per cent of the total sample. The remaining cases were distributed among the new or war entrants (4.4 per cent), the new entrants on the Scottish Register from 1952 (2.5 per cent), and those who were issued new numbers on their loss of identity cards and/or ration books (4.6 per cent). This reflected the distribution of the sample according to year of registration where the biggest proportion was registered in 1939. The period of war and the following years till 1949 were characterized by higher rates of registration than other years.

The highest of these rates was recorded in 1946 with the number of registration mounting to 1062 cases (5.3 per cent).

The remaining registrations were evenly distributed among the rest of the years with an average equal to approximately 1.3 per cent of the total sample.

The distribution of individuals according to their area of registration reflects the effect of the area size as measured by the size of its population. The largest number of first registrations were recorded in Glasgow which also had the largest size of population. The next highest scores both in terms of the total number of population and the number of registrations were made by Lanark area and Edinburgh city respectively (about ten and nine per cent of the total registrations). Renfrew, Ayr and Fife were next providing 6.5, 6.2, and 5.6 per cent of the total registrations. On the other hand, five of the 25 local executive areas contributed by less than one per cent of all registrations. These areas are: Ross and Cromarty, Caithness, Orkney, Zetland, and Sutherland. They are among the smallest size areas in terms of the total number of population as found in 1961 census for Scotland.

Table 3:1

Frequency and Percentage Distributions of the Whole Sample With Regards to Basic Variables

(Total Cases = 19886)

Variable	No. of Cases	Percentage
i: Type of Registration:		
1. New registers, 1939	11061	55.6
2. War entrants	877	4.4
3. Demobilisation from A.F.	1282	6.4
4. Birth registers	5264	26.5
5. Entrants on S.C.R. 1952	493	2.5
6. New entrants, loss of identity	907	4.6
Unstated	3	-
ii: Year of Registration:		
1939	11127	56.0
1940	343	1.7
1941	407	2.0
1942	458	2.3
1943	393	2.0
1944	380	1.9
1945	699	3.5
1946	1062	5.3
1947	476	2.4
1948	461	2.3
1949	345	1.7
1950	257	1.3
1951	241	1.2
1952	233	1.2
1953	280	1.4
1954	271	1.4
1955	256	1.3
1956	249	1.3

Table 3:1 Contd.

Variables	No. of Cases	Percentage	Population 1961*
1957	242	1.2	
1958	258	1.3	
1959	259	1.3	
1960	263	1.3	
1961	258	1.3	
1962	266	1.3	
1963	256	1.3	
1964	143	0.7	
i: Area of Registration:			
1. Argyle and Bute	340	1.7	72236
2. Aberdeen and Kincardine	671	3.4	165077
3. Angus	378	1.9	97417
4. Ayr	1238	6.2	361785
5. Banff, Moray and Nairn	430	2.2	107043
6. Caithness	103	0.5	27915
7. Dunbarton	634	3.2	237682
8. Dumfries	376	1.9	88274
9. Fife	1121	5.6	327817
10. Kirkcudbright and Wigtown	253	1.3	55069
11. Inverness	347	1.7	89409
12. Lothians and Peebles	955	4.8	311393
13. Lanark	2029	10.2	628111
14. Orkney	86	0.4	17254
15. Perth and Kinross	540	2.7	132045
16. Ross and Cromarty	176	0.9	58770
17. Renfrew	1296	6.5	363151
18. Roxburgh, Selkirk and Berwick	363	1.8	83993
19. Sterling and Clackmannan	861	4.3	254903
20. Sutherland	46	0.2	13140
21. Zetland	82	0.4	17567
22. Aberdeen City	710	3.6	178441
23. Dundee City	680	3.4	182467
24. Edinburgh City	1794	9.0	448895

Table 3:1 Contd.

	Variables	No. of Cases	Percentage	Population 1961*
	25. Glasgow City	4374	22.0	897848
	Unknown	3	-	
iv:	Sex			
	Males	10383	52.2	
	Females	9491	47.8	
	Unstated	12	-	
v:	Marital Status			
	Single	12929	66.6	
	Married	5665	29.2	
	Widowed	821	4.2	
	Divorced	12	0.1	
	Unstated	459	2.3	
vi:	Change in Marital Status			
	1. No Change	18062	90.9	
	2. Single to married	1445	7.3	
	3. Change of name	344	1.7	
	4. Change of name from adoption	29	0.1	
	5. Unknown	6	-	
vii:	Living or Dead During 1939-64			
	Living	16524	83.1	
	Dead during 1939-64	3166	15.9	
	Unknown year of death	196	1.0	
viii:	Total Number of Moves			
	No moves	11052	55.6	
	One move	4238	21.3	
	Two moves	2038	10.3	
	Three moves	970	4.9	
	Four moves	621	3.1	
	Five or more moves	962	4.8	
	Unstated	5	-	

x The figures are extracted from 1971 census for Scotland, volume 1, table 3,

(3) Sex and Marital Status:

The sample consists of 10383 males (52.2 per cent) and 9491 females (47.8 per cent) with a sex ratio of 109.4. In a further twelve cases the sex was unrecorded. The highly dominating figure for males may reflect the fact that males are likely to re-register themselves owing to the changes that may more often occur to males rather than to females, such as the demobilization from the armed forces.

Regarding the marital status at registration, we notice the following facts. Almost two thirds of individuals are single persons and more than a quarter of them are married while the widowed and the divorced constitute smaller proportions of the total sample (five and one per cent respectively). Several hundreds of cases are of unknown marital status.

If we consider the changes that might have occurred in marital status during the whole period we may conclude that the majority of the sample maintained their original marital status at registration throughout the whole period while about seven per cent of all cases changed their statuses at some stage during the period through marriage. Slightly fewer cases changed their names without a real change in the marital condition. Further cases recorded similar changes of names but from adoption. Analysis of such changes in the marital status with time seems impossible since almost ninety per cent of those who changed their marital statuses did not specify the dates when such changes occurred.

(4) Deaths:

Over eighty per cent of the total sample (16524 cases) were under observation since their first registration till the end of the whole period, while seventeen per cent (3166 cases) of them were counted as deaths during the period. For as many as six per cent of those who were reported dead (196 cases), the year of death is unknown. Otherwise the deaths were evenly distributed among the years of observation, with only a few exceptions. For example, the number of deaths was underestimated for the years 1940 (5 deaths) and 1953 (97 deaths) while no deaths at all were reported in 1939. An interpolation procedure for assigning values for the unknown years of death and the unknown years of moves was introduced to rectify this deficiency. This is contained in section 3.4. An excess in the number of deaths, on the other hand, was observed in some other years. For example, the deaths recorded in 1946 and 1956 (162 and 199 respectively) were relatively higher than those recorded in other years.

(5) Occupation:

Occupation is one of the most important variables in the study of migration. Unfortunately, the information available about occupation in the present sample is not sufficient to draw representative conclusions. For nearly seventy per cent of the total number of cases the occupation is unknown. Even within each of the classified occupations, there are further cases where the individuals were retired persons in which case the previous occupation is of little interest. Within these known occupations, however, the largest proportions were classified either as service, sport, and recreation workers or as sales workers. These constitute 3.4 and 3.3 per cent of the total sample respectively. In the next place come labourers; farmers, foresters, and fishermen; and engineers and allied trade workers with corresponding proportions of 3.1, 2.7 and 2.4 per cent of all cases.

(6) Mobility:

For each individual the total number of moves as well as the year and areas of all subsequent moves up to thirteen were recorded. About fifty six per cent of the whole sample (11052 cases) have zero frequency of movement and may be classified as non-movers. On the other hand, the mobile sector consists of approximately forty per cent of the total population. These are those persons who made one or more moves during the period since first registration till death or 1964. Half of the mobile persons and fifth of the whole sample (4238 cases) made just one move. A second move was made by almost a quarter of the movers (10 per cent of the total sample). The third and fourth moves were exercised by ten and seven per cent of the movers (five and three per cent of the total population). Still about five per cent of the total sample made five or more moves. This indicates the high potentiality of the present sample as a valuable source of information on repeated migration. A more detailed study of the frequency of movement will follow in a subsequent chapter.

3.4: Interpolating Missing Dates of Migration Events:

It has already been shown that there is a considerable number of missing observations for most of the variables considered in the sample. In some instances they are so widespread that no gains could be achieved by interpolating values for these cases (e.g. occupation). In other instances, the removal of the undefined cases will make the analysis of the variables under consideration incomplete with a resulting loss of generality. An interpolation procedure for filling in the undefined values for these variables is seen to be useful. The most desirable variables that require the development of such a procedure are the year of death for those who were reported dead during the study period and the years of subsequent moves for those who did move.

Hollingsworth in the first of his two studies showed that the National Register was obviously badly kept for both death and migration registration for the first period of 18 months (Hollingsworth, 1968). He later proposed a theory[1] for interpolating the missing dates of death and movements according to which as many of the unknown dates as possible should be defined in that period.

For the unknown dates of death, the procedure begins with defining for every case the last definite event (L.D.E.) ever occurred. Then it proceeds to assuming probabilities of dying in the years 1939, 1940, or 1941 in the proportion 1:4:1 if L.D.E. is 1939; probabilities of dying for 1940 or 1941 in the proportion 4:1 if L.D.E. is 1940; and probability of dying = 1 for 1941 if it is the date of the last definite event.

If L.D.E. is 1942 or later, equal probabilities will be given for the years from L.D.E. up to 1964. For each case a random integer/number 'R' in the range 1 to the sum of weights given is chosen and the unknown date of death is assigned a value according to 'R' and the corresponding weight.

For the unknown dates of movement, the same technique of correction is adopted. Similar assumption as in the case of death is made, that is, the weakness of migration registration in the first period of 18 months of the systems' existence. A slight difference, however, is that the unknown date of movement D_2 is being interpolated between two known dates D_1 and D_3. In some cases D_1 may be the year of registration, if

1. T.H. Hollingsworth, Migration Research, 1939-64: Theory of Interpolation of Missing Dates, an unpublished note (August, 1974).

the unknown date is that of the first move. On the other hand, D_3 will be either the year of death (if the person is reported dead) or 1964 (if he was still living at the end of the period) if the unknown date is that of the last move. Varying probabilities of moving are assumed for the years between and including D_1 and D_3. The proportions of these probabilities are as follows:

1. 1:4:1 for the years 1939, 1940 or 1941 if D_1 is 1939 and D_3 is 1941 or later.
2. 1:4 for the years 1939 or 1940 if D_1 is 1939 and D_3 is 1940.
3. Probability equal to 1 for 1939 if D_1 and D_3 are 1939.
4. 4:1 for 1940 or 1941 if D_1 is 1940 and D_3 is 1941 or later.
5. Probability equal to 1 for 1940 if both D_1 and D_3 are 1940.
6. Probability equal to 1 for 1941 if D_1 is 1941 and D_3 is 1941 or later.
7. If $1942 \leqslant D_1, D_3 \leqslant 1951$, the weight 2 is given to the years corresponding to D_1 and D_3 and the weight 4 is given to each of the intervening years.
8. If $1942 \leqslant D_1 \leqslant 1951$ and $D_3 \geqslant 1952$, the weight 2 is given to the year corresponding to D_1, the weight 4 is given to each of the years following D_1 up to 1951, the weight 2 is given to each of the years following 1951 up to the year just before D_3 and the weight 1 is given to the year corresponding to D_3.
9. If $D_1 \geqslant 1952$ the weight 1 is given to both the years corresponding to D_1 and D_3 and the weight 2 is given to each of the years between them.

For each of the above nine cases and similar to the case of death, an integer random number 'R' between 1 and the sum of weights given is chosen and the unknown date of movement D_2 is assigned a value according to the value of 'R' and the corresponding weight.

The results of the application of the above procedure are contained in table 3.2. It shows the interpolated values for the unknown years of death and the unknown years of successive moves. For each event, the number of cases before and after interpolation along with the differences between them are given.

The Interpolated Values of Missing Years of Death and Moves

Event	Death			Move 1			Move 2			Move 3			Move 4			Move 5			Moves 1-
Year	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(3)
1939		31	31	3	189	186	1	5	4								1	1	191
1940	5	140	135	216	872	656	81	162	81	25	32	7	4	5	1				745
1941	120	140	20	945	1080	135	408	434	26	148	155	7	45	46	1				169
1942																			0
1943										163	164	1				50	51	1	2
1944				404	406	2							85	86	1				3
1945				386	388	2	223	226	3	114	116	2				50	51	1	8
1946				568	573	5	227	228	1										6
1947				416	420	4	245	247	2	144	145	1				61	63	2	9
1948				450	456	6	301	304	3	140	141	1	103	105	2				12
1949				352	360	8	266	269	3	140	144	4	90	91	1	51	52	1	17
1950	145	146	1	297	301	4	230	232	2	149	152	3	104	105	1				10
1951	149	150	1	289	294	5	260	273	13	173	179	6	101	105	4				28
1952				159	163	4	84	91	7	81	85	4	62	69	7	39	40	1	23
1953				177	182	5				68	69	1	34	35	1	31	32	1	8
1954				182	188	6	101	103	2	71	72	1				25	26	1	10
1955				175	179	4	108	111	3	59	60	1	38	40	2	24	27	3	13
1956				178	179	1													1
1957				160	164	4	134	136	2	73	79	6							12
1958							124	125	1	57	61	4	59	60	1				6
1959	143	144	1	144	147	3	118	120	2	65	66	1				27	29	2	8
1960				194	196	2	107	109	2	68	72	4	49	50	1				9
1961				173	179	6	138	142	4							29	31	2	12
1962				163	164	1	112	113	1	70	71	1	30	31	1				4
1963	128	130	2	178	182	4	96	97	1							30	31	1	6
1964	70	72	2	96	97	1													1
Missing Cases	196	3	193	1056	2	1054	163		163	55		55	24		24	17		17	1313

(1) = Number of cases before interpolation.

(2) = Number of cases after interpolation.

(3) = Number of cases being interpolated.

Table (3.2) Contd.

Event	Move 6			Move 7			Move 8			Move 9			Move 12			Moves 1-12
Year	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(3)
1939																191
1940																745
1941																169
1942																0
1943																2
1944																3
1945							18	19	1							9
1946																6
1947																9
1948																12
1949							20	21	1							18
1950							22	24	2							12
1951	50	52	2	44	45	1				15	17	2				33
1952				15	16	1	8	9	1							25
1953																8
1954														1	1	11
1955																13
1956																1
1957																12
1958																6
1959				9	10	1										9
1960										9	10	1				10
1961	15	16	1							4	5	1				14
1962																4
1963																6
1964	8	9	1													2
Missing Cases	5	1	4	3		3	5		5	4		4	1		1	1330

Chapter (4)

THE FREQUENCY OF MOVEMENT: DETERMINANTS AND IMPACT ON FUTURE MOBILITY

4.1: Introduction:

The classical approach to the study of internal migration through comparison of residence at two fixed points of time is mainly concerned with population redistribution. It puts more emphasis on specific moves while overlooking the others. Duration of residence and frequency of movement approach, on the other hand, is a much more useful device for documenting migration. It allows a longitudinal analysis of the social and environmental changes that affect the migration behaviour and makes possible the differentiation of movers by frequency of residential changes and by type of move (Taeuber et.al., 1961; Taeuber, 1966; Eldridge, 1965).

The study of the frequency of movement is specifically important in the determination of whether the high mobility rates suggested by the official statistics in some communities are the outcome of repeated movement of the same elements of population from place to place or single moves by a larger number of elements from one place to another (Goldstein, 1954 and 1964; Morrison, 1971).

Frequency of movement, therefore, is a key subject in the study of internal migration in general and repeated migration in particular. Recognizing the particular set of factors that affect and is affected by the experience of a specific record of movement will enrich our knowledge about repeated migration and other related phenomena. This is the main focus of the present chapter. It is particularly designed to fulfill the following two objectives. The first is to provide a measure for the distribution of the Scottish population, as represented by the S.C.R. data, by the frequency of movement maintained during 1939-1964. Related to this objective is to show how such a distribution is influenced by specific characteristics of individuals such as sex, age, etc., or by specific events or changes in their life histories. The second objective is to investigate the effect of past frequency of movement on the propensity to move in the future with the aim of discovering any patterned relationship between residential mobility experience in the past and mobility behaviour in the future. Again as with the first objective, it is hoped to delineate the effect on such a relationship of the variables that relate to the characteristics of individuals or to changes in their life histories.

4.2: Differentials in the Average Frequency of Movement:

A simple and straightforward measure of the distribution of population by the frequency of movement is the mean number of moves exercised by individuals in a specific period of time. Comparison of the calculated means for different segments of population indicates whether repeated migration is a common characteristic of the whole population or a characteristic of specific segments of it who show a stronger tendancy towards mobility than others. For the S.C.R. data, the mean number of moves and standard deviation were computed for the total sample during 1939-1964. For useful comparisons, such statistics were also computed for the mobile sector, those who made at least one move during the same period. Within each classification, the average frequency of movement was differentiated by sex, marital status, change in the marital status, age at registration, and age at 1964 or death. An analysis of variance was carried out to investigate whether the observed differences in the frequency of movement between different groups for the mentioned variables are statistically significant or whether they are merely chance variations. The results are shown in tables (4:1) through (4:5) from which we conclude the following. First, the possible existence of repeated migration incident is revealed at once by the fact that, within the total population an individual made on the average 1.038 moves during the period studied while a migrant made more than double that total average (2.315 moves) during the same period. Such a proportion between the average for the total population and that for the migrant group prevailed for all subclasses considered, i.e., for males and females, for all categories of marital status and all types of change in this status, and for all age groups. This finding confirms the proposition that repeated movement accounts for the observed high mobility rates based on the number of moves and not on the number of movers.

Second, whether the total population or the migrant segment of it is considered, females seem more migratory than males since in both classifications, females had a higher average of movement than had the males. However, the differences between the two sexes are not statistically significant for the total sample while they are significant for the mobile group (table 4:5). This suggests that the two sexes in general are equally disposed to mobility with only accidental variations but if movers were differentiated from non-movers, females show a greater disposition to mobility than males. In most migration studies, sex differentials has been shown to be insignificant and the balance in favour of one sex to the other was not sustained.

Third, a clear pattern of mobility behaviour emerges with regards to marital status. Considering the population as a whole, single and divorced categories moved more frequently than married or widowed categories. The average number of moves for the former group was above the total average (table 4:1). For the movers, single and widowed persons scored the highest frequencies of movement among all groups. For both classifications, the differences between all groups are significant. The significance is more pronounced for the migrant-nonmigrant classification (table 4:5) than for the total population. These results are in accordance with the existing knowledge on migration. The fact that single persons with less family ties and household commitments are more mobile than married persons is well documented.

Similar patterns are observed if the marital status variable was furtherly disaggregated by sex. When marital status is controlled for, females tend to score higher averages of movement than males. Exceptional cases are total singles and divorced, and divorced movers. The differences, however, are not so large.

Fourth, the frequency of movement is highly correlated with changes in the marital status. Again, this result supports the general view that migration propensity is largely influenced by changes in the life cycle which starts from family formation through marriage and ends with family dissolution through divorce. The highest frequency of movement was recorded by people whose marital statuses were altered during the period of study whether such alterations were caused by marriage or were merely changes in the names of the individuals. The latter, however, are likely to be accompanied by movement of some kind.

These patterns are observed for the movers as well as for the total sample. The significance of the differences in the average frequency of movement between the different groups are higher in the change in marital status variable than in the marital status variable itself.

Once again, the balance in favour of females over males for all categories of this variable (change in marital status) is observed when further disaggregation by sex is maintained.

Finally, age is well demonstrated as an important variable in determining the individual's frequency of movement. It has been often argued that migrants tend to be selective of the younger age groups implying that young people are more inclined to move than the old. People in the former group consist mainly of adults completing their education and entering the labour force for their first time. Such events as well as other events predispose adults to migration more appreciably than old people. The evidence here substantiates this well-accepted fact. For the total sample, the frequency of movement is higher for people at young ages (5-34) than for people at older ages (35 and over). The average for the former group is well above the overall average while that of the latter group is below it. In general the average number of moves decreases as age increases. The relatively high average for the 5-9 and 10-14 age groups compared with the average for the other young age groups is hard to explain.

The above age patterns are also observed for the mover group (table 4:3). Presumably one may notice some exceptions to these general patterns.

Another way of assessing the extent of age selectivity is to consider the age of respondents at the end of the period or at death whichever is relevant. This time the frequency of movement curve with age is reversed. That is, people who are quite young at the end of the period had lower average number of moves than people in older ages (table 4:4). Those of ages between twenty and sixty had a long record of migration summarized in average frequency above the total average while those of ages under twenty or ages over sixty had averages below the total average. The peak in the average frequency curve is virtually at the age group 30-34. This pattern is a characteristic of the total sample as well as a characteristic of the movers.

There are differences, though slight, in the magnitude of the average frequency of movement between males and females in the different age groups.

Table (4:1)

Mean Number of Moves and Standard Deviation

For the Movers and For the Total Sample

by Sex and Marital Status

Sex	Males		Females		Total	
Marital Status	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
The Movers:						
Single	2.193	1.902	2.545	2.206	2.348	2.048
Married	2.147	2.038	2.399	2.093	2.260	2.066
Widowed	2.118	1.915	2.532	2.852	2.429	2.652
Divorced	2.00*	1.414	1.667	1.033	1.750	1.035
Total	2.175	1.925	2.488	2.192	2.315	2.055
The Total Sample:						
Single	1.153	1.761	1.148	1.949	1.150	1.853
Married	0.793	1.615	0.873	1.711	0.829	1.659
Widowed	0.443	1.224	0.677	1.850	0.607	1.691
Divorced	1.333**	1.528	1.111	1.167	1.167	1.193
Total	1.019	1.707	1.038	1.874	1.028	1.788

* Based on two observations.

** Based on three observations.

Table (4:2)

Mean Number of Moves and Standard Deviation

For the Movers and for the Total Sample

by Sex and Change in Marital Status

Change in the Marital Status	Males		Females		Total	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
The Movers:						
No change of status	2.174	1.929	2.230	2.027	2.195	1.966
Single to married	1.900	1.287	3.292	2.506	3.275	2.499
Change of name	2.875	1.408	3.052	2.315	3.039	2.261
Change of name (adoption)	1.444	0.882	1.462	1.127	1.455	1.011
Total	2.175	1.925	2.488	2.192	2.315	2.055
The Total Sample:						
No change of status	1.016	1.707	0.846	1.652	0.943	1.686
Single to married	1.357	1.393	1.858	2.492	1.853	2.484
Change of name	1.769	1.796	2.035	2.375	2.015	2.336
Change of name (adoption)	1.182	0.982	1.056	1.162	1.103	1.081
Total	1.019	1.707	1.038	1.874	1.028	1.788

Table (4:3)

Mean Number of Moves for the Movers
and for the Total Sample by
Sex and Age at Registration

Age Groups	The Movers			The Total Sample		
	Males	Females	Total	Males	Females	Total
0-4	2.220	2.158	2.191	0.756	0.712	0.735
5-9	2.785	2.833	2.806	2.320	1.741	2.026
10-14	2.007	2.813	2.351	1.699	1.757	1.728
15-19	1.736	2.672	2.181	1.490	1.855	1.683
20-24	2.138	2.783	2.384	1.387	1.678	1.502
25-29	2.165	2.878	2.440	1.309	1.503	1.391
30-34	2.266	2.708	2.444	1.202	1.245	1.220
35-39	2.240	2.425	2.315	0.877	0.934	0.902
40-44	2.286	2.053	2.181	0.817	0.695	0.759
45-49	2.167	2.133	2.148	0.554	0.640	0.600
50-59	2.309	2.161	2.228	0.632	0.696	0.665
60-69	1.943	2.192	2.081	0.405	0.569	0.486
70-79	1.452	2.108	1.871	0.255	0.455	0.373
80+	1.286	1.286	1.286	0.225	0.111	0.149
Total	2.175	2.488	2.315	1.019	1.038	1.028

Table (4:4)

Mean Number of Moves for the Movers and for the Total Sample by Sex and Age at 1964 or Death

Age Group	The Movers			The Total Sample		
	Males	Females	Total	Males	Females	Total
0-4	1.270	1.164	1.222	0.139	0.121	0.131
5-9	1.563	1.439	1.502	0.321	0.323	0.322
10-14	1.742	1.734	1.738	0.461	0.499	0.481
15-19	2.009	1.940	1.977	0.740	0.679	0.711
20-24	2.342	2.581	2.468	1.059	1.290	1.176
25-29	2.758	2.621	2.702	1.998	1.528	1.780
30-34	2.777	2.832	2.801	2.145	1.705	1.930
35-39	2.065	2.758	2.329	1.568	1.758	1.648
40-44	2.038	2.790	2.362	1.490	1.893	1.672
45-49	2.067	2.773	2.352	1.312	1.572	1.424
50-59	2.129	2.702	2.371	1.070	1.268	1.157
60-69	2.154	2.249	2.200	0.702	0.715	0.708
70-79	2.112	2.133	2.124	0.489	0.585	0.540
80+	2.013	2.220	2.132	0.654	0.672	0.665
Total	2.175	2.488	2.315	1.019	1.038	1.028

Table (4:5)

One-Way Analysis of Variance of Total Number of Moves for the Movers and for the Total Sample by Sex, Age*, Marital Status and Change in Marital Status Variables

Variable	Sum of Squares			Degrees of Freedom			Mean Sum of Squares		F-Statistic
	Between Groups	Within Groups	Total	Between Groups	Within Groups	Total	Between Groups	Within Groups	
The Movers:									
Sex	218.2	37050.1	37268.3	2	8826	8828	109.107	4.198	25.991***
Marital Status	82.4	37185.9	37268.3	4	8824	8828	20.588	4.214	4.886***
Change in M. Status	1001.7	36266.6	37268.3	4	8824	8828	250.430	4.110	60.932***
Age at Registration	317.8	36950.5	37268.3	14	8814	8828	22.697	4.192	5.414***
Age at 1964 or Death	790.9	36477.4	37268.3	14	8814	8828	56.495	4.139	13.651***
The Total Sample:									
Sex	3.9	63562.5	63566.4	2	19878	19880	1.964	3.198	0.614
Marital Status	590.2	62976.2	63566.4	4	19876	19880	147.561	3.168	46.572***
Change in M. Status	1448.7	62117.7	63566.4	4	19876	19880	362.172	3.125	115.885***
Age at Registration	4222.9	59343.5	63566.4	14	19866	19880	301.639	2.987	100.978***
Age at 1964 or Death	5808.5	57757.9	63566.4	14	19866	19880	414.895	2.907	142.704***

* Grouped Age Variables are Used (14 Groups each)
*** Significant at p = 0.01 signigicance level.

Tabulated Values of F-Statistic

p	$F(2,\alpha)$	$F(4,\alpha)$	$F(15,\alpha)$
0.10	2.30	1.94	1.49
0.05	3.00	2.37	1.67
0.01	4.61	3.32	2.04

Age differences, in general, are highly significant (table 4:5). This significance, however, is more marked in age at the end of the period or at death than in age at registration and for the total population than for the movers. In explaining the second finding, we may appreciate that there is a high degree of homogeneity in the data if the movers were taken as a separate group since they are commonly concentrated in the young age groups. A partial explanation of the first, on the other hand, is that age at the end of the period measures the length of time the individuals were at risk which varies from only a few years to twenty five years (being at risk is equivalent to being under observation).

4.3: Predicting the Frequency of Movement: A Linear Regression Model:

The findings of section (4.2) were particularly useful in determining the variables which affect the frequency of movement experienced by the individuals under observation. They indicated how each of these variables or combinations of them operated on the average frequency during 1939-1964.

The analysis can be further extended so as to allow the prediction of the number of moves a person is likely to experience in a specific period of time in the future given a particular set of individual's characteristics. This may be conventially effected by multiple regression analysis. Multiple regression has two main objectives. The first is to produce a linear combination of some independent variables which will correlate as highly as possible with the dependent variable. Examination of the regression coefficients enables us to understand the relationship between the variables. The second objective is to predict values of the dependent variable according to some known values of the independent variables. In our example, the total number of moves will be regressed as a dependent variable on some selected independent variables. The independent variables are those identified in the previous section, namely, sex, age, marital status, and change in the marital status. In addition, some other variables extracted from the sample can be added as explanatory variables. These are, for example, the size of the area in which the individual was first registered, the length of time during which the individual was under observation, and the duration of residence since registration. The size of the area of first registration, measured here by the total number of population of the area, is chosen as an indicative factor to the relative attraction of areas. It is argued that the larger the size of the area the more likely that people will be attracted to the area as inmigrants and the less likely people will be leaving it as outmigrants. The duration of residence variable has always proved to be an important factor in determining the propensity to migrate. The longer the stay in the present place of residence the less likely the individual will move in the future. Duration of residence prior to the first move is not the ideal duration variable to be used for explaining the bulk of future migration experience. Intermediate durations surely have some effect not of less importance than first duration. Moreover, there are parts of the first duration (prior to registration) which are unknown and hence are not included in the analysis. The latter effect is less serious, however, than the former. Nevertheless, duration of residence since registration is still a useful indication to residence history. For the distribution of individuals according to duration of residence prior to the first move will resemble to a large extent the distribution prior to any other move.

In the regression analysis which follows, we shall utilize the dummy variables technique. The use of dummy variables is a valuable device when some or all of the variables are of qualitative nature such as sex, marital status, area, etc. or variables that are measured in a non-metric manner. It also has the advantage of studying the numeric variables which have a non-linear effect on the dependent variable such as age. The procedure involves a splitting of the scale(s) of a conventionally measured variable(s) into a set(s) of intervals and defining dummy variables for each of these intervals. For a particular variable, if 'k' dummy variables are defined, only 'k-1' of them are introduced in the regression as separate independent variables thus constraining the 'k' th. variable to have a coefficient equal to zero (Suits, 1957). The split of variables into different categories is by no means optimal and must be founded on theoretical basis. The interpretation of regression coefficients of the dummy variables, however, differs from that of the ordinary coefficients. Dummy variables coefficients are deviations from the omitted variable coefficient.

The following variable classifications are utilized. The total number of moves, the dependent variable, preserves its original classification, i.e. 0,1,2,.....n moves. The length of risk period, the exposure time, variable is classified as one for those who was under observation for ten years or more and zero otherwise. Sex variable is coded as one for males and zero for females. Marital status is broken into two subclasses:singles (single, widowed, and divorced) against married and are given the codes 1 and 0 respectively. The change in the marital status consists also of two groups. The first comprises those who changed their status from single to married and are given the code 1. The second consists of cases of other types of change or no change at all. They are given the code 0. Three duration of residence categories are defined as three separate independent variables. These are durations of one or less year, 2-3 years, and 4-9 years. Duration of ten or more years is taken as a reference group. The criterion for splitting the area variable is whether the population of the area is equal to 250 thousand individuals or more. Finally eleven independent variables define eleven age groups. The reference age group is that of people sixty years old and over.

The regression analysis using the least squares method was carried out. The results are shown in table (4:6) from which the following remarks are concluded.

(1) The explanatory power of the regression equation represented by the value of the F-statistic is fairly high. However, only thirty nine per cent of the variation in the dependent variable is explained by the regression equation. This may be considered as a high value relative to similar studies which deal with individual rather than aggregate-level behaviour (Morrison, 1971). Nevertheless, it can be shown that the frequency of movement varies systematically with most of the independent variables considered although we lack many of the important variables like educational attainment, occupation, etc.

(2) Examination of the regression coefficients shows that repeated migration is highly correlated with duration of residence since registration, the length of observation period, changes in the marital status, age, and area of registration variables. The importance of each of the above variables is preserved by the above order. The regression coefficients corresponding to these variables are statistically significant at all conventional levels of probability. Regression coefficients corresponding to sex, marital status, and ages 25 and over variables, on the other hand, are insignificant.

Table (4:6)

Regression Analysis of the Frequency of Movement

During 1939-1964 for the S.C.R. Sample

Variable	Regression Coefficient	Standard Error	F Statistic
ea size; population = 250,000 or more	-0.023	0.009	6.711
x; males = 1	-0.034	0.021	2.644
rital status; single = 1	-0.016	0.031	0.256
ange in the marital status;			
single to married = 1	0.651	0.043	224.506
posure time; 10+ years = 1	0.422	0.028	222.318
ration of residence since registration;			
0 - 1 years	2.481	0.032	6150.982
2 - 3 years	2.090	0.033	3902.961
4 - 9 years	1.826	0.034	2917.028
e at registration;			
0 - 4 years	0.163	0.042	15.288
5 - 9 years	0.974	0.060	263.265
10 - 14 years	0.106	0.061	3.019
15 - 19 years	-0.205	0.057	12.804
20 - 24 years	-0.094	0.053	3.209
25 - 29 years	0.003	0.052	0.004
30 - 34 years	0.045	0.054	0.694
35 - 39 years	-0.044	0.055	0.620
40 - 44 years	-0.053	0.059	0.828
45 - 49 years	-0.067	0.063	1.134
50 - 59 years	0.019	0.053	0.125
mmary statistics:			
F-statistic 644.284			
R^2 0.387			
Standard error 1.401			
Constant term -0.163			

(3) The signs and magnitudes of most of the coefficients are as expected. It may be noticed, however, that although the coefficients of the age groups 15-19 and 20-24 are significant the signs of the corresponding coefficients are not as expected. There is no direct explanation of such divergence other than that no effort has been made to remove the effect of interaction terms especially those between age and other variables.

(4) Duration of residence variables, though not ideally defined, played a prominent role in explaining the variation in the observed frequency of movement. They accounted for almost eighty per cent of the total 'R^2'. Such a finding should presumably have a significant bearing on the construction of a suitable model for repeated migration. This will receive further attention in the next chapter.

4.4: Frequency of Past Movement as a Factor in Determining the Probability of Moving in the Future.

4.4.1: Problem Definition and Hypotheses Tested:

One way of studying repeated migration is to see to what extent the propensity of an individual to migrate is affected by the number of moves he made during a given period in the past. Specifically, we are concerned to find out what kind of relationship, if any, might exist between the frequency of past movement and the probability of moving again in the future, and with the investigation of what variables might have an influence on any such relationship. The idea that residential mobility in a given time period is highly related to residential mobility experience in the past has received widespread support, but the precise mode in which such past mobility experience should be expressed has been subject to disagreement among researchers. Some works have demonstrated that frequency of past movement is a powerful expression of such an experience (Goldstein, 1954 and 1964; Rowntree, 1957; Morrison, 1971).

This part of the research investigates the importance of mobility experience in the past in determining the probability of moving in the future. Such mobility experience is here expressed in terms of the number of moves a person made in a given time in his past history. Our first task is to test the null hypothesis that regards knowledge of past movement as an irrelevant factor in the determination of the probability of future movement. This would be the condition of a Markov process, in which the transition probabilities governing change from time n to time n+1 are determined entirely by the state occupied at n, and the states occupied before n are not involved in the subsequent transition probabilities. The alternative hypothesis assumes that the probability of moving in the future is related to the number of moves already made by a person in the past. This probability is expected to increase as the number of previous moves increases. This is equivalent to the proposition that people who were more mobile in the past are more likely to move again in the future than those who were stable in their place of residence. Moreover, if the available set of data on migration in Scotland were to support the alternative hypothesis, which acknowledges the existence of a positive relationship between migration risk and the frequency of past movement, would other variables influence this relationship? And if the answer to this question was yes, could we discover these variables and the nature of their influence?

4.4.2: Procedure of Analysis:

In order to test the above hypotheses, the whole period 1939-1964 was divided into overlapping combinations of equal time intervals. Presumably, the choice of the length of a time interval may influence the type of the expected relationship. The data set related to the period September 1939 to August 1964 and for convenience 1939 and 1964 were excluded. The remaining 24 whole years yielded 20 five-year time-intervals, allowing overlapping. These time intervals were then defined as (1940-1944), (1941-1945, (1942,1946),... and (1959-1963).

Each interval was further divided into sub-intervals of varying length (one year, two years, three years, ... etc.) according to the number of years during which the person was at risk. For example, the people who were registered in the year 1940 and lived throughout the period 1940-1944 will be considered at risk for five years and will be counted in a subcategory of five year length within the whole period 1940-1944. Those who were registered in 1941 will be at risk for four years and will be counted in a subcategory of four year length in the same period 1940-1944. Similarly, people who were registered in or before 1940 and were reported dead in 1943 are assumed to have been at risk for three years (1940, 1941, 1942) and will be counted in the same subcategory of three year length as those who registered in 1942 and survived the whole period. The supplementary partitioning of the time intervals helps in testing the same hypotheses within shorter periods of time. The people at risk in a specific time interval were classified into eleven categories (0,1,2,...10 + Moves) according to the number of moves they made during the interval. Each of the above categories then was held to comprise two distinct groups of the population; those who moved in the year next to the end of the chosen period and those who did not move in that year. The probability of moving was calculated as the number of people who moved divided by the number of people at risk during the interval (the sum of the two groups). In addition, average probabilities of moving for all periods in the system were calculated as follows:

Let N_{ijk} define the number of people who were at risk for k years (k = 1,2, ...5) in the j^{th} period (j = 1, ..., 20) and i moves during that period. Let also R_{ijk} define the number of people out of that N_{ijk} who moved in the year following the end of the interval j. Then the average probability of moving for the category of people who made i moves in the previous k years is P_{ik}, defined as

$$P_{ik} = \frac{\Sigma_j \; R_{ijk}}{\Sigma_j \; N_{ijk}}$$

This way of averaging the probability of moving was preferred to averaging through the arithmetic mean of probabilities of all periods. The latter device seems to be much affected by the irregularities that occurred due to the insufficiency of data, especially at the far ends of tabulation (the case of high numbers of previous moves). However, the above procedure still suffers the same deficiency, but notably to a lesser extent.

The procedure was carried out first for the sample as a whole and then repeated for each sex and age group separately.

4.4.3 The Findings:

1. General

We now state the main results obtained from the application of the above procedure. Table (4:7) contains average probabilities of moving in the year following previous periods of five years. The most important conclusion that can be drawn from this table is that the probability of moving in the future does indeed appear to be a function of the number of moves already made in the past. For all periods of risk, the probability

Table (4.1)

Probability of Moving in the Year Following a 5-year Time Interval According to the Number of Years At Risk and the Number of Moves Made During the Interval

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
One Year At Risk											
Number of Cases	8588	580	122	15	6	1	-	-	-	-	-
Moved Next Year	599	108	28	2	3	0	-	-	-	-	-
Did Not Move	7989	472	94	13	3	1	-	-	-	-	-
Probability of Moving	.0697	.1862	.2295	.1333	.5000	.0000	-	-	-	-	-
Two Years At Risk											
Number of Cases	7819	966	288	74	22	10	3	1	-	2	-
Moved Next Year	313	144	57	26	4	6	0	0	-	1	-
Did Not Move	7506	822	231	48	18	4	3	1	-	1	-
Probability of Moving	.0400	.1491	.1979	.3513	.3636	.6000	.0000	.0000	-	.5000	-
Three Years At Risk											
Number of Cases	7499	1170	410	132	53	25	15	5	-	2	1
Moved Next Year	240	137	57	25	16	11	5	1	-	1	0
Did Not Move	7259	1033	353	107	37	14	10	4	-	1	1
Probability of Moving	.0320	.1171	.1390	.1894	.3019	.4400	.3333	.2000	-	.5000	.0000
Four Years At Risk											
Number of Cases	7219	1304	506	173	87	34	21	20	3	2	3
Moved Next Year	157	120	57	43	21	9	3	8	2	0	1
Did Not Move	7062	1184	449	130	66	25	18	12	1	2	2
Probability of Moving	.0217	.0920	.1126	.2485	.2414	.2647	.1428	.4000	.6667	.0000	.3333
Five Years At Risk											
Number of Cases	215519	27171	9439	3189	1408	585	329	153	104	52	52
Moved Next Year	4473	2190	1017	496	240	145	82	38	30	18	11
Did Not Move	211046	24981	8422	2693	1168	440	247	115	74	34	41
Probability of Moving	.0207	.0806	.1077	.1555	.1704	.2479	.2492	.2484	.2885	.3462	.2115

of moving generally tends to increase as the number of moves made in the past increases.

The relationship between future and past movement seems to be fairly regular for the longer periods of risk, but it fluctuates somewhat for the shorter periods, especially at high numbers of previous moves, but, as only a few people had made large numbers of moves in short periods, this is to be expected.

2. Sex Differences

We now consider sex as a factor in determining the relationship between frequency of past movement and the propensity to move again in the future. Table (4:8) shows the average probability of moving for males and females. The same patterns of movement that were observed for the sample as a whole also apply for each sex. For both males and females, the probability of moving in the near future increases with the number of previous moves. This increasing trend of probability holds only up to a certain point, however, after which the probability begins to vary rather wildly, and does not increase generally.

Although they are not very big in magnitude, we may notice some differences between the figures for the two sexes. For all periods of risk, for example, the probability of moving in the future, given that there had been no previous moves in the past, is higher for males than for females. Where there was no recent mobility experience, it appears that males are more inclined to migrate than females. This greater inclination of males to migrate extends to most of the categories of previous experience if we only consider the shorter periods of risk. When full periods of risk are considered, however, the picture is rather different. These have much more data, and are also likely to be more typical. A foreigner just arriving, for example, would not have a whole period of risk - nor would a baby or a man returned from abroad.

Table (4.6)

Probability of Moving, By Sexes, In the Year Following a 5-Year Time Interval According to the Number of Years At Risk and the Number of Moves Made During The Interval

(1) Males

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
One Year At Risk											
Number of Cases	4961	328	70	6	5	1	-	-	-	-	-
Moved Next Year	366	65	16	1	3	0	-	-	-	-	-
Did Not Move	4595	263	54	5	2	1	-	-	-	-	-
Probability of Moving	.0738	.1982	.2286	.1667	.6000	.0000	-	-	-	-	-
Two Years At Risk											
Number of Cases	4470	552	180	51	9	6	2	1	-	2	-
Moved Next Year	180	75	30	20	3	5	0	0	-	1	-
Did Not Move	4290	477	150	31	6	1	2	1	-	1	-
Probability of Moving	.0403	.1359	.1667	.3921	.3333	.8333	.0000	.0000	-	.5000	-
Three Years At Risk											
Number of Cases	4263	687	243	76	31	13	12	3	-	2	1
Moved Next Year	141	88	33	12	9	7	5	1	-	1	0
Did Not Move	4122	599	210	64	22	6	7	2	-	1	1
Probability of Moving	.0331	.1281	.1358	.1579	.2903	.5385	.4167	.3333	-	.5000	.0000
Four Years At Risk											
Number of Cases	4099	757	299	105	51	22	10	13	3	2	3
Moved Next Year	95	72	37	30	15	8	1	5	2	0	1
Did Not Move	4004	685	262	75	36	14	9	8	1	2	2
Probability of Moving	.0232	.0951	.1237	.2857	.2941	.3636	.1000	.3846	.6667	.0000	.3333
Five Years At Risk											
Number of Cases	109195	14949	4960	1489	609	257	165	81	55	30	36
Moved Next Year	2293	1191	465	194	117	59	31	23	15	10	8
Did Not Move	106902	13758	1295	1295	492	198	134	58	40	20	28
Probability of Moving	.0210	.0797	.0991	.1303	.1921	.2296	.1979	.2840	.2727	.3333	.2222

Table (4:8) Contd.

(2) Females

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
					One Year At Risk						
Number of Cases	3616	251	52	9	1	-	-	-	-	-	-
Moved Next Year	232	43	12	1	0	-	-	-	-	-	-
Did Not Move	3384	208	40	8	1	-	-	-	-	-	-
Probability of Moving	.0642	.1713	.2308	.1111	.0000	-	-	-	-	-	-
					Two Years At Risk						
Number of Cases	3344	412	108	23	13	4	1	-	-	-	-
Moved Next Year	133	59	27	6	5	1	0	-	-	-	-
Did Not Move	3211	343	81	17	8	3	1	-	-	-	-
Probability of Moving	.0398	.1675	.2500	.2609	.3846	.2500	.0000	-	-	-	-
					Three Years At Risk						
Number of Cases	3231	481	167	56	22	12	3	2	-	-	-
Moved Next Year	98	49	24	13	7	14	0	0	-	-	-
Did Not Move	3133	432	143	43	15	8	3	2	-	-	-
Probability of Moving	.0303	.1019	.1437	.2321	.3182	.3333	.0000	.0000	-	-	-
					Four Years At Risk						
Number of Cases	3116	545	207	68	36	12	11	7	-	-	-
Moved Next Year	62	48	20	13	6	1	2	3	-	-	-
Did Not Move	3054	497	187	55	30	11	9	4	-	-	-
Probability of Moving	.0199	.0881	.0966	.1912	.1667	.0833	.1818	.4286	-	-	-
					Five Years At Risk						
Number of Cases	106306	12214	12214	1700	799	328	164	72	49	22	16
Moved Next Year	2179	998	552	302	123	86	51	15	15	8	3
Did Not Move	104127	11216	4197	1398	676	242	113	57	34	14	13
Probability of Moving	.0205	.0817	.1162	.1776	.1539	.2622	.3110	.2083	.3061	.3636	.1875

The probability of moving for females is above that for males for anyone who had moved one, two or three times in the previous five years. This implies that females who had moved recently, but not very often, have a greater disposition toward further migration than similar males. Once they have made between one and three moves they seem more liable to move in the near future. With four or more previous moves (in five years of risk), there is no consistent difference between the probability of moving for the sexes, however.

3. Age Differences

We next consider the effect of age on the probability of moving, given the past record of migration. The shorter periods of risk are not now considered. Tables (4:9) and (4:10) contain the average probabilities by broad age groups and sex for periods of five years' risk.

We first consider both sexes grouped together. For those with no previous moves, or only one, the middle age-group (15-44) had the highest probabilities of moving, with children under fifteen next. The lowest probabilities were for old people (65 years and over). This largely reflects the well-known fact that adults have a higher inclination to move than youngsters and elderly people, and even if the aged people had moved once in the recent past they were less likely to move again in the immediate future.

About 94 per cent of the sample had only moved once or not at all. As the number of previous moves increases, people in the age-group 45 to 64 years begin to take the lead after three moves, and the probabilities for 0-14 and 65+ age-groups approach (and in some cases exceed) those of the 15-44 age-group. This again partly reflects the fact that adults (15-44 years old) have the highest general propensity to migrate, since some of the people in the 45-64 age-group were the same as those previously aged 15-44 who were identified earlier.

Table (4:9)

Probability of Moving by Age Groups in the Year Following a Whole 5-Year Time Interval

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
Number of Cases 0-14	31654	3687	1514	463	154	66	31	21	2	2	-
Moved Next Year	628	327	130	61	25	18	9	5	1	0	-
Did Not Move	31026	3360	1384	402	129	48	22	16	1	2	-
Probability of Moving	.0198	.0887	.0859	.1317	.1623	.2727	.2903	.2381	.5000	.0000	-
Number of Cases 15-44	99356	17782	6248	2190	696	416	220	104	82	40	37
Moved Next Year	2914	1528	719	352	168	97	51	26	23	15	7
Did Not Move	96442	16254	5529	1838	828	319	169	78	59	25	30
Probability of Moving	.0293	.0859	.1151	.1607	.1687	.2332	.2318	.2500	.2805	.3750	.1892
Number of Cases 45-64	57919	3954	1237	389	191	78	46	20	10	7	9
Moved Next Year	636	249	125	65	37	23	13	6	2	1	2
Did Not Move	57283	3705	1112	324	154	55	33	14	8	6	7
Probability of Moving	.0110	.0630	.1011	.1671	.1937	.2949	.2826	.3000	.2000	.1429	.2222
Number of Cases 65+	26563	1740	438	147	67	25	32	8	10	3	6
Moved Next Year	295	86	43	18	10	7	9	1	4	2	2
Did Not Move	26266	1654	395	129	57	18	23	7	6	1	4
Probability of Moving	.0111	.0492	.0982	.1224	.1493	.2800	.2812	.1250	.4000	.6667	.3333

For high frequencies of previous movement (more than five moves in five years), the people of sixty-five or more become the most likely age group to move again in the future. Their long record of movement in the past presumably accounts for this great propensity to move again. Perhaps the population has separated into "movers" and "stayers" by this age, and there really are certain classes of people who make frequent moves. These chronic movers have a large amount of mobility experience in the past, and a high propensity to migrate in the future. This would conform with the proposition of Goldstein (1954).

4. Interaction Effects

The preceding analysis of the effect of sex and age on the relationship between the probability of moving in the future and the frequency of past movement may be inadequate, because there may be interaction between the variables. To study the effect of interaction, the probability of moving was tabulated by age and sex simultaneously. Table (4:10) shows this tabulation for five years' previous risk.

If we control for age we see that, contrary to the original tentative finding, the females are more likely to move than the males if there was no previous experience of movement, but the differences between the two sexes are slight. This is true for all age groups except 15-44, which signifies that adult males have a higher propensity to migrate than adult females.

For up to three previous moves, females of all ages were more prone to migrate than males (with only a few exceptions). This result is consistent with earlier findings.

When the past history involved larger numbers of moves than three, it became obvious that there is no clear distinction between the patterns of movement of each sex. The position of having the higher probability of further migration mostly alternates between the sexes, with perhaps a tendency for young males and old females to take the lead. One may repeat again that the irregularities at later moves are due to the fact that only a few cases apply and thus there is not sufficient information to give a more precise judgement.

Now we control for sex. The results already obtained for the effect of age found when the sexes were combined are given support. Adults are the most likely group to migrate of those who had made no previous movement or who had only recorded a few moves. At larger numbers of previous moves, the people in the 45-64 age group were more likely to move than people either in young or old ages, while at the highest numbers of moves, the elderly people come out ahead of all the other age groups. In general terms, it is a well-established fact that mobility declines with age in adult life, and is relatively low after the age of 45. But the evidence above indicates that there may be a group of individuals who are mobility prone in the sense that they exhibit above average mobility throughout their life cycle. They certainly do so in their later years.

Adult males consistently record higher levels of probability of moving than adult females, but aged males have lower probabilities of subsequently moving than aged females, particularly those with a large number of previous moves. It is not clear what are the causes or consequences of such a phenomenon.

Age differences are generally greater than sex differences, and thus have more influence upon mobility behaviour in the future in the light of mobility in the past.

Table (4:10)

Probability of Moving, by Sex and Age Groups, in the Year Following a Whole 5-year Time Interval

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
(1) Males											
Number of Cases 0-14	16332	1844	742	225	69	33	15	13	1	-	-
Moved Next Year	306	16	61	30	13	8	4	4	1	-	-
Did Not Move	16026	1683	681	195	56	25	11	9	0	-	-
Probability of Moving	.0187	.0873	.0822	.1333	.1884	.2424	.2667	.3077	1.0000	-	-
Number of Cases 15-44	51515	10537	3326	1025	420	173	108	47	41	23	27
Moved Next Year	1593	895	340	127	75	35	18	12	11	9	5
Did Not Move	49922	9642	2886	898	345	138	90	35	30	14	22
Probability of Moving	.0309	.0849	.1054	.1239	.1786	.2023	.1667	.2553	.2683	.3913	.1882
Number of Cases 45-64	29549	1817	554	199	94	41	29	14	10	5	6
Moved Next Year	264	101	49	32	26	13	8	6	2	0	2
Did Not Move	29285	1716	505	167	68	28	21	8	8	5	4
Probability of Moving	.0089	.0556	.0884	.1600	.2766	.3171	.2759	.4286	.2000	.0000	.3333
Number of Cases 65+	11774	746	167	40	26	10	13	7	3	2	3
Moved Next Year	130	34	15	5	3	3	1	1	1	1	1
Did Not Move	11644	712	152	35	23	7	12	6	2	1	2
Probability of Moving	.0110	.0456	.0898	.1250	.1154	.3000	.0769	.1429	.3333	.5000	.3333

Table (4:10) Contd.

Number of Moves	0	1	2	3	4	5	6	7	8	9	10+
(2) Females											
Number of Cases 0-14	15322	1843	772	238	85	33	16	8	1	2	-
Moved Next Year	322	166	69	31	12	10	5	1	0	0	-
Did Not Move	15000	1677	703	207	73	23	11	7	1	2	-
Probability of Moving	.0210	.0901	.0894	.1303	.1412	.3030	.3125	.1250	.0000	.0000	-
Number of Cases 15-44	47833	7238	3012	1165	576	243	112	57	41	17	10
Moved Next Year	1320	632	379	225	93	62	33	14	12	6	2
Did Not Move	46513	6606	2643	940	483	181	79	43	29	11	8
Probability of Moving	.0276	.0873	.1254	.1931	.1615	.2551	.2946	.2456	.2927	.3529	.2000
Number of Cases 45-64	28360	2136	683	190	97	37	17	6	-	22	3
Moved Next Year	372	148	76	33	11	10	5	0	-	1	0
Did Not Move	27988	1988	607	157	86	27	12	6	-	1	3
Probability of Moving	.0131	.0693	.1113	.1737	.1134	.2703	.2941	.0000	-	.5000	.0000
Number of Cases 65+	14789	994	271	107	41	15	19	1	7	1	3
Moved Next Year	165	52	28	13	7	4	8	0	3	1	1
Did Not Move	14624	942	243	94	34	11	11	1	4	0	2
Probability of Moving	.0112	.0523	.1033	.1215	.1707	.2667	.4211	.0000	.4286	1.0000	.3333

Chapter (5)

A SIMULATION PROBABILITY MODEL FOR REPEATED MIGRATION: PRELIMINARY ASSESSMENT

5.1: Introduction:

In chapter four, the probability of future migration as a function of past migration experience, when such an experience is measured by the number of moves made by the individual in a given time in the past, has been considered. In general, the proposition which acknowledges previous mobility experience as an important factor in determining future mobility level has been strongly supported. The study also stimulated the need to formulate a simple, practical model that efficiently describes the phenomenon of repeated migration as one of the prominant features of the migration process.

The formalization of the relationship between residential mobility in the future and mobility experience in the past has often been made by means of models that regard the length of maintaining residence in a specific place as the key variable in determining the probability of future migration. The basic assumption of such models is that the propensity to migrate decreases as the time of being resident in the same place increases. Many of these models have utilized the simple Markov chain as a stochastic process to represent social, industrial, or geographical mobility (Prais, 1955a and 1955b; Blumen et. al., 1955; Tarver and Gurley, 1965). In most applications, however, the simple Markov chain did not prove to be a good representation of these phenomena. The reason for this failure lies in its basic assumption, which argues that the occurrence of an event at time t is entirely dependent on the state at time t-1 and not on any other state at earlier time.

The Cornell model, originated by McGinnis and his associates, employs a more elaborate Markov chain for the sake of reducing the effect of such an unrealistic assumption. This elaboration is achieved by the introduction of an additional axiom, which takes into account both the location of an element at time t-1 and the duration of residence at that location. This axiom of cumulative inertia, as termed by McGinnis, suggests that the probability that an individual remains at the same location increases as the length of prior residence at that location increases. The changes of elements in the system from one state to another are then not governed by the same probabilities. Rather, people who have been longer in some state have a greater probability of remaining there than those of short stay. (McGinnis, 1968). Empirical studies by McGinnis and his associates and by others have given support to the axiom of cumulative inertia (Myers, et. al., 1967; Morrison, 1967; Land, 1969).

However, the classification of a population into movers and stayers, each with varying prior durations of residence, and the need to assign different transition probabilities to each sub-class with a specific duration, have added much to the complexity of the model both in analytical and empirical terms. Furthermore, if the need arises (and it presumably will) to introduce other relevant variables, such as age, into the model, even more complications would have to be introduced and the usefulness of the model would be greatly lessened. A simpler model, with only a few parameters, may therefore have some practical value.

5:2 <u>The Model</u>:

It is highly desirable that we should gain some clearer understanding of the phenomenon of repeated migration than simply admitting that the propensity to migrate diminishes as the time increases since the previous move. While this is probably true in a general way, it seems improbable that the maximum propensity to migrate occurs immediately after a move. It is more likely that some delay occurs, during which the likelihood of a further migration is low. Moreover, whatever we believe is the underlying pattern of repeated migration, we shall need a quantitative analysis of our data in order to test the validity of our ideas. This in turn implies that we must formulate a precise model of repeated migration. The model developed below basically stems from the idea that the propensity to migrate is disproportionately related to the time of being resident in the same place since the last move. It also incorporates a further postulate about the individual's migration behaviour. That is, a probable delay time, immediately after a move, might elapse during which a person's propensity to make a further move is negligible. The model involves three parameters which respectively define the general level of probability (A), the delay time between a move and a subsequent move (B), and some factor (C) used to identify the mode of decline in the probability of moving as duration in the same place of residence increases. If (C) approaches zero, the model will reduce to a simple inverse relationship between the probability of moving and duration of residence allowing only for the delay time to operate.

The model is formalized as:

$$\lambda(t) = \begin{cases} \dfrac{A}{(t-t_0) + C} & \text{if } t > B + t_0 \\ 0 & \text{otherwise} \end{cases}$$

where λ is the instantaneous probability of moving,

A, B, C are constants ≥ 0, defined as above,

t is time, and

t_0 is time of last move.

A preliminary analysis of the properties of the model and the characteristics of the parameters involved was carried out using simple mathematical approximation[1]. The concentration however, was made on the average probability of moving. Empirically, the model has proved to yield promising results. In view of the few applications attempted, some general knowledge about the levels at which the model parameters should be set for further analysis has been established. More detailed study of the model and the generalized levels of parameters seems necessary.

Among the lessons gained from this previous analysis was that the derivation of model properties of interest and the estimation of the appropriate levels of parameters are too complicated to be handled only through ordinary analytical techniques. Alternatively, computer simulation is believed to be a more helpful technique in such a situation. This will be the main theme of this chapter and subsequent chapters. It is in order, therefore, to give a brief discussion on simulation as one of the most recent techniques in scientific research; what is meant by simulation? why we resort to it? and what are the methodological considerations in a simulation experiment, in general?

5.3. A Simulation Experiment:

5.3.1. What is Simulation and Why?

The term "Simulation" has been used in a number of different ways most of which acknowledge it as a technique of performing experiments upon a model (which can take the form of a physical, mathematical or computer model) that represents a real world situation. Given a system

---

1. For a detailed discussion of these properties, see T.H. Hollingsworth, and M.G. El. Rouby, "Models for Repeated Migration", a paper to be published.

that is composed of a group of entities (variables) which act and interact together, a model can be constructed to represent these variables and the interactions among them (Schmidt and Taylor, 1970; p.4). The construction of such a model requires theoretical and empirical guidance from inside as well as from outside the system to be studied.

Simulation has proved to be a valuable technique in solving a large variety of problems and in studying many of the systems with varying degrees of complexity. The following is a brief account of the advantages and disadvantages of simulation as has been frequently stated in the literature.

Advantages:

(1) Simulation is the best solution of the problem if the system can be described mathematically but the correct analysis of the model representing it is beyond the level of the existing mathematical sophistication or if the system itself is too complex to be described adequately in a mathematical form.

(2) A simulation model once constructed can be utilized as often as it is required to analyse different situations.

(3) The analyst has full control over all variables and can replicate conditions as he desires by varying the levels of these variables at any stage.

(4) Simulation is useful for analysing systems in which information is inadequate or lacking. For data needed for further analysis can be generated from a simulation model with less effort and expense than it can be obtained from the real world.

(5) Simulation permits the study of stochastic or random processes that can not be studied by classical analytic tools either because of the lack of these tools or the lack of ability to use them.

(6) Time, as an important variable in describing dynamic systems, can be easily represented through simulation rather than using a mere symbolic representation in mathematics.

Disadvantages:

(1) Simulation models, especially those using computers, are very costly to construct and to validate. The running of a simulation programme can involve a great deal of computer time and storage which are very costly.

(2) Simulations provide individual case results representing particular situations and fail to give general solutions to the problem under consideration.

(3) A less serious deficiency, in fact it may not be considered as a deficiency of simulation itself, is that people as they become familiar with simulation tend to use it in every situation, even in cases where simulation is not the best method of analysis.

5.3.2: Methodological Considerations of Simulation:

Naylor, in his article on methodological considerations in simulating social and administrative systems has outlined a procedure of six steps for experimenting using computer models of social systems which may be valid for any other type of systems (Naylor, 1972). These considerations are briefed as follows.

First, computer simulation experiments should begin with the formulation of the problem to be studied. This includes defining precisely the objectives of the experiment, the hypotheses to be tested and the effects to be estimated.

Second, the investigator has to formalize theprecise mathematical model that best describes the behaviour of the system under investigation. This implies defining the basic variables which act upon the system and the relationships which relate these variables to each other.

Third, after formalizing the mathematical model, comes the computer programming phase of the experiment. This includes writing a computer programme in whatever convenient language is, setting up the appropriate initial conditions, and selecting the appropriate data generation rountines (including generation of random numbers),

Fourth, once the model is constructed and programmed on the computer, it is necessary to establish appropriate criteria for validating it and its basic assumptions. A common criterion for model verification is to show the degree of conformity between the data generated by the simulation model and the observed data (extracted from real world situation). This is statistically known as the goodness of fit criterion. There is a variety of measures for the goodness of fit, each has its own merits and is designed to meet specific situation.

Fifth, simulation experiments like biological or chemical experiments, need to be set in a framework of one of the well-known experimental designs. The most commonly used experimental designs are factorial and fractional factorial designs. Special designs were developed by Box and associates to explore response surfaces of the form $Y = f(X_1, \ldots, X_k)$ where Y is some dependent variable and the X's are the independent variables (Box and Wilson, 1951; Box et. al., 1953; and Box, 1954).[2] They suggested least squares regression analysis as a method for fitting such surfaces. Response surface designs are favoured over factorial designs in that the former reduce the size of the experiment without reducing the amount of information obtained. The problems encountered in simulation experiments, for example the size of experiment and the concern about reducing the random error effect are similar to those encountered in standard experimental designs. Presumably, the choice of a specific experimental design is determined by the objective of the experiment, whether the aim is to achieve

---

2. Response surface designs are used in chapters six and seven with reference to the present model. A detailed description of the methods is given in chapter six.

an optimum solution or just a general understanding of the problem. Finally, as the last step of a simulation experiment, the investigator has to analyse the output data obtained from the simulation with reference to some observed data. The analysis may support the hypotheses set by the experimenter, or may suggest further modifications of these hypotheses, or may possibly lead to a complete rejection of them and the entire abandonment of the model.

A wide variety of standard techniques for analysing simulation output is now available. The most common of these techniques are the analysis of variance, correlation and regression analysis, and the non-parametric methods for statistical inference.

For the purpose of simulating the above suggested model, a specific simulation experiment has been designed. It describes how the model is operationally adjusted to fit computer simulation requirements, how the initial values of the basic model parameters and additional simulation parameters are defined, and what are the logical steps considered for simulating individuals. Such a computer simulation model is not included, however, in this chapter. It is rather contained in the appendix along with a flow diagram of the computer programme utilized.[3]

5.4 Criteria employed for model verification:

The problem of verifying a computer model as a special type of the general problem of verifying any type of model is the toughest problem the investigator is confronted with. So many theoretical and practical considerations are involved. However, two major questions are to be asked the answers of which indicate whether or not the model under investigation is suitable to represent the problem in question. The first of these questions is, do the generated data from the simulation conform with theoretical data if available? The second

3. The proposed model is fitted to a subset of the original data, those who lived twenty years between 1945 and 1964, constituting 6469 cases. The appendix contains the actual distribution of movement for these cases and the corresponding distributions generated from selected successful simulations of the model. For more details see the appendix.

question is, what is the relative power of the model as far as prediction of future behaviour of the system is concerned? (Naylor, 1966; pp.39-40). For the time being, we shall confine ourselves to the answering of the first of these questions. Specifically, given an observed distribution of individuals according to the number of moves they made during a specific time period, how far a similar distribution generated from the simulation agrees with that observed distribution, and whether the agreement, if it exists, between these two sets of data is not attributed merely to chance.

The measures taken in this respect are known as measures for the "goodness of fit" of process models to real world situations. One of the statistically well known measures of goodness of fit is the Chi-Squared test. In some applications however, it may occur that (for some particular set of conditions) some undesirable distributions, which make the calculation of the statistic difficult and unreliable, are produced. For example, we may have a distribution with most of the frequency cells equal to zero and confounding these cells together will affect the efficiency of the test. In addition, the particular disadvantage of the Chi-Squared statistic that it is relatively sensitive to non-normality, which is almost certain in our case, added to the existing doubts about its usefulness. Alternatively, the Kolmogorov-Smirnov two sample test which has no computation problems similar to the Chi-Squared test, was used (Siegel, 1956; Gibbons, 1971).

The test is concerned with the comparison between two empirical distribution functions of two different samples. Following Gibbons' notation, the test is described as follows (Gibbons, 1971; pp.127-128). Define two random samples of sizes m and n drawn from continuous populations F_x and F_y. If the two samples were ordered as $X_{(1)}, X_{(2)}, \ldots, X_{(m)}$ and $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$, their respective empirical distribution functions will be defined as

$$S_m(x) = \begin{cases} 0 \text{ if } x < X_{(1)} \\ \frac{k}{m} \text{ if } \quad X_{(k)} \leqslant x < X_{(k+1)} \text{ for } k = 1, 2, \ldots, m-1 \\ 1 \text{ if } x \geqslant X_{(m)} \end{cases}$$

and

$$T_n(y) = \begin{cases} 0 \text{ if } x < Y_{(1)} \\ \frac{k}{n} \text{ if } \quad Y_{(k)} \leqslant x < Y_{(k+1)} \text{ for } k = 1, 2, \ldots, n-1 \\ 1 \text{ if } x \geqslant Y_{(n)} \end{cases}$$

We are testing the null hypothesis:

H_o : $F_Y(x) = F_x(x)$ for all x against the alternative hypothesis H_1 : $F_Y(x) \neq F_x(x)$ for some x. If the null hypothesis was true, the two samples are said to be drawn from the same population. The Kolmogorov-Smirnov two samples test, denoted by $D_{m,m}$ is the maximum absolute difference between the two distributions; $D_{m,n} = \max_x \left| S_m(x) - T_n(x) \right|$. The value of $D_{m,n}$ is compared with a tabulated value representing the sampling distribution to indicate whether the divergence between the two distributions is merely due to chance. If the calculated value was greater than or equal to the tabulated value associated with specific sample size and specific significance level, the fit is said to be not good at that level of probability.

The main advantage of this test is that no information is lost through the combining of categories as it is the case in the Chi-Squared test where zero frequencies occur. It is also appropriate than the Chi-Squared test, especially for small samples, a case which is needed in preliminary experimentation. Moreover, Kolmogorov-Smirnov statistic is sensitive to all types of differences between the cumulative distribution functions. The primary application of the test should then be for preliminary studies of the data (Gibbons, 1971, p.131).

Apart from the above-mentioned measures of overall goodness of fit, there are other simple standard statistics which might be useful to determine whether or not the proposed model explains, with reasonable efficiency, the actual patterns of movement. These include the mean, standard deviation, and coefficient of variation, for both the observed and simulation generated distributions. By comparing these statistics in the two distributions, one should be able to conclude how far they are in agreement or disagreement.

5.5: Simulation Output, Preliminary Findings:

5.5.1: "Crude" Search Procedure:

In an exploratory phase where we have not sufficient information about the model or the levels at which the parameters involved should be set, the most suitable course of action for the estimation of these parameters is to explore as large a range of values as possible. This may be achieved by a, "Crude", search method in which only one variable at a time is changed. According to this method and given a model of three parameters, A, B, and C any two of these parameters, say A and B are held constant at some arbitrary values while the third parameter C is taken every possible value within a specified range and considering a specific increment. Next, one of these two parameters, say B, is moved from its initial level to a new constant level and the second, A is kept at its initial level while the third C is again taken the same set of values as in the first step. This next step is repeated until all possible values of the second parameter B, within a predetermined range, are considered. Throughout these two steps, the first parameter A was kept unchanged. It is in a new step that this parameter assumes a new value and the previous steps are repeated in a new cycle of computation. For the present exercise, as many simulations as the number of parameter combinations introduced in the method are required, each producing a different value of the criterion variable employed (Kolmogorov-Smirnov statistic). It is unwise at this stage that the whole number of cases at risk

(6469)[3] should be simulated in such a large number of trials. The computer time and cost are perhaps major deterrents.

Moreover, since the objective at such an exploratory stage is to shed some light on the range of values for those parameters which contribute much to the reduction of the criterion statistic, it is more sensible to simulate a much smaller number of cases. A sample of thirty cases (any other number could have been used) is believed reasonable to fulfill this objective.

Following this, "Crude", search method, five simulation experiments were performed, each comprising a large number of simulations with varying levels for A, B, and C in each simulation. The first four experiments used $(^1/_{20})$ of a year as a simulation step while the simulations in the fifth experiment used $(^1/_5)$ of a year. Moreover, experiment five used the same range of values as experiment four but was run on sixty cases instead of thirty. The main purpose of experiment five is to show how far the simulation step and the sample size affect the result of the experiment.

Table (5:1) contains the conditions and output of these five experiments. It includes the ranges of values experimented upon for the three parameters and the combinations of parameter values which produced the lowest value of Kolmorov-Smirnov (K-S) statistic in each experiment. From this table we notice at once the big fall in the statistic. For it went down from 0.2667 in the first experiment to .0667 in the fourth experiment. When sixty cases instead of thirty were simulated and when longer time interval was used as a simulation step, as the conditions of

---

3. Those are people who were first registered in 1945 and survived a whole period of 20 years from 1945 to 1964. For more details of the specific simulation experiment, reference should be made to the appendix.

experiment five, the value of the statistic increased to .0833. On the other hand, the table clearly indicates that the variable combinations which yielded lower values of the criterion function were much like the combinations of values suggested by the preliminary analytical work (those were A = 0.159, B = 0.065, and C = 0.017).

5.5.2: Use of Minimization Routines:

Although the search method described above was useful in getting the Kolmogorov-Smirnov statistic at a reasonably low level, it had the disadvantage that the mere "trial and error" procedure may require the accomplishment of hundreds and hundreds of simulations before we reach a satisfactory result. A more efficient technique might help to reduce the number of simulations and consequently reduce the amount of time and cost needed to carry out these simulations. Since the main target is to produce as low a value of (K-S) statistic as possible, a routine of function minimization seems relevant. The routine employed here is a computer algorithm based on the Simplex method.[4] It is designed to minimize a general function f(x) of n independent variables $x = (x_1, x_2, \ldots x_n)$ in the frame of an iterative procedure. First, a simplex of (n+1) points is set up in the n-dimensional variable space. The starting point which is based on the initial values of the variables is the first vertex of the simplex. The remaining vertices are generated by the routine. The vertex of the simplex with the largest function value is reflected in the centre of gravity of the other vertices. This function value is compared with the other function values. The result of this

---

4. The routine used is E04CCF, one of Nottingham Algorithms Group, ICL 1900 System, N.A.G. Library Mannual, Doc. No. 187, (Sept. 1971). This routine was based on an article by J.A. Nelder and R. Mead, "A Simplex Method for Function Minimization", Computer Journal, Vol. 7, (1965), pp. 308-13.

Table (5:1)

Conditions and Output of Five Exploratory

Simulation Experiment

	Levels of Model Parameters			Value Of Kolmogorov-Smirnov Statistic	No. Of Simulations
	A	B	C		
	Experiment (1)				
Lowest Level	0.0001	0.0	0.0		
Highest Level	0.25	0.125	0.025		
Increment	0.05	0.025	0.005		
Minimum At	0.20	0.0	0.005		
	0.20	0.075	0.070		
	0.20	0.100	0.010		
	0.20	0.100	0.020	0.26667	125
	Experiment (2)				
Lowest Level	0.20	0.0	0.0		
Highest Level	0.20	0.100	0.015		
Increment	0.0	0.005	0.00075		
Minimum At	0.20	0.005	0.0008		
	0.20	0.010	0.0		
	0.20	0.015	0.011		
	0.20	0.040	0.0113		
	0.20	0.055	0.0		
	0.20	0.055	0.0053		
	0.20	0.070	0.00075		
	0.20	0.075	0.0113	.13333	400
	Experiment (3)				
Lowest Level	0.20	0.075	0.005		
Highest Level	0.30	0.100	0.020		
Increment	0.01	0.005	0.003		
Minimum At	0.23	0.079	0.011		
	0.24	0.090	0.011		
	0.24	0.095	0.011		
	0.24	0.095	0.017		
	0.27	0.080	0.017		
	0.28	0.085	0.014		
	0.29	0.080	0.014	.13333	250

Table (5:1) Contd.

	Levels of Model Parameters			Value Of Kolmogorov-Smirnov Statistic	No. Of Simulations
	A	B	C		
	Experiment (4)				
Lowest Level	0.25	0.065	0.005		
Highest Level	0.30	0.097	0.021		
Increment	0.01	0.004	0.002		
Minimum At	0.28	0.081	0.011	.06667	320
	Experiment (5)				
	0.25	0.065	0.005		
	0.30	0.097	0.021		
	0.01	0.004	0.002		
	0.25	0.065	0.017		
	0.25	0.077	0.019		
	0.26	0.085	0.013		
	0.27	0.089	0.015	.08333	320

test determines whether the new point is accepted or rejected and accordingly the simplex is expanded or reduced in length. The method is repeated until a satisfactory value is obtained. Although the method tends to be slow, it is sufficient for functions that are subject to inaccuracies.

As a criterion for ending the routine, a scalar value "E" must be provided by the user which should be set on entry to the required accuracy criterion. If F_i, i = 1, 2, ..., n+1 are the individual function values at the n+1 vertices of the simplex, and $\bar{F}$ is the function mean value, the routine will terminate when $\left[\sum_i (F_i-\bar{F})^2/(n+1)\right] < E$. Since the levels of variables (other than the initial levels) to be experimented upon using the above routine are not under the control of the investigator, the routine might result in undesirable levels even when such levels produce a minimum. In our case, for example, the routine may end with negative values for the parameters in general, or value of more than one for the parameter A which denotes a probability. For this reason, a penalty function may be established to constitute a sort of function constraint. This may be done by setting the function to be minimized at a very high level when such undesirable values occur so that the set of parameters at this specific iteration can be rejected. When using the above routine, the sample size was increased to a hundred cases. Table (5:2) contains the conditions and output of five of the successful runs of this routine. The first three values of the criterion function are by no means better than the lowest value obtained by the previous crude search method, although the sample size is different in the two methods. The last two values, however, show an appreciable reduction in the value of (K-S) statistic. This signifies that the minimum value of the function will be better found in the direction of the parameter levels corresponding to these values.

Table (5:2)

Conditions and Output of Five Selected Simulations Based on A Minimization Routine

Sequence No.	Model Parameters			Kolmogorov-Smirnov Statistic
	A	B	C	
1	.329	.206	.229	.222
2	.322	.204	.238	.233
3	.479	.465	.429	.133
4	.387	.459	.131	.060
5	.322	.209	.191	.040

The other striking result is that the levels of parameters extracted by this routine are different, in magnitude, from the levels suggested by the preliminary analytical work. The new values have more than double the size of the old ones.

5.6: Effect of Sample Size and Length of Simulation Step:

Although the results obtained in the exploratory stage described above are by no means conclusive, we have managed to bring about the value of (K-S) statistic at least to a reasonably low level.[5] Moreover, we did gain more knowledge about the levels that the model parameters should assume, not the exact levels but at least the right directions as where these levels might be located. Before we proceed to a further step in the simulation process, the acquisition of some good points enables us to show how far the size of the sample affects the value of our target function. Moreover, does it matter if we consider a simulation step other than the one that has been used, i.e., a simulation step of length $\frac{1}{5}$, $\frac{1}{2}$, or a full year instead of $\frac{1}{20}$ of a year. The following paragraphs are devoted to answer these queries.

Analysis of the (K-S) statistic as well as other standard statistics based on ten different points (five from each of the "Crude" search and minimization procedures) was carried out by simulating different sample sizes, i.e., 100, 300, and 6469 (the whole sample at risk) and considering values of $\frac{1}{20}$, $\frac{1}{5}$, $\frac{1}{2}$ and 1 full year as alternative simulation steps. Table (5.3) contains a summary of such analysis. It includes the above

---

5. Up to now, we are only interested in obtaining a value of (K-S) statistic as low as we can. We will be more interested, in later stages, in obtaining some specific low values of this statistic, that is, values at which the differences between the hypothetical and observed distributions of movement are not statistically significant.

statistics as generated from the simulation along with similar statistics obtained from the sample. The results were shown for different sample sizes and different simulation step lengths. Inspection of the entries of this table reveals the following remarks.

First, experimentation using large samples does produce a better overall fit of the theoretical distribution of movement, as hypothesized by the model, to the observed distribution representing the actual experience of individuals. This is indicated by the fact that, for most cases, generally lower values for the (K-S) statistic accompanied large samples. Moreover, the fluctuations in the value of this statistic (for the step lengths considered) are more observed in the small samples than in the large samples.

As for the average number of moves and other related statistics, we notice that the actual figures as obtained from the sample are usually underestimated, below the simulation generated figures, in the first set of simulations (resulted from the "Crude" search routine) whereas they are overestimated, above the simulation generated figures, in the second set of simulations (minimization oriented simulations). In addition, the differences in the simulated statistics are much higher in the first set of simulations, than they are in the second.

As far as the simulation step length is concerned, the following is observed. In general, the smaller the length of time taken as a simulation step, the lower is the value of the (K-S) statistic and the better are the estimates for the average number of moves and other statistics. The best of all possible lengths in most simulations is the twentieth of a year. Again, this is more marked in the first set of simulations (the "Crude" search routine) than in the second.

A decreasing trend of the simulated averages with increasing length of time increment can be clearly observed in the former, while a rather fluctuating trend is characteristic of the latter.

Table (5.3)

Conditions and Summary Statistics for Two Selected Sets of Simulations Compared with the Actual Statistics

(a) Exploratory Simulations

Seq. No.	Sample Size	Length of Simul-ation Step	Number of Simul-ation Steps	Simulation Statistics: Mean Number of Moves $\bar{X}_S$	Simulation Statistics: Stan-dard Devi-ation σ_S	Simulation Statistics: Coeff-icient of Vari-ation $\sigma_S/\bar{X}_S$	Actual Statistics: $\bar{X}_A$	Actual Statistics: σ_A	Actual Statistics: $\sigma_A/\bar{X}_A$	Kolmo-gorov Smirnov Statis-tic
		1): A = 0.25 B = 0.065 C = 0.017								
1	100	1/20	400	0.750	1.749	2.330	0.930	1.373	1.470	0.2600
2	100	1/5	100	0.540	1.438	2.660	0.930	1.373	1.470	0.3500
3	100	1/2	40	0.280	0.753	2.690	0.930	1.373	1.470	0.3300
4	100	1	20	0.420	1.057	2.560	0.930	1.373	1.470	0.3200
5	300	1/20	400	0.903	2.116	2.340	0.757	1.340	1.770	0.1100
6	300	1/5	100	0.737	1.980	2.680	0.757	1.340	1.770	0.1767
7	300	1/2	40	0.497	1.144	2.300	0.757	1.340	1.770	0.1633
8	300	1	20	0.457	1.140	2.490	0.757	1.340	1.770	0.1800
9	6469	1/20	400	0.731	1.735	2.370	0.736	1.471	1.999	0.1124
10	6469	1/5	100	0.701	1.719	2.452	0.736	1.471	1.999	0.1201
11	6469	1/2	40	0.581	1.355	2.332	0.736	1.471	1.999	0.1105
12	6469	1	20	0.477	1.097	2.300	0.736	1.471	1.999	0.1235
		2): A = 0.25 B = 0.077 C = 0.019								
13	100	1/20	400	0.740	1.733	2.360	0.930	1.373	1.470	0.2700
14	100	1/5	100	0.840	1.973	2.350	0.930	1.373	1.470	0.2900
15	100	1/2	40	0.620	1.427	2.300	0.930	1.373	1.470	0.2700
16	100	1	20	0.350	0.833	2.380	0.930	1.373	1.470	0.3100
17	300	1/20	400	0.720	1.777	2.470	0.757	1.340	1.770	0.1633
18	300	1/5	100	0.717	1.744	2.430	0.757	1.340	1.770	0.1567
19	300	1/2	40	0.677	1.469	2.170	0.757	1.340	1.770	0.1300
20	300	1	20	0.353	0.923	2.160	0.757	1.340	1.770	0.2067
21	6469	1/20	400	0.747	1.758	2.350	0.736	1.471	1.999	0.1060
22	6469	1/5	100	0.695	1.706	2.454	0.736	1.471	1.999	0.1204
23	6469	1/2	40	0.616	1.402	2.276	0.736	1.471	1.999	0.1036
24	6469	1	20	0.508	1.143	2.250	0.736	1.471	1.999	0.1132

Table (5.3)
(continued)

Seq. No.	Sample Size	Length of Simulation Step	Number of Simulation Steps	Simulation Statistics $\bar{X}_S$	σ_S	$\sigma_S/\bar{X}_S$	Actual Statistics $\bar{X}_A$	σ_A	$\sigma_A/\bar{X}_A$	Kolmogorov Smirnov Statistic
	3): A = 0.260 B = 0.085 C = 0.013									
25	100	1/20	400	0.940	1.728	1.840	0.930	1.373	1.470	0.2000
26	100	1/5	100	0.920	2.058	2.240	0.930	1.373	1.470	0.2500
27	100	1/2	40	0.370	0.928	2.510	0.930	1.373	1.470	0.3300
28	100	1	20	0.400	1.092	2.730	0.930	1.373	1.470	0.3200
29	300	1/20	400	1.000	2.214	2.210	0.757	1.340	1.770	0.1033
30	300	1/5	100	0.673	1.545	2.290	0.757	1.340	1.770	0.1333
31	300	1/2	40	0.623	1.301	2.080	0.757	1.340	1.770	0.1200
32	300	1	20	0.700	1.469	2.090	0.757	1.340	1.770	0.1133
33	6469	1/20	400	0.774	1.789	2.312	0.736	1.471	1.999	0.1057
34	6469	1/5	100	0.779	1.816	2.331	0.736	1.471	1.999	0.1056
35	6469	1/2	40	0.636	1.480	2.327	0.736	1.471	1.999	0.1062
36	6469	1	20	0.535	1.223	2.286	0.736	1.471	1.999	0.1101
	4): A = 0.270 B = 0.089 C = 0.015									
37	100	1/20	400	0.790	2.007	2.540	0.930	1.373	1.470	0.2500
38	100	1/5	100	0.980	2.322	2.370	0.930	1.373	1.470	0.2500
39	100	1/2	40	0.510	1.020	2.000	0.930	1.373	1.470	0.2600
40	100	1	20	0.230	0.548	2.380	0.930	1.373	1.470	0.3400
41	300	1/20	400	0.590	1.477	2.500	0.757	1.340	1.770	0.1767
42	300	1/5	100	0.687	1.673	2.420	0.757	1.340	1.770	0.1567
43	300	1/2	40	0.713	1.629	2.280	0.757	1.340	1.770	0.1367
44	300	1	20	0.597	1.251	2.090	0.757	1.340	1.770	0.1401
45	6469	1/20	400	0.798	1.914	2.390	0.736	1.471	1.999	0.1066
46	6469	1/5	100	0.818	1.926	2.354	0.736	1.471	1.999	0.1019
47	6469	1/2	40	0.699	1.549	2.216	0.736	1.471	1.999	0.0886
48	6469	1	20	0.573	1.244	2.244	0.736	1.471	1.999	0.0970

Table (5.3)
(continued)

Seq. No.	Sample Size	Length of Simulation Step	Number of Simulation Steps	Simulation Statistics $\bar{X}_s$	σ_s	$\sigma_s/\bar{X}_s$	Actual Statistics $\bar{X}_A$	σ_A	$\sigma_A/\bar{X}_A$	Kolmogorov Smirnov Statistic
				5): A = 0.280 B = 0.081 C = 0.011						
49	100	1/20	400	0.860	1.853	2.155	0.660	1.350	2.045	0.0500
50	100	1/5	100	1.100	2.195	1.996	0.660	1.350	2.045	0.1100
51	100	1/2	40	0.790	1.748	2.213	0.660	1.350	2.045	0.0400
52	100	1	20	0.650	1.381	2.381	0.660	1.350	2.045	0.0600
53	300	1/20	400	0.827	1.880	2.273	0.757	1.340	1.770	0.1333
54	300	1/5	100	0.757	1.870	2.470	0.757	1.340	1.770	0.1600
55	300	1/2	40	0.810	1.716	2.118	0.757	1.340	1.770	0.1000
56	300	1	20	0.503	1.252	2.487	0.757	1.340	1.770	0.1633
57	6469	1/20	400	0.919	2.048	2.267	0.736	1.471	1.999	0.0904
58	6469	1/5	100	0.886	2.050	2.314	0.736	1.471	1.999	0.0949
59	6469	1/2	40	0.709	1.587	2.238	0.736	1.471	1.999	0.0889
60	6469	1	20	0.575	1.259	2.188	0.736	1.471	1.999	0.0943
				(b) Simulations Based on Minimization Routine						
				1): A = 0.329 B = 0.2064 C = 0.229						
61	100	1/20	400	0.670	1.240	1.850	0.930	1.373	1.470	0.2200
62	100	1/5	100	0.650	1.373	2.110	0.930	1.373	1.470	0.2500
63	100	1/2	40	0.530	1.167	2.200	0.930	1.373	1.470	0.2800
64	100	1	20	0.770	1.523	1.980	0.930	1.373	1.470	0.2100
65	300	1/20	400	0.640	1.512	2.350	0.757	1.340	1.770	0.1200
66	300	1/5	100	0.893	1.702	1.910	0.757	1.340	1.770	0.0699
67	300	1/2	40	0.760	1.616	2.120	0.757	1.340	1.770	0.1000
68	300	1	20	0.743	1.439	1.940	0.757	1.340	1.770	0.0900
69	6469	1/20	400	0.748	1.534	2.050	0.736	1.471	1.999	0.0771
70	6469	1/5	100	0.708	1.490	2.110	0.736	1.471	1.999	0.0794
71	6469	1/2	40	0.732	1.546	2.110	0.736	1.471	1.999	0.0671
72	6469	1	20	0.652	1.323	2.030	0.736	1.471	1.999	0.0693

Table (5.3)
(continued)

Seq. No.	Sample Size	Length of Simulation Step	Number of Simulation Steps	Simulation Statistics			Actual Statistics			Kolmogorov Smirnov Statistic
				$\bar{X}_s$	σ_s	$\sigma_s/\bar{X}_s$	$\bar{X}_A$	σ_A	$\sigma_A/\bar{X}_A$	
2): A = 0.322 B = 0.2045 C = 0.238										
73	100	1/20	400	0.650	1.336	2.060	0.930	1.373	1.470	0.2300
74	100	1/5	100	0.900	1.738	1.930	0.930	1.373	1.470	0.2100
75	100	1/2	40	0.770	1.530	1.990	0.930	1.373	1.470	0.2100
76	100	1	20	0.580	1.121	1.930	0.930	1.373	1.470	0.2100
77	300	1/20	400	0.787	1.620	2.060	0.757	1.340	1.770	0.0867
78	300	1/5	100	0.480	1.020	2.130	0.757	1.340	1.770	0.1433
79	300	1/2	40	0.697	1.328	1.910	0.757	1.340	1.770	0.0933
80	300	1	20	0.673	1.324	1.960	0.757	1.340	1.770	0.0933
81	6469	1/20	400	0.718	1.465	2.040	0.736	1.471	1.999	0.0748
82	6469	1/5	100	0.674	1.432	2.120	0.736	1.471	1.999	0.0767
83	6469	1/2	40	0.697	1.472	2.110	0.736	1.471	1.999	0.0719
84	6469	1	20	0.627	1.284	2.050	0.736	1.471	1.999	0.0688
3): A = 0.479 B = 0.465 C = 0.429										
85	100	1/20	400	1.020	1.688	1.650	0.930	1.373	1.470	0.1300
86	100	1/5	100	1.240	1.870	1.510	0.930	1.373	1.470	0.1299
87	100	1/2	40	1.120	1.945	1.740	0.930	1.373	1.470	0.1200
88	100	1	20	1.010	1.778	1.760	0.930	1.373	1.470	0.1700
89	300	1/20	400	0.973	1.711	1.760	0.757	1.340	1.770	0.0700
90	300	1/5	100	1.013	1.762	1.740	0.757	1.340	1.770	0.0900
91	300	1/2	40	0.990	1.761	1.780	0.757	1.340	1.770	0.0800
92	300	1	20	1.077	1.846	1.710	0.757	1.340	1.770	0.0867
93	6469	1/20	400	1.036	1.722	1.660	0.736	1.471	1.999	0.0929
94	6469	1/5	100	1.016	1.726	1.750	0.736	1.471	1.999	0.0863
95	6469	1/2	40	1.046	1.800	1.720	0.736	1.471	1.999	0.0861
96	6469	1	20	1.074	1.896	1.770	0.736	1.471	1.999	0.0894

Table (5.3)
(continued)

Seq. No.	Sample Size	Length of Simulation Step	Number of Simulation Steps,	Simulation Statistics $\bar{X}_S$	Simulation Statistics σ_S	Simulation Statistics $\sigma_S/\bar{X}_S$	Actual Statistics $\bar{X}_A$	Actual Statistics σ_A	Actual Statistics $\sigma_A/\bar{X}_A$	Kolmogorov Smirnov Statistic
4): A = 0.387 B = 0.459 C = 0.131										
97	100	1/20	400	0.590	1.074	1.820	0.660	1.350	2.045	0.0600
98	100	1/5	100	1.240	2.001	1.614	0.660	1.350	2.045	0.1600
99	100	1/2	40	0.890	1.626	1.826	0.660	1.350	2.045	0.0800
100	100	1	20	0.660	1.436	2.177	0.660	1.350	2.045	0.0600
101	300	1/20	400	0.807	1.482	1.836	0.757	1.340	1.770	0.0700
102	300	1/5	100	0.863	1.795	2.079	0.757	1.340	1.770	0.0800
103	300	1/2	40	1.003	1.730	1.725	0.757	1.340	1.770	0.0967
104	300	1	20	0.933	1.776	1.903	0.757	1.340	1.770	0.0700
105	6469	1/20	100	0.863	1.618	1.875	0.736	1.471	1.999	0.0519
106	6469	1/5	100	0.836	1.591	1.903	0.736	1.471	1.999	0.0448
107	6469	1/2	40	0.878	1.673	1.901	0.736	1.471	1.999	0.0510
108	6469	1	20	0.896	1.715	1.914	0.736	1.471	1.999	0.0546
5): A = 0.322 B = 0.209 C = 0.191										
109	100	1/20	400	0.560	1.113	1.987	0.660	1.350	2.045	0.0400
110	100	1/5	100	1.120	1.805	1.612	0.660	1.350	2.045	0.1600
111	100	1/2	40	0.810	1.716	2.118	0.660	1.350	2.045	0.0500
112	100	1	20	0.620	1.369	2.208	0.660	1.350	2.045	0.0500
113	300	1/20	400	0.697	1.356	1.946	0.757	1.340	1.770	0.0833
114	300	1/5	100	0.777	1.621	2.086	0.757	1.340	1.770	0.1200
115	300	1/2	40	0.777	1.599	2.058	0.757	1.340	1.770	0.1000
116	300	1	20	0.587	1.320	2.248	0.757	1.340	1.770	0.1367
117	6469	1/20	400	0.738	1.511	2.047	0.735	1.471	1.999	0.0597
118	6469	1/5	100	0.655	1.454	2.186	0.736	1.471	1.999	0.0852
119	6469	1/2	40	0.707	1.507	2.131	0.736	1.471	1.999	0.0792
120	6469	1	20	0.635	1.311	2.064	0.736	1.471	1.999	0.0748

Chapter (6)

A SIMULATION PROBABILITY MODEL FOR REPEATED MIGRATION

FINAL ASSESSMENT

6.1: A more precise objective:

So far, the analysis has been mainly directed to explore the possibility that the model suggested in Chapter 5 can describe with reasonable efficiency the phenomenon of repeated migration as experienced by Scottish migrants. A natural consequence to this was that if the model does not fail completely to explain the observed patterns of movement, what are the possible ranges of values that the concerned parameters should assume in order to provide such an explanation. In other words, the aim in the earlier stages was not the search for exact or optimum solutions to the problem investigated, rather it was a search for the areas of potential interest where these solutions may exist. Now that we have enetered the domain of such areas, we ought to move a step forward and define more precisely the target of the analysis.

The main criterion for judging the validity of the proposed model was the agreement or disagreement between two different distributions of movement, one representing the actual experience of individuals and the other generated by the model. The questions to be asked now are: How far we can accept the agreement or reject the disagreement between these two different distributions? Would the observed discrepancy be considered as statistically significant or would it be attributed to mere chance? The Kolmogorov-Smirnov (K-S) two sample statistic was chosen as an appropriate criterion for testing the above hypotheses.[1] Basically we have been concerned with diminishing the observed value of this statistic as far as we could without paying attention to specific

---

1. Chapter 5 provides a description of this statistic and its merits relative to the Chi-squared statistic.

probability level of significance. We are in a position now to define a specific value of this statistic as our target, a value which if compared with the theoretical value at certain level of probability, would indicate that the variations between the above-mentioned distributions are chance variations and the two samples (actual and simulated) are drawn from the same population. The critical value of (K-S) corresponding to 6469, the size of the sample at risk and to $p = 0.01$ significance level, is 0.0282. This was chosen as an intermediate probability level in table (6.1) which shows the theoretical value of (K-S) at different probability levels.

The implication of this is that any observed value equal to or greater than 0.0282 will be significant with probability of occurrence under H_0 (the null hypothesis) equal to or less than $p = 0.01$ (Siegel, 1956; pp.127-136). Accordingly, we can formulate the objective of this chapter as follows. Given the limited knowledge about the model and the relevant characteristics, what are the values of the parameters which produce a cumulative probability distribution of movement that deviates from the corresponding actual distribution at most with an absolute value less than 0.0282.

Since the objective is to provide a final testing of the model, experimentation should be performed on the whole sample at risk. It has been mentioned earlier that it is almost impossible for a "hunting" procedure or a single computer minimization routine of chapter 5 to carry out such experimentation process without consuming huge amounts of computer time and cost which would seem an unwise waste of resources.[2] More efficient techniques are then required to fulfil the target with the minimum of these considerations.

---

2 A single simulation of 6469 individuals at risk for 20 years with 1/20 of a year as a simulation step, required 158.75 seconds of computer time and cost about £15.56.

Table (6.1)**

Some Selected Critical Values of D* in the
Kolmogorov-Smirnov Two Sample Test:
Large Samples; Two-Sided Test

Level of Significance	Formula for any n_1, n_2	$n_1 = n_2 = n$	
		n = 300	n = 6469
0.10	$1.22\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.0999	0.0211
0.05	$1.36\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.1114	0.0235
0.025	$1.48\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.1212	0.0256
0.01	$1.63\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.1335	0.0282
0.005	$1.73\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.1417	0.0299
0.001	$1.95\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	0.1597	0.0337

* $D = \text{maximum} \left| S_{n1}(X) - S_{n2}(X) \right|$

** Adapted from Sidney Siegel, Nonparametric Statistics for the Behavioural Sciences, McGraw-Hill Series in Psychology (1956), p.279

The following sections are devoted to the introduction and application of such techniques. Firstly, an outline of the methods utilized, theoretical background as well as practical considerations, is presented. Secondly, the application of the outlined methods to the present problem is considered and the results of such an application are analysed. Thirdly, a final assessment of the model and the relevant characteristics of the parameters is given.

6.2: Methods for locating optimum conditions:

In this section we shall be concerned with presenting some methods that deal with the problem of finding the best conditions which operate upon a process by maximizing or minimizing specific features of it. The methods were developed by Box and Wilson for the study of chemical processes (Box and Wilson, 1951; Box et. al., 1953; and Box, 1954). The following is a brief description of the practical aspects of the procedures involved in these methods.

The problem of determining the conditions that give an optimum result for a certain process within a system of any kind may be theoretically handled by the application of such ascertainable laws that govern the behaviour of the system (Box et. al., 1953; p.495). In very complicated systems, however, this approach is not easy to apply. Rather, an empirical approach based on experimentation is more practical. Following Box's notation, the problem can be stated as follows. Given that a response U depends on the level of k quantitative variables (factors) $x_1, x_2, \ldots x_k$ such that

$$U = f(x_1, x_2 \ldots x_k) \ldots\ldots (6.1)$$

is the underlying response function; there exists a region R, in the whole K-dimensional factor space, that corresponds to factor combinations of potential interest. The problem is to find, within that region and using the smallest number of experiments, the point $(x_1^o, x_2^o, \ldots\ldots x_k^o)$ at which the response U is a maximum or a minimum[3] (Box and Wilson, 1951; pp.1-2).

A possible way of tackling the problem is to explore the whole experimental region by carrying out experiments on every possible point in R. But if the response surface is too complicated, in which case the required number of experiments is exceedingly large, the full exploration of the experimental region is impractical. An alternative way is to explore a small sub-region of R, performing only a limited number of experiments. In such a way, it may be possible to use the results of an experiment, carried out in one sub-region, as input to a second experiment for the exploration of another sub-region in which the response is thought to be higher. By successive application of the procedure, we could reach a maximum (minimum) or at least a near-stationary region at which the maximum may be located. The method described here involves two distinct phases. First, if the initial conditions are remote from the maximum, the surface can be represented by a plane, by carrying out some suitably arranged experiments in the sub-region and estimating first order coefficients of that plane. From the magnitudes and signs of these coefficients one can calculate the

---

3. For convenience, the description of the method will concentrate on the maximization aspect. The problem of maximization can be, without serious difficulty, converted to one of minimization. For more detailed discussion of the methods themselves, reference should be made to the original papers.

direction of the steepest ascent (descent) up (down) the plane. This direction indicates the amounts by which the factors are to be varied in order to get the greatest gain in the response path. That can be accomplished by taking a convenient increment in one variable and calculating the proportionate changes to be made in the other variables (Box et. al., 1953; p.502).

It should be borne in mind that this phase of procedure alone does not accurately lead to an absolute maximum or minimum. It does however bring the experimenter rapidly to a near-stationary region, a region where the slopes of the surface flatten and become relatively small compared with the error of estimation.

The second phase of the method entails the determination of the nature of the surface in the near-stationary region and the location of the optimum, using a few more extra points. In the near-stationary region, various possibilities are likely to occur. The experimenter may be in the neighbourhood of a true maximum and he will be wishing to locate its position. It may happen that the surface indicates a ridge of some sort and the desire then is to determine the direction of this ridge. It is also possible that the point reached is a minimax (saddle point) where the response is a minimum for one direction and a maximum for another. The priority in this case is to find out the right direction to climb out of this minimax (Box et. al., 1953; p.505).

In any case, the surface in the near-stationary region is represented by a polynomial of higher degree than the first. The nature of the surface can be fairly accurately determined by the study of the coefficients of that polynomial.

Sequential experimentation, obviously, is a characteristic feature of the above method. If the analysis of the fitted polynomial indicates a true minimum or maximum, one would have to calculate its co-ordinates and perform a confirmatory experiment based upon the point represented by these co-ordinates. The result of this further experiment is added to the previous results and a new regression is fitted if necessary. On the other hand, if the analysis shows instead a ridge or col, additional experiments are carried out at points along the path of greatest gain. Again the results of these experiments are added to the previous ones and a new analysis is performed to picture more clearly the surface under investigation. In either case, the process is repeated until no further improvement is gained (Box, 1954; p.30).

It is not necessarily important, however, that both phases should be implemented. For in some situations, where the aim is to improve a process which has already received attention or where the region of interest is a near-stationary region, only the second phase of the method is relevant. In what follows, more attention is given to this phase, i.e. the determination of the nature of the surface at the near-stationary region.

<u>Exploration and analysis of a response surface in a near-stationary region:</u>

Suppose that we have already been in a near-stationary region and we want to know whether we have reached a true maximum, a ridge or a col, etc.

Consider the case of three variables, x_1, x_2 and x_3.[4]

For a suitably arranged set of points, a polynomial of second degree may be adequately fitted. Such a polynomial assumes the form:

$$U = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_{11}x_1^2+\beta_{22}x_2^2+\beta_{33}x_3^2+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{23}x_2x_3+e \ldots\ldots (6.2)$$

where, U is the true response,

β_0 is the level of response at the origin,

$\beta_1,\beta_2,\beta_3,\beta_{12},\beta_{13},\beta_{23}$ are first order and interaction effects, and

$\beta_{11},\beta_{22},\beta_{33}$ are second order or quadratic effects.

An experimental arrangement that yields efficient estimates of both first and second order effects is required. Furthermore, a convenient method is needed to determine the nature and characteristics of the fitted surface.

For the first requirement, designs that are suitable for determining the estimates of the coefficients in the quadratic equation are factorial and fractional factorial designs. In the case of three variables, first and second order effects can best be determined using what Box and associates called a composite design (Box et. al., 1953; p.532). This design is built up from a complete two level factorial design. Specifically, a 2^3 factorial design is first used to estimate first order and interaction effects. Seven further points are added, one at the centre of the design and the remaining six points are located in pairs along the co-ordinate axes at $\pm\alpha_1$, $\pm\alpha_2$, and $\pm\alpha_3$. The

---

4. The principles outlined here may be applied equally well to cases of k variables whatever k is. The difference is that each case requires the appropriate experimental arrangement that provides efficient estimates of the coefficients.

addition of these points allows the estimation of the quadratic effects. Taking the levels of the factors at either the upper or lower limits designated by plus or minus ones, the design matrix for such experiment is shown in table(6.2). The experimental points of the design are located at the vertices of a cube which is represented in figure(1). An othogonal composite design can be obtained by taking $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$ where α is a suitably chosen value.[5]

Using the above arrangement and considering the values of response at the different points of the design, the desired effects can be easily determined due to the fact that the Least squares estimates of the coefficients of equation(6.2)are orthogonal linear functions of the observations.

The fitted quadratic equation will thus take the form

$$Y = b_0+b_1x_1+b_2x_2+b_3x_3+b_{11}x_1^2+b_{22}x_2^2+b_{33}x_3^2+b_{12}x_1x_2+b_{13}x_1x_3+b_{23}x_2x_3 \ldots\ldots (6.3)$$

where Y is the response as calculated from the regression equation. The regression equation should account for a large amount of variation in the data. In other terms, some measure of goodness of fit must confirm that there is no reason to suggest that the quadratic representation of the surface was not adequate.

The interpretation of the quadratic equation (6.3) is not so easy and does not give a clear idea about the nature of the surface,

---

5. Values of which ensure orthogonality in cases of 2, 3, 4 and 5 variables are given by Box et al as 1, 1.215, 1.414, and 1.547 respectively. See Box G.E.P. et al, "The Determination of Optimum Conditions", p.534.

Table 6.2

A THREE-FACTOR COMPOSITE DESIGN

Trial	Factor level x_1	x_2	x_3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1
9	$-\alpha_1$	0	0
10	α_1	0	0
11	0	$-\alpha_2$	0
12	0	α_2	0
13	0	0	$-\alpha_3$
14	0	0	α_3
15	0	0	0

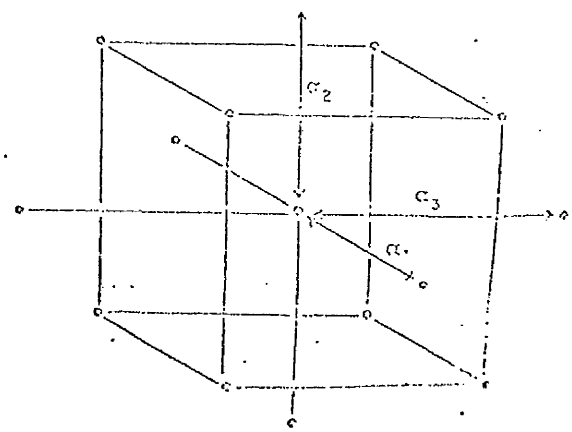

Fig.(1). A three factor composite design

Source: C.E.P. Box et al, "The Attainment of Optimum Conditions", in The Design and Analysis of Industrial Experiments, O.L. Davies, ed. (1953), p.533.

particularly if the number of variables was more than three. Fortunately, this difficulty could be overcome by reducing the quadratic equation to its Canonical form which consists only of a few terms. The Canonical form of equation (6.3) is

$$Y - Y_s = B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 \ldots\ldots(6.4)$$

where Y, Y_s denote the response as measured from the origin and the centre of the design respectively; B_{11}, B_{22} and B_{33} represent the changes in response on moving away from the centre of the design in the directions of the new co-ordinates X_1, X_2 and X_3.

Box argues that different shapes of the response surface are recognized according to the signs and magnitudes of the canonical coefficients (Box, 1954; pp.36-38). If the B's were all negative, the centre of the system is said to be a point maximum. There is always a loss in the response on moving away from the centre at any direction. On the other hand, if all the B's were positive, the centre of the system is a point minimum. If some of the coefficients were negative while the others were positive, we should have what is known as a saddle-point or a minimax. When one or more of the coefficients are zero and the others are negative, we have either a line, a plane or a space miximum, i.e. a stationary ridge of one, two....., or (k-r) dimensions. In these cases, there is a wide range of alternative optimum conditions. If one or more of the coefficients were small compared with the others, the X - axes belonging to these coefficients define the line, the plane or the space ridges and the corresponding equations of these axes determine the direction(s) of these ridges that are to be followed by the

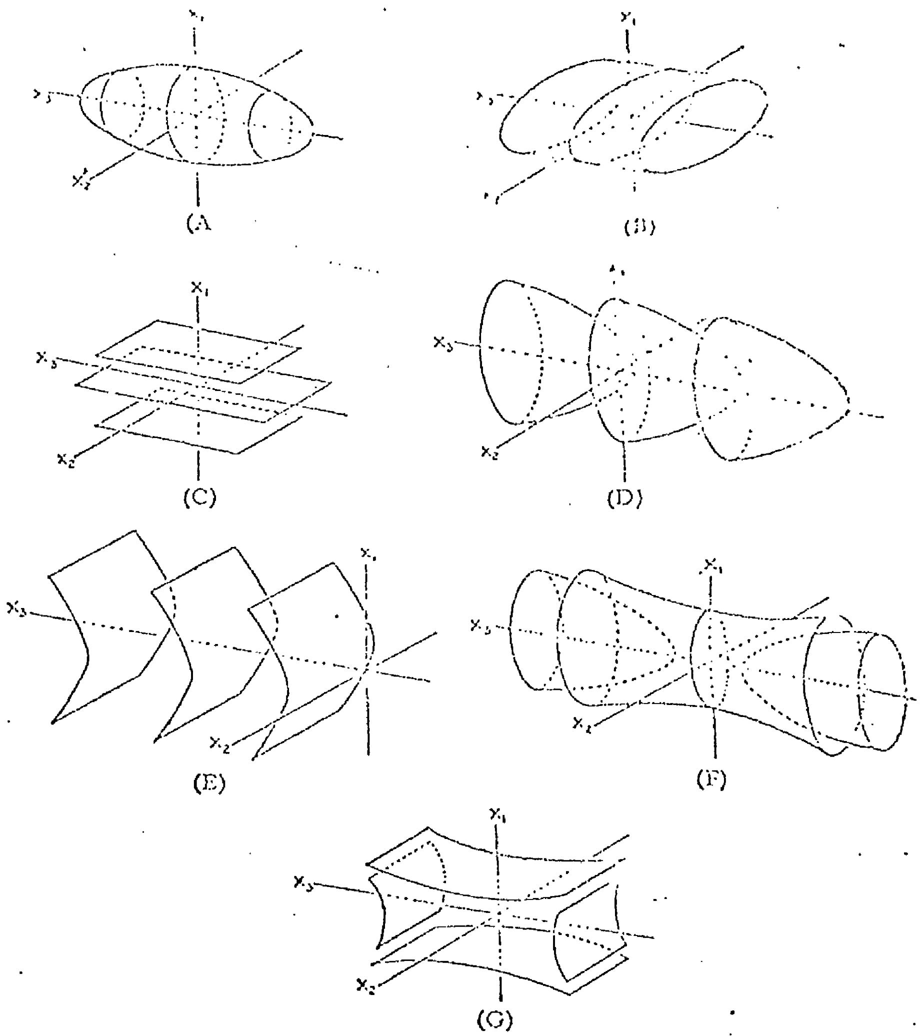

Fig.(2). Some possible three-dimensional contour surfaces in a near-stationary region

Source: G.E.P. Box, "The Exploration and Exploitation of Response Surface: Some General Considerations and Examples", Biometrics, 10 (1954), pp.16-60.

steepest descent (ascent) method. As mentioned above there are other possible shapes of the response surface which can be met in practice depending upon the signs and magnitudes of the canonical coefficients. Some of these surfaces are exposed in figure (2).

The rest of this section is devoted to the calculation of the canonical form of the quadratic equation (6.3). First, we have to define the centre of the system 's', the point at which the response is stationary. This can be obtained when $\frac{\partial Y}{\partial x_1}$, $\frac{\partial Y}{\partial x_2}$, and $\frac{\partial Y}{\partial x_3}$ are all equal to zero. By differentiating the fitted quadratic equation with respect to the independent variables x_1, x_2 and x_3 and equating each of them to zero we should have

$$\begin{aligned} 2b_{11}x_1 + b_{12}x_2 + b_{13}x_3 &= -b_1 \\ b_{12}x_1 + 2b_{22}x_2 + b_{23}x_3 &= -b_2 \qquad \ldots\ldots(6.5) \\ b_{13}x_1 + b_{23}x_2 + 2b_{33}x_3 &= -b_3 \end{aligned}$$

If the above system of simultaneous equations is solved, x_{1s}, x_{2s}, x_{3s} the co-ordinates of the centre point 'S' can be determined. Substituting these values in the quadratic equation gives 'Y_s' the response at this stationary point. Secondly, the canonical coefficients B_{11}, B_{22}, B_{33} are found by solving the following characteristic equation.

$$f(B) = \begin{vmatrix} b_{11} - B & \frac{1}{2}b_{12} & \frac{1}{2}b_{13} \\ \frac{1}{2}b_{12} & b_{22} - B & \frac{1}{2}b_{23} \\ \frac{1}{2}b_{13} & \frac{1}{2}b_{23} & b_{33} - B \end{vmatrix} = 0 \qquad \ldots\ldots(6.6)$$

The diagonal elements of the determinant on the right hand side of (6.6) are the quadratic effects minus B, an unknown quantity. The off-diagonal elements are half the interaction effects (Box et. al., 1953; pp.527-529).

One way of solving (6.6) is to multiply out the determinant and then find the roots of the resulting cubic equation in B. For k variables, the characteristic equation is a polynomial of degree k whose k roots are the values B_{11}, $B_{22} \ldots B_{kk}$ respectively. Finally, the new axes of the fitted quadratic surface are defined by the following orthogonal transformations.

$$
\begin{aligned}
x_1 &= m_{11}(x_1 - x_{1s}) + m_{12}(x_2 - x_{2s}) + m_{13}(x_3 - x_{3s}) \\
x_2 &= m_{21}(x_1 - x_{1s}) + m_{22}(x_2 - x_{2s}) + m_{23}(x_3 - x_{3s}) \quad \ldots\ldots(6.7) \\
x_3 &= m_{31}(x_1 - x_{1s}) + m_{32}(x_2 - x_{2s}) + m_{33}(x_3 - x_{3s})
\end{aligned}
$$

The coefficients m_{11}, m_{12} and m_{13} of any of the above orthogonal transformations are the solution of the set of equations whose variables (unknowns) are m_{11}, m_{12} and m_{13} and whose coefficients are the elements of the determinant in (6.6) after the substitution of the corresponding value of B_{11} as B. For example, m_{11}, m_{12} and m_{13} are the solutions of the following set of equations (when solved simultaneously).

$$
\begin{aligned}
(b_{11} - B_{11})m_{11} + \tfrac{1}{2}b_{12}m_{12} + \tfrac{1}{2}b_{13}m_{13} &= 0 \\
\tfrac{1}{2}b_{12}m_{11} + (b_{22} - B_{11})m_{12} + \tfrac{1}{2}b_{13}m_{23} &= 0 \\
\tfrac{1}{2}b_{13}m_{11} + \tfrac{1}{2}b_{23}m_{12} + (b_{33} - B_{11})m_{13} &= 0
\end{aligned}
$$

Once the above calculations have been carried out and the quadratic equation has been reduced to its canonical form, the investigator can easily recognize the nature of the response surface of interest. He would be able to know whether or not he has reached the minimum or maximum secured and/or what further steps he should follow in order to decide on the conditions of his experiment that yield the optimum result.

6.3: Applications of the methods:

The problem of determining the levels of the model parameters A, B and C that provide the best fit of the model to the observed data may be viewed as typical of the general problem of defining the factor levels that yield a maximum or minimum response. The response here being the value of the Kolmogorov-Smirnov statistic, chosen as the measure of goodness of fit, and the factors being the set of parameters which, when applied to the model, produce a specific level of this statistic. As we have seen in section (6.2) the methods of locating an optimum response comprise two distinct phases. The first is to define a near-stationary region by eliminating first order effects using the steepest ascent (descent) method. The second is to explore the nature of the response surface in this near-stationary region by determining second or higher order effects. It has also been shown that if the problem is to improve still further a process which has already received attention, and the region reached may be considered as a near-stationary region, the first of these phases is less necessary while the second is more relevant. The preliminary analysis carried out in chapter 5 transfers us safely to the second phase, i.e. the analysis of the response surface at the near-stationary region. In this case, the aim is to reduce further the value of (K-S) statistic and to locate the parameters' levels that affect such a reduction.

Choice of experimental design:

When a quadratic representation of the response surface is utilized and effects up to second order of three factors x_1, x_2 and x_3[6] are to be estimated, a suitable experimental arrangement, such as the one explained

---

6. For convenience, we shall follow the notation of section 6.2; x_1, x_2 and x_3 will denote the parameters A, B and C. The response Y will correspond to a value of Kolmogorov-Smirnov statistic as the measure of goodness of fit.

in section 6.2, must be used to yield efficient estimates of these effects. The levels of the factors to be employed in such an arrangement are always determined by the judgment of the experimenter and his own experience.

What we have obtained through simulation in the early exploration phase of the problem is a diverse set of points each associated with a specific set of variable levels. Some of these points were in the direction of a minimum response. Not all the points, however, used the same conditions of simulation. Rather, a variable sample size and a variable length of time increment as the simulation step were utilized. As the final testing of the model is our target in this stage, we ought to concentrate on the points obtained from simulation of all individuals at risk (6469) at time intervals each equal to one-twentieth of a year. These are the conditions which provided the best results in the preliminary testing phase. Before establishing the experimental designs that will be used for locating the optimum conditions, we have to summarize this last collection of points (simulations using 6469 cases, and one-twentieth of a year as time increment) in an intermediate point that takes into account all combinations of factor levels associated with these points. This may be done by fitting a parabolic regression equation of the form 6.2 in which the (K-S) is a function of the parameters. Differentiating the fitted equation with respect to the independent variables, equating all the derivatives with zero and solving the resulting equations simultaneously, we obtain the stationary point 'S'. This point defines the centre of the system whose co-ordinates x_{1s}, x_{2s} and x_{3s} are the solution of the above system of differential equations.[7]

7. See section 6.2.

The points used in such analysis are included in table (6.3).

The quadratic equation fitted to these points is:

$$Y = 0.713 + 6.482x_1 + 1.709x_2 - 4.529x_3 - 12.471x_1^2 + 3.329x_2^2 - 1.765x_3^2$$
$$- 8.457x_1x_2 + 17.396x_1x_3 - 1.942x_2x_3 \ldots\ldots (6.8)$$

Differentiating (6.8) with respect to x_1, x_2 and x_3 and equating the derivatives to zero, we have a system of three simultaneous equations the solution of which is $x_{1s} = 0.322$, $x_{2s} = 0.209$ and $x_{3s} = 0.191$ and $Y_s = 0.0597$. This new point is the second lowest point ever obtained as far as the level of response is concerned (the lowest was Y = 0.0519).

Two alternative experiments:[8]

We now have two different points associated with two different combinations of variable levels. The first of these points is based on a single simulation while the second summarizes a group of fourteen points including this last one. Both of them scored the lowest values of response so far. Each of these points or a combination of them, along with the information gained in the preliminary testing of the model, can be used as a basis for a suitable experimental design to approximate a quadratic response surface. Accordingly, two alternative experiments are established. In the first, the factor levels are derived from those levels which provided the lowest of these points (0.0519), while in the second they are derived from the average levels obtained from both points. These derived levels are listed in table (6.4). The two experiments are set out in a frame of an experimental design of the type mentioned in section 6.2 i.e. a "composite design". The levels chosen for variables

8. An experiment is defined here as the application of the model through a number of simulations using different combinations of factor levels. The outcome of the experiment is the (K-S) value denoting the degree of goodness of fit. The experiment includes regression and canonical analyses of the fitted response surface.

Table(6.3)

Conditions and Output of Fourteen Exploratory Simulations using 6469 as a Sample Size and 1/20th of a Year as a Simulation Step

No.	x_1(A)	x_2(B)	x_3(C)	K-S*
1	0.250	0.065	0.010	0.1124
2	0.270	0.0	0.017	0.1238
3	0.250	0.0	0.007	0.1260
4	0.255	0.065	0.0	0.1136
5	0.250	0.065	0.017	0.1124
6	0.260	0.085	0.013	0.1057
7	0.270	0.077	0.019	0.1060
8	0.270	0.089	0.015	0.1066
9	0.280	0.081	0.011	0.0904
10	0.332	0.205	0.238	0.0748
11	0.329	0.206	0.229	0.0771
12	0.479	0.465	0.429	0.0929
13	0.387	0.018	0.116	0.1266
14	0.387	0.459	0.131	0.0519

* (K-S) denotes Kolmogorov-Smirnov statistic.

Table 6.4

Factor Levels for Two Alternative 2^3 Factorial Experiments

Factor	Factor Level -1	Factor Level +1	Base Level	Unit
Experiment (1)				
x_1	0.200	0.445	0.323	0.123
x_2	0.120	0.320	0.220	0.100
x_3	0.071	0.311	0.191	0.120
Experiment (2)				
x_1	0.305	0.405	0.355	0.050
x_2	0.184	0.484	0.334	0.150
x_3	0.111	0.211	0.161	0.050

Note: -1 corresponds to the lower level of a factor, and +1 corresponds to its upper level. The factor level associated with zero in the design is called the base-level. Finally, the unit is the change in the factor level corresponding to a change from zero to one in the design.

Table (6.5)

Values of $-\alpha$ and $+\alpha$ Required for Orthogonality in Experiments (1), (2)

	$-\alpha$	$+\alpha$
Experiment (1)		
x_1	0.174	0.471
x_2	0.099	0.342
x_3	0.045	0.339
Experiment (2)		
x_1	0.294	0.416
x_2	0.152	0.516
x_3	0.100	0.222

x_1 and x_3 in the first experiment were somewhat wider in range than the corresponding levels used in experiment (2). On the other hand, the range of x_2 was narrower in the first experiment than in the second. This was suggested by the fact that first order effects corresponding to x_1 and x_3 in regression equation (6.8) were large compared with their standard errors while that of x_2 was not as significant as the others. Therefore, a reduction in the step length for x_1 and x_3 contrasted with an increase in this step for x_2 was maintained in experiment (2). It must be emphasised, however, that there is no strictly satisfactory way of choosing the variable levels on an a priori basis and that the choice depends solely on the intuitive judgement of the investigator.

To obtain an orthogonal composite design, we take $\alpha_1 = \alpha_2 = \alpha_3 = 1.215$ the value given by Box et. al. for a three factor design (Box et. al., 1954; p.534). Seven extra points, six of them associated with this value of α and one at the centre that does not depend at all on α were then determined and added for each of the above experiments. The following is an example of the calculation of these points. For x_1 in experiment (1), the lower and upper levels are 0.200 and 0.445. The base level (the level corresponding to 0) and the unit (the change from 0 to 1) are 0.323 and 0.1225. $-\alpha_1$ and α_1 are finally computed as:

$$-\alpha_1 = 0.323 - (0.1225 \times 1.215) = 0.174$$

$$\alpha_1 = 0.323 + (0.1225 \times 1.215) = 0.471.$$

These two values along with the base levels of x_2 and x_3 constitute the points $(-\alpha_1, 0, 0)$ and $(+\alpha_1, 0, 0)$ respectively. The values of α corresponding to x_1, x_2 and x_3 as computed for experiment (1) and experiment (2) are shown in table 6.5.

After defining the variable levels and the extra points to be employed in the simulation, experiment (1) was first carried out followed by experiment (2). The two experiments were performed in terms of the natural values of the variables. The value of response associated with each set of factor combinations was recorded. A polynomial of second degree, in the form of equation (6.2) was fitted for each of the two experiments using the method of Least Squares. Following the procedures of section 6.2, each of these fitted equations was reduced to its canonical form to enable the recognition and analysis of the response surface represented by it. The results are analysed in the next paragraphs.

Response surfaces corresponding to experiments (1) and (2):

The combinations of factor levels and the corresponding response values for both experiment (1) and (2) are shown in table 6.6 (table 6.6 contains all the simulation results of which the first 30 trials constitute experiments (1) and (2)). The regression and canonical analyses of these two experiments are contained in table 6.7. Both tables reveal the following facts. First, the ranges of factor levels considered in experiment (2) proved to be more relevant than those utilized in experiment (1). This is evidenced by the fact that the former levels were associated with responses (K-S values) lower than those associated with the latter. No improvement has been made in the response level as a result of experiment (1). For all the fifteen points used in this experiment produced a level of response above .0519, the lowest level yet attained. In fact, nine of these points resulted in a response above 0.1, a level that is almost double that of .0519. On the other hand, no matter how slight the improvement in the response was, the points included in experiment (2) were generally associated

Table (6.6)

Conditions and Output of Simulations Using 6469 Sample Size and 1/20th of a Year Simulation Step

Trial	Input			Output*			
	x_1(A)	x_2(B)	x_3(C)	$\bar{x}$	σ	$\sigma/\bar{x}$	K-S**
1	0.445	0.320	0.311	1.066	1.855	1.740	0.0880
2	0.445	0.320	0.071	1.215	2.220	1.827	0.1079
3	0.445	0.120	0.311	1.196	2.216	1.853	0.1014
4	0.445	0.120	0.071	1.779	3.189	1.793	0.1589
5	0.200	0.320	0.311	0.314	0.810	2.579	0.1700
6	0.200	0.320	0.071	0.336	0.860	2.559	0.1636
7	0.200	0.120	0.311	0.350	0.874	2.498	0.1574
8	0.200	0.120	0.071	0.387	1.024	2.647	0.1685
9	0.174	0.220	0.191	0.291	0.778	2.674	0.1779
10	0.471	0.220	0.191	1.332	2.388	1.793	0.1221
11	0.323	0.099	0.191	0.772	1.654	2.654	0.0750
12	0.323	0.342	0.191	0.669	1.386	2.071	0.0672
13	0.323	0.220	0.045	0.789	1.680	2.131	0.0742
14	0.323	0.220	0.337	0.625	1.275	2.041	0.0739
15	0.323	0.220	0.191	0.716	1.524	2.129	0.0740
16	0.405	0.484	0.211	0.859	1.597	1.859	0.0471
17	0.405	0.484	0.111	0.861	1.632	1.895	0.0513
18	0.405	0.184	0.211	1.029	1.959	1.901	0.0753
19	0.405	0.184	0.111	1.228	2.280	1.857	0.1040
20	0.305	0.484	0.111	0.580	1.210	2.088	0.0812
21	0.305	0.484	0.211	0.536	1.153	2.150	0.0999
22	0.305	0.184	0.211	0.676	1.438	2.127	0.0770
23	0.305	0.184	0.111	0.691	1.526	2.208	0.0927
24	0.294	0.334	0.161	0.592	1.258	2.126	0.0878
25	0.416	0.334	0.161	0.978	1.859	1.902	0.0671

Table (6.6)

(continued)

Trial	Input			Output*			
	x_1(A)	x_2(B)	x_3(C)	$\bar{x}$	σ	σ/$\bar{x}$	K-S**
26	0.355	0.152	0.161	0.894	1.818	2.034	0.0526
27	0.355	0.516	0.161	0.714	1.382	1.935	0.0444
28	0.355	0.334	0.100	0.772	1.592	2.063	0.0651
29	0.355	0.334	0.222	0.721	1.458	2.023	0.0592
30	0.355	0.334	0.161	0.770	1.529	1.985	0.0502
31	0.454	0.128	0.367	1.169	2.143	1.833	0.0972
32	0.380	0.498	0.143	0.828	1.566	1.892	0.0441
33	0.395	0.697	0.149	0.733	1.381	1.885	0.0386
34	0.405	0.696	0.143	0.765	1.412	1.845	0.0337
35	0.381	0.672	0.113	0.747	1.396	1.870	0.0329
36	0.441	1.066	0.087	0.746	1.330	1.781	0.0317
37	0.414	1.093	0.092	0.753	1.280	1.701	0.0397
38	0.513	1.882	0.004	0.750	1.176	1.568	0.0461
39	0.506	2.038	0.0001	0.765	1.156	1.512	0.0536
40	0.508	2.035	0.001	0.701	1.123	1.602	0.0280
41	0.502	2.000	0.002	0.732	1.137	1.553	0.0390
42	0.506	2.154	0.00001	0.739	1.116	1.511	0.0453
43	0.493	1.959	0.009	0.703	1.115	1.587	0.0331
44	0.506	2.154	0.0001	0.695	1.105	1.589	0.0295
45	0.403	0.801	0.135	0.793	1.407	1.773	0.0437
46	0.403	0.803	0.135	0.790	1.421	1.800	0.0383
47	0.414	0.744	0.036	0.846	1.503	1.777	0.0509
48	0.402	0.747	0.130	0.777	1.386	1.784	0.0385
49	0.403	0.764	0.128	0.733	1.367	1.866	0.0329
50	0.405	0.781	0.126	0.742	1.352	1.821	0.0281
51	0.396	0.669	0.138	0.784	1.434	1.829	0.0366
52	0.397	0.689	0.136	0.744	1.414	1.899	0.0373

Table (6.6)
(continued)

Trial	Input			Output*			
	x_1(A)	x_2(B)	x_3(C)	$\bar{x}$	σ	σ/$\bar{x}$	K-S**
53	0.399	0.709	0.134	0.731	1.356	1.855	0.0311
54	0.400	0.728	0.132	0.772	1.418	1.837	0.0343
55	0.414	0.894	0.114	0.787	1.359	1.728	0.0450
56	0.416	0.909	0.113	0.725	1.334	1.840	0.0286
57	0.417	0.923	0.111	0.743	1.320	1.776	0.0315
58	0.418	0.938	0.110	0.764	1.326	1.736	0.0390
59	0.419	0.952	0.108	0.772	1.347	1.745	0.0416
60	0.406	0.799	0.124	0.744	1.369	1.840	0.0279
61	0.408	0.815	0.123	0.732	1.338	1.828	0.0275
62	0.409	0.832	0.121	0.761	1.362	1.791	0.0377
63	0.410	0.848	0.119	0.771	1.385	1.796	0.0373
64	0.420	0.966	0.107	0.773	1.338	1.731	0.0394
65	0.422	0.979	0.105	0.718	1.314	1.829	0.0250
66	0.412	0.863	0.118	0.791	1.382	1.748	0.0428
67	0.413	0.879	0.116	0.728	1.330	1.828	0.0260
68	0.400	1.032	0.010	0.684	1.234	1.803	0.0221
69	0.400	1.032	0.002	0.723	1.290	1.785	0.0238
70	0.399	1.046	$.47 \times 10^{-8}$	0.715	1.269	1.775	0.0252
71	0.398	1.084	$-.32 \times 10^{-8}$	0.702	1.251	1.781	0.0261
72	0.400	1.131	$-.49 \times 10^{-8}$	0.655	1.212	1.851	0.0363
73	0.401	1.177	-.067	0.670	1.196	1.784	0.0246
74	0.402	1.221	-.083	0.693	1.217	1.757	0.0247
75	0.391	1.311	-.135	0.683	1.189	1.742	0.0210
76	0.403	1.309	-.138	0.675	1.187	1.759	0.0170
77	0.404	1.265	-.099	0.635	1.214	1.772	0.0193

* The actual statistics as calculated from the sample are:
$\bar{x}_A = 0.736$, $\sigma_A = 1.471$, and $\frac{\sigma A}{x_A} = 1.999$

** K-S denotes Kolmogorov-Smirnov statistic.

Table (6.7)

Regression and Canonical Analyses of the Fitted Surfaces for Four Consecutive Experiments

	Experiment (1) n = 15	Experiment (2) n = 15	Experiment (3) n = 44	Experiment (4) n = 59
Coefficients of Quadratic Equation:*				
b_o	0.549	1.010	0.966	0.977
b_1	-2.476(±.210)	-5.021(±.564)	-5.249(±.581)	-5.305(±.536)
b_2	-0.188(±.227)	0.327(±.095)	0.246(±.074)	0.238(±.068)
b_3	-0.195(±.155)	-0.994(±.331)	0.285(±.196)	0.278(±.190)
b_{11}	3.959(±.303)	7.950(±.777)	8.035(±.980)	8.116(±.907)
b_{22}	0.579(±.453)	0.019(±.087)	0.082(±.021)	0.083(±.020)
b_{33}	0.536(±.314)	3.838(±.777)	0.500(±.244)	0.512(±.239)
b_{12}	-0.735(±.273)	-1.538(±.192)	-1.133(±.243)	-1.126(±.239)
b_{13}	-0.617(±.228)	-1.795(±.576)	-1.773(±.648)	-1.789(±.236)
b_{23}	0.574(±.279)	0.982(±.192)	0.414(±.183)	0.447(±.143)
Regression Analysis:				
R^2	0.983	0.985	0.908	0.901
F	31.241	35.940	38.537	50.699
S.e.	0.009	0.004	0.007	0.007
Canonical Analysis:				
B_{11}	0.270	-0.097	-0.009	-0.017
B_{22}	0.773	3.678	0.446	0.466

Table (6.7)
(continued)

	Experiment (1) n = 15	Experiment (2) n = 15	Experiment (3) n = 44	Experiment (4) n = 59
B_{33}	4.031	8.226	8.181	8.262
Y_s	0.059	0.055	0.064	0.055
x_{1s}	0.357	0.344	0.338	0.361
x_{2s}	0.268	0.050	-0.936	-0.249
x_{3s}	0.243	0.204	0.700	0.469
Transformation Coefficients:				
m_{11}	-0.007	0.083	0.030	0.026
m_{12}	-0.685	0.991	0.943	0.936
m_{13}	0.729	-0.105	-0.331	-0.351
m_{21}	-0.149	0.219	-0.134	-0.134
m_{22}	-0.720	0.084	-0.324	-0.345
m_{23}	-0.678	0.972	-0.937	-0.929
m_{31}	0.989	-0.972	0.991	0.991
m_{32}	-0.113	0.104	-0.072	-0.071
m_{33}	-0.097	0.210	-0.116	-0.116

with lower levels of response than those of experiment (1). Only one point out of fifteen in experiment (2) resulted in a response above 0.1 while four points in the same experiment had responses below 0.0519. The lowest response of these points is 0.0444(trial 27).

Second, the explanatory power of the fitted quadratic regression equations in both experiments is fairly high. It is higher, however, in experiment (2) relative to experiment (1). This is indicated by the big values of the proportion of explained variance (R^2) and the F-statistic, the measures of an overall goodness of fit. Moreover, most of the regression coefficients in the two experiments are large compared with the corresponding standard errors of estimates. This implies that all the coefficients, with only few exceptions, are significantly different from zero. Also, the standard deviation of the experimental error (the standard error of the equation) is fairly low in both experiments. It is smaller, however, in the second experiment than in the first.

Finally, the canonical analysis of the corresponding response surfaces reveals the following characteristics. The surface fitted in experiment (1) indicates that the stationary point Y_s (0.059) whose coordinates are: $x_{1s} = 0.357$, $x_{2s} = 0.268$ and $x_{3s} = 0.243$ is a point minimum. The fact that we had positive values for all the canonical coefficients B_{11}, B_{22} and B_{33} implies that there will be an almost certain increase in the response (K-S) in moving away from the centre of the design in the direction of any of the variables. However, the response at this point is still too much higher than the target value "0.0282" and one would have to concentrate the analysis on response surface fitted by experiment (2). For such a surface, the following is remarked. Since not all the canonical coefficients are of the same sign, the centre of the system Y_s can be considered neither a point minimum nor a point maximum. Rather, it indicates a minimax (a saddle-point or a col) at which the response is a minimum for one direction and a maximum for the other. This may suggest the existence of two alternative regions of minimum response, with two different peaks on both sides. The main issue in this case is to discover the right direction to climb out of this minimax. The fact that B_{11} is negative would imply that there will be a further reduction in the response by moving away from the centre in either direction along the X_1-axis. The smallness of this coefficient relative to the other coefficients points out to the attenuation of the fitted response surface along this axis. The most advisable means of climbing out of this col is to follow the line of steepest descent along the X_1-axis by calculating new points on this line in either direction. These further points may be added to the previous points to constitute a new experiment to determine the conditions which produce the minimum response on this surface. The procedure to calculate points on the steepest descent line is as follows. The X axes which have positive coefficients, i.e. X_2 and X_3 are put equal to zero, while the X_1 -axis which has the negative coefficient is assumed to have a predetermined value. The equations defining these three axes are solved simultaneously to yield values for the x's. This predetermined value of X_1 is computed as follows. Putting X_2 and X_3 equal to zero in the canonical equation

$$Y - Y_s = B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2, \text{ gives}$$

$$X_1 = \sqrt{\frac{(Y - Y_s)}{B_{11}}}$$

Assuming some value for Y so that $(Y-Y_s)$ in the LHS of the canonical form will correspond to the desired level of response measured from the centre of the design,[9] X_1 will be the square root of the quotient $\frac{Y - Y_s}{B_{11}}$. Substituting this value of X_1 and zero for both X_2 and X_3, inserting the values of m_{ij} and x_{is}, i, j = 1, 3 in equation (6.7), and solving these equations simultaneously, should produce new values for x_1, x_2 and x_3. These values, in turn, can be utilized in a new set of simulations whose results along with the previous results may define a new surface.

The path of steepest descent was calculated, following the above explained procedure, first on the negative side of X_1. Although the points calculated on this side were designed to give values of response approaching zero they can hardly be accepted as genuine points. For the value of x_2, which corresponds to the delay time between two subsequent moves, came always negative. Therefore, the negative direction of the steepest descent line was abandoned. On the other hand, points on the positive side of this line, calculated by the same technique, were acceptable. They are considered in experiment (3).

Experiment (3):

Since our aim at this stage of analysis was to gain as much information as possible about the response surface in the neighbourhood of the present experimental design, a large collection of the above calculated new points (24 points) was chosen and a number of simulations corresponding to these points were carried out. The results of these simulations, as far as the

---

9. According to the target set out in this phase of model testing, the value of $Y - Y_s$ must be less than 0.0282.

Kolmogorov-Smirnov statistic is concerned, were recorded. They are listed, alongside the previous points, in table 6.6 and denoted as trials 31-54. However, ten of the whole set of points (54) for which the response was above 0.1 were excluded to avoid any possible distortion of the shape of the surface when regions far from minimum are included. The remaining 44 points were used to approximate a new response surface, following exactly the same techniques as experiments (1) and (2). The regression and canonical analyses of the fitted quadratic equation for experiment (3) are contained in table 6.7.[10] Inspection of the results of such an experiment leads to the following conclusions.

First, a marked improvement in the level of response (K-S) has been achieved. Not only that the value of the statistic measuring the model goodness of fit dropped from 0.0444, the lowest level so far, but also the fall below this level was sustained in most of the twenty four new points. For example, nineteen out of these points had responses below 0.0444, and three of these nineteen scored a response below 0.03. In fact we could have stopped at this stage of analysis since two of these last points (trials 40, 50) resulted in responses equal to 0.0280 and 0.0281 which are apparently below the target value of 0.0282. But as with any simulation experiment, it is possible that these two points could constitute purely random occurrences and hence further experimentation is necessary

---

10. An intermediate experiment combining the points of experiments (1) and (2) (30 points) was carried out to estimate a new response surface. The fitted surface, however, was much similar to the one fitted in experiment (1). All the canonical coefficients were positive, suggesting that the stationary point Y_s (.048) is a point minimum. Also, all the coefficients had the same magnitudes ($B_{11} = 0.041$, $B_{22} = 0.781$ and $B_{33} = 4.044$) with B_{11} much smaller than in experiment (1). The co-ordinates of Y_s were $x_{1s} = 0.357$, $x_{2s} = 0.268$ and $x_{3s} = 0.243$ respectively.

to establish the validity of our findings. Moreover, having shown that the differences between the sample and the simulation are not significant at a probability level equal to 0.01 there is a renewed hope of showing the insignificance of these differences at other levels of probability, i.e. at $p = 0.05$ or $p = 0.10$. Therefore, it would be useful if we could extend the analysis to a few more confirmatory experiments to consolidate the findings of experiment (3).

Second, it is becoming obvious now that the values of model parameters that raised the model power of predictivity, as evidenced by lower values of K-S statistic, have completely different magnitudes from what was estimated earlier, either from the analytical work or from the preliminary testing procedures. There is not much change however, in the value of the parameter "A", which denotes the general level of the propensity to migrate; it is some quantity around 0.4. On the other hand, enormous changes characterized the levels of parameters B and C, which denote the time elapsed immediately after a move and before any subsequent move, and the factor used to identify the mode of decline in the propensity to migrate with increasing duration in the same place of residence. The change in the two parameters took opposite directions. "B" tends to be quite a large amount of time; between one and two years before any new move could occur while "C" tends to be small ranging from .08 to .14. The implications of such values are that a repeated migrant, as estimated from the Scottish experience, usually allows for one to two years to elapse after his last move before he thinks of any further move, and the likelihood of future movement is inversely related to the length of stay in the last place of residence (the smallness of the 'c' parameter confirms this inverse relationship). Yet there is little to be said at this stage about the ultimate levels of these parameters and the analysis to follow may or may not agree with what has been concluded so far.

Third, regarding the quadratic representation of the response surface, we may notice the following. The assumption of the existence of appreciable second order and interaction effects compared with linear effects is confirmed. The estimates in general are relatively large compared with the standard errors, implying that they are significantly different from zero. The regression equation accounts for a large and significant part of the variation in the data. The high values of R^2 and the F-statistic confirm the high significance of the quadratic equation. There is also a large similarity between the fitted regression equations in experiments (2) and (3) as far as the signs and magnitudes of the coefficients are concerned. We may notice, however, that b_3, the regression coefficient corresponding to x_3, has changed sign in the last experiment. The implication of this is that this variable has been moved away from its base level.

Finally, the new fitted response surface is much analogous to the surface fitted in experiment (2). This is much explained by the corresponding canonical coefficients. B_{11} is negative and has the smallest magnitude. It is, however, smaller in this experiment than it was in experiment (2) signifying a smaller possibility of reducing further the value of the response. On the other hand, B_{22} and B_{33} are both positive, which with negative B_{11} suggests that we still have a saddle-point surface whose contours are elongated along the X_1-axis. As has been mentioned earlier, this may indicate the existence of two regions of minimum response with a two peak system. In this case, the issue will be to find out the right direction to climb out of this minimax towards the minimum. This is again possible by moving away from the centre of the design along the X_1 -axis in either direction, following the path of the steepest descent which is further investigated in experiment (4). It may also be noticed that one of the co-ordinates of Y_s the stationary point, that is x_{2s}, has an absurd negative sign. Of course a time variable cannot assume negative values. It is hoped that such absurdity does not appear in further experimentation.

Experiment (4):

We now turn to the calculation of the steepest descent line as predicted from the surface fitted in experiment (3). New points on this line are extrapolated following the same techniques explained in the transfer from experiment (2) to experiment (3). Again, the points calculated on the negative side of the X_1-axis may be rejected on the basis of negative values of x_2. On the other hand, fifteen further points were chosen from the set of points calculated on the positive side of the axis. These were used in a new set of simulations using the same conditions of previous experiments (i.e. 6469 cases and one-twentieth of a year as simulation step). The input as well as the output of these simulations are contained in table 6.6 and are denoted as trials 45-69. These results were added to the results of the previous 44 trials, and a new quadratic regression equation based on a total of 59 observations was estimated. The fitted equation was reduced, as usual, to its canonical form for recognizing the new fitted response surface. The particulars of this surface (the regression and canonical analyses) are shown in table 6.7 under the label of experiment (4). Examination of results in this table indicates the following.

Moving away from the centre of the design established in experiment (3) in the positive direction of X_1 did contribute to a further improvement, though slight, in the value of response. The value of the (K-S) statistic dropped from 0.028 the lowest level attained so far, to 0.022, a value that implies insignificant differences between the simulation sample and the actual sample at the five percent level.[11] Moreover, four out of the fifteen new points were below 0.028 and twelve out of fifteen resulted in response levels below 0.04. It may also be concluded that these extra fifteen points emphasize the earlier assessment about the levels of model parameters which was concluded in experiment (3). That is, a value of the parameter "A" around = 0.4, a large value of "B" mounting to over a year as a delay time and a diminishing value of "C" indicating that the behaviour of the probability of moving in the future is determined reciprocally by the length of duration in the last place of residence. The fitted quadratic regression equation looks much the same as the one fitted in experiment (3), as far as the signs and magnitudes of the coefficients are concerned. There are also appreciable second order and interaction effects as compared with first order effects. Most of the coefficients are statistically significant. High values for R^2 and the F-statistic were again observed assuring an overall goodness of fit.

As for the fitted response surface, there is little to be added about the shape of the surface in this experiment as opposed to the surface

---

11. See table 6.1.

fitted in experiment (3). The canonical coefficients have similar signs and magnitudes in both experiments. B_{11} again is negative, and both B_{22} and B_{33} are positive indicating a saddle-point surface. B_{11} is the smallest of the coefficients suggesting the elongation of the surface contours on the X_1-axis. The fact that B_{11} is negative means that there is still a possibility of reducing the level of response further. The method of steepest descent can still be useful. The negative side of the X_1-axis was explored and points containing negative values for x_2 occurred as anticipated. The positive side was alternatively considered but a new source of inconsistency was identified. That is, the points calculated on this side of the axis contained negative values for x_3 which corresponds to the parameter "C" that defines the mode of decline in the probability of future migration. It is not easy to interpret such negative values of "C" although it is of less absurdity than the case of having "B" negative. However, being small in magnitude, the parameter "C" seems to be far less important, than the parameter "B", if it was necessary to include it in the model.

Despite the negative values of "C", a curious investigator might hope that these rather strange values can contribute to the efforts aiming at reducing the response function to a lower level. Accordingly, eight of these new points were utilized in a new set of simulations whose results are shown in the bottom of table 6.6. Surprisingly, three of these points (trials 75-77) resulted in lower responses than 0.022, the lowest response yet obtained. The drop in the response in these three points constituted respectively 9.1, 17.3 and 23.7 percent below that level. The response in trial 76 (K-S= 0.017) indicate non-significant differences between the two sets of data (observed and expected) at all levels of probability considered in table 6.1.

Once again, one could add these results to the previous results and fit a new response surface from which further analysis could be performed. But, given the cost and time constraints imposed on this work, it is impossible to continue with this sequential testing procedure indefinitely.

Nevertheless, the results of the experimentation process pursued in this chapter seem to fully justify our faith in the value of the suggested model as evidenced by the dramatic fall in the response from 0.08, as obtained from the preliminary analysis in chapter five, to the present level of 0.017. The indications are clearly that continued experimentation along these lines with the incorporation of more disaggregated variables such as age, sex, occupation,.... etc., could eventually produce a much lower level of response and consequently add more confirmation to the underlying premises of the model.

6.4: The choice of optimum conditions:

Having already proved that the model in general succeeds in representing the actual patterns of movement as experienced by Scottish migrants, we ought to conclude the levels of parameters which are more likely to contribute to such a successful representation. In other words, the purpose here is to discover the set of conditions, or at least the region where these conditions exist, that achieve the best fit of the model to the observed data. This can be accomplished in various possible ways. One can simply choose the set of parameters which produced the lowest value of the Kolmogorov-Smirnov statistic, the goodness of fit criterion, as the optimum parameters. Although there is a certain logic in this choice it has the serious deficiency that chance might have played a great role in producing such a value of the statistic. This is due to the fact that a good deal of random variability is inherited by nature in the simulation as a validating technique.[12] A second possible way of choosing the optimum conditions (which is consistent with the method of sequential experimentation developed in this chapter) is to fit a new response surface combining all the points together and take the co-ordinates of Y_s, the centre of the design, as the set of desired optimum conditions. The advantage of this procedure is that the resultant set of conditions will be intermediate ones that take into account all possible conditions considered. Moreover, the fitted surface (with a bit of luck) may lead to a recognizable point minimum whose co-ordinates can be taken for granted to represent the required optimals. There are, however, disadvantages to this method. The most serious of these was identified in the previous section. That is, the indefinite continuation in this sequential experimentation is impossible while time and cost factors constitute considerable constraints. The other deficiency is that a further repetition of the response surface fitting technique may not easily lead to a satisfactory result. This has already been felt from the results of experiments (3) and (4) where we had unacceptable values for one or two of the co-ordinates of the stationary point. The least disadvantageous way of

---

12. As an evidence to such random variability, one of the good points (trial 68) was simulated a number of 20 times using 20 different initial random integers for generating the required sequence of real random numbers. For cost and computer time considerations, a random sample of only 300 cases were used in these simulations. The value of K-S enjoyed a wide range of variability. The minimum and maximum values for this point were 0.033 and 0.110 with average of 0.073 and standard deviation = 0.023. Applying the same procedure to other sets of points led to similar conclusions.

defining optimum values for the parameters is to take arithmetic averages of some of the good points, good in the sense that they produce a satisfactory level for the criterion function (K-S). It is a fairly simple procedure. It avoids the danger of relying on only one set of points which may have come accidentally, and in the meantime takes into consideration all different points included. However, two major defects are encountered in this method. The first is that there is no definite criterion for averaging one particular set of points rather than another. This may be overcome by considering all the points which achieved the model significance at the lowest level permitted (in our case at $p = 0.01$). The other drawback of the averaging process is that it may produce averages that are far from the individual best point (the point which achieved the highest level of goodness of fit). Again, we can limit the seriousness of this effect by considering only the points which satisfy the minimum of the above requirements, i.e. the points which meet the $p = 0.01$ significance level. For all these considerations, the third of these alternative procedures is preferred.

Accordingly, the combination of variable levels associated with a value of K-S below 0.0282, namely the points corresponding to trials 40, 50, 60, 61, 65, 67-71 and 73-77, were considered. The average, standard deviation and standard error based on 15 observations were computed for the three parameters A, B and C and for the related output, i.e. the mean number of moves, the standard deviation, the variation coefficient and the Kolmogorov-Smirnov statistic. In addition, confidence limits for estimating population means for these variables associated with ninety-five and ninety-nine percent probabilities were established. The results are shown in table 6.8. The basic findings of this table are as follows.

First, the estimated average for the goodness of fit criterion (K-S) shows still insignificant differences between the simulated patterns of movement and the actual patterns at $p = 0.01$ probability level. We may notice however that among the points which produced this average are points which achieved higher levels of significance than did the average itself.

Second, while the model in general proved to fit the actual data, only two of the parameters (A and B) are significantly different from zero whereas the third parameter C is insignificant. The implication of this is that, repeated movement can best be described by a simple inverse relationship between the probability of moving in the future and the duration of time spent in the last area of residence. This relationship proves effective when we allow some delay time to have elapsed immediately after a move and before any subsequent move. However, for prediction purposes only we could use a value of "C", which was a small quantity approaching zero and within the confidence limits of the corresponding population mean (between -0.059 and 0.083 at $p = 0.01$).

Third, the levels of parameters which produce the best fit of the model seem to agree with what has been concluded earlier in this chapter. A fairly high value of "A", corresponding to the propensity to migrate for repeated migrants as measured from the experience of Scottish migrants, is recorded. The average level is 41.1 per cent ranging from 39 to 43.2 per cent with probability equal to 99%. As for the estimated average for the delay time "B", during which the occurrence of any further move for those who are frequently moving is lagged, it is about thirteen months.

That is, a repeated mover on the average is likely to retard his next move for a period of just over a year after his last move, allowing himself a chance to assess the success or failure of his previous move. From table 6.7 we can see that the value of this time lag ranges between 0.881 and 1.329 of a year.

Finally, the average number of moves as calculated from the model does not significantly differ from that produced from the actual sample at all significance levels. This is a further confirmation of the high predictivity power of the model in general. This average is 0.705 contrasted with 0.736, the actual figure.

Table (6.8)

Summary Statistics for the Optimum Levels of Parameters and the Corresponding Output

	$\bar{x}$	σ	S.e.	Confidence Limits 95%		Confidence Limits 99%	
				Lower Limit	Upper Limit	Lower Limit	Upper Limit
Parameter A	0.411**	0.027	0.007	0.396	0.426	0.390	0.432
Parameter B	1.105**	0.305	0.076	0.943	1.267	0.881	1.329
Parameter C	0.012	0.097	0.024	-0.039	0.063	-0.059	0.083
Mean no. of moves	0.705*	0.028	0.007	0.690	0.720	0.684	0.726
St. Dev. of moves	1.261*	0.074	0.018	1.223	1.299	1.208	1.314
Variation coefficient	1.788*	0.058	0.015	1.756	1.820	1.744	1.832
Kolmogorov-Smirnov statistic	0.025	0.003	0.001	0.023	0.027	0.022	0.028

** Significant at both 0.05 and 0.01 significance levels.

* For comparison with the corresponding actual figures, reference should be made to table 6.5.

Chapter (7)

IDENTIFICATION OF SEX AND AGE VARIABLES' EFFECTS ON THE CONSTRUCTION OF REPEATED MIGRATION MODELS

7.1: Introduction:

In chapter six, it has been concluded that the model suggested to describe repeated migration, as reported in Scotland between 1939-1964, showed a great deal of relevance. This was evidenced by the non-existence of significant differences between the distribution of movement as generated by the model and the corresponding actual distribution at a fairly high level of probability. It has been contemplated, however, that the power of model predictivity could have been improved had a further disaggregation of the relevant variables been considered.

Among the variables that seem to affect the relationship between the propensity to migrate in the future and the duration of residence since last move are sex and age variables. It is the aim of this chapter to assess such variables' effects through the application of the model to different sex and age subgroups. The purpose being to show whether the proposed theory fits all sets of groups (with same or varying degree of fitness) or whether it is more applicable to some sets than to the others. Moreover, for the cases that prove applicability, we hope to show what sorts of levels that the model parameters should assume and whether or not the estimated parameter levels for these particular groups differ significantly from the aggregate levels concluded for all groups combined together.

7.2: Fitting the model of chapter five to some selected sex and age categories:

The 6469 cases (the whole sample at risk) identified as having lived throughout the period 1945-1964, were first disaggregated by sex and four broad age groups (0-14, 15-44, 45-64 and 65+). Each of the above age groups was further disaggregated by sex, giving a total of fourteen age/sex subgroups. For each of these groups the actual probability distribution of movement and some related statistics (the mean and standard deviation of the number of moves made by the group) were computed. For measuring the goodness of fit of the model to the actual data for all the subclasses considered, it was quite impossible, for time and cost considerations, to apply the same search routines utilized in chapters five and six when experimentation was performed on the whole sample. Alternatively, the points concluded in chapter six (the conditions and output of experiments (1) through (4)) may constitute a useful basis for further analysis, thus avoiding the trouble of starting from scratch. The probability distributions of movement, already obtained from the application of the model to the whole sample using 87 different sets of parameter level combinations, were matched against the corresponding actual distributions. As many values of Kolmogorov-Smirnov statistic (goodness of fit measure) as these combinations were computed for each of the fourteen sex and age groups separately.

For a particular subgroup, only the cases where the value of the statistic is less than the theoretical value that corresponds to a probability of occurrence $p = 0.001$ (the lowest probability level in table (7.1)) are considered. They indicate that the differences between the simulated and actual distributions are only chance variations and the probability of having differences of greater magnitude than these theoretical values is $p = 0.001$. The values of A, B and C, the model parameters and the corresponding values of (K-S) statistic were arithmetically averaged over the number of cases for which the value of (K-S) was below that theoretical level. Table (7.2) contains these averages as well as the number of cases to which the averages were computed. Also shown in the same table are the appropriate levels of significance for the average value of (K-S) for the investigated age and sex groups. But before any further analysis can be made we ought to draw attention to the following. The evaluation of the appropriateness of the model to any of the studied subgroups is restricted to the above chosen ranges of values. For there may be some other ranges that contain sets of parameter combinations which reveal greater applicability of the model to all or some of the categories than do the selected combinations. Nevertheless, the following remarks are revealed by table (7.2).

Consider first the sex variable. Within the above specified ranges of values, there are sets of values which clearly prove the applicability of the model for both sexes. Such an applicability, however, is more obvious for males than for females. The level of significance for the average value of (K-S) in the former group is double that level in the latter group. As for the average values of the parameters, the differences between the two sexes are not so large. Nevertheless, males show a higher general level of probability of future movement (parameter A) and they tend to stay immobile after their last move for longer time (parameter B) than do females. Whereas the parameter (C) is of similar magnitude and of less importance than (A) or (B) for both groups.

If both sexes are combined while age variable is considered, the only age group that emerges relevant to the model is that of children under

Table (7.1)

Calculated values of D* in the Kolmogorov-Smirnov Two Sample Test: Large Samples, Two-Sided Test for Selected Age and Sex Groups**

	n_1 @	Significance Level					
		0.10	0.05	0.025	0.01	0.005	0.001
Males	3849	0.0247	0.0275	0.0300	0.0330	0.0350	0.0395
Females	2613	0.0284	0.0316	0.0344	0.0379	0.0402	0.0453
Males 0-14	1618	0.0339	0.0377	0.0411	0.0452	0.0480	0.0541
Females 0-14	1524	0.0347	0.0387	0.0421	0.0464	0.0492	0.0555
Total 0-14	3142	0.0265	0.0295	0.0321	0.0353	0.0375	0.0423
Males 15-44	1421	0.0358	0.0399	0.0434	0.0478	0.0507	0.0572
Females 15-44	845	0.0447	0.0498	0.0542	0.0597	0.0633	0.0714
Total 15-44	2266	0.0299	0.0333	0.0363	0.0399	0.0424	0.0478
Males 45-64	734	0.0476	0.0530	0.0577	0.0636	0.0675	0.0761
Females 45-64	168	0.0954	0.1063	0.1157	0.1274	0.1352	0.1524
Total 45-64	902	0.0433	0.0483	0.0525	0.0579	0.0614	0.0692
Males 65+	71	0.1456	0.1623	0.1766	0.1945	0.2064	0.2327
Females 65+	74	0.1426	0.1590	0.1730	0.1906	0.2023	0.2280
Total 65+	145	0.1024	0.1142	0.1243	0.1369	0.1453	0.1637

@ $n_2 = 6469$ for all the 14 cases.

* $D = \text{maximum} \left| S_{n1}(X) - S_{n2}(X) \right|$

** The entries of the table were calculated according to the formula

$D = C\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$ where C is 1.22, 1.36, 1.48, 1.63, 1.73 and 1.95 for 0.10, 0.05, 0.025, 0.01, 0.005, and 0.001 significance levels respectively.

Table (7.2)

Average Values of Model Parameters and Kolmogorov-Smirnov Statistic for Cases Satisfying the 0.001 Significance Level of Table (7.1)

Group	Number of cases	Model Parameters A	B	C	K-S	Corresponding level of significance*
Males	40	0.412	1.151	0.056	0.0320	0.014
Females	22	0.396	0.932	0.043	0.0393	0.007
All ages 0-14	5	0.195	0.220	0.191	0.0252	0.120
" " 15-44	-	-	-	-	-	-
" " 45-64	-	-	-	-	-	-
" " 65+	-	-	-	-	-	-
Males 0-14	7	0.202	0.200	0.144	0.0299	0.140
Females 0-14	7	0.202	0.200	0.144	0.0349	0.095
Males 15-44	-	-	-	-	-	-
Females 15-44	2	0.506	2.096	0.000	0.0594	0.011
Males 44-64	5	0.500	1.645	0.081	0.0649	0.008
Females 44-64	-	-	-	-	-	-
Males 65+	20	0.464	0.977	0.087	0.2140	0.003
Females 65+	-	-	-	-	-	-

* The values in this column were linearly interpolated from table 7.1 using a nomogram for interpolating statistical tables for other probability levels suggested by G.E.P. Box et al. in Statistical Methods in Research and Production, O.L. Davies, ed., (1958), pp.388-90.

fifteen years of age with a surprisingly high level of significance (0.120). The values of the parameters, however, differ significantly from those levels concluded for either sex of all ages together. This fact is well illustrated by comparison of the size of the parameter estimates for the total sample (taken as the averages of the figures of total males and total females) with those of the 0-14 age subgroup. For parameter A, the subgroup estimate was approximately half that of the total group, while for B the subgroup estimate was some twenty per cent of that for the total group. In contrast, the subgroup estimate for parameter C was nearly four times the size of the whole group.

When age and sex were examined simultaneously, the picture is apparently different. Both males and females in the youngest age group had values for the parameters that proved the model goodness of fit to the observed data. However, the fit in that specific age group, was appreciably closer for males than for females. This was evidenced by the higher level of significance for the former group than for the latter.

As far as the average levels of parameters are concerned, no significant differences between the two sexes were observed. The levels for either sex were of similar magnitudes to those for the total age group 0-14. For the rest of the age groups, adult females (15-44) and males of age groups 44-64 and 65+ appear to show some relevance to the model, with varying degrees of significance. The parameter levels for all these subgroups are within the expected ranges of levels for the total sample.

The above analysis, in broad terms, suggests that sex and age variables were very important variables to have affected the process of fitting the model of chapter five to the observed data. It appears that the age variables were of greater importance than the sex variable. It is quite evident now that the outcome of the fitting process, manipulated in chapters five and six, with the inevitable neglect of such variables in a preliminary testing phase, was more than satisfactory. It could eventually be improved with the inclusion of these and other relevant variables such as occupation, skill level, education, etc.

7.3: Response surfaces for sex and age groups:

In section 7.2, the importance of sex and age variables in formalizing a more comprehensive theory for describing the patterns of multiple movement has been identified. However, the extent of such an importance and the particular conditions that bring about such a comprehensive formalization are not fully recognized. In the present section, it is not aimed at concluding a final and complete identification of these conditions. A contribution to the achievement of this purpose is rather attempted by fitting response surfaces which describe the relationship between the target fuction (K-S) and the corresponding combinations of parameter levels for the sex and age groups investigated. The aim being to recognize, for each group, the nature of the fitted surface and consequently locate the conditions that yield the minimum response on this surface. If a true minimum is found, the adequacy of the proposed theory to describe the phenomenon under study for that particular group is well demonstrated.

The methods of chapter six are applied once more to the fitting of quadratic response surfaces for the above fourteen age and sex subgroups. The same 87 combinations of variable levels utilized in the analysis of section 7.2 are re-employed here to fit a quadratic regression equation of the form (6.3) for each of these groups. The resulting equations are then reduced to the standard canonical form to enable the recognition and analysis of the fitted surfaces. The coefficients of the regression equation and the regression and canonical analyses are shown in table (7.3) for all the defined subcategories. Examination of the results in this table reveals the following points.

First, the quadratic regression mode seems to fit the observed data for almost all cases fairly well. The values of the F-statistic are highly significant, and the amounts of explained variances (R^2) are over 95 per cent of the total variances. The standard errors of the estimated regression coefficients are low relative to the estimates themselves, implying that the estimates are significantly different from zero. (Presumably there are some exceptions to these general remarks). In addition, the standard deviation of the experimental error (standard error of Y) is fairly low for all cases (0.009).

Second, the response surfaces, however, as shown by the canonical analysis, do not indicate that we are having a minimum response in most of the fourteen age and sex groups. Rather, they are more likely to show points of maximum. The only case which produced a true minimum is that of males of all ages grouped together. For this case, all the coefficients in the canonical equation are positive, indicating that there will be an almost certain increase in the response in moving away from the centre of the design in the direction of any of the variables. The value of response at the centre (although being negative is not acceptable in real terms) is at a fairly low level (-0.0196). For other cases, females of all ages, total 0-14 age group, males 0-14 and 45-64 age groups, and females 15-44 age group, the coefficient B_{11} is negative and the other two coefficients (B_{22} and B_{33}) are positive. This suggests that the function can be further minimized by moving away from the centre along the X_1 -axis, following the steepest descent line in either side of the axis. The corresponding values of the stationary point for these subclasses are low too, especially for the first and last of these classes.

For the rest of the cases, either two or all of the canonical coefficients are negative indicating saddle or rather more complicated shapes of surface. In the meantime, the values of the response function at the centre of the system are rather large except those of total 15-44 and 45-64 age groups (0.028 and 0.0703), and males in the 15-44 age group (0.063).

Third, we may notice huge differences between the two sexes for all ages together as regards the values of model parameters which produce the minimum. It may be thought that this finding is not in accordance with earlier findings that showed only slight differences between the two sex groups. But the fact that the quadratic analysis for females is not showing a true minimum as contrasted with that for males provides us with a fair justification of not taking such a finding as a conclusive result.

Finally, as a concluding remark, we could follow the usual procedure of chapter six to calculate the path of the steepest descent line for each of the above categories, add the calculated points to the already existing points, and fit and identify the nature of the new response surface which in turn enables us to recognize the appropriate steps to be followed next. Repetition of such a procedure, as demonstrated earlier, could eventually lead to the conditions that approximate the minimum value of response. This in turn would undoubtedly increase our knowledge about the applicability of the model to the particular age/sex subcategories under investigation. However, as we mentioned earlier, this task requires considerable amounts of time and cost which are beyond the resources at our disposal. Nevertheless, the analysis above reinforces the argument that sex and age variables are important factors in determining a significant and effective relationship between any future movement and the recent history of migration. It has paved the way to fully identifying the effects of these and all other relevant variables such as skill level, occupation,.... etc. Extending the research along these lines should ultimately produce a deeper understanding of the problem under consideration.

Table 7.3

Coefficients of Quadratic Equations and Regression and Canonical Analyses based on 87 Observations for Selected Sex and Age Groups

	Total	Total	Age Groups				0-14		15-44		45-64		65+	
	Males	Females	0-14	15-44	45-64	65+	Males	Females	Males	Females	Males	Females	Males	Female
Regression Coefficient*														
b_0	0.502	0.520	-0.093	0.474	0.495	0.649	-0.098	-0.081	0.487	0.450	0.441	0.703	0.578	0.717
b_1	-2.309	-2.722	0.647	-1.050	-1.055	-1.052	0.680	0.562	-1.052	-1.047	-1.035	-1.052	-1.052	-1.052
	(0.150)	(0.144)	(0.129)	(0.144)	(0.146)	(0.145)	(0.144)	(0.127)	(0.145)	(0.146)	(0.142)	(0.145)	(0.145)	(0.145
b_2	0.121	0.119	-0.010	-0.007	-0.013	-0.013	-0.011	0.002	-0.013	0.006	0.008	-0.013	-0.013	-0.013
	(0.037)	(0.035)	(0.032)	(0.035)	(0.036)	(0.036)	(0.033)	(0.031)	(0.036)	(0.036)	(0.035)	(0.036)	(0.036)	(0.036
b_3	-0.103	-0.099	0.003	0.016	0.019	0.020	-0.022	0.024	0.020	0.008	0.005	0.020	0.020	0.020
	(0.076)	(0.072)	(0.065)	(0.072)	(0.073)	(0.073)	(0.068)	(0.064)	(0.073)	(0.073)	(0.071)	(0.073)	(0.073)	(0.073
b_{11}	3.642	4.540	0.292	0.567	0.515	0.511	0.228	0.427	0.511	0.680	0.658	0.511	0.511	0.511
	(0.233)	(0.220)	(0.198)	(0.220)	(0.225)	(0.223)	(0.207)	(0.194)	(0.223)	(0.220)	(0.217)	(0.223)	(0.223)	(0.223
b_{22}	0.101	0.071	0.053	-0.005	-0.011	-0.011	0.050	0.054	-0.011	0.008	0.010	-0.011	-0.011	-0.011
	(0.013)	(0.013)	(0.011)	(0.013)	(0.013)	(0.013)	(0.012)	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013
b_{33}	0.567	0.502	0.085	-0.018	-0.068	-0.071	0.097	0.047	-0.071	0.089	0.080	-0.071	-0.071	-0.071
	(0.179)	(0.171)	(0.154)	(0.171)	(0.174)	(0.173)	(0.161)	(0.151)	(0.173)	(0.174)	(0.169)	(0.173)	(0.173)	(0.173
b_{12}	-0.876	-0.764	-0.319	0.038	0.089	0.090	-0.296	-0.360	0.090	-0.065	-0.075	0.090	0.090	0.090
	(0.123)	(0.117)	(0.106)	(0.118)	(0.119)	(0.119)	(0.110)	(0.104)	(0.119)	(0.119)	(0.116)	(0.119)	(0.119)	(0.119
b_{13}	-0.786	-0.768	-0.398	0.007	0.103	0.104	-0.341	-0.417	0.104	-0.187	-0.166	0.104	0.104	0.104
	(0.170)	(0.162)	(0.146)	(0.162)	(0.165)	(0.164)	(0.152)	(0.143)	(0.164)	(0.164)	(0.160)	(0.164)	(0.164)	(0.164
b_{23}	0.379	0.380	0.117	-0.011	-0.049	-0.051	0.105	0.128	-0.051	0.067	0.067	-0.051	-0.050	-0.051
	(0.084)	(0.080)	(0.072)	(0.080)	(0.081)	(0.081)	(0.075)	(0.071)	(0.081)	(0.081)	(0.079)	(0.081)	(0.081)	(0.081

Table 7.3

(continued)

	Total Males	Total Females	Age Groups 0-14	15-44	45-64	65+	0-14 Males	0-14 Females	15-44 Males	15-44 Females	45-64 Males	45-64 Females	65+ Males	65+ Females
Regression Analysis														
F	216.195	148.548	279.309	305.355	279.980	300.280	268.445	277.579	300.187	295.742	310.092	300.210	300.321	300.172
R^2	0.962	0.946	0.970	0.973	0.972	0.972	0.969	0.970	0.972	0.972	0.973	0.972	0.972	0.972
S.e.	0.009	0.009	0.008	0.009	0.009	0.009	0.008	0.008	0.009	0.009	0.009	0.009	0.009	0.009
Canonical Analysis														
x_{1s}	0.637	0.332	1.462	0.825	0.737	0.734	1.483	1.671	0.735	0.889	0.902	0.735	0.742	0.73
x_{2s}	3.113	0.046	3.198	3.061	4.286	4.383	3.865	4.834	4.362	0.896	0.203	4.356	4.177	4.35
x_{3s}	-0.508	0.336	1.203	-0.361	-0.858	0.884	0.618	0.603	-0.874	0.553	0.823	-0.872	-0.784	-0.87
Y_s	-0.020	0.054	0.365	0.028	0.070	0.225	0.378	0.402	0.063	-0.010	-0.024	0.279	0.152	0.29
B_{11}	0.009	-0.013	-0.057	-0.020	-0.084	-0.087	-0.055	-0.050	-0.087	-0.004	-0.003	-0.087	-0.087	-0.08
B_{22}	0.553	0.515	0.008	-0.003	-0.003	-0.003	0.020	-0.009	-0.003	0.085	0.079	-0.003	-0.003	-0.00
B_{33}	3.749	4.611	0.478	0.568	0.523	0.519	0.410	0.587	0.519	0.696	0.672	0.519	0.519	0.51

* The figures in parenthesis are the standard errors of the estimates.

Chapter (8)

SUMMARY AND CONCLUSIONS

A systematic random sample of 1 in 440 of the registered population drawn from the Scottish Central Register (S.C.R.) and constituting nearly 20,000 cases was used to study repeated migration in Scotland between 1939 and 1964. Earlier studies by Hollingsworth in 1968 and 1970 have utilized the same sample for analysing migration patterns in Scotland in the same period 1939-64. Such studies, however, did not cover the topic of repeated migration, which the present study intended to do.

The study had two main objectives. The first was to establish the existence of repeated migration phenomenon in Scotland, thus contributing to the existing knowledge of Scottish migration. The second objective was to suggest a simple practical model that describes efficiently the actual patterns of repeated migration as evidenced by the S.C.R. data for the period 1939-1964. Such a model could eventually be used as a general explanation of repeated migration phenomenon regardless of the particular country under investigation.

The statistics based on the S.C.R. are far from complete and may need correction before analysis. Only a partial correction of some of the entries of the sample was considered. However, the data set was still regarded as a valuable source of information on internal migration in general and on repeated migration in particular. It was particularly useful in permitting a longitudinal approach to the analysis of migration. A record of migration in a sufficiently long period of time, ranging between one and twenty-five years of the individual's life history, could be traced and analysed. This allowed the study of migration not only as a function of the individual's current characteristics, but also as a function of his past behaviour and changes in his characteristics. Specifically, the frequency of past migration and duration of residence variables were studied as latent factors in determining future migration.

As for the achievement of the first objective, the possible existence of repeated migration phenomenon in Scotland was evidenced by the fact that a migrant, on the average, made approximately 2.315 moves compared with an overall average of 1.038 moves between 1939 and 1964. This is further evidence to Goldstein's proposition that migration is characteristic of a rather small segment of the population (habitual movers) who move more frequently from place to place within a relatively small time interval than do the others.

When repeated migration was differentiated among different subgroups of the total sample, the following patterns emerged.

No significant differences were found between males and females, although females had a higher frequency of movement than males.

Single persons were more migratory than married persons. The differences between the two groups were more significant than those between the two sexes.

There is a high correlation between the recorded frequency of movement and changes in the marital status of the individual. The change from single to married accounted for the biggest proportion of the variation in the observed frequency.

The hypothesis that individuals in the early adult ages (before getting married and finally settling down) are more likely to make more than one move and in doing so the interval between successive moves tends to be short was, therefore, given support.

Age appeared to be the most important of all individual's characteristics in determining repeated migration behaviour. For instance an age differential was more significant than either sex or marital status differentials. Repeated migrants were found to be mostly of the young ages, where the highest frequency of movement was recorded by people in these ages.

These patterns, in general, strongly support the view that changes in the life cycle from family formation to family dissolution are among the major stimuli for migration.

The effects of some other variables on the frequency of movement were also examined. Duration of residence, the size of area of registration, and the length of time during which the individual was observed (exposure risk) were found to be of particular importance. Duration of residence was the most important of these variables.

The importance of past mobility experience in determining future mobility behaviour was also investigated. The probability of moving in the near future was obviously found a function of the number of moves already made in the past. It increased as the number of previous moves increased. The proposition that people who were more mobile in the past are more likely to move again in the future was then tested and validated. The effects of sex and age variables on the relationship between past and future mobility sequences were also examined. Once again, age differences were shown to be generally greater than sex differences, and thus have had greater influence upon future mobility behaviour as related to mobility experience in the past.

The above findings, as evident in the S.C.R. data, had considerable implications for the development of a general model of repeated migration. For it appeared that residential past mobility experience, either expressed in terms of the number of previous moves or the duration of residence since last move, strongly affected any future mobility. However, a more quantitative approach to the analysis of the data was needed to validate such findings. This was accomplished by the formalization of a precise, mathematical model that explains the patterns of repeated migration in Scotland, and in the meantime could be used to explain a more general type pattern.

The model suggested is of a stochastic nature. Stochastic models are those class of models which describe processes that develop with time according to probabilistic laws and whose future behaviour cannot be predicted with great certainty. The model cannot be considered, however, as a member of the class of models which treat migration as a simple Markov process. For the conditions of such a process as far as migration is concerned are not met in practice. Specifically, the constancy of migration probabilities over time and the irrelevance of migration knowledge prior to the period of estimation are unrealistic assumptions. The model was intended to relax these simple Markov process restrictions.

Two postulates constitute the basic assumptions of the model. The first implies that the propensity of an individual to migrate is not constant over time. Rather, it decreases as the individual continues to reside in the same place of residence thus signifying a linkage between previous residential mobility experience and future mobility prospects.

The second postulate suggests that the propensity to migrate does not build up immediately after a move has occurred. It is more likely that some delay time elapses during which the likelihood of any further move is negligible.

The model involves only three parameters providing a fairly high degree of generality. Only one of the parameters is enough to specify the simple Markov process of migration in which there is no dependence on the past history of migration. The three parameters are respectively defined as the general level of the probability of moving 'A', the delay time between a move and a subsequent move 'B', and some constant 'C' used to identify the mode of decline in the probability of moving in relation to the duration of residence since last move. If 'C' did not exist, the underlying relationship would simply assume a truncated reciprocal form.

In mathematical terms, the model was expressed as:

$$\lambda(t) = \begin{cases} \dfrac{A}{(t-t_0) + C} & \text{if } t > B + t_0. \\ 0 & \text{otherwise.} \end{cases}$$

where λ is the instantaneous probability of moving,

A, B, C are constants $>$ 0 defined above,

t is time, and

t_0 is time since last move.

Testing the validity of such a model and deriving the properties of interest through ordinary analytical techniques appeared impractical, for a good deal of mathematics were required. Experimentation via a computer simulation model was alternatively favoured as a useful technique in such a situation.

The basic criterion for model verification was set as follows. Given a distribution of individuals defined according to the number of moves they made in a given period of time (representing their actual experience of movement), how far could a similar distribution generated by the model conform with that actual distribution and how far could the differences between the two distributions be attributed to mere chance?

The Kolmogorov-Smirnov two sample statistic was selected as a measure for the goodness of fit of the suggested model to the actual data. It was preferred to the standard Chi-squared statistic for practical considerations, cited in the text. The main concern was to obtain a level of (K-S) statistic as low as to indicate non-significant differences between the simulated and actual distributions of movement. Related to this was to identify the values of the model parameters A, B and C which contributed to that level.

The process of fitting the model to the observed data was carried out in two stages. The first meant to explore, for each parameter, as large a range of values as possible in order to define the region of potential interest in which the desired optimum values can be located and a further investigation can be effected. In such a stage, 'crude' search procedures (in which only one parameter at a time was changed) were utilized. Owing to the lack of information at this stage, a necessary large number of simulations was required and experimentation, therefore, could only be based on small samples, chosen at random from the universe, in order to reduce the required amounts of computer time and cost. The result of the analysis at this stage was that the value of (K-S) statistic was brought down to markedly low levels (ranging between 0.044 and 0.083).

The implication of this was that the proposed model is showing a great deal of promise and more rigorous investigation is required to confirm it.

In the second stage of model testing, a more specific target was defined. That was, the attainment of a maximum absolute deviation between the cumulative probability distributions of movement for the simulated and the actual samples less than or equal to 0.0282, the theoretical value of K-S statistic at $p = 0.01$ significance level. Experimentation at this stage had to be based on the whole sample at risk (those who were registered before or at 1945 and living throughout the whole period, constituting 6469 cases of the original sample). This necessitated the use of more efficient methods that require a fairly small number of experiments in the hope of minimizing the time and the cost of the experimentation process.

The methods introduced for this purpose were methods of locating the optimum conditions operating upon a process, of some kind, which maximize or minimize specific features of it. They were originally developed by Box and associates for the study of chemical processes (Box and Wilson, 1951; Box et. al., 1953; Box, 1954). The following is a brief account of the techniques employed in this method.

First, the surface of a response function of the form $U = f(X_1, X_2 \ldots\ldots X_k)$, where U is some response that depends on the levels of k independent variables, is represented by a plane by carrying out some suitably arranged experiments in a subregion of the whole experimental region and calculating first order coefficients of that plane. The signs and magnitudes of these coefficients indicate the direction of steepest ascent (descent) up (down) the plane according to which the variables are changed proportionately in order to produce the greatest gain in the response. The purpose being to reach a near-stationary region, a region where the slopes of the surface flatten and become relatively small compared with the error of estimation.

Second, the response surface in the near-stationary region is represented by a polynomial of higher degree than the first. By studying the coefficients of the fitted polynomial, the nature of the surface can be fairly accurately determined and the minimum or maximum response can be easily located using a few more extra points.

Sequential experimentation is one of the characteristic features of the above method. If a true minimum (maximum) is found, a confirmatory experiment based on the point represented by the co-ordinates of this minimum is performed. The result of this experiment is added to the previous results and a new surface is defined, if necessary. If, on the other hand, a surface of another kind is determined (a saddle point or a minimax for example), additional experiments are performed along the path of greatest gain (steepest descent or ascent). The results of these experiments are similarly added to the previous results and a new shape of the response surface is recognized. The process is repeated until no further improvement is achieved.

For estimating the coefficients of first and second order effects, Box and associates suggested a 'composite design' as a suitable experimental arrangement. Such a design is built up from a complete 2^k factorial design (k is number of factors) from which first order and interaction effects are estimated. Further (2k + 1) points are added, one at the centre of the design and the remaining points are located in pairs along the co-ordinate X-axes at $\pm \alpha_1, \pm \alpha_2, \ldots \pm \alpha_k$, to estimate quadratic or higher order effects.

In order to facilitate analysis of the fitted polynomial, the convention of reducing the polynomial to the standard canonical form which has only few terms is employed.

The above explained methods were used to minimize the value of (K-S) statistic (the goodness of fit measure) at a level equal to or below 0.0282, and to determine the values of A, B and C, the model parameters, which yield this minimum. The conditions and output of the simulations carried out in the exploratory stage constituted the basis for further experimentation using these methods.

Four consecutive experiments comprising a total of 77 additional simulations of the whole sample at risk were performed. The net outcome of these experiments was that the value of (K-S) statistic appreciably dropped to 0.017, a value far below the target level 0.0282. Such a value implied that, with a fairly high level of probability, the differences between the simulation and actual samples, as far as the distribution of movement is concerned, are random variations and that the two samples are drawn from the same population.

However, none of the surfaces fitted in these experiments indicated a true minimum. The canonical coefficient corresponding to the parameter 'A' was always negative and small in magnitude compared with the other coefficients. This implied that the response surfaces all had a saddle point shape elongated along the X_1-axis. It also indicated the possibility of reducing the value of (K-S) further by moving away from the centre of the design in the direction of X_1-axis on either side following the path of steepest descent. Further analysis including the new calculated points had to be performed until we could establish a true minimum. This was virtually the usual procedure for the transfer from one experiment to another.

But as a result of time and cost constraints, it was impossible to continue with this sequential testing procedure indefinitely. Nevertheless, the results obtained appeared to fully justify our belief in the applicability of the suggested model to the actual data. That was evidenced by the dramatic fall in the goodness of fit measure from a value of 0.083 in the exploratory stage to a final value of 0.017.

For concluding the values of parameters which produced the model best fit, the points corresponding to a value of (K-S) statistic < 0.0282 were arithmetically averaged, for each parameter and for the output. This was practically preferred to considering either the parameter combination which produced the lowest value of the criterion function or that corresponding to the centre of the design of a new experiment which combines all the points together. The average value of (K-S) statistic is 0.025 in the range (0.022, 0.028), a value still exhibiting non-significant differences at $p = 0.01$ significance level. The parameters 'A' and 'B', denoting the general level of probability and the delay time respectively, were shown to be significantly different from zero, while the parameter 'C' was insignificant. This implied that the probability of further migration bears a disproportionate relationship to the length of time spent in the last place of residence, permitting a delay time to operate immediately after a move and before any subsequent move. However, the value of 'C' could simply be used for prediction purposes.

The value of 'A' ranged between 0.390 and 0.432 with an average of 0.411. That of 'B' ranged between 0.881 and 1.329 of a year with an average of nearly thirteen months. This indicated that a repeated mover is not likely to move again within a year or so of his last move. During such a time he is likely to make an assessment of the success or failure of his latest move before considering any further move. The parameter 'C' on the other hand assumed a relatively small value ranging from -0.059 to 0.083 with an average of 0.012. The simulated average number of moves using such sets of parameter values was 0.705 as compared with an actual figure of 0.736.

To assess the effect of disaggregating the data by variables of interest, sex and age for example, the model was fitted to some selected sex and age subgroups (fourteen). The points already established in the application of the model to the whole data set were the basis for such disaggregate fitting processes. However, the analysis did not go much deeper in this regard since most of the fitted surfaces did not indicate a true minimum and one would have to perform further experimentation along the steepest descent line for each of the fourteen sex and age subcategories which is beyond the limited resources of this research. Nevertheless, the analysis reinforced the earlier argument that sex and age variables as well as other relevant variables are important factors in formalizing a significant relationship between the probability of moving and the past history of movement. The implication is that the degree of applicability of the model to the actual data differed from one migrant subgroup to another.

Concluding remarks:

The analysis undertaken in this study has led to the following general conclusions.

(1) It is apparent, at least from the Scottish experience, that if internal migration increases in the future then it is likely that there would be an increased number of first movers, but at the same time repeated migration is likely to be a phenomenon of equal or even greater significance. Repeated migrants are going to move away again, not new migrants. They seem to account for the majority of the recorded moves at any particular point of time while the rest of population remain in their usual place of residence.

(2) The influence of duration of residence since last move on any subsequent move is overwhelming. The notion that the propensity to migrate is inversely related to the length of stay in the same place of residence is confirmed by the Scottish migration experience as exhibited from the S.C.R. data. People who migrated recently are more disposed to further migration than people who continued residing in one place for a longer stay. However, the relationship between the probability of moving and duration of residence needs to be formalized in a framework of a dynamic stochastic process; not a static one such as represented by a first order Markov chain. The simple, practical model utilized in this thesis appears to constitute a useful first step in this direction.

(3) Age also has proved to play a dominant role in the construction of models that describe repeated migration. It is more influential than other customary variables, sex or marital status for example. Because of age effects on repeated migration, which have been identified in the text, children who are currently moving (presumably accompanying the movement of an adult member of the family, often the parents) are going to act differently when they grow up with respect to mobility behaviour.

(4) The study in general gives credit to two of the recently elaborated techniques as particularly valuable devices in the study and analysis of complex phenomena such as migration. The first of these techniques is simulation. Specifically, simulation seemed to be the most suitable technique for analysing repeated migration, a phenomenon which has a great deal of random variability and which has been described by a model the verification of which required a high degree of mathematical sophistication. Without such a technique, it would have been much more difficult to go through the tedious experimentation process necessary for such a large sample. The other technique is that of the experimental attainment of optimum conditions which was originally developed by Box and associates for the study of chemical processes. The application of these methods to the problem of fitting and verifying the suggested model, where the measure of goodness of fit was treated as the desired response and the model parameters as independent variables that yield this response, was an evidence to their success and usefulness. Using a relatively small number of experimental points, we have been able to obtain a good deal of information about the model, and the characteristics and behaviour of its parameters.

Suggestions for further research:

The model suggested to describe repeated migration patterns in Scotland between 1939-64 has generally been shown to yield satisfactory results. It has also been suggested that the power of model predictability could have been improved had a further disaggregation of the data by sex, age, and other relevant variables been maintained. It would be more profitable if such variables were considered from the outset of the model construction process. Moreover, some constraints have been imposed on the sample chosen for experimentation such as the condition of being at risk (under observation) throughout the simulation period (20 years). Such a restriction was laid to guard against the possible random disturbances which might be caused by birth and death variables. Such variables surely have important influences on the behaviour of the system under investigation. It would be of interest if such a constraint had been relaxed, so that the new-born could be included and the dead be included only for the period they were alive but excluded afterwards.

Furthermore, there was a remarkable change in the registration system in Britain since 1952. The model was fitted to data concerning the whole time period, before and after 1951. It may be a useful exercise to investigate the applicability of the model to disaggregated periods of time in order to assess the impact (if any) of such a change in the registration system. One would expect, at least on an a priori basis, different magnitudes of the model parameters, particularly for the parameter 'A' which denotes the general level of the probability of moving.

Return migration (moving back to where the person migrated from) is one of the important features of the migration process. It may be considered as a special case of repeated migration. However, return migrants may have different characteristics from non-return migrants. The hypothesis of the homogeneity of movers has not been tested yet. It would be interesting to show how far the patterns of age and sex differentials, as well as other possible differentials, with respect to repeated migration vary significantly among different migration categories: non-movers, first movers, returners, and repeated migrants. Moreover, are the models constructed to describe repeated migration phenomenon (the model suggested in the thesis is an example of such models) suitable for describing return migration as well? or should return migration have its own models? Further research may hopefully provide answers to these questions.

Appendix

COMPUTER SIMULATION MODEL FOR REPEATED MIGRATION

(1) Objectives:

The problem investigated, the construction and testing of a probability model for repeated migration, is a typical situation in which simulation is necessary. A mathematical model has been formulated to describe the probability of moving as a function of previous duration of residence and some delay time. Preliminary testing of this model using analytical techniques seemed laborious and required a great deal of mathematics. A full testing of the model and a deep understanding of its properties may be effectively completed by simulation. The methodological considerations of a simulation experiment described in chapter five are all relevant. Some of them have been utilized already, the others are to be utilized in their proper order.

Our main objective of the simulation experiment described below is to test whether or not the proposed model is suitable for describing repeated migration phenomenon, as measured by S.C.R. data. If the patterns of migration obtained from the simulation conform with those observed patterns, the validity of the model is demonstrated. If, on the other hand, there is disagreement between the two sets of patterns, the model has to be adjusted or replaced by a more appropriate model that provides a better approximation to the phenomenon investigated.

A second major objective, though indistinguishable from the first, is to find out the most suitable range of values of the model parameters A, B and C which produce the best fit to the observed data. In addition, it might be quite interesting to investigate the characteristics of the model as well as the behaviour of the system represented by it under various circumstances or with different sets of conditions operating upon this system.

(2) Sample size:

Experimentation upon the model, using the whole data set (almost 20,000 cases) seems impractical. Rather, it may be more suitable, especially in the preliminary phase of model testing to experiment only upon a fraction of the whole sample. This helps to a large extent to reduce computer time and cost which in turn enables to handle a wider variety of testing measures. Once we have gained more insight into the model, we could proceed to generalise it using larger sets of data.

For the purpose of producing useful comparisons, different sample sizes have been tried. Specifically, samples of 30, 60, 100, 300 or 6469 (the whole sample at risk) cases were employed. In all but the last of these cases, the sample chosen was drawn at random from people who lived at least twenty years among the original sample. Those are the people who were first registered in or before 1945 and were reported alive throughout the whole period (1945-1964).[1]

(3) Initial conditions:

3.1: Simulation step, simulation period:

By simulation step we mean the time interval during which the probability of moving for a specific individual is computed and the

---

1. Birth and death are variables that affect population change. They are not however considered in the above migration model. The restriction that an individual should be at risk the whole simulation period (20 years) is to abstract from the influences of these variables and eliminate the random disturbances that might occur if they were included.

decision as to whether this individual has moved or not is taken. The simulation period on the other hand is the total number of simulation steps multiplied by the length of the step.

So far, the analysis of the probability of moving was carried out with reference to a period of one year length. While a full test of the model on the same basis is necessary, it is not of less importance that we test it with reference to periods of varying length. One should expect that the applicability of the model will be much affected by the choice of an appropriate length of the simulation step. Furthermore, one would expect that as the change (increment) in time becomes smaller the model becomes more and more workable.

To test these hypotheses, fractions of a year as well as a full year were alternatively employed as possible simulation step lengths. Such fractions are : a twentieth, a fifth and a half year.

The whole period of simulation is also critical. A very long period of time allows the effects of other uncontrollable variables to act upon the model and consequently causes simulation generated patterns to largely deviate from the observed patterns. On the other hand, a very short period of time is not sufficient to produce a conclusive assessment as far as the model validity is concerned. Of course, a period of an intermediate length is a compromise that helps to avoid deficiencies of either very long or very short periods. Practically, a period of twenty-year length is believed to be such a compromise for the present simulation exercise.

3.2: Initial duration of residence:

The basic variable in the above suggested model is the length of time during which an individual resided in the same place since he last moved.

Duration of residence in this case is calculated as the difference between the date at which the probability is measured t and the date of previous move. If a move was decided in the previous simulation step, the duration of residence will be equal to the length of the step. It will be increased by the same length each time a further move is not yet decided.

As hypothetical cases are considered in the simulation and since we have no knowledge about any previous migrations, an estimation of duration of residence prior to simulation is required. The sample data may be utilized to approximate such prior duration of residence. A frequency distribution of individuals according to their duration of residence since they first entered the S.C.R. until they moved to a new location can be established. Using such a distribution along with a random number generation procedure, an initial duration of residence can be assigned for each individual. Subsequent output generated from successful simulations can be utilized for further elaboration of this initial duration of residence distribution.

3.3: Initial values of model parameters:

Estimation of the levels at which the parameters should be ultimately set requires the provision of some initial values for these parameters. The more these initial values are close to the actual values the quicker we could reach the final optimal values and with the least effort. The analytical and empirical analysis conducted in the early stages of setting up the model has suggested 0.1593, 0.0650 and 0.0170 as approximate values for the parameters A, B and C respectively.[2] Such values were used as starting conditions for further simulation experimentation.

---

2. T.H. Hollingsworth and M.G. El-Rouby, 'Models for ......', Op.Cit.

(4) Operational forms of the model:

The basic step in the simulation process is to assign for each individual at each time point a certain probability of moving and to decide whether this particular individual has moved at this particular point of time or not. The model was originally defined in terms of the instantaneous probability of moving. For simulation purposes however, it would be more suitable if such a probability was defined within a specific time interval and not instantaneously. If λ is the instantaneous probability of moving and t is time, the probability of moving in the time interval $(t, t + \delta t)$ will be $\lambda \delta t$, where δt is a small quantity denoting the length of the time interval.

If we consider a larger time interval, say (t_1, t_2) the probability of moving will be equal to the product of probabilities corresponding to all possible time intervals of length δt within (t_1, t_2). An equivalent expression to this product is the integral:

$$p = \int_{t_1}^{t_2} \lambda dt, \text{ where}$$

p is the probability of moving between t_1 and t_2.

Recalling that λ was defined as

$$\lambda = \begin{cases} \dfrac{A}{(t-t_0) + C} & \text{if } t > t_0 + B \\ 0 & \text{otherwise;} \end{cases}$$

either of three situations may be met in practice. First, the delay time B entirely occurs before the beginning of the time interval during which the probability of moving is defined, i.e. $t_1 > t_0 + B$. In this case we have

$$p = \int_{t_1}^{t_2} \lambda \ dt.$$

$$= \int_{t_1}^{t_2} \frac{A}{(t - t_0) + C} \ dt.$$

$$= A.\text{Log}_e \ \frac{t_2 - t_0 + C}{t_1 - t_0 + C}.$$

Second, the delay time is still in operation at t_1 so that the probability of moving starts operating from $t = t_0 + B$. In this case, the expression will be:

$$p = \int_{t_0 + B}^{t_2} \lambda \quad dt.$$

$$= A.\text{Log}_e \ \frac{t_2 - t_0 + C}{B + C}$$

Finally, the delay time applies throughout the whole period of interest, i.e. there is not any real chance for the individual to move in (t_1, t_2). The probability of moving in this case will be zero anywhere.

For any of the above three cases and given specific values for A, B and C and other simulation parameters, the appropriate formula for calculating the probability of moving is applied. The probability for the individual in question is then defined and the decision as to whether or not he moved at this particular point of time is taken.

(5) <u>The simulation process</u>:

The simulation process itself develops as follows. First of all, the starting time point for simulation t_1 is defined. Since the range of values for the theoretical distribution of the duration of residence as estimated from the sample lies between zero and twenty years, we shall assume that the simulation starts after an experience of twenty years, i.e. t_1 will be equal to 20. We shall be simulating individuals for another twenty years ahead. One individual at a time will be considered. For each individual the following eight steps are performed.

1. An initial duration of residence 'd' is assigned by virtue of a generated random number in the range (0,1) along with the approximated probability distribution of residence duration.

2. Given that $t_1 = 20$ and $t_2 = t_1 + \delta t$, where δ is the time length used as simulation step, i.e. 1/20, 1/5, 1/2, or one full year; t_0, the time of last move is determined.

3. A new variable 'Z' is created as $Z = t_0 + B$, where B is the delay time during which no movement is initiated. Inspection of the value of Z would indicate which formula to apply for the calculation of the probability of moving. Such a probability is now determined.

4. A second random number in the range (0, 1) is generated. If the probability calculated in step 3 is less than or equal to that generated random number, the individual is considered to have moved at this particular time point; otherwise he would be counted in the same place of residence. If a move is decided, the record of movement for the simulated individual is augmented by one. Other particulars of the move such as date, age of individual, and duration of residence since last move could be recorded as well.[3] If, on the other hand, the person is reported to have not moved,

---

3. In a preliminary phase of model testing, we are only interested in counting the number of moves performed by the individual.

nothing is added to his movement records and he is progressed farther to a new time interval.

5. In case of a move in step 4 the values of t_0, t_1 and t_2 are changed as follows. A new value of t_0 is taken as the average of t_1 and t_2 as defined above. The average is taken because the exact time at which the move did occur is unknown. The new value of t_1 would be the old value of t_2 and the new value of t_2 would be the new value of t_1 plus the time increment δt. In case of no move, the new values of t_1 and t_2 will be defined in a similar fashion as in the case of a move while t_0 will be unchanged as it is the date of last move.

6. The steps from 3 to 5 are repeated as many times as the number of simulation steps in the whole simulation period, i.e. 400, 100, 40 or 20.

7. Considering a new individual, the steps from 1 to 6 are repeated a number of times equal to the number of individuals simulated.

8. Finally, summary statistics are produced as output from a particular simulation. Examples of such statistics are the frequency distribution of movement, the mean number of moves and standard deviation, the distribution of individuals according to their duration of residence in a specific place, i.e. between two consecutive moves, etc. Such statistics may be computed on a yearly basis or they may be aggregated to summarize the experience of individuals throughout a longer period of time. In addition, similar statistics summarizing the actual experience could be derived from the sample and some measures of goodness of fit between the two sets of data can be computed and printed out for analysis. The above steps are shown in a flow diagram, to be used with a computer programme, in figure (3).

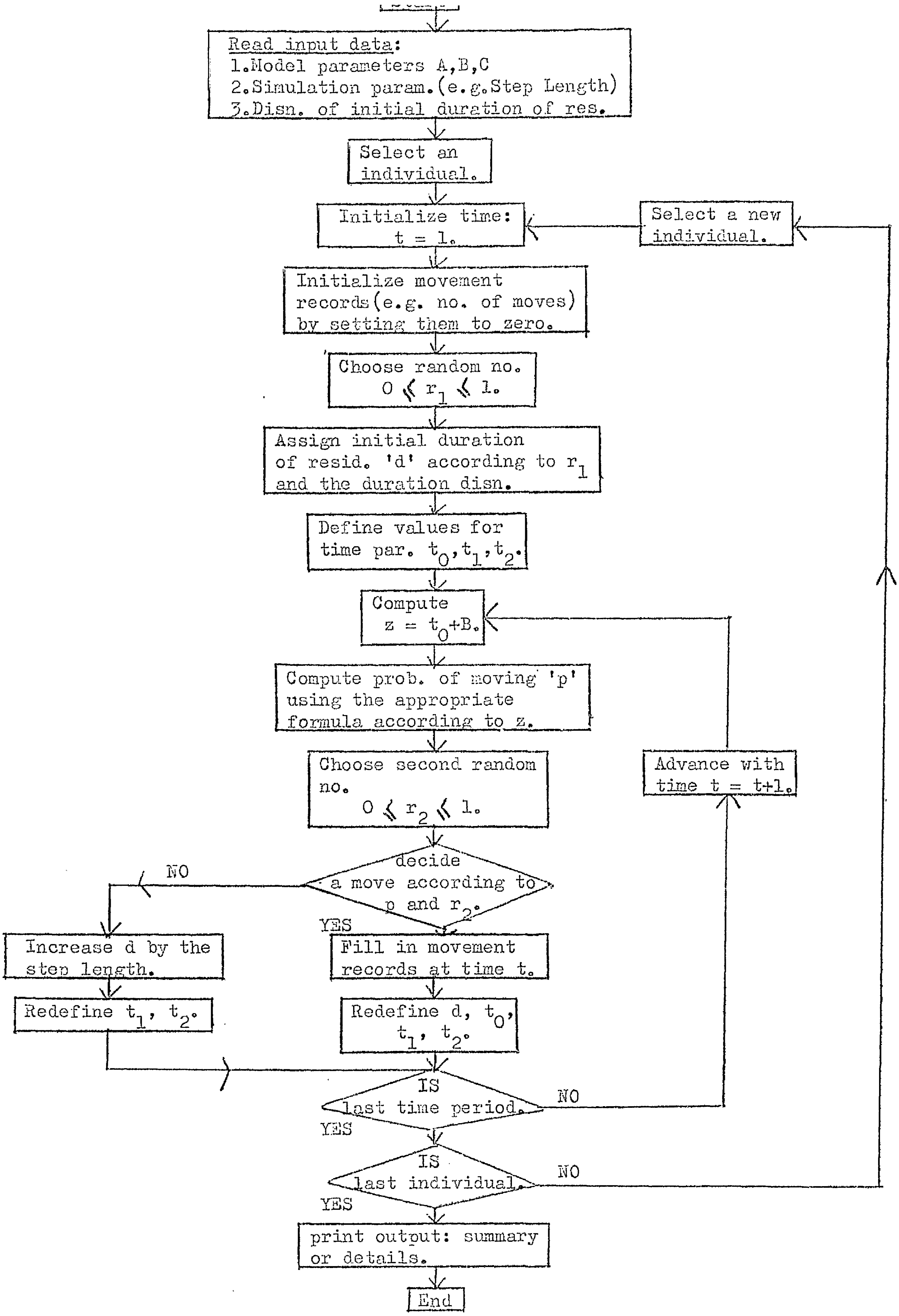

Fig.(3)
Flow diagram of a computer simulation model for repeated migration.

Table (A)

The Actual Distribution of Movement for a Sample of 6469 Cases and the Corresponding Distributions Generated from Successful Simulations of the Model

Number of Moves	Actual	Number of Cases			
		Simulation 68	Simulation 75	Simulation 76	Simulation 77
0	4209	4352	4319	4299	4334
1	1173	923	960	947	929
2	520	585	576	603	563
3	235	300	339	358	177
4	147	180	179	164	73
5	71	82	65	66	26
6	45	33	22	22	7
7	26	7	7	8	4
8	13	7	2	2	0
9	10	0	0	0	0
10	10	0	0	0	0
11	1	0	0	0	0
12	2	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	3	0	0	0	0
16+	4	0	0	0	0
Kolmogorov-Smirnov Statistic		0.0221	0.0170	0.0210	0.0193

Bibliography

1. Anderson, Theodore R. "Inter-Metropolitan Migration: a Comparison of the Hypotheses of Zipf and Stouffer," American Sociological Review, 20 (1955), pp.287-92.

2. Balan, Jorge et. al. "A Computerized Approach to the Processing and Analysis of Life Histories Obtained in Sample Surveys," Behavioural Science, 14 (March, 1969), pp.105-20.

3. Beaumont, P.B. "Assisted Labour Mobility Policy in Scotland 1973-4," Urban Studies, 13 (February, 1976), pp.75-9.

4. ___________. "The Problem of Return Migration Under a Policy of Assisted Labour Mobility: an Examination of some British Evidence," British Journal of Industrial Relations, 14, 1 (March, 1976), pp.82-8.

5. Beshers, James M. and Eleanor N. Nishiura, "A Theory of Internal Migration Differentials," Social Forces, 39 (March, 1961), pp.214-8.

6. Beshers, James M. "Computer Models of Social Processes: the case of Migration," Demography, 4, 2 (1967), pp.838-42.

7. Blumen, I., M. Kogan, and P.I. McCarthy, "The Industrial Mobility of Labor as a Probability Process," Cornell Studies of Industrial and Labor Relations, Vol.VI (The New York State School of Industrial and Labor Relations, Cornell University, Ithaca, New York, 1955).

8. Bogue, Donald J. "Internal Migration," The Study of Population: an Inventory and Appraisal, (Philip Hauser & Otis Dudley Duncan, eds., Chicago, 1959), pp.486-509.

9. Box, G.E.P. and K.B. Wilson, "On the Experimental Attainment of Optimum Conditions," Journal of the Royal Statistical Society, B, 13, 1 (1951), pp.1-45.

10. Box, G.E.P. et. al. "The Design and Analysis of Industrial Experiments", O.L. Davies, ed., (Oliver and Boyd, Edinburgh, 1953).

11. Box, G.E.P. "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," Biometrics, 10 (1954) pp.16-60.

12. Brown, L.A. and J.C. Belcher, "Residential Mobility of Physicians in Georgia," Rural Sociology, 31 (December, 1966), pp.439-48.

13. Browning, Harley L. and Waltraut Feindt, "Selectivity of Migrants to a Metropolis in a Developing Country," Demography, 6, 4 (November, 1969), pp.347-58.

14. Cebula, Richard J. and Richard K. Vedder, "A Note on Migration, Economic Opportunity, and the Quality of Life," Journal of Regional Science, 13, 2 (1973), pp.205-11.

15. Christian, William et.al. "Rural Migration and the Gravity Model," Rural Sociology, 31 (1966), pp.73-80.

16. De Cani, John S. "On the Construction of Stochastic Models of Population Growth and Migration," Journal of Regional Science, 3, 2 (1961), pp.1-13.

17. Dodd, Stuart C. "The Interactance Hypothesis: A Gravity Model Fitting Physical Masses and Human Groups," American Sociological Review, 15, 2 (April, 1950), pp.245-56.

18. Duncan, Otis Dudley and Robert W. Hodge, "Education and Occupational Mobility: a Regression Analysis," American Journal of Sociology, 58, 6 (May, 1963), pp.629-44.

19. Eldridge, Hope T. "Primary, Secondary and Return Migration in U.S.A., 1955-1960," Demography, 2 (1965), pp.444-55.

20. Elizaga, Juan C., "Internal Migration: an Overview," The International Migration Review, (1970), pp.121-46.

21. Fabricant, Ruth, "An Expectational Model of Migration," _Journal of Regional Science_, 10, 1 (1970), pp.13-24.

22. Feindt, Waltraut and Harly L. Browning, "Return Migration: Its Significance in an Industrial Metropolis and Agricultural Town in Mexico," _International Migration Review_, 6 (1972), pp.158-65.

23. Folger, John, "Some Aspects of Migration in the Tennessee Valley," _American Sociological Review_, 18 (1953), pp.253-60.

24. Friedlander, D. and R.J. Roshier, "A study of Internal Migration in England and Wales, Part I: Geographical Patterns of Internal Migration, 1851-1951," _Population Studies_, 19, 3 (March, 1966), pp.239-79.

25. ______________________, "A Study of Internal Migrants - Their Movements and Characteristics," _idem_, 20, 1 (July, 1966), pp.45-59.

26. Gallaway, Lowelle, "Industry Variations in Geographic Labour Mobility," _Journal of Human Resources_, 2 (1967), pp.461-74.

27. Galle, Omer R. and Karl E. Taeubler, "Metropolitan Migration and Intervening Opportunities," _American Sociological Review_, 31 (1966), pp.5-13.

28. Garrison, William L., "Toward Simulation Models of Urban Growth and Development," in _Proceedings of the IGU Symposium in Urban Geography_, Knut Norborg, ed., (Lund Studies in Geography, Ser. B, Human Biology, 1960), pp.91-108.

29. Geiser, Peter, "Some Differential Factors Affecting Population Movement: The Nubian Case," _Human Organization_, 25, 4 (Winter, 1966), pp.334-51.

30. George, P. "Types of Migration of the Population According to the Professional and Social Composition of Migrants," in Readings in the Sociology of Migration, Clifford J. Jansen, ed., (Pergamon Press, 1966), pp.39-47.

31. Gibbons, J.D. Nonparametric Statistical Inference, (McGraw-Hill Book Company, 1971).

32. Goldstein, Sidney, "Repeated Migration as a Factor in High Mobility Rates," American Sociological Review, 19 (1954), pp.536-41.

33. __________, "Migration and Occupational Mobility in Norristown, Pennsylvania," American Sociological Review, 20, 4 (August, 1955), pp.402-8.

34. __________, "The Extent of Repeated Migration: an Analysis Based on the Danish Population Register," Journal of the American Statistical Association, 59 (December, 1964), pp.1121-32.

35. Gould, W.T.S. and R.M. Prothero, "Space and Time in African Population Mobility," People on the Move, L.A. Kosinski and Prothero, eds., (Methuen & Co. Ltd., London, 1975), pp.39-49.

36. Greenwood, M.J. "The Determinants of Labor Migration in Egypt", Journal of Regional Science, 9, 2 (1969), pp.283-90.

37. __________, "Lagged Response in the Decision to Migrate," Journal of Regional Science, 10, 3 (1970), pp.375-84.

38. __________, "A Comparison of the Determinants of White and Non-White Interstate Migration," Demography, 8, 4 (November, 1971), pp.141-55.

39. Greenwood, M.J. and D. Sweetwood, "The Determinants of Migration Between Standard Metropolitan Statistical Areas," idem, 9 (1972), pp.665-81.

40. Greenwood, M.J. "Research on Internal Migration in the United States: a Survey," Journal of Economic Literature, 13, 2 (1975), pp.397-433.

41. Haenszel, William, "Concept, Measurement, and Data in Migration Analysis," Demography, 4, 1 (1967), pp.253-61.

42. Hamilton, Horace, "Practical and Mathematical Considerations in Formulation and Selection of Migration Rates," idem, 2 (1965), pp.429-43.

43. Hitt, Homer L. "The Role of Migration in Population Change Among the Aged," American Sociological Review, 19 (1954), pp.194-200.

44. Hodge, Robert W. "Occupational Mobility as a Probability Process," Demography, 3, 1 (1966), pp.19-34.

45. Hollingsworth, T.H. "Internal Migration Statistics from the Central Register for Scotland of the National Health Service," Journal of the Royal Statistical Society, ser. A, 131 (1968), pp.340-83.

46. ___________, "Migration: a Study Based on Scottish Experience Between 1939 and 1964," (Oliver and Boyd, Edinburgh, January, 1970).

47. Hunter, Laurence C. and Graham L. Reid, Urban Worker Mobility, (O.E.C.D., 1968).

48. Illsley, R., A. Finlayson, and B. Thompson, "The Motivation and Characteristics of Internal Migration: a Socio-Economic Study of Young Migrants in Scotland," Milbank Memorial Fund Quarterly, 41, 2 (April, 1963), pp.115-44 and 3 (July, 1963), pp.217-48.

49. Isard, Walter and David F. Bramkall, "Gravity Potential, and Spatial Interaction Models," in Methods of Regional Analysis: an Introduction to Regional Science, by Walter Isard et. al., Chap.11 (The M.I.T. Press, 1960).

50. Isbell, Eleanor C. "Internal Migration and Intervening Opportunities," in Readings in the Sociology of Migration, C.J. Jansen, ed., (1966), pp.341-64.

51. Jansen, Clifford J. "Some Sociological Aspects of Migration," in Migration, Sociological Studies, 2, J.A. Jackson, ed., (University Press, Cambridge, 1969), pp.60-73.

52. Kosinski, L.A. and R.M. Prothero, "The Study of Migration," Introduction to People on the Move, L.A. Kosinski and R.M. Prothero, eds., (1975), pp.1-16.

53. Kulldorf, Gunnar, "Migration Probabilities," Lund Studies in Geography, Series B, Human Geography, 14 (Royal University of Lund, Lund, 1955).

54. Kyd, J.G. "Scottish Population Statistics, Including Webster's Analysis of Population, 1755." Publications of the Scottish History Society, Third Series, 43 (1951).

55. Laber, Gene and Richard X. Chase, "Interprovincial Migration in Canada as a Human Capital Decision," Journal of Political Economy, 79, 4 (1971), pp.795-804.

56. Ladinsky, Jack, (a) "Occupational Determinants of Geographic Mobility Among Professional Workers," American Sociological Review, 32 (1967), pp.253-64.

57. ___________, (b) "Sources of Geographic Mobility Among Professional Workers: a Multivariate Analysis," Demography, 4, 1 (1967), pp.293-309.

58. ___________, (c) "The Geographic Mobility of Professional and Technical Manpower," Journal of Human Resources, 2 (1967), pp.431-48.

59. Land, Kenneth C. "Duration of Residence and Prospective Migration: Further Evidence," Demography, 6, 2 (1969), pp.130-40.

60. Lansing, John B. and James N. Morgan, "The Effect of Geographic Mobility on Income," Journal of Human Resources, 2, (1967), pp.449-60.

61. Lee, Everett S. "Negro Intelligence and Selective Migration: a Philadelphia Test of the Klineberg Hypothesis," in Demographic Analysis: Selected Readings, J.J. Spengler, and O.D. Duncan, eds., (1963), pp.425-31.

62. ____________, "A Theory of Migration," in Migration, Sociological Studies, 2, J.A. Jackson, ed., (1969), pp.282-97.

63. Leslie, G.R. and A.H. Richardson, "Life Cycle, Career Pattern and Decision to Move," American Sociological Review, 26 (December, 1961), pp.894-902.

64. Long, Larry H. "On Measuring Geographic Mobility," Journal of the American Statistical Association, 65 (1970), pp.1195-203.

65. ____________, "Migration Differentials by Education and Occupation: Trends and Variations," Demography, 10, 2 (May, 1973), pp.243-58.

66. Lowry, Ira, Migration and Metropolitan Growth: Two Analytical Models, (Chandler, San Francisco, 1966).

67. McGinnis, Robert, "A Stochastic Model of Social Mobility," American Sociological Review, 33 (1968), pp.712-22.

68. Macisco, J.J., Jr., "Some Directions for Further Work on Internal Migration in Latin America," International Migration Review, 6 (1972), pp.216-22.

69. Mangalam, J.J. and Harry K. Schwarzweller, "General Theory in the Study of Migration: Current Needs and Difficulties," The International Migration Review, 3 (1968), pp.3-18.

70. Masnick, George, "Employment Status and Retrospective and Prospective Migration in the United States," Demography, 5, 1 (1968), pp.79-85.

71. Massey, Frank J., Jr. "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of the American Statistical Association," 46 (March, 1951), pp.68-78.

72. Merton, Robert K. "Patterns of Influence: Local and Cosmopolitan Influentials," in Social Theory and Social Structure, Chapter 10, by Robert K. Merton (The Free Press, Glencoe, 1957), pp.387-420.

73. Miller, Ann R. "Migration Differentials in Labor Force Participation: United States, 1960," Demography, 3, 1 (1966), pp.58-67.

74. Morril, Richard L. "Simulation of Central Place Patterns over Time," in Proceedings of the IGU Symposium in Urban Geography, Knut Norborg, ed., Lund Studies in Geography, Series B, Human Geography 24, (1960), pp.109-20.

75. ____________, "Migration and the Spread and Growth of Urban Settlement," Lund Studies in Geography, Series B, Human Geography, 26, (1965).

76. ____________, "Marriage, Migration, and the Mean Information Field: a Study in Uniqueness and Generality," Annals of the Association of American Geographers, 57, (1967), pp.401-22.

77. Morrison, Peter A. "Duration of Residence and Prospective Migration: the Evaluation of a Stochastic Model," Demography, 4, 2 (1967), pp.553-61.

78. ____________, "Chronic Movers and the Future Redistribution of Population: a Longitudinal Analysis," idem, 8, 1 (1971), pp.171-84.

79. Morrison, Peter A. "Population Movements and the Shape of Urban Growth: Implications for Public Policy," in Regional Policy: Readings in Theory and Applications, John Friedmann and William Alonso, eds., (1975), pp.221-43.

80. Myers, George C., Robert McGinnis, and George Masnick, "Preliminary Assessment of a Stochastic Process Model of Internal Migration," Abstract of Paper Presented at 1965 Annual Meeting of the Population Association of America, Population Index, 31 (July, 1965), pp.256-7.

81. ____________, "The Duration of Residence Approach to a Dynamic Stochastic Model of Internal Migration: a Test of the Axiom of Cumulative Inertia," Eugencis Quarterly, 14, 2 (June, 1967), pp.121-6.

82. Naylor, Thomas H., "Methodological Considerations in Simulating Social and Administrative Systems," in Simulation in Social and Administrative Science: Overviews and Case Examples, H. Guetzkow, P. Kotler, and R.L. Schultz, eds., (Prentice, Hall Inc., Englewood Cliffs, New Jersey, 1972), pp.647-72.

83. Naylor, Thomas H. et. al. Computer Simulation Techniques, (John Wiley & Sons, Inc., 1966).

84. Nelson, Philip, "Migration, Real Income, and Information," Journal of Regional Science, (Spring, 1959), pp.43-62.

85. Newton, Mary P. and James R. Jeffrey, "Internal Migration, Some Aspects of Population Movements Within England and Wales," Studies on Medical and Population Subjects, 5 (London, 1951).

86. Olsson, Gunnar, "Distance and Human Interaction: a Migration Study," Geographiska Annaler, 47, B (1965), pp.3-43.

87. Pack, Janet Rothenberg, "Determinants of Migration to Central Cities," Journal of Regional Science, 13, 2 (1973), pp.249-60.

88. Petersen, William, "A General Typology of Migration," American Sociological Review, 23 (1958), pp.256-66.

89. Pilhbland, C.T. and C.L. Gregory, "Selective Aspects of Migration Among Missouri High School Graduates," American Sociological Review, 19, (June, 1954), pp.314-24.

90. Prais, S.J., "Measuring Social Mobility," Journal of the Royal Statistical Society, Series A, 118, 1 (1955), pp.56-66.

91. ____________, "The Formal Theory of Social Mobility," Population Studies, 9, (July, 1955), pp.72-81.

92. Price, Daniel O. "A Mathematical Model of Migration Suitable for Simulation on an Electronic Computer," Proceedings of the International Population Conference, Wien, 1959, (Wien, International Union for the Scientific Study of Population, 1959), pp.665-73.

93. Ravenstein, E.G., "The Laws of Migration," Journal of the Royal Statistical Society, 48, 2 (June, 1885), pp.167-227.

94. ____________, "The Laws of Migration," idem, 52 (june, 1889), pp.241-301.

95. Read, D.R. "The Design of Chemical Experiments," Biometrics, 10 (1954), pp.1-15.

96. Rider, R.V. and G.F. Badger, "A Consideration of Issues Involved in Determining Migration Rates for Families," Human Biology, 15, 2 (may, 1943), pp.100-26.

97. Rogers, Andrei, "Estimating Internal Migration from Incomplete Data Using Model Multi-regional Life Tables," Demography, 10, 2 (1973), pp.227-87.

98. Rose, A.M. "Distance of Migration and Socio-Economic Status of Migrants," in Readings in the Sociology of Migration, Clifford J. Jansen, ed., (Pergamon Press, 1966), pp.85-91.

99. Rowntree, John A., "Internal Migration: a Study of the Frequency of Movement of Migrants," Studies on Medical and Population Subjects, 11 (London, 1957).

100. Sahota, Gian S. "An Economic Analysis of Internal Migration in Brazil," Journal of Political Economy, 76 (1968), pp.218-85.

101. Sanders, John, "The Depressed Area and Labor Mobility: the Eastern Kentucky Case," Journal of Human Resources, 4, 4 (1969), pp.437-50.

102. Schmidt, J.W. and R.E. Taylor, Simulation and Analysis of Industrial Systems, (Irwin Series in Quantitative Analysis for Business, R.D. Irwin Inc., January, 1970).

103. Schwartz, Aba, "Interpreting the Effect of Distance on Migration," Journal of Political Economy, 81, 5 (1973), pp.1153-69.

104. ____________, "On Efficiency of Migration," Journal of Human Resources, 6, 2 (Spring, 1971), pp.193-205.

105. Shryock, Henry, Jr. and Elizabeth A. Larmon, "Some Longitudinal Data on Internal Migration," Demography, 2 (1965), pp.579-92.

106. Shryock, Henry S., Jacob S. Siegel and Elizabeth A. Larmon, "Internal Migration and Short Distance Mobility," in The Methods and Materials of Demography, by Henry Shryock et. al., Vol.2, Chap. 21 (May, 1973), pp.616-72.

107. Siegel, Sidney, Non-parametric Statistics for the Behavioural Sciences, (McGraw-Hill Koga-Kusha Ltd., 1956).

108. Simmons, Alan B. and Ramiro Cardona G. "Rural Urban Migration: Who Comes? Who Stays? Who Returns? The Case of Bogota, Colombia, 1929-1968," International Migration Review, 6 (1972), pp.166-81.

109. Sjaastad, L.A. "The Costs and Returns of Human Migration," Journal of Political Economy, 70 (Supplement, October, 1962), pp.80-93.

110. Speare, A., Jr. "A Cost-Benefit Model of Rural to Urban Migration in Taiwan," Population Studies, 27 (1971), pp.117-31.

111. Stewart, Charles T., Jr. "Migration as a Function of Population and Distance," American Sociological Review, 25, 3 (June, 1960), pp.347-56.

112. Stinner, William F. and Gordon F. de Jong, "Southern Negro Migration: Social and Economic Components of an Ecological Model," Demography, 64 (1969), pp.455-71.

113. Stouffer, Samuel A. "Intervening Opportunities: A Theory Relating Mobility and Distance," American Sociological Review, 5, 6 (1940), pp.845-67.

114. __________, "Intervening Opportunities and Competing Migrants," Journal of Regional Science, 2, 1 (1960), pp.1-20.

115. Strodtbeck, Fred, "Population, Distance, and Migration from Kentucky," Sociometry, 13 (May, 1950), pp.123-30.

116. Suits, Daniel B. "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, 52 (1957), pp.548-51.

117. Taeuber, Karl E. "Duration of Residence Analysis of Internal Migration in the U.S.A.," Milbank Memorial Fund Quarterly, 39, 1 (January, 1961), pp.117-31.

118. Taeuber, Karl E., William Haenszel, and Monroe G. Sirken, "Residence Histories and Exposure Residences for the United States Population," Journal of the American Statistical Association, 56 (1961), pp.828-34.

119. Taeuber, Karl E. "Cohort Migration," Demography, 3, 2 (1966), pp.416-22.

120. Tarver, James D. and William R. Gurley, "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Divisions in the U.S.," Demography, 2 (1965), pp.134-9.

121. Tarver, James D. et. al., "Vector Representation of Migration Streams Among Selected State Economic Areas During 1955 to 1960," Demography, 4, 1 (1967), pp.1-18.

122. Tarver, James D. "A Test and Modification of Zipf's Hypothesis for Predicting Interstate Migration," Demography, 10, 2, (May, 1973), pp.259-75.

123. Taylor, R.C. "Migration and Motivation," in Migration, Sociological Studies 2, J.A. Jackson, ed., (University Press, Cambridge; 1969), pp.60-73.

124. Ter Heide, H. "Migration Models and Their Significance for Population Forecasts," Milbank Memorial Fund Quarterly Bulletin, 41, 1 (January, 1963), pp.56-76.

125. Thomas, Dorothy Swaine, "Selective Internal Migration: Some Implications for Mental Hygiene," in Demographic Analysis; Selected Readings, J.J. Spengler and O.D. Duncan, eds. (1963), pp.425-31.

126. Thomlinson, Ralph, "A Model for Migration Analysis," Journal of the American Statistical Association, 56 (1961), pp.675-86.

127. U.N. Methods of Measuring Internal Migration, Manuals of Estimating Population, Manual 6, Sales No. 70 x III.3 (1970).

128. Van Arsdol, Maurice D., Jr. "Metropolitan Growth and Environmental Hazards: an Illustrative Case Study, Ekistics, 21 (1966), pp.48-50.

129. Van Arsdol, Maurice D., Jr. et. al. "Retrospective and Subsequent Metropolitan Residential Mobility," Demography, 5, 1 (1968), pp.249-67.

130. Vanderkamp, John, "Migration Flows, their Determinants and the Effects of Return Migration," Journal of Political Economy, 79, 5 (1971), pp.1012-31.

131. Wadycki, Walter J. "A Note on Opportunity Costs and Migration Analysis," Annals of Regional Science, 8, 1 (February 1974), pp.85-108.

132. Walshaw, R.S. "The Time Lag in Migration Movements," American Sociological Review, 30 (1938), pp.278-87.

133. Whitney, Vincent H. and Charles M. Grigg, "Patterns of Mobility Among a Group of Families of College Students," American Sociological Review, 23 (1958), pp.643-52.

134. Willis, Kenneth G. Problems in Migration Analysis, (Saxon House, D.C. Heath Ltd. 1974).

135. Wolpert, J. "Migration as an Adjustment to an Environmental Stress," Journal of Social Issues, 22 (1966), pp.92-102.

136. Woofter, T.J. "Migration in the South East," Demography, 4, 2 (1967), pp.479-96.

137. Yoesting, Dean R., and Joe M. Bohlen, "A Longitudinal Study of Migration Expectations and Performances of Young Adults," Journal of Human Resources, 4, 4 (1969), pp.485-98.

138. Zachariah, K.C. "Bombay Migration: a Pilot Analysis of Migration to an Asian Metropolis," Demography, 3, 2 (1966), pp.378-92.

139. ____________, "Estimation of Return Migration from Place-of-Birth and Duration of Residence Data," Contributed Papers, IUSSP Sidney Conference, (1967), pp.615-22.

140. Zelinsky, Wilbur, "The Hypothesis of the Mobility Transition," The Geographical Review, 61 (1971), pp.219-49.

141. Zipf, George K. "The P_1P_2/D Hypothesis: On the Intercity Movement of Persons," American Sociological Review, 11 (1946), pp.677-86.