This is to certify that the

thesis entitled

Pre-Retirement Exit of Farm Operators:
Conceptual Framework and Empirical Study

presented by

Claude Falgon

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Agricultural Economics

[signature]
Major professor

Date Aug. 11, 1978

O-7639

PRE-RETIREMENT EXIT OF FARM OPERATORS:
CONCEPTUAL FRAMEWORK AND EMPIRICAL STUDY

By

Claude Falgon

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1978

ABSTRACT

PRE-RETIREMENT EXIT OF FARM OPERATORS: CONCEPTUAL FRAMEWORK AND EMPIRICAL STUDY

By

Claude Falgon

The number of farm operators has declined drastically in most industrialized countries during recent decades. This phenomenon has often been considered necessary to maintain or improve incomes of farm operators who remained in farming. The number of farm operators is the aggregate and cumulative result of individuals' entry and exit decisions. Causes affecting these decisions will, therefore, bear on farm demography. In order to have the capability of influencing the evolution of farm demography to fulfill some pre-determined social goals, one must know the factors which affect the entry into farming, the pre-retirement exit, and the retirement of farm operators.

Three main objectives were set for this study:

1. To provide a conceptual framework for the analysis of the decision to leave farming before retirement.

2. To appraise the ability of the Canadian Census of Agriculture Match to identify entrants and exiters, and to evaluate its usefulness as a major data source for analytical studies of entry and exit from farming.

3. To test as many hypotheses concerning factors affecting off-farm movement, as the available data would permit.

A review of literature on off-farm movement and migration summarizes theories which have been proposed to explain off-farm movement and migration, empirical studies conducted at either the macro-level or the micro-level, and policy recommendations which have been formulated.

A theoretical framework for the pre-retirement decision to leave farming is developed. The standard economic model of off-farm movement, based on a maximizing behavior, is critically reviewed. Elements for an expanded framework, which are drawn from the economic, psychological, and sociological literatures, are presented. Thus, farming is seen as a form of habitual behavior whereas pre-retirement decision to leave farming is considered as a genuine decision. Farm operators' behavior is looked upon as being of a satisficing nature. The decision to leave farming before retirement is considered as a group (family) decision, where family members often hold conflicting goals or aspirations. These goals are constantly adjusted upwards or downwards according to successes or failures in meeting past goals. Farm operators' decisions are based on reality as it is perceived. Perception of reality is both a screening process, by which some elements are known and some remain unknown, and the organizing of known elements into a meaningful whole. Attachments to the community and to farming as an occupation are important factors affecting pre-retirement decision to leave farming.

A newly available longitudinal data base, the Canadian Census of Agriculture Match, is presented and assessed in relation to its use for this study. Methodological considerations on inference and

data analysis are discussed and hypotheses to be confronted to empirical evidence are stated.

In preparation for the empirical study, a review of statistical methods for the analysis of qualitative dependent variables is performed. It covers the standard linear model, logit/probit-type models, discriminant analysis, measures of association in contingency tables and a multivariate log-linear model with exogenous variables.

The empirical study consists of an exploratory and confirmatory analysis of data drawn from the Census of Agriculture Match for the province of Saskatchewan. Pre-retirement exit of farm operators is shown to be positively related to age, value of land and buildings (expressed as a percentage of total capital value) and negatively related to residence on the farm. Inconclusive results are obtained concerning the hypothesized negative relationship of pre-retirement exit to total acreage and total capital value of the farm. Pre-retirement exit of farm operators is shown to be unrelated to involvement in off-farm work, degree of ownership of farm land, total sales of agricultural products, productivity of land and of capital, degree of mechanization, and distance to towns.

These results prompted a proposal for a longitudinal survey which would fruitfully continue this study.

ACKNOWLEDGMENTS

A number of people contributed to the development and fulfillment of this study. I express my appreciation to all.

Professor Manderscheid provided, despite the distance between us, the necessary guidance and encouragement during this study. Professor Obst, by remaining critical of my ideas, helped me in making them more precise.

Drs. W. V. Candler, K. J. McKenzie, D. McClatchy, J. Nash, G. T. MacAulay, and Mr. M. Mouelhi, all of Agriculture Canada at some time or another, contributed in different ways and at different stages to this study.

Statistics Canada and R. D. Bollman made available to me the data used in this study. Professors M. Nerlove and H. H. Stokes provided me with one of the computer programs used in the empirical analysis.

Finally, I would like to express my gratitude to Agriculture Canada, for granting educational leave to undertake graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

TABLE		Page

TAB

2

2

TABLE | Page

LIST OF FIGURES

of f

pla

Sas

Tab

Sc

fa

ar

13

13

13

(a

13

CHAPTER I

INTRODUCTION

Farm Demography in Canada in the Recent Decades

In Canada, as in most other industrialized countries, the number of farmers has rapidly decreased during recent decades. Table 1 displays the changes in the number of farm holdings in Canada and Saskatchewan from 1951 to 1976.

Table 1. Farm Holdings in Canada and Saskatchewan

Year	Canada	Saskatchewan
1951	617,722	111,586
1961	476,125	93,924
1966	428,794	85,686
1971	366,128	76,970
1976	338,578	70,958

Source: Census of Agriculture, Statistics Canada.

For Canada, the average annual rate of decrease in the number of farms was 2.3 percent between 1951 and 1961, 2.0 percent between 1961 and 1966, 2.9 percent between 1966 and 1971, and 1.5 percent between 1971 and 1976. For Saskatchewan, this rate was 1.6 percent between 1951 and 1961, 1.8 percent between 1961 and 1966, 2.0 percent between 1966 and 1971, and 1.6 between 1971 and 1976. The number of farms (and farmers) has been declining at an increasing rate from 1951 to 1971 and thereafter at a lower rate.

Dec

age di

farmer

vinces

In Sa

47.2

In Ca

in 19

Tabl

Age
Unde
25 -
35
45
55
60
65
70

So

Ta

Ag
U
2
3
4
5
6
6
7

S

Decreasing numbers of farmers have been accompanied by changing age distributions. Tables 2 and 3 display the age distributions of farmers from 1951 to 1976, for Saskatchewan and Canada. In most provinces the mean age of farmers had been increasing slightly until 1971. In Saskatchewan the mean age of farmers increased from 45.1 in 1951 to 47.2 in 1961, 47.4 in 1966, 48.6 in 1971, and decreased to 47.0 in 1976. In Canada the mean age of farmers increased from 46.9 in 1951 to 48.3 in 1961, 48.8 in 1966, 48.8 in 1971, and declined to 48.4 in 1976.

Table 2. Percentage of Farmers by Age Class, 1951 to 1971, Saskatchewan

Age Class	1951	1961	1966	1971	1976
Under 25	5.5	3.6	3.1	3.5	6.8
25 - 34	22.2	15.7	13.9	12.6	16.0
35 - 44	25.4	25.9	24.0	21.6	19.3
45 - 54	19.9	25.8	28.1	29.0	25.9
55 - 50	9.1				12.5
60 - 64		17.8	19.5	22.8	9.7
65 - 69	14.3				5.4
70 and over	3.6	11.2	10.4	10.5	4.4

Source: Census of Agriculture, Statistics Canada.

Table 3. Percentage of Farmers by Age Class, 1951 to 1971, Canada

Age Class	1951	1961	1966	1971	1976
Under 25	3.5	2.6	2.2	2.4	3.8
24 - 34	18.3	14.2	13.1	12.8	15.3
35 - 44	25.3	24.7	23.8	22.8	19.9
45 - 54	23.4	26.6	27.7	29.1	27.3
55 - 59	10.1				12.0
60 - 64		20.2	21.5	22.1	9.3
65 - 69	14.4				5.4
70 and over	5.0	11.7	11.7	10.8	5.0

Source: Census of Agriculture, Statistics Canada.

T

labor e

estima

farm w

nearly

farmer

farmer

commu

recen

healt

in di

most

major

to b

reso

sect

tran

deve

from

Wher

be c

The decrease in the number of farmers implies that the amount of labor expended in agriculture has decreased. This, however, underestimates the decrease in farm labor since, concurrently, the number of farm workers has declined similarly.

The decrease in the number of farms and farmers used to be seen nearly unanimously as a necessity by the main parties, including farmers' organizations. External effects of the decreasing number of farmers, mainly as increased cost of providing public services in rural communities with dwindling population came to be recognized only recently. Since then, it has been realized that the two objectives, healthy rural communities and an efficient agriculture, have become in direct conflict.

Low Income in Farming

Low rates of return on farm resources have been considered by most agricultural economists and in many different countries as the major "farm problem." Low rates of returns are generally considered to be the direct consequence of a farm resources disequilibrium: resources are not properly allocated between the farm and nonfarm sectors nor are they within the farm sector itself.

Low rates of return on farm resources (including farm labor) translate into low income for farm families. The emergence and wide development of multiple job-holding imply, however, that low income from farming does not necessarily mean low income for farm families. When low farm family income is studied, nonfarm sources of income must be considered.

Resources disequilibrium is thought to be generated by three fa

ta

p

T

i

c

c

factors: (1) economic growth,(2) inflation, and(3) output increasing technology.[1] Tweeten reviewed critically three theories which have purported to explain the persistence of this resource disequilibrium. These three theories, namely, the fixed resource theory, the decreasing cost theory, and the imperfect competition theory lead to contradictory conclusions as to the persistency and permanency of low rates of return on farm resources. The fixed resource theory implies that low rates of return are temporary. The increasing returns to size theory implies that low returns will last as long as a majority of farms operate under an economic size. The imperfect competition theory yields the conclusion that low rates of returns are permanent.

A minority of economists have argued that in fact there is no resources disequilibrium and no low rates of returns on farm resources: the opportunity cost of farm resources (especially labor) has been overestimated by simply equating it to the rates of returns on nonfarm resources. Perkins,[2] for example, maintains that low income in farming is a marginal phenomenon reflecting less the labor market disequilibrium than poor career decisions, lack of resources when starting farming, and preference for farm life. The income problem appears, then, to be one of low absolute income rather than low relative income.

The general viewpoint has been, however, that low returns prevail; in the short run, these can be alleviated by transfer payments

---

[1] See Luther G. Tweeten, "Theories Explaining the Persistence of Low Resource Returns in a Growing Farm Economy," American Journal of Agricultural Economics 51 (November 1969): 799-801.

[2] Brian B. Perkins, "Farm Income and Labour Mobility," American Journal of Agricultural Economics 55 (December 1973): 914-916.

or h

sive

econ

impli

opera

farme

and w

to ha

incom

Agric
Heady

Farm

Agricu
Heady

or high support prices but a long run solution must be found in massive transfer of labor out of agriculture. Some leading agricultural economists strongly expressed such a view:

> Since agriculture obviously has a surplus labor force, it would seem that returns on resources in agriculture in the long run can be best put on a part with those in other industries by maintaining a growing number of nonfarm employment opportunities and by reducing the total farm input and population in agriculture.[3]

> If farm people are to share in the fruits of technical progress and economic growth in the next decade or two, it appears that the rate of labor transfer from agriculture must be increased or the rate of technical advance must be decreased.[4]

> There is no satisfactory alternative to greater mobility of labor if agricultural incomes are to be increased relative to nonagricultural. Labor must be made more expensive by making it scarce.[5]

As mentioned previously, transfer of labor out of the farm sector implies, but is not equivalent to, a decrease in the number of farm operators. It was shown in the preceding section that the number of farmers decreased rapidly in recent decades, according to the forecasts and wishes of most economists. This phenomenon does not seem, however, to have alleviated the farm income problem.

Darcovich et al. have attempted to estimate the extent of low income in the farm sector in Canada, drawing on newly available income

---

[3]Earl O. Heady, "Adjusting the Labor Force to Agriculture," Agricultural Adjustment Problems in a Growing Economy, eds. Earl O. Heady et al. (Ames: Iowa State College Press, 1958), p. 145.

[4]Karl A. Fox, "Guiding Agricultural Adjustments," Journal of Farm Economics 34 (December 1957): 1099.

[5]D. Gale Johnson, "Labor Mobility and Agricultural Adjustment," Agricultural Adjustment Problems in a Growing Economy, eds. Earl O. Heady et al. (Ames: Iowa State College Press, 1958), p. 171.

tax data.[6] They found that 29 percent of farm taxfilers families in Canada in 1974 were in the low income category, with a high of 52 percent in Newfoundland and a low of 22 percent in Saskatchewan. Conceptual difficulties related to the definition of low income cut-offs and the questionable reliability of the data prohibit the consideration of these figures as definitive estimates. These findings support the hypothesis that low income in farming is a real problem, even after nonfarm income has been considered.

In conclusion, it must be said that despite its failure to alleviate low income in farming, massive transfer of labor out of agriculture is still considered as the solution provided one forgets the external effects on rural communities:

> This solution is still advanced as the most obvious one, even though the high exodus rates of the past apparently have not resulted in changing either the income distribution within agriculture or the relative income position of farm and nonfarm people.[7]

Programs Bearing on Farm Demography

In this section programs bearing in some way on farm demography are presented. Direct increase or decrease in the number of entries or exits may not be the main objectives of these programs but they all make reference to changing these flows.

---

[6]W. Darcovich, J. Gellner and Z. Piracha, "Estimates of Low Income in the Farm Sector, 1974," Working Paper, Agriculture Canada, November 1977.

[7]Dale E. Hathaway and Brian B. Perkins, "Farm Mobility, Migration and Income Distribution," American Journal of Agricultural Economics 50 (May 1968): 342.

Small Farm Development Program (S.F.D.P.)

The S.F.D.P. is a joint federal/provincial program, operated, with some variations, in all provinces of Canada. It consists of two major parts: first, a Land Transfer Program (L.T.P.) and, second, Rural Counselling and Farm Management Services.

The objectives of the L.T.P. is to enable small farm owners to purchase additional land and to financially assist those who want to enter nonfarm occupations or retire. In other words, the goals are to incite some farm operators to leave farming, either to retire or take up nonfarm occupations, thereby freeing land to enable other farms to reach an economic size. The L.T.P. is thus perfectly consistent with, first, the hypothesis that there is redundant labor in agriculture and, second, the decreasing cost theory according to which the majority of farms operate at a size below the optimal economic size. Under the L.T.P., grants are offered to farmers leaving farming, with certain conditions restraining the eligibility. Do these grants constitute a sufficient incentive? It seems this question was not considered at the outset of the program.

The objectives of the Rural Counselling Service are to help the individual farmer obtain information and do the analysis required to enable him to decide his future and to assist him in the adjustment to nonfarm employment or retirement. These activities constitute a lever of a different kind from the grant under the L.T.P. These counselling activities have, however, been very limited. It seems that the S.F.D.P. will put more emphasis in the future on counselling services as opposed to grants. An evaluation of the S.F.D.P. is in p

o

f

f

F
A

es

vi

in

fa

La

ne

se

Es
Fo

pu

in

/

Pr
Pr

progress and should shed some light on the respective effectiveness of grants and services.

The objective of the Farm Management Service is to help small farmers who remain in agriculture to develop a commercially viable farm business by providing them with farm management advice.

Farm Development Loan Board and Capital Assistance Program, Newfoundland

The objectives are to "provide low interest loans for the establishment or improvement of farms"[8] and to "increase productivity and efficiency throughout the agricultural sector by encouraging individuals to enter the farming profession and assist existing farmers in becoming more productive."[9]

Land Development Corporation, Prince Edward Island

The Land Development Corporation assists existing farmers and new farmers who want to acquire agricultural land, by buying and selling land and making land improvements.

Establishment of New Farmers--Interest Forgiveness, Nova Scotia

The objectives of this program are to "assist new farmers in purchasing and establishing a farm operation by eliminating the interest payable on borrowed capital over the first two years."[10]

---

[8]A. R. Jones, Policies and Programs for Agriculture--Atlantic Provinces, Publication No. 76/13 (Ottawa: Agriculture Canada, 1976), program Nfld.-37.

[9]Ibid., program Nfld-38.

[10]Ibid., program N.S.-41.

Farm Enlargement and Consolidation Program, Ontario

One of the objectives of this program is to "enable farmers wishing to retire, relocate or adjust out of agriculture the opportunity to do so."[11]

Junior Agriculturalist Program, Ontario

The objective is to "provide a practical learning experience during the summer for young people from nonfarm homes who are interested in agriculture."[12]

Land Lease Program, Manitoba

The objective is to "purchase farm land and lease it to eligible farm operators who do not have the necessary security to establish or maintain a viable farm unit."[13]

Farmstart, Saskatchewan

One of the objectives of Farmstart is to "assist farmers and potential farmers who face difficulties in developing economic farm units."[14]

---

[11] A. R. Jones, Policies and Programs for Agriculture--Ontario--Quebec, Publication No. 76/14 (Ottawa: Agriculture Canada, 1976), program 0-86.

[12] Ibid., program 0-89.

[13] A. R. Jones, Policies and Programs for Agriculture--Western Provinces, Publication No. 76/12 (Ottawa: Agriculture Canada, 1976), program M-63.

[14] Ibid., program S-51.

Green Certificate Program, Alberta

This is a joint federal/provincial program whose objective is to "provide a necessary on-farm and institutional training to young people choosing farming as a vocation."[15]

Beginning Farmer's Program, Alberta

A provincial program whose objective is to "assist young potential farmers to become established in farming operations"[16] through loans for land, machinery, and improvements.

Land Use and Farm Adjustment--ARDA, British Columbia

A partial objective is to assist in the withdrawal from agriculture of low agricultural capability land by buying land from farmers operating nonviable farms who are willing to sell.

Conclusion

From the brief foregoing review one can realize that objectives of these programs can be seen as contradictory: some programs tend to accelerate the rate of exit and others to encourage entry. This can only be reconciled if the clients of these programs differ in some relevant way (e.g., age, education, progressiveness, etc.). Reconciliation of these programs is certainly a task remaining to be done.

Definition of the Problem

Different theories purporting to explain low income in farming have been mentioned earlier in this chapter; they stress in different

---

[15]Ibid., program A-06.

[16]Ibid., program A-57.

ways the importance of redundant labor. The resource fixity and the imperfect competition theories identify causes which would influence the decision of a farm operator with respect to leaving farming and would induce him to stay. In doing so, these theories rely on the conventional model of individual behavior; thus, they look for "external" causes of the failure of farmers to adjust out of farming. It is postulated here, that the failure of farmers to leave farming as they are expected according to standard theory has "internal" causes, or more precisely, that the behavioral model commonly used in economic theory does not describe adequately the decision making process actually taking place.

Several programs aiming at modifying flows of farm operators into and out of agriculture were briefly presented. These programs have been developed in reaction to political concerns and were designed with little study of their possible effects which depend largely on how each individual farm operator reacts to the different stimuli. Without a clear knowledge of how the decision to enter farming or to leave farming is made, it is impossible to forecast with precision the effects of these programs. This research attempts to shed some light on one of these decisions, namely the decision to leave farming before retirement.

Objectives of the Study

The first objective of this study was to provide a conceptual framework for the analysis of the decision to leave farming before retirement. Because of the many theoretical difficulties, mainly due to the efforts in bringing together elements and results of different disciplines, no pretence is made to propose a well developed and consistent theory. The framework will provide, however, concepts to work with and will indicate directions in which further research would be valuable.

A second objective was to evaluate a new data base, the 1966-1971 Census of Agriculture Match, with respect to its ability to identify individual entrants and exiters and, consequently, to allow an estimate of gross flows into and out of farming.

A third objective was to test some of the propositions and hypotheses implied by the conceptual framework. This was attempted using the 1966-1971 Census of Agriculture Match. Because of the relative inadequacy of this data base, this objective was only partially met, and this prompted a proposal for further empirical research.

Relevance of the Study

This study is relevant from a policy and program development point of view and from a methodological point of view.

First, it is thought that a good knowledge of factors affecting the decision of farm operators to leave farming is necessary to the proper design of policies and programs aimed at affecting the rate of exits from agriculture. It was shown that political willingness presently exists, even though consensus on which direction the programs should act is absent.

Second, major methodological issues are tackled. The adequacy of various models of human behavior is evaluated; results and concepts drawn from economics, psychology and sociology are assembled into an expanded theoretical framework of the decision to leave farming.

Problems in statistical methodology are also tackled: the empirical analysis rely on both exploratory and confirmatory analysis; statistical methods suitable to the analysis of qualitative dependent variables are reviewed and used either in an exploratory mode or a confirmatory mode.

Area Chosen for the Empirical Study

The province of Saskatchewan was chosen to test some of the hypotheses which evolved from the conceptual framework. The reasons for this choice were:

1. Saskatchewan is essentially a rural agricultural province where no city is of such importance as to upset and distort what can be called a normal pattern of off-farm movement and migration.

2. The Saskatchewan Department of Agriculture has shown a marked interest in the evolution of farm demography and, more specifically, in factors preventing or prompting the decision to leave farming.

Definition of Terms

For the sake of clarity, definitions of the terms which are central to this study are presented in this section. No pretence is made that these definitions are the best or the only ones. The value of a term lies in the existence of a precise definition (whatever it is) and its consistent use thereafter.

<u>Off-farm mobility</u>. Potential ability of a farm operator or a farm worker to cease working in the farm sector.

<u>Off-farm movement</u>. The act by a farm operator or a farm worker to cease working on a farm.

Off-farm migration. The act by a farm operator or a farm worker to leave a farm area. Off-farm migration is related to change of geographical location, to change of residence. Off-farm migration can be considered to imply off-farm movement, but off-farm movement may occur without off-farm migration.

Exit from farming. A farm operator's cessation of farming activities. In this study, a farm operator is considered to have ceased farming activities when he is not reported to be an operator of a census farm.

Entry into farming. The act of starting to operate a farm. In this study this is construed as meaning to start being reported as a farm operator of a census farm.

Pre-retirement exit from farming. A farm operator's off-farm movement followed by his involvement in a nonfarm occupation. Pre-retirement exit from farming may or may not be accomplished by off-farm migration.

Exiter. A farm operator who exits from farming.

Stayer. A farm operator who remains in farming over the considered period. In this study this will be construed as meaning a farm operator who is reported as a farm operator of a census farm both at the beginning and at the end of an intercensal period.

Pre-retirement exiter. A farm operator who performs a pre-retirement exit from farming. In this study pre-retirement exiters are identified according to their age, a method which, admittedly, is not without shortcomings.

Organization of the Thesis

This thesis is divided into eight chapters; this first chapter serves as an introduction.

Chapter II is a review of literature on off-farm movement and migration. It summarizes theories which have been proposed to explain off-farm movement and migration, empirical studies conducted at either the macro-level or the micro-level, and policy recommendations which have been formulated.

Chapter III develops a theoretical framework for the pre-retirement decision to leave farming. First, the standard economic model of off-farm movement is critically reviewed. Second, elements for an expanded framework, which are drawn from the economic, psychological, and sociological literatures, are presented.

Chapter IV includes a description of the data base used in the empirical analysis, some methodological considerations on inference and data analysis, and a statement of the hypotheses to be confronted to empirical evidence.

Chapter V is a review of statistical methods for the analysis of qualitative dependent variables; it covers the standard linear model, logit/probit-type models, discriminant analysis, measures of association in contingency tables and a multivariate log-linear model with exogenous variables.

Chapter VI reports on empirical findings ensuing the use of exploratory and confirmatory data analysis methods.

Chapter VII outlines a longitudinal survey which could fruitfully continue the present empirical study.

Chapter VIII provides a summary of this research and an overview of the methodological problems and issues raised in the course of this work.

CHAPTER II

A REVIEW OF LITERATURE ON OFF-FARM MOVEMENT AND MIGRATION

This chapter is divided into five sections. The first section delimits the scope of the review of literature and provides a framework through which various studies can be related to each other and to the topic of this research. The second section is a brief review of theories of off-farm movement and migration which have appeared in the literature. In the third and fourth sections, the main empirical studies of off-farm movement and migration, conducted at the macro-level or at the micro-level, are summarized. The fifth section presents some of the policy recommendations made by economists in relation to off-farm movement and migration.

General Scope of the Literature

From the following review the reader will realize that numerous studies of off-farm movement and migration have been conducted. Their topics are closely related to pre-retirement exit from farming, but the overlap is far from being complete. Approaches vary greatly and, consequently, the degree of relationship to pre-retirement exit should be kept in mind.

Most of these studies have been prompted by one of the following general trends observed in recent decades: (1) decreasing number of farm operators,(2) decreasing farm labor force, (3) decreasing farm popu

popu

tion

real

ali

be de

These

farm

farm

of

amou

into

labo

numbe

pers

the

measure

agricult

cities a

not occ

at the

of the s

threaten

population,(4) decreasing rural population as a proportion of total population. Number of farm operators, farm labor force, farm population and rural population are different concepts covering different realities. Nevertheless, they have something in common. This commonality consists of the ability to measure a latent variable which can be described as the social and economic importance of the farm sector. These measures differ substantially one from the other. The number of farm operators, besides giving information on the magnitude of the farm industry, provides information on its structure and its degree of economic concentration. The farm labor force is a measure of the amount of labor utilized in agriculture. It does not, however, take into account the number of days or hours that each individual in the labor force actually works. The farm population, defined as the number of persons living on farms, is a measure of the number of persons related sociologically to agricultural production; it embodies the concept of farming as a "way of life." Rural population is a measure of the number of persons living directly or indirectly from agriculture only if nonagricultural industries are located only in cities and if daily commuting of rural residents to city-work does not occur. These preconditions can hardly be considered satisfied at the present time and the validity of rural population as a measure of the socio-economic importance of the agricultural sector is threatened.

Number of farm operators, farm labor force, farm population and rural population are all the cumulative and aggregate result of decisions made by individuals. They result, however, from different "elementary decisions," namely: to operate a farm or not to operate a farm, to work in agriculture or to work in another sector, to live on a farm or to live elsewhere, to live in a rural area or to live in an urban area. Decisions, as they are actually made by individuals, consist of a single or several elementary decisions: a farm operator may decide to leave farming and go to work in the nearest city, in which case an exit decision and a migration decision are taken simultaneously as a single decision. This composite nature of actual decisions renders empirical analysis more complex.

Authors of the macro-level studies reviewed in this chapter have tended to concentrate on the analysis of changes in one of the following quantities: (1) the number of farm operators,(2) farm population,(3) rural population. In micro-level studies, authors have tended to concentrate on one of the corresponding elementary decisions, without clearly understanding that such elementary decisions are rarely observable in isolation; it follows that such micro-level studies are often looking at the <u>same actual</u> micro-behaviors, but through different glasses because they focus on one particular aspect of this behavior.

It is contended that the failure to analyze and clarify the methodological issues arising because the off-farm movement and migration phenomena can be viewed from different angles has hindered an orderly development of empirical research. This failure explains also the lack of consistency in the terminology used in the litarature.

The topic of this research concerns a specific individual decision and its consequences on the farm industry: the pre-retirement decision of farm operators to leave agriculture. Studies surveyed in this chapter are directly related to pre-retirement exit from farming, but they usually deal with several of the aforementioned elementary decisions.

Theories Explaining Off-Farm Movement and Migration

Several theories have been proposed to explain, at different levels, off-farm movement or off-farm migration. The treadmill theory purports to explain the general forces which seem to act on the entire farm industry and to elicit a rapid structural adjustment. The career theory proposes a general pattern for each individual behavior in relation to occupational choice. The "push-pull" theory proposes a classification of the factors which affect individuals in their decision to leave farming. The benefit-cost theory proposes a general model for the individual decision concerning occupational mobility and geographical migration. Each of these theories is discussed below.

The Treadmill Theory

The general consensus among economists is to consider that the drastic reduction in the relative importance of agriculture in the recent decades is due to economic growth and output-increasing technology.

For McDonald, farm out migration is to be analyzed as an integrative adjustment to economic growth: "the approach is to view farm outmigration as a major aspect of a long, historical process of integration, through which agriculture is gradually becoming at one with a growing industrial economy."[1]

---

[1]Stepehn L. McDonald, "Farm Outmigration as an Integrative Adjustment to Economic Growth," Social Forces 34 (December 1955): 119.

The introduction of output-increasing technology together with a slowly increasing demand for agricultural products has entailed generally depressed prices for these products. Low returns due to depressed product prices increase the incentive to adopt further output-expanding technology as the only way to maintain income. Farmers find themselves caught into a general movement of forced adoption of new technology and ever depressed prices. This situation is typically what Platt calls a social trap.[2] Tweeten, while accepting this treadmill theory as an explanation of past economic disequilibrium until the 1950s, discarded it for the more recent years on the grounds that, since 1960, productivity in agriculture rose slower than demand for agricultural products. He concluded:

> The treadmill theory gives very useful insight into the dynamic context in which the farm problem of low returns arise, but is less helpful in explaining the persistence of the problem in growing farm economy in which demand grows faster than supply.[3]

The Career Theory

The career theory holds that there is a typical pattern of occupational choices made by individuals during their life span. Based on his tastes and ability, an individual chooses a certain occupation-type at the time he enters the labor force. This individual is expected to stay in this occupation until retirement.

---

[2] John Platt, "Social Traps," American Psychologist 28 (August 1973): 641-651.

[3] Luther G. Tweeten, "Theories Explaining the Persistence of Low Resource Returns in a Growing Economy," American Journal of Agricultural Economics 51 (November 1969): 800-801.

The career theory emphasizes rigidity and irreversibility of occupational choice. This is justified by the existence of forces which tend to maintain an individual in the occupation he entered. These forces can be classified into two categories: economic and sociological forces. First, occupation is an important determinant of the social status of an individual and casts him into a role. As Weerdenburg states: "the transition to another occupation mostly requires radical social re-orientation and very considerable adaptation on the part of the person concerned."[4] Second, because specific skills and knowledge are required for each occupation, mobility from one to another occupation entails that some previously acquired skills are rendered useless and others need to be acquired. Some previous investments in human capital are wasted and some new investments are necessary. A human capital loss is involved in all transfers from one occupation to another. Consequently, an individual who has to move into a new occupation finds himself at an economic disadvantage compared to others who have been in this occupation since they entered the labor force. In summary, the necessity of social re-adjustment and human capital losses are justifications for the rigidity and irreversibility bias embodied in the career theory of individual occupational choice.

The career theory, when applied to farm operators, holds that, in general, individuals enter farming at an early age and leave farming

---

[4] L.J.M. Weerdenburg, "Farmers and Occupational Change," Sociologia Ruralis 13 (First quarter 1973): 29.

in order to retire; career choices become more and more irreversible and farm operators' mobility decreases as age increases. In this context, pre-retirement exit from farming appears as a departure from a normal pattern of behavior.

For Sofranko and Pletcher there is no complete irreversibility of the original occupational choice, but when this decision is to be reversed, transition toward nonfarm employment follows a general pattern:

> The evidence suggests that a typical pattern of mobility for many low income, full-time farmers is from a full-time to part-time status, to complete off-farm employment. From this perspective, part-time farming is regarded as a stage in a 'career,' a step to eventual full-time, off-farm employment.[5]

Part-time farming, thus is considered to be a transitional state.

The "Push-Pull" Theory

This theory was developed by demographers in relation to research on population adjustments to changes in social and economic conditions. Sofranko and Pletcher[6] applied it to the analysis of off-farm movement. This theory postulates that individuals are simultaneously under the effect of two actions: a "pull" action and a "push" action. The "pull" action is caused by better employment opportunities, higher income, more enjoyable living conditions, and new or different activities. The "push" action is caused by the actual

---

[5] A. J. Sofranko and W. R. Pletcher, "Factors Influencing Farmers' Expectations for Involvement in Agriculture," Illinois Agricultural Economics 14 (July 1974): 6.

[6] Ibid.

hardship people are enduring in their situation; then, movement of individuals can "occur as a flight from undesired social or economic situations."[7]

The "push" variables proposed by Sofranko and Pletcher to explain why farmers leave farming are farm size, as measured by the total number of acres operated, and net farm income (excluding income from custom farm work). The "pull" variables chosen were age of operator, level of education of operator, involvement in off-farm work, and distance from the farm to the nearest large urban center.

Sofranko and Pletcher concluded their study as follows:

> . . .it is not easy to assign primacy to either 'pull' or 'push' factors when explaining farmers' expectations, or to discuss 'pull' factors apart from consideration of farm income level.
>
> . . .it is difficult to differentiate between 'push' and 'pull' influences.[8]

Two main criticisms can be leveled at the "push-pull" theory. First, it does not rely on a precise model of individual behavior. Second, the dichotomy push factors versus pull factors, which is the very crux of the theory, does not seem to have any empirical relevance. Reference to factors and to their effect on individual decision, without qualification of the way this effect is thought to be brought about, is sufficient.

---

[7]Donald J. Bogue, Principles of Demography (New York: John Wiley and Sons, 1969), p. 157, quoted in Sofranko and Pletcher, "Farmers' Expectations," p. 7.

[8]Sofranko and Pletcher, "Farmers Expectations," p. 11.

A Benefit-Cost Model

The decision of a farm operator to leave farming very often includes the decision to migrate; in this case the decisions regarding off-farm movement and off-farm migration are directly connected. Aside from the simultaneity, the decisions to leave farming and to migrate are related in the sense that the consequences they entail are of the same nature, namely these consequences are lasting and imply drastic changes in all aspects of one's individual life, social, professional, familial. This explains that the economic model of the individual decision used in studies on geographical migration and occupational mobility have been very similar. Most studies on off-farm movement and off-farm migration conducted by economists have relied implicitly or explicitly on this benefit-cost model.

The most formalized version is due to Sjaastad who described his approach as follows:

> This treatment places migration in a resource allocation framework because it treats migration as a means in promoting efficient resource allocation and because migration is an activity which requires resources.[9]

The model can most easily be presented taking as example the simplified case of a farm operator who is faced with two alternatives: to keep operating his farm or to leave farming and seek nonfarm employment. This farmer owns certain amounts of different resources namely labor, land, fixed capital, and liquid assets. These resources are expected to yield a certain income stream in the future. If the nonfarm

---

[9] Larry A. Sjaastad, "The Costs and Returns of Human Migration," Journal of Political Economy 70 (October 1962): 80.

alternatives were to be chosen these resources would be employed in different uses and expected to yield another income stream. The transfer of resources from one set of uses to another would entail a stream of costs. The net income stream after the transfer is the difference between the stream of incomes and the stream of transfer costs. The stream of benefits to be gained by leaving farming is the difference between the stream of net incomes when resources are employed in nonfarm uses and the stream of incomes when resources are used in farming.

The analogy between this decision of production abandonment and the decision to invest is clear.[10] The decision criterion and decision rule chosen can be any of those discussed in the capital theory literature.[11] The most common denominator used to compare income streams is their present value; the decision rule is, then, the maximization of this present value.

Income can be construed in many different ways. If income is construed as money income, adoption of the above decision criterion and decision rule is equivalent to assuming that the individuals are money income maximizers or profit maximizers. If income is construed as satisfaction, derived either from money income or "psychic" income, adoption of the maximization of present value decision rule is

---

[10]Early bibliographical search revealed the lack of theoretical work on production abandonment. The analogy with investment decision was rediscovered later in Marshall R. Colberg and James P. King, "Theory of Production Abandonment," Rivista Internationale die Scienze Economiche 20 (October 1973): 961-971.

[11]See for example, J. Hirschleifer, Investment, Interest and Capital (Englewood Cliffs, N.J.: Prentice Hall, Inc., 1970).

equivalent to assuming that individuals are satisfaction, or utility, maximizers. All authors have acknowledged, in different ways, the relevance of nonmonetary benefits and costs in the individual decision to achieve occupational mobility and geographical migration; these nonmonetary benefits and costs, have often been dubbed "psychic" income or costs, but the problem of translating them into monetary units has been evaded.

Sjaastad broke down the private costs of human migration into money costs and nonmoney costs.[12] The money costs include the expenses incurred by migrants in the course of moving. The nonmoney costs include foregone earnings (or opportunity costs) and "psychic" costs. Private returns are also divided into money returns and nonmoney returns. Sjaastad specified:

> Money returns so defined are sufficiently general to encompass not only those returns stemming from earnings differential between places, but also the return accruing to the migrant in his capacity as a consumer.[13]

Nonmoney returns reflect the preference of the individual for the new place of residence as compared to his former one. When the model is applied in the context of occupational change nonmoney returns reflect the preference for the new occupation as compared to the former one.

Sjaastad acknowledged that psychic costs and psychic returns affect an individual's decision. Nevertheless, he argued that "the psychic costs of migration involve no resources for the economy and should not be included as part of the investment in migration," and that "likewise we should ignore nonmoney returns arising from locational

---

[12] Sjaastad, "Human Migration," p. 83.

[13] Ibid., p. 86.

preferences to the extent that they represent consumption which has a zero cost of production."[14]

Maddox[15] inventoried and investigated the private costs associated with the movement out of agriculture. The cash costs include moving expenses for family and belongings and increase in living expenses during the process of moving. Nonmoney costs are the opportunity cost of income which would have been received during the time the migrant stays unemployed had he remained in agriculture and the psychic cost of leaving occupation or residence. Maddox suggested the following hypotheses regarding the various costs:

> First, the opportunity cost is of such minor importance that it can be ignored. Second, the cash costs of transportation are of minor significance. Third, the costs of food and lodging during the transition period, above the level of such costs on the farm, are not of great importance, provided the transition period between farm and nonfarm employment is not unduly long. Fourth, the subjective or psychic costs, though difficult to define and measure have both personal and social implications which cannot safely be ignored.[16]

To the inventory of costs by Maddox should be added cash costs of search for urban housing and nonfarm jobs, whether these costs are incurred before or after actual occupational transfer or migration has taken place.

There is no agreement in the literature on which income should be entered for the farm alternative. This issue is closely related to

---

[14]Ibid., pp. 85-86.

[15]J. G. Maddox, "Private and Social Costs of the Movement of People out of Agriculture," American Economic Review 50 (May 1960): 392-402.

[16]Ibid., p. 393.

whether the decision unit is the farm operator or the family and to whether the decision to leave agriculture necessarily entails geographical migration or not. The issue related to the choice of the decision unit will be discussed further in Chapter III; at this point, it suffices to state that the family should be considered as the basic decision unit. From a theoretical point of view it is clear that, in this case, expected family income should be entered for each of the alternatives. This way, any income earned by the operator's wife must be added to farm income considered as return to the operator's labor. If migration is not necessary when the operator ceases to farm, the wife's earnings will be entered again for the nonfarm alternative. Even though the issue is settled theoretically, practical problems will often remain because of the difficulties in collecting data on income earned and expected to be earned by family members in the different alternatives.

From this strictly economic model a number of hypotheses have been derived and many attempts at testing them have been conducted. These hypotheses are:

1. The main reason for leaving farming is income differential between the farm and nonfarm alternatives;

2. For most off-farm movers and migrants the actual gains exceed the costs;

3. The rate of unemployment in urban areas, which is a measure of the lack of nonfarm employment alternatives, is negatively related to off-farm movement and migration;

4. Experience in nonfarm employment will increase farm operators' mobility as it increases expected nonfarm income, facilitates search

for a nonfarm job, and, consequently, reduces costs associated with job-search and migration;

5. Rates of off-farm movement and migration among adult operators decrease with age since money and psychic costs associated with movement and migration are lower for young adults and since potential benefits can be expected to be captured by young adults over a longer span of time;

6. Rates of off-farm movement and migration decrease with distance to nearest large urban center since costs associated with job-search and migration increase with distance;

7. Off-farm movement and migration need not equalize money income because of the existence of fixed costs in off-farm migration and of differential in psychic income between occupations and locations.[17]

Summary

Four theories explaining off-farm movement and migration were reviewed in this section: (1) the treadmill theory, (2) the career theory, (3) the "push-pull" theory, and (4) the benefit-cost model. These theories have different scopes. The treadmill theory mainly sheds some light on the long-term dynamic context in which off-farm movement and migration occur. The career theory presents a general pattern of occupational choices as made by farmers and is more a simplifying description than an explicative theory. The "push-pull" theory distinguishes two types of causes of off-farm movement and

[17]These hypotheses, except Hypothesis 4, are stated in lesser details in Stephen L. McDonald, "Economic Factors in Farm Out-Migration: A Survey and Evaluation of the Literature," Austin, Texas, n.d. (Mimeographed), p. 6.

migration according to the way they exert their action. Finally, the benefit-cost model is the most refined theory, and accommodates any hypothesized factor; it has been underlying, explicitly or implicitly, the majority of the empirical studies which are reviewed in the next two sections.

Empirical Studies at a Macro-Level

This section reviews empirical studies of off-farm movement and migration conducted at a macro-level. In these studies emphasis is put on the analysis of relationship between aggregate variables such as number of farm operators, farm population, unemployment rate, aggregate farm income, etc. Data used are usually from secondary sources where they are already aggregated: the main source is census publications. In some cases individual data are re-aggregated in a more suitable way. In any case, aggregate data are analyzed.

D. G. Johnson,[18] in one of the first studies devoted to migration out of agriculture, analyzed data from the 1950 U.S. Census of Population. He concluded that: (1) most farm to nonfarm migrations were for relatively short distances,(2) a large proportion of migrants were family members, i.e., the assumption that most migration from farm to nonfarm are made by young people and bachelors was not consistent with the data analyzed,(3) jobs obtained by off-farm migrants had earnings ranging from 85 to 90 percent of those obtained by members of the nonfarm population of the same age and sex.

---

[18] D. Gale Johnson, "Policies to Improve the Labor Transfer Process," American Economic Review 50 (May 1960): 403-418.

Cohort analyses of farmers have been conducted in the U.S.A. by Kanel, Tolley and Hjort, and Clawson.[19] A cohort of farmers consists of all farmers born during the same period, usually a decade. The number of farmers in each cohort can be followed over long periods in different censuses. Patterns of different cohorts have been found to be very similar: the number of farmers in a chohort rises till middle age as young people keep entering farming and falls thereafter as more farmers leave farming due to off-farm movement, retirement or death. Most cohort analyses have emphasized the fixity of cohort patterns; thus one, or both, of the following two hypotheses hold: (1) young people and established farmers do not respond to changing economic conditions and, more specifically, to decreasing opportunities in farming,(2) relative opportunities in agriculture and in the nonfarm sector remain the same. Both hypotheses are very strong. Time is considered as being the only factor of the number of farmers in each cohort; thus, the analysis is very mechanistic.

Cohort analysis relies on aggregate data which display only net changes in the number of farmers: the net change in the number of farmers in each cohort is the algebraic sum of exits and entries. Consequently, the same cohort pattern can result from very different combinations of gross entries and exits.

---

[19]Don Kanel, "Age Components of Decrease in Numbers of Farmers, North Central States, 1890-1954," Journal of Farm Economics 43 (May 1961): 247-263; Don Kanel, "Farm Adjustment by Age Groups, North Central States 1950-1959," Journal of Farm Economics 45 (February 1953): 47-60; George S. Tolley and H. W. Hjort, "Age-Mobility and Southern Farmer Skill: Looking Ahead for Area Development," Journal of Farm Economics 45 (February 1963): 31-46; Marion Clawson, "Aging Farmers and Agricultural Policy," Journal of Farm Economics 45 (February 1963): 13-30.

Kanel found that cohort patterns were actually very similar. The only apparent change in cohort patterns appeared in the 1930s which corresponded to a decade of massive nonfarm unemployment. During this period fewer farmers seemed to have left farming or retired. This seemed to imply a qualification of his previous statement on the constancy of cohort patterns. The main finding of Kanel was:

> . . . the differences between decades, shown by aggregate rates of entry and withdrawal, were primarily the consequences, in successive periods, of the decreasing size and constant pattern of successive cohorts. . . the total number of farmers had decreased primarily because few young people had been able to enter farming.[20]

Clawson[21] also observed similarity in cohort patterns and stressed the importance of the decreasing size of entering cohorts as the cause for the decreasing number of farmers. The main consequence of these smaller entering cohorts was a shift of the age distribution of farmers toward older age classes. Clawson advocated the conducting of cohort analyses at lower level of aggregation, e.g. state and county levels, because social and economic conditions can differ drastically from one region to another and because these differences are masked through the process of aggregation. Clawson concluded from his study that, since farmers do not withdraw from farming but refuse to enter, any drastic decrease in the number of farmers, as was being advocated at that time is impossible; the current adjustment in the number of farmers was going to be continuing adjustment. Finally, Clawson emphasized that the adjustment in agriculture is not just only a matter of number of

---

[20]Kanel, "Farm Adjustment," p. 53.

[21]Clawson, "Aging Farmers," p. 16.

farr

and

com

The

Jon

thi

the

pen

fro

inc

wer

equ

was

rat

far

be

cou

rat

Eco

farmers but also of quality: the quality of young farmers entering and of those staying in farming and the quality of the rural-agricultural communities affected by this adjustment of the farm sector.

Johnston and Tolley developed an extension of cohort analysis. They rejected the usual assumptions embodied in cohort analysis:

> Most of the analysis has been descriptive rather than econometric and has tended to assume that cohort patterns are completely fixed.
>
> A closer examination of cohorts suggests that the differences found between cohort patterns are in part systematic. While it is true that each cohort has a generally similar pattern, in a decade when there is sharp change in total number of farms, all cohorts are deflected in the same direction from their characteristic pattern.[22]

Johnston and Tolley specified and estimated a model which incorporated this effect of exogenous factors on cohort patterns. In their model, the number of farm operators in any one age class was assumed to depend on the number of rural farm males surviving to this age class from the previous period and the unobserved ratio of farm to nonfarm income facing career choosers. Given that coefficients of this model were assumed to be different for each age class, there were as many equations to estimate as there were age classes. As the income ratio was not observed, the model was used to make inferences about this ratio. As a consequence the absolute elasticities of the number of farmers with respect to the ratio of farm to nonfarm income could not be estimated, but only the ratio of the elasticities of the age groups could be estimated. Results showed that response to changes in the ratio of farm to nonfarm income declined as age increased.

---

[22]W. E. Johnston and G. S. Tolley, "The Supply of Farm Operators," Econmetrica 36 (April 1968): 366.

r
h
T
f
f
i
l
i
co

La
es
di
th
th
ch

Per

cul

in
ver
Sta

Winkelman[23] aimed at determining factors affecting the rate of reduction of the labor force employed in agriculture. To do so, he hypothesized several linear models, which he tested on county data. The dependent variable was the rate of reduction of the farm labor force. The independent variables were: expected per capita incomes from farm labor, from farm assets and from work off the farm holding, information about nonfarm opportunities, average age of the county farm labor force, dispersion of farm income and expected level of employment in nonfarm work. Winkelman was well aware of the methodological shortcomings of using the net rate of change in the labor force:

> The rate of a change is influenced by entrants into the farm work force, retirements, deaths, and those quitting farming for other reasons. Each of these categories is undoubtedly influenced in a different way [by the independent variables]. . . .[24]

Lack of data on gross entries and exits compels one to specify and estimate aggregate models which may differ substantially from the disaggregated models. The most interesting finding of the study was that increased farm income tended to reduce the rate of decrease of the farm labor force. The sensitivity of the rate of decrease to changes in farm income was, however, small.

Sjaastad[25] studied off-farm migration relying on U.S.D.A. data pertaining to changes in residence (from a farm residence to a nonfarm

---

[23]Don Winkelman, "A Case Study of the Exodus of Labor from Agriculture: Minnesota," Journal of Farm Economics 48 (February 1966): 12-21.

[24]Ibid., p. 14.

[25]Larry Sjaastad, "Occupational Structure and Migration Patterns," in Labor Mobility and Population in Agriculture, ed. Iowa State University Center for Agricultural and Economic Adjustment (Ames: Iowa State University Press, 1961), pp. 8-27.

resi

ther

infe

must

the

This

cant

rate

he fo

capit

deter

be an

force

less

for s

ticat

appea

Econom
Geogra

residence and vice-versa); the concept central to his analysis is, therefore, farm population. This feature of the study implies that inferences about off-farm movement, as defined in the present study, must be taken with great care. Sjaastad's main contention was that the rate of off-farm migration is related to the business cycle:

> Percentage unemployment rates are perhaps the best single index to reflect the manner in which swings of the [business] cycle are likely to affect off-farm opportunities for potential migrants and hence the rate at which they abandon agriculture.[26]

This was shown clearly by correlation analysis, where a very significant negative relationship appeared for all time periods, between the rate of net off-farm migration and the unemployment rate. Even though he found that the relative farm income variable (ratio of regional per capita farm to nonfarm income) was not a significant variable in the determination of off-farm migration, Sjaastad concluded that "it would be an error to conclude that income differences are not a relevant force influencing off-farm migration"[27] and suggested that the use of less aggregated data, improvement in the measure of income, controlling for skill and educational levels, and the development of a more sophisticated permanent income concept would allow relative farm income to appear as a significant factor of off-farm migration.

Szabo[28] studied the decrease in farm population on the Canadian

---

[26]Ibid., p. 12.

[27]Ibid., p. 16.

[28]Michael L. Szabo, "Depopulation of Farms in Relation to the Economic Conditions of Agriculture on the Canadian Prairies," Geographical Bulletin 7 (Fourth Quarter, 1965): 187-202.

pr

fr

as

st

se

hy

be

ne

wit

to

is

pre

tive

rela

is "

rela

of of

Costs
1-11.

prairies during the period 1961-1971 on the basis of data generated from the 1961 and 1971 Census of Agriculture. His study was dubbed as being "areal" by which it is meant that each "observation" used for statistical analysis consisted of the average values of a number of selected variables for each area. Szabo tested a certain number of hypotheses using regression techniques with the ratio of depopulation being the dependent variable. The following hypotheses, which were neither clearly stated nor well justified, were found to be consistent with the data analyzed:

1. The proportion of rented land in an area is inversely related to the ratio of farm depopulation, as a high proportion of rented land is considered to indicate that off-farm migration occurred massively in previous periods, thereby leading to a more stable farm population.

2. The proportion of farmers reporting off-farm work is positively related to the ratio of farm depopulation.

3. The total capital value of farms in an area is negatively related to the ratio of farm depopulation.

4. High level of investment in machinery and equipment, which is "indicative of good conditions for agriculture,"[29] is negatively related to the ratio of farm depopulation.

Diehl,[30] starting from a crude benefit-cost micro-economic model of off-farm movement and migration, hypothesized a macro-level where

---

[29] Ibid., p. 193.

[30] William D. Diehl, "Farm-Nonfarm Migration in the Southeast: A Costs-Returns Analysis," Journal of Farm Economics 48 (February 1966): 1-11.

net

char

pote

ment

expe

vari

The

On the

invers

ment c

tween

suggest

Diehl's

income

I

able se

variabl

prising

3

net off-farm migration and movement is a function of the following characteristics of the farm population in the area considered: earning potential outside agriculture adjusted for age and educational attainment, nonfarm occupational experience, race composition, farm income, expected future income from assets and distance of migration. Several variations of the basic model were tested using regression analysis. The results showed that the rate of net migration was:

1. negatively related to farm income,
2. positively related to age,
3. negatively related to capital gains,
4. positiviely related to the proportion of negroes in the population, and
5. positively related to level of skill.

On the other hand, the hypothesis that the rate of migration is inversely related to distance from the farm area to a nonfarm employment center was rejected. Furthermore, a positive relationship between rate of migration and distance to nonfarm employment center was suggested, in complete contradiction to the original hypothesis. Diehl's general conclusion was that "people do migrate in response to income incentives."[31]

In conclusion, macro-level studies which have to rely on available secondary data cannot test macro-models which include all the variables thought to affect the micro-behavior; it should not be surprising, therefore, to experience some disillusion with the results.

---

[31]Ibid., p. 11.

0

g

or

se

ic

ti

of

are

cit

for

the

cal

bec

sis

stud

thes

leave

ing t

tudes
Static

One particularly limiting factor, has been the lack of estimate of gross off-farm movement and migration. Micro-level studies can rely on data pertaining to individuals who have moved or migrated and, consequently, it should be expected that factors of the decision can be identified more easily at the micro-level.

Empirical Studies at a Micro-Level

As it was mentioned before, numerous studies have aimed at identifying, at a micro-level, the various factors affecting farmers' off-farm movement or migration decisions. Theoretical contributions are few and most studies reviewed in this section, usually rely implicitly, on a benefit-cost model similar to the one formalized by Sjaastad[32] for the decision to migrate. Theoretical issues will be discussed further in Chapter III; in this section studies with an emphasis on empirical analysis are reviewed. These studies are labeled "micro-level" because they rely on data on individuals and because their main emphasis is on the factors affecting the individuals' decisions.

Bowring and Durgin[33] reported some surprising results from a study of a sample of farmers in New Hampshire. The following hypothesis were tested and rejected:

1. Lower income farmers are more prone to considering to leave farming.

2. The age of the operator bears on his attitude towards leaving the farm (the direction of this effect was not specified).

---

[32] Sjaastad, "Human Migration."

[33] J. R. Bowring and O. B. Durgin, Factors Influencing the Attitudes of Farmers Towards Migration Off Farms, Agricultural Experiment Station Bulletin 458 University of New Hampshire, n.d.

i

s

i.

tc

mi

fa

ac

fa

anc

stu

thu

hyp

thr

inc

exit

that

Agric
State

3. Farmers with higher level of formal education are more likely to consider off-farm movement or migration.

4. Farmers not considering leaving farming are more progressive in their farm management practices.

5. Farmers with strong attachment to their community, as represented by participation in church and farm organizations, are less inclined to consider leaving farming.

Bowring and Durgin ascribed these negative results to the failure to consider off-farm income as an intrinsic determinant of off-farm migration on the same footing as farm income and other farm related factors.

Baird and Bailey[34] addressed the questions of what are the characteristics of farmers who leave farming, of whether the shift from farm to nonfarm occupations is a sudden one or proceeds by stages, and of what happens to the land when people have left. The empirical study was performed on a very small sample in a Mississippi county; thus presenting a serious threat to the validity of the results. The hypothesis that off-farm movement or migration is achieved through three stages, namely, obtaining a nonfarm income to supplement farm income, increasing the income share from nonfarm employment, and actual exit from farming, was supported by the data. Results also showed that a majority of farmers having left farming had retained their land.

Roy also used a small sample to study the "differential aspiration

---

[34] Andrew W. Baird and Wilfrid C. Bailey, Farmers Moving Out of Agriculture, Agricultural Experiment Station Bulletin 568, Mississippi State University, 1958.

a

i

w

mo

re

a

as

Farm

Reaso
Farm

made
in the
"mobil

of farmers to seek a better paying job,"[35] as measured on a Gutman-type scale based on the answers to a question on which factors, among the ten which were suggested, would stop the farmer to grasp an opportunity of earning a substantially higher money income. Results showed that "differential aspiration" was not related to any of five measures of performance of the farm, to educational attainment, to family income, to level of living, nor to operator's nonfarm experience. On the other hand, "differential aspiration" was inversely related to age and years in farming. Surprisingly, Roy concluded that he did not pick up the adequate area to study factors related to level of aspiration; according to Roy, very low incomes and low nonfarm opportunities explained why results did not fit the expected economic and sociological patterns.

Baumgartner[36] defined an attitude scale, called "potential mobility,"[37] which purported to measure an individual's degree of readiness to leave farming. The empirical analysis was conducted on a sample of 100 full-time farmers. "Potential mobility" was considered as the dependent variable. Baumgartner summarized his study as follows:

> The statistical analysis supported the central concept of this study that potential mobility among farmers is influenced by both economic and noneconomic variables. The most important factors proved to be age, income, and nonfarm work experience, as shown in the following findings: (1) under a variety of personal, economic, and social psychological

---

[35]Prodipto Roy, "Factors Related to Leaving Farming," <u>Journal of Farm Economics</u> 43 (August 1961): 668.

[36]H. W. Baumgartner, "Potential Mobility in Agriculture: Some Reasons for the Existence of a Labor-Transfer Problem," <u>Journal of Farm Economics</u> 47 (February 1965): 74-82.

[37]The term "potential mobility" is a pleonasm which has been made necessary because the term "mobility" has been used extensively in the literature to mean "movement." In this study, "movement" and "mobility" have been used in their original meaning.

> conditions, age exerted a more extensive influence than any other independent variable; (2) in about half the relationships tested, potential mobility was significantly greater among farmers under 45 than among those aged 45 or over; (3) income was associated inversely with potential mobility in the younger group; (4) present consideration of nonfarm work had a positive relationship to potential mobility in the older group; and (5) nonfarm work experience was associated positively with potential mobility among farmers irrespective of age.[38]

Hill,[39] instead of looking at mobility, surveyed a small sample of farmers who had actually left the farm to take urban employment. Low income, health, nonavailability of a farm, poor facilities, and lack of available credit were stated as the main factors which had influenced their decision to leave farming. The income factor was the most important. But money income after migration had not increased in proportion to cost of living off the farm; in other words, real income had actually decreased with migration. Surprisingly, however, when asked whether they would take the same decision again, a large majority answered that they would. This seems to be in contradiction to the statements that income was the major determinant of the decision to leave farming and that real income had actually decreased; expectation of higher future incomes could, however, reconcile these contradictory statements. Difficulty in finding nonfarm employment was shown to be the most important hindrance to off-farm movement and migration. A comparison of movers and nonmovers showed that these two groups did not differ in age, farm income, or farm size. Hill concluded from these results that, contrary to a widespread

---

[38] Ibid., p. 668.

[39] Lowell D. Hill, "Characteristics of the Farmers Leaving Agriculture in an Iowa County," Journal of Farm Economics 44 (May 1962): 419-426.

belief, those who are leaving agriculture are not "those with the lowest incomes, the least efficient, the poorest farmers, or the physically disabled."[40] Finally, the high proportion of tenants in the sample of movers seemed to suggest the relevance of land tenure as a factor in off-farm movement and migration.

Guither[41] interviewed two hundred Illinois farm operators who left farming. Characteristics of farmers who left farming were compared to those of the total farm population. Exiters were found to be more concentrated in the lowest and highest age classes and to have had more off-farm employment experience. Most exiters responded that they liked farming and cited as major reasons for leaving: income, tenure problems, physical health, and retirement. Most of the farm operators leaving farming showed a strong desire to remain and live in their community. This reluctance to migrate was reflected in the distance of migration: only ten percent migrated more than twenty-five miles away. Guither stressed that, when farmers did leave, they did it in a minimal way, trying as much as possible to minimize the social and economic adjustments they foresaw.

Bennett[42] performed three surveys on a small sample of farmers over a time span of 15 years. His concern was movement from full-time to part-time farming. This can be considered as partial movement if,

---

[40]Ibid., p. 426.

[41]Harold D. Guither, "Factors Influencing Decisions to Leave Farming," Journal of Farm Economics 45 (August 1963): 567-576.

[42]Claude F. Bennett, "Mobility from Full-time to Part-time Farming," Rural Sociology 32 (June 1967): 154-164.

following Sofranko and Pletcher,[43] part-time farming is hypothesized to be a transitional stage towards complete disengagement from agriculture. Farm income, farm income per family member, and farm operator age were hypothesized to be inversely related to subsequent movement from full-time to part-time farming. Proximity to nonfarm jobs and farm operator's level of education were hypothesized to be directly related to movement from full-time farming. Multivariate analysis showed that none of the factors considered were good single predictors of full-time to part-time movement. In other words, interaction between the above factors was an important determinant of farmers' mobility.

Kaldor and Edwards[44] surveyed a sample of 304 Iowa pre-retirement exiters, i.e., farm operators who left farming and took nonfarm jobs. Characteristics of farm operators leaving farming were compared to those of the total population of Iowa farmers; the following facts were evidenced:

1. Exiters had significantly more formal schooling;

2. Exiters comprised a higher proportion of tenants;

3. The size of the farm of exiters, as measured by acreage, was not significantly different; and

4. The proportion of part-time farmers among exiters was higher.

Income was considered by pre-retirement exiters as the major factor in their decision. Contrary to common belief that farmers migrate to large cities, it was found that most of the exiters found jobs in small

---

[43] Sofranko and Pletcher, "Farmers Expectations."

[44] Donald R. Kaldor and William M. Edwards, "Occupational Adjustment of Iowa Farm Operators Who Quit Farming in 1959-1961," Agricultural and Home Economics Experiment Station Special Report 75, Iowa State University, 1975.

to

fa

em

in

th

si

had

to

mov

tog

stu

lon

far

vid

emp

of e

Reco
Iden
in t
Pres
Labo
of A
Moci
Unit
on R
Bria
of A
and
Jobs,
State

to medium size towns. Occupational movement of farm operators from farm to nonfarm employment coincided with a large increase in wives' employment. A majority of respondents thought that the real family income had increased since they left farming. A small proportion of the respondents (fifteen percent) indicated that their postmovement situation had not matched their expectations; these unsatisfied exiters had thought of re-entering farming, but very few actually had plans to do so.

The major contribution to the micro-level analysis of off-farm movement and migration is due to Hathaway and Perkins who authored together or separately several publications.[45]

From an empirical point of view, the major innovation of their studies was the use of a new data set: the one percent Social Security longitudinal sample; from this sample each individual who had covered farm employment was selected. Information available for each individual included covered earnings, industry of employment, location of employment and characteristics of the originating or receiving areas of employment. Using these data, Hathaway and Perkins studied off-farm

---

[45]Dale E. Hathaway, "Migration from Agriculture: The Historical Record and its Meaning," American Economic Review 50 (May 1960): 379-391; Idem, "Occupational Mobility from the Farm Labor Force," in Farm Labor in the United States, ed. C. E. Bishop (New York: Columbia University Press, 1967), pp. 71-96; Dale E. Hathaway and Brian B. Perkins, "Farm Labor Mobility, Migration, and Income Distribution," American Journal of Agricultural Economics 50 (May 1968): 342-53; Idem, "Occupational Mobility and Migration from Agriculture," in Rural Poverty in the United States, A Report by the President's National Advisory Commission on Rural Poverty (Washington: Government Printing Office, 1968): 185-237; Brian B. Perkins, "Farm Income and Labour Mobility," American Journal of Agricultural Economics 55 (December 1973): 913-920; Brian B. Perkins and Dale E. Hathaway, The Movement of Labor Between Farm and Nonfarm Jobs, Agricultural Experiment Station Research Bulletin 13, Michigan State University, 1966.

mo
th
sc
re
to
da
is
af
us
ef
th
cr
wi
tha

app
tha
was
sel
mob
hyp
agr
the
by H

ence
Perk
pati

movement and migration, both separately and from the point of view of their relationship to each other. Even though, Hathaway and Perkins sometimes aggregated individual data to study rate of movement with regional variables, their studies are reviewed in this section devoted to studies performed at the micro-level; the main justification is that data used are individual and longitudinal (i.e., the same individual is followed over time), and the main emphasis is on identifying factors affecting the individual decision to leave farming. Data were analyzed using univariate or multivariate methods which allowed control for the effects of other variables. In most cases, results obtained through the use of multivariate analyses confirmed those obtained through cruder univariate analyses. The main findings are reviewed below, with heavier emphasis given to those pertaining to off-farm movement than to those pertaining to off-farm migration.[46]

Age and mobility were found to be inversely related. Negroes appeared to be more mobile in analyses unadjusted for other factors, than when adjustments were made for other factors. This discrepancy was due to the spurious effect of employment status (wage earner versus self-employed) and age. Multiple jobholders were significantly more mobile than single jobholders. This result strongly supports the hypothesis that part-time farming is a means towards adjustment out of agriculture rather than a stable situation. Income while employed in the farm sector was not related to off-farm movement; this was taken by Hathaway and Perkins as supporting the hypothesis that people with

[46]In order to save space, results reviewed here will not be referenced individually; they all come from the already cited publications: Perkins and Hathaway, Movement of Labor; Hathaway and Perkins, "Occupational Mobility"; Idem, "Farm Labor Mobility."

high earnings in farming have also high potential earnings in other sectors; this also contradicted the thesis that off-farm movement tends to reduce income dispersion. Proximity to nonfarm employment centers, measured by the distance to nearest Standard Metropolitan Statistical Area (SMSA), affected positively the rate of off-farm movement; this relationship, however, was not simple nor linear as evidenced by occurrence of the highest off-farm movement rates in counties which are farthest from any employment center. No clear results were obtained concerning the effect of nonfarm unemployment on the rate of off-farm movement. It appeared, however, that high rate of nonfarm unemployment reduced off-farm movement by a greater amount in regions with low farm income. This result cast further doubt on the commonly held thesis that off-farm movement is conducive to income equalization. Farm employment status was shown to be the most important determinant of the rate of off-farm movement; farm operators were significantly less mobile than farm workers, whether they were single job holders or multiple jobholders. This result supported the hypothesis that ownership of assets reduces off-farm mobility. Hathaway and Perkins, gave a snapshot description of people leaving farming: ". . .the probability of off-farm mobility [movement] is highest for young, multiple-job farm wage workers located in counties within 50 miles of an SMSA."[47]

Data available allowed Hathaway and Perkins only indirect and partial tests of the theory that occupational change is decided on the basis of expected incomes adjusted for differences in cost of living.

---

[47]Hathaway and Perkins, "Occupational Mobility," p. 191.

H

A
fa
of
in
to
to

tha
much
work
foun
inco
compa
farms
sayin

Hathaway and Perkins, however, found enough evidence to conclude:

> In general, it appears that mobility [movement] from the farm sector is largely a function of the income expectations of movers and the extent to which these expectations are achieved. The fact that mobility [movement] rates for persons with various characteristics is highly consistent with the income experience of such persons suggests that farm people may have a realistic evaluation of their non-farm employment opportunities.[48]

A decrease in income after off-farm movement seemed to be the major factor in the decision to return to farming. Long term earning after off-farm movement were directly related to earnings in the year following movement, and more importantly, they were also positively related to earnings in the farm sector. This finding led Hathaway and Perkins to the following conclusion:

> farm-nonfarm occupational mobility [movement] does not seem to close the income gap between the poor and those better off--indeed it may widen it.[49]

The study of factors affecting migration from agriculture showed that the proportion of migrants decreased with increasing age, was much higher for negroes than for non-negroes, and was higher for farm workers than for farm operators. Surprisingly little difference was found between the proportion of migrants from counties with high family income and counties with low family income; the same result was obtained comparing counties with high proportion and low proportion of commercial farms. Hathaway and Perkins concluded their study of migration by saying:

---

[48]Perkins and Hathaway, Movement of Labor, p. 31.

[49]Hathaway and Perkins, "Occupational Mobility," p. 207.

a

o

a

in

mi

The

bef

(3)

fam

stuc

on t

from

impo

cause

Depar

> . . .the significant fact is that neither low income areas nor areas lacking a commercialized agriculture have high migration rates, and consequently that farm workers in such areas rely primarily on local labor markets for non-farm employment opportunities.[50]

Analyzing the relationship between off-farm movement and migration, Hathaway and Perkins found that most off-farm movers do not migrate when they leave farming and that "the proportion of migrants is highest among off-farm movers who are young, Negroes, or farm wage workers, and tends also to be high among movers from prosperous farm areas, in close proximity to employment centers."[51]

Abrahamson[52] conducted a survey of a hundred farm exiters who migrated to Saskatoon (Saskatchewan) and sought nonfarm employment. The survey attempted to collect information on (1) migrants' situation before off-farm migration took place, (2) causes for off-farm migration, (3) the migration decision process, (4) migrants' situation after off-farm migration. Recourse to recall data is a major weakness of this study.

The major cause for off-farm migration was material deprivation on the farm. Other factors such as isolation from medical services, from secondary schools, and from advanced training played also an important role. Poor health or crippling accidents were the immediate causes in several cases.

---

[50]Ibid., p. 198.

[51]Ibid., p. 201.

[52]Jane A. Abrahamson, Rural to Urban Adjustment (Ottawa, Department of Regional Economic Expansion, 1968).

i

g

th

mo

of

vi

mi

wi

had

soc

ing

dif

plo

far

appe

migr

were

inte

case

pover

nativ

to mi

The majority of migrants relied on farming only as a source of income. There was no indication that off-farm migrants lacked progressiveness in adopting new farm technology and practices; rather, the economic base of the farm was not sufficient to support such modern technology.

Off-farm migrants tended to recall farming and the social life of the farm neighborhood in an idealized version. Such an idealized vision may not correspond to the situation prevailing at the time of migration since "the disintegration of farm communities associated with farm consolidation and the trend away from the small family farm had served frequently to undermine the traditional basis for rural social life."[53]

Off-farm migrants' low level of education and vocational training put them at a disadvantage in the urban labor market. Consequently, difficulties of adjustment to urban employment, such as recurring unemployment, inadequate wages, and dislike of working conditions, were by far the migrants' major adjustment problems. Migrants, as a group, appeared to be at a disadvantage: rate of unemployment among off-farm migrants was three times the average rate for Saskatoon and migrants were mainly employed in unskilled jobs. Problems related to social integration and psychological stress symptoms were mentioned in many cases.

The decision to migrate seemed to have been made as a result of poverty on the farm rather than because of attractive nonfarm alternatives. Social and psychological factors also affected the decision to migrate.

---

[53]Ibid., p. 115.

ir

sc

mo

an

ad

for

as

gen

and

far

the

are

Acco

clas

and

progr

more

aim a

reduc

farmin

In conclusion, micro-level studies have provided some valuable insight into the factors of off-farm movement and migration, though some contradictory results have been obtained. Differences in methods most likely account for a large part of these inconsistencies.

Policy Recommendations

A constantly recurring policy issue related to off-farm movement and migration is the choice of actions to be taken to ensure the adequacy of the number of farm operators or the size of the farm labor force. Most economists have regretted the excessive number of farms as well as the excessive size of the farm labor force. These are generally seen as reflecting resource misallocation in the economy and as causes of low returns to resources, especially labor, in the farm sector.

Johnson[54] saw the accelerated transfer of labor from the farm to the nonfarm sector as desirable. Policy decisions, in this context, are a matter of devising ways to improve and accelerate this transfer. According to Johnson, this could be accomplished by having three classes of programs concerning respectively information, assistance, and educational improvement in rural areas. Information disseminating programs should be devised so as to make farm operators and workers more aware of nonfarm job opportunities. Assistance programs would aim at increasing the number of off-farm movers and migrants, and at reducing the number of "decision mistakes" leading to re-entry into farming. Finally, the setting up of programs geared at raising the

[54] Johnson, "Labor Transfer."

i

b

f

r

c

e

d

l

co

try

whi

inf

fil

fac

the

Peop

level of educational attainment in the farm population was advocated by Johnson as the best long-run policy to improve transfer of labor from the farm to the nonfarm sector. Because farm people had shown reluctance to migrate over long distances, Johnson recommended encouraging the creation of job opportunities in rural areas.

Sjaastad downplayed the severity of the "farm problem" (popular expression referring to the problem of low returns to farm resources due to the excessive number of operators) by stressing its self-liquidating nature:

> The conclusion I would venture is that long-run prospects of American agriculture are improving with each passing year. As the relative size of the farm sector diminishes a greater ease of adjustment to even further revolutions of supply should result. It is conceivable that with a conscientious maintenance of full employment and no blunders in government farm policy an alleviation of the farm income problem may reasonably be anticipated.[55]

Despite Sjaastad's denials, this seems to imply that the best course of action for policy makers is to sit and wait, while only trying to ensure low nonfarm unemployment.

Smith[56] came to conclusions regarding the role of information which are substantially different from Johnson's. Smith showed that information that discouraged potential off-farm movers or migrants filtered back from disenchanted movers and migrants. Consequently, facilitating and easing the flow of information may very well reduce the rate of off-farm movement and migration. As Smith puts it:

> The possibility of stimulating continued migration through improved information services is obviously questionable unless the experience of urban employment and urban life is

[55]Sjaastad, "Occupational Structure," p. 27.

[56]Eldon D. Smith, "Nonfarm Employment Information for Rural People," Journal of Farm Economics 38 (August 1956): 813-827.

Fr

in

tha

and

all

sec

off-

then

and

a su

to f

Prog

the

natur

agric

the p

> such that it can be accepted in full knowledge of its consequences. Indeed, migration would probably be socially undesirable if its consequences were so personally calamitous as to result in return migration, or the desire to return to farm life.[57]

From this analysis, Smith concluded that emphasis should be on preparing potential movers and migrants to their new setting:

> The evidence suggests that for some fairly large groups of underemployed rural people the baffling problem of how to prepare them for urban life must be solved before informational services can be very effective.[58]

Hathaway and Perkins added the missing link to Smith's argument, that is, the empirical evidence of the importance of return movement and migration.[59] They inferred from this discovery that, if an overall objective is to reduce the number of people employed in the farm sector, it can be better achieved by increasing the proportion of off-farm movers and migrants who stay in nonfarm occupations, that there is no need to give incentives to more people to leave farming, and that it suffices to ensure that a higher proportion of them make a successful move to nonfarm employment. Research should be devoted to factors affecting the success or failure of movement and migration. Programs should be designed to improve the success rate by changing the characteristics of these movers and migrants, or by altering the nature of the labor market these movers and migrants must enter.

Hill, before Hathaway and Perkins, showed that farmers leaving agriculture are not "those with lowest incomes, the least efficient, the poorest farmers, or the physically disabled."[60] Consequently, he

---

[57] Ibid., p. 820.   [58] Ibid., p. 825.

[59] Hathaway and Perkins, Movement of Labor, pp. 32-35.

[60] Hill, "Farmers Leaving Agriculture," p. 426.

warned against a policy of low farm prices as a means to accelerate off-farm movement and to increase productivity of the farm sector; such a policy could have the opposite effect. Consistently with his identifying nonfarm unemployment as the major impediment to off-farm movement, Hill recommended policies geared to ensure a low unemployment rate in the whole economy.

Conclusion

The above review of literature on off-farm movement and migration is comprehensive, but no claim is made that it is exhaustive. It exemplified, however, the diversity of approaches encountered in the literature and, more important, the often conflicting results which have been obtained.

Agreement is reached on money income being the prime factor or motivation affecting off-farm movement and migration; but the relative importance of nonmonetary factors remains to be established. Lumping together these factors, and naming them "psychic income," appears to lead to imprecise empirical work. At a macro-level, the adverse effect of nonfarm unemployment on the rate of off-farm movement and migration is generally acknowledged, but its selective and differential effect on various types of potential movers and migrants is unknown. Uncertainty still exits, as to whether movers actually increase their real income as a result of off-farm movement or migration. Distance to nonfarm employment center, assumed to represent location of alternative nonfarm jobs, is generally considered to be negatively related to off-farm movement and migration; conflicting results, however, were obtained, which suggest that the effect of this distance factor is not simple; further empirical and theoretical work is needed in this area.

Ca

in

a

br

ta

in

th

an

ob

It

so

re

and

tha

ger

run

on

impo

psyc

of c

the

tica

Conflicting evidence appears, also, as to the role of operators' involvement in off-farm work in off-farm movement and migration; is it a transitional state or a permanent state corresponding to a new equilibrium?; does it facilitate exit from farming? Another major uncertainty pertains to the effect of off-farm movement and migration on income differentials, between the farm and nonfarm sectors, and within the farm sector itself.

Most of the studies which were reviewed pertained to the U.S.A. and only a few to Canada. It is difficult to say whether results obtained in the U.S.A. can be generalized to Canada, or other countries. It should be noted, however, that the U.S.A. and Canada have similar social and economic structures. Furthermore, differences between two regions in Canada are often greater than between one region in Canada and its closest neighbor states. Thus it is reasonable to consider that results which were obtained for a region of the U.S.A., can be generalized to a region of Canada, if the regional agricultural and rural socio-economic structures are similar.

A major conclusion to be drawn from this review of literature on off-farm movement and migration is that, though economic factors are important in explaining the decision to leave farming, social and psychological factors must also be considered. The benefit-cost model of off-farm movement and migration seems to be ill-adapted to pursue the investigation of social and psychological factors. Further theoretical work is necessary.

CHAPTER III

AN EXPANDED THEORETICAL FRAMEWORK FOR THE PRE-RETIREMENT DECISION TO LEAVE FARMING

The intent of this chapter is to provide a theoretical framework which can serve as a guide in studying the pre-retirement decision to leave farming. This is achieved by: (1) investigating the applicability of some recent theoretical advancements in the economic theory of individual behaviour which, unfortunately, have not been taken up by applied economists in their work, and (2) by bringing together concepts and results of economics, psychology, and sociology.

This chapter is divided into three sections. In the first section, the standard economic model of the decision to leave farming is presented, and behavioral assumptions on which this model is based are reviewed and examined critically, both in a general context and as they relate to off-farm movement. In the second section, the elements of the expanded theoretical framework are introduced. In the last section a summary is provided and implications for empirical study are examined.

The Standard Economic Model of Off-Farm Movement

The Model

By "standard economic" model it is meant the maximizing model of individual behavior with simplifying assumptions which is used throughout the applied economic literature. Thus, it is not contended that economic theory is limited to this model, but that this model is the one commonly used in empirical studies. Therefore, the adequacy of this model for an empirical study of off-farm movement is to be examined.

Two versions of the standard economic model will be briefly presented: 1) an income-maximizing model and 2) a utility-maximizing model.

The Income-Maximizing Model

The case where two occupational opportunities are open to the farm operator and his family provides a useful simplification for the sake of presentation. Each alternative yields a stream of net income (gross income minus cost involved in choosing this alternative); if nonmonetary incomes are received, they are assumed to be convertible to monetary equivalents. The criterion of choice between alternatives is, then the present value of these income streams. The decision rule is one of maximization of this present value. A rate of discount, representing the individual's time preference, is used to convert income-streams to present values:

$$PV_F = Y_1 + \frac{Y_2}{1+r} + \frac{Y_3}{(1+r)^2} + \ldots, \frac{Y_T}{(1+r)}\ T\text{-}1,$$

$$PV_{NF} = Z_1 + \frac{Z_2}{1+r} + \frac{Z_3}{(1+r)^2} + \ldots, \frac{Z_T}{(1+r)}\ T\text{-}1,$$

where

PV_F = present value of income-stream while farming,

PV_{NF} = present value of income-stream in nonfarm occupation,

Y_i = net income while farming for year i, $i = 1 \ldots T$,

Z_i = net income in nonfarm occupation for year i, $i = 1 \ldots T$,

r = discount rate, assumed constant overtime,

T = years of working life remaining.

It is assumed that the length of the income-stream is the same for both alternatives,or, in other words, duration of life is the same; this may not be true since some occupations are more dangerous than others. It is also assumed that the rate of discount (i.e., time preference) is the same for both alternatives.

If time is taken to be continuous rather than discontinuous,

$$PV_F = \int_0^T Y_t \, e^{-rt} \, dt,$$

$$PV_{NF} = \int_0^T Z_t \, e^{-rt} \, dt.$$

Whether time is continuous or discontinous, the income maximizing model specifies that the individual chooses the alternative which maximizes the present value of his income stream; in mathematical form:

$$\text{Maximize } (PV_F, PV_{NF}).$$

This model of off-farm movement is formally the same as the usual model of choice of investment; all results derived for the choice of investment carries over to off-farm movement.

The Utility-Maximizing Model

In this model a multi-period utility function must be assumed. Then, the model can be written as follows:

$$\text{Max } U\ (x_{11},\ \ldots,\ x_{it},\ldots,\ x_{IT},\ z_{111},\ldots,z_{jtk},\ldots,\ z_{JTK})$$

$$\text{Subject to} \quad \sum_i p_{it} \cdot x_{it} \leq Y_{tk}\ ,$$

where

x_{it} = quantity of good or service i purchased in period t, $i = 1 \ldots I,\ t = 1 \ldots T$,

z_{jtk} = free good or service j available in period t, when occupational opportunity k is chosen,

y_{tk} = monetary income in period t when occupational alternative k is chosen.

The utility maximizing model lead to an integer programming problem. The high level of formalization reached is satisfying but the empirical use of such a model is limited by the difficulties encountered in specifying the utility function.

A Critical Review of the Behavioral Assumptions

Outline of the Behavioral Assumptions

Economic theory rests on the assumed behavior of basic economic units: firms and consumers. Stringent assumptions are usually imposed to the motivational and cognitive aspects of this behavior. Since, on a farm, consumption and production decisions are closely interrelated both sets of assumptions concerning firms and consumers need to be reviewed.

According to standard economic theory, firms and consumers share a common decision rule: the principle of maximization: they choose the course of action which ranks highest in term of some criterion. Firms and consumers are assumed to differ, though, on the criterion they maximize: consumers maximize utility (or satisfaction) and firms maximize profits. Firms and consumers are assumed to be rational; given the other assumptions imposed, rationality does not add anything to the model but merely stresses that firms and consumers systematically maximize profits and utility.

Beside the motivational assumptions described above, some assumptions related to cognitive aspects of behavior are imposed: perfect and costless information, as well as complete foresight, are assumed. These assumptions ensure that no doubt can persist on what criterion should be maximized. If uncertainty is introduced in the model many criterions can possibly be maximized and rationality which had an unambiguous meaning becomes ambiguous.[1] Thus the concept of rationality has evolved through time as simplifying assumptions of the standard model were relaxed.

Within this framework of maximizing behavior in a world of perfect knowledge and foresight, economic agents appear as mere automatons whose actions can be precisely forecast.

---

[1] Herbert A. Simon, "Theories of Decision Making in Economics and Behavioral Science," American Economic Review 49 (June 1959): 256.

These behavioral assumptions have been the cause of bitter feuds in the economic profession. Agreement is not even reached on whether the verisimilitude of these assumptions is relevant to the validity of economic theory. Friedman's well known position is that realism of the assumptions is completely irrelevant and that economic theory should be judged only on its predictive power. Hypotheses derived on the basis of these assumptions are to be confronted to facts. Furthermore, Friedman argues:

> ... the relation between the significance of a theory and the "realism" of its "assumptions" is almost the opposite.... Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions....[2]

The long, and seemingly never ending, methodological and philosophical debate will not be augmented here.[3] The position taken here lies on Samuelson's side: whatever value a theory has is owed to the realism of its assumptions, and whatever are the shortcomings of a theory, they originate from the unrealism of the assumptions. In other words: realism of assumptions is desirable. Such a position is not in conflict with Friedman's less provocative statement:

---

[2] Milton Friedman, Essays in Positive Economics (Chicago: University of Chicago Press, 1953), p. 14

[3] The interested reader is referred to: Fritz Machlup et al., "Problems of Methodology," American Economic Review 53 (May 1963): 204-236; Fritz Machlup, "Professor Samuelson on Theory and Realism," American Economic Review 54 (September 1964): 733-736; Paul A. Samuelson, "Theory and Realism: A Reply," American Economic Review 54 (Sept. 1964): 736-739; Gerald Garb, "Professor Samuelson on Theory and Realism: Comment," American Economic Review 55 (December 1965): 1151-1172; Stanely Wong, "The 'F-Twist' and the Methodology of Paul Samuelson," American Economic Review 63 (June 1973): 312-325; Jack Melitz, "Friedman and Machlup on the Significance of Testing Economic Assumptions," Journal of Political Economy 73 (February 1965): 37-60.

> ... the relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic", for they never are, but whether they are sufficiently good approximations for the purpose at hand.[4]

Profit Maximizing Assumption

This assumption is precise and objective, since a generally accepted and tangible measure of profits exist. It is widely recognized that firms do not behave according to this assumption; the disagreement is mainly on whether this justifies the dropping of the assumption, and economists' opinions on that matter depend mainly on their main interest. Some have been tempted to propose a utility maximizing assumption for businessmen's behavior which would include both monetary and non-monetary profits. This practice is not free of drawbacks as Machlup remarked:

> If whatever a businessman does is explained by the principle of profit maximization - because he does what he likes to do, and he likes to do what maximizes the sum of his pecuniary and non-pecuniary profits - the analysis acquires the character of a system of definitions and tautologies and loses much of its value as an explanation of reality. It is preferable to separate the non-pecuniary factors of business conduct from those which are regular items in the formation of money profits.[5]

Utility Maximizing Assumption

The most severe charge leveled at the utility maximizing assumption is its "unfalsifiability"; in other words, a model based on utility maximizing behavior cannot be proven wrong since one variable is always free, namely, tastes. Any departure from prediction made with such a model can be explained by a change in tastes.

---

[4]Friedman, Positive Economics, p. 15.

[5]Fritz Machlup, "Marginal Analysis and Empirical Research," American Economic Review 36 (September 1946): 526.

Utility theory has been debated with respect to two fundamental issues: (1) the existence of a utility function, and (2) the nature of this utility function, i.e., ordinality or cardinality. At the present time, the second issue is solved. The works of Von Neumann and Morgenstern and later Savage[6] showed that if a utility function can be defined for an individual facing uncertain prospects, its cardinality follows. The existence of this utility function, though, rests on whether the axioms of individual behavior imposed to derive these results are verified or not. Consistency in choice is the key requirement imposed on individual behavior. It appears that though consistent choices are observed when alternatives are amounts of money, inconsistent behavior is the rule when alternatives are goods with multi-dimensional payoffs. Thus, it is concluded that, for economic relevant choices, no utility function can be constructed from observed choices.[7]

---

[6] J. Von Neumann and O. Morgenstern, The Theory of Games and Economic Behavior (Princeton: Princeton University Press, 1944); L.J. Savage, Foundations of Statistics (New York: John Wiley and Sons, 1954).

[7] For a review of works on existence and nature of utility function see: Herbert A. Simon and Andrew C. Stedry, "Psychology and Economics," in The Handbook of Social Psychology, eds. Gardner Lindzey and Elliot Aronson, 2nd ed., Vol. 5 (Reading, Mass.: Addison-Wesley, 1969), pp. 269-314; John L. Dillon, "An Expository Review of Bernoullian Decision Theory in Agriculture: Is Utility Futility?" Review of Marketing and Agricultural Economics 39 (March 1971): 3-80.

Rationality in Economic Decisions

In the standard economic model, with perfect knowledge and foresight, the postulate of rationality merely emphasizes the postulate of maximizing behavior: individuals weigh <u>all</u> opportunities which are accessible to them and choose the one which ranks highest on a fixed a priori scale. This postulate has been challenged ever since it was proposed and Marshall had little faith in its validity though he thought it was a useful abstraction:

> When we speak of the measurement of desire by the action to which it forms the incentive, it is not to be supposed that we assume every action to be deliberate, and the outcome of calculation. For in this, as in every other respect, economics takes man just as he is in ordinary life: and in ordinary life people do not weigh beforehand the results of every action, whether the impulses to it come from their higher nature or their lower.[8]

Despite Marshall's contention that economics "takes man as he is in ordinary life", economists have not used very much the positive empirical studies of behavior conducted by other social scientists. As Hicks put it: "Pure economics has a remarkable way of producing rabbits out of a hat - apparently a priori propositions which apparently refer to reality".[9]

Although the actual behavior of an individual is not exactly described by the "rational" maximizing economic man, it cannot be dubbed as being "irrational". Thus, the concept of rationality had to be revised. Simon introduced the notion of "global rationality" as embodied in the economic man and set to himself the task

---

[8] Alfred Marshall, <u>Principles of Economics</u>, 8th ed. (London: MacMillan Press, 1920), p. 17.

[9] J.R. Hicks, <u>Value and Capital</u>, 2nd ed. (Oxford: Clarendon Press, 1946), p. 23.

> ... to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist.[10]

Thus, global rationality is seen has being not descriptive of actual behavior of individuals. Actually, rational individual behavior has to account for two set of constraints: (1) constraints arising because of the environment in which the individual evolves, and (2) constraints arising from the limited information-processing capacities of the individual.[11]

---

[10] Herbert A. Simon, "A Behavioral Model of Rational Choice," Quarterly Journal of Economics 69 (February 1955): 99.

[11] Idem, "Theories of Bounded Rationality," in Decision and Organization, eds. C.B. McGuire and Roy Radner (Amsterdam: North Holland, 1972), p. 162.

The Standard Economic Model and Off-Farm Movement

It was mentioned before that economic theory acknowledged two sub-species of economic man: a profit maximizing man and a utility maximizing man. These sub-species of economic man, are abstractions of real economic decision units: a firm or its owner-manager and a consumer. Thus, whenever one uses the abstraction of the economic model, one should decide of what is the real economic decision-unit to be represented by this model. Once the decision unit has been ascertained, a choice of model is to be made: profit maximizing or utility maximizing? These two consecutive questions seem especially relevant in the case of a study of off-farm movement.

With respect to the choice of decision-unit, two main alternatives exist: (1) the operator decides by himself, and (2) the family, possibly construed in various ways, decides. Support for the second alternative is drawn from the following arguments:

1. Though the decision to leave farming and to engage in a full-time nonfarm occupation is primarily affecting the farm operator, other members of the family are also concerned since they very often participate in the operation of the farm-business. This participation may be a part-time or a full-time activity. It can take various forms, but in most cases it is very real.

2. On most farms, no clear-cut separation between farm-business finances and family finances exist. It is well known that income from different sources (from the farm and from part-time nonfarm activities of the operator or a family member) are pooled.

Similarly, all expenses, whether related to the farm-business or to private consumption, are paid from the same account. Even though some farm accounting system may be used, this does not bear on the way finances are managed. This situation results in continuous competition between farm and familial uses of the same resources. Trade-offs between familial consumption, farm expenses, farm investments and saving are made constantly. Making these trade-offs allows family farms to dampen the effects of price and income instability. This ability to cope with instability is sometimes considered as the major advantage of the family farm over the larger corporate farm, thereby explaining the development of the family farm sector of agriculture.

3. In most cases, the operator's family resides on the farm, thus developing an attachment to land and farming of a non-economic nature. Furthermore, off-farm movement implies very often, though not always, migration and entails drastic consequences for all aspects of other family members' lives. Since all members of the family will be affected by the decision to leave farming, though not necessarily from a narrow economic point of view, it is likely that they will participate in the decision-making process and bear on the final decision.

Support for a utility maximizing rather than a profit or money income maximizing model arises from the necessity to take into account non-monetary factors of the decision to leave farming. Thus, within the framework of classical economic theory, the adequate model is one based on the family unit maximizing its utility.

This, however, leaves some problems unresolved, the main one being how such a family utility function is derived from individual utility functions (forgetting for awhile the problems related to the existence of such individual utility functions). The theoretical problem is one of the aggregation of utility functions or preferences. Arrow's famous "Possibility Theorem" states that such "acceptable"[12] aggregation is impossible unless some special assumptions are made on individual orderings. A family-farm utility function can be derived if, for example, dictatorship or identical ranking of all alternatives by all family members are present. These conditions are not acceptable premises for any empirical analysis.

In this first section, two versions of the standard economic model of off-farm movement were presented (an income maximizing and a utility maximiznng), the assumptions underlying this model were critically reviewed, and some difficulties in applying this model were identified.

---

[12] See Kenneth J. Arrow, "A Difficulty in the Concept of Social Welfare," in Readings in Welfare Economics, eds. K.J. Arrow and T. Scitovsky (Homewood, Illinois: Richard D. Irwin, 1969), pp. 147-168. Social preferences are considered to be acceptable if they satisfy the five following axioms; complete ordering; responsiveness to individual preferences, nonimposition, nondictatorship, and independence of irrelevant alternatives.

Elements for an Expanded Framework

In the previous section it was argued that the standard economic model of individual behavior was an oversimplification of actual individual behavior. In this section, several avenues of potential improvement will be explored.

One might wonder whether such refinement is necessary. In fact, it has often been argued that economics is not interested in individual behavior, perse, but in the aggregate and that though individuals vary widely in their motives, attitudes, time perspective, etc., all these differences cancel out when aggregation is performed. Thus, precision in the model of individual behavior would not be necessary or even desirable. If the analyst is concerned only with aggregates and if differences in motives, attitudes, time perspective, etc., are randomly distributed among individuals, it is true that a crude model of individual behavior can be used. If, however, these intervening variables are related to some exogenous factors and are susceptible to vary in the same direction for all, or even many, individuals, a study of aggregates must be based on a more realistic model of individual behavior.[13] When knowledge and understanding of individual decisions, per se, is an objective in itself, then, the standard model of behavior becomes patently insufficient; such a knowledge is desirable when government programs are created to advise individuals in their decision making process.

Simon explained that different levels of detail concerning individual behavior are required depending on whether one studies equilibria or the adaptation to a new equilibrium:

---

[13]George Katona, Psychological Economics (New York: Elsevier, 1975) pp. 49-58.

> . . .The equilibrium behavior of a perfectly adapting organism depends only on its goal and its environment; it is otherwise completely independent of the internal properties of the organism.
>
> . . .[but] to predict the short-run behavior of an adaptive organism, or its behavior in a complex and rapidly changing environment, it is not enough to know its gaols. We must know also a great deal about its internal structure and particularly its mechanisms of adaptation.[14]

The importance of nonrandom changes of such intervening variables as motives, attitudes, and time perspective, the very dynamic nature of off-farm movement, conceived as an adaptation to some disequilibrium, and the existence of programs assisting farmers in their decision to leave farming are the three main reasons for trying to provide an individual model of off-farm movement with more realism.

This attempt, presented in this section, is in the line of an already, old, but still nascent discipline, called sometimes psychological economics and sometimes economic psychology. Few authors have contributed to its development: George Katona, Herbert A. Simon, Samuel P. Hayes, Jr. and Pierre Reynaud; their ideas permeate this whole section.

Single-Motive and Multiple-Motives Theories of Action

Katona explained that the postulate of maximizing behavior embodies a single-motive theory of action.[15] To this single motive theory, psychology opposes a multiple motives theory. If one recognizes the multiplicity of motives, the patterns into which these motives are

---

[14]Simon, "Theories of Decision Making," p. 225.

[15]George Katona, "Rational Behavior and Economic Behavior," Psychological Review 60 (May 1953): 313.

articulated become relevant; this opens up a new field for empirical observation and experimentation: the changes in the way these different motives are articulated to each other. Tastes cease to be considered constant, and changes in tastes are viewed as a change in relative importance of the several motives of the individual. The consideration of different or changing motivational patterns is especially relevant for several reasons.

First, individuals may find themselves subject to the same stimuli at different points in time and because their motivational pattern may have changed, their behavior may differ in the two situations. Such differences are inexplicable if global rationality is assumed together with constant tastes. Explicit recognition of a multiple-motives behavior (as distinct from maximization) provides some avenues for empirical research on changes that may have occurred. A single-motive theory of behavior overlooks potential conflicts among motives and, consequently, ignores the resolution of these conflicts and the issuing changes in the motivational pattern.

Second, some conflict may occur between different individuals belonging to a group having to make a decision. Economists sometimes assume away the problem of the aggregation of individual preferences in the case of group decision either by assuming a group utility function or by evading conflict via the Pareto criterion. Assuming a group utility function ignores an important factor of the decision to come: the aggregation of individual motivational patterns. The explicit recognition of conflict between motivational patterns leads to empirical research on how these conflicts are resolved. Game theory is an attempt to handle problems of individual behavior in conflictual situation. It is however, normative, rather than descriptive, and assumes that individuals have very high reasoning power.

Finally, different individuals may have different motivational patterns, i.e., they may have arranged the same basic motives in a different whole. Thus, in the same situation they will behave differently. This implies that the way different motives are articulated must be studied if differences between individuals' behavior patterns are to be explained.

Thus, Katona concluded that two empirically relevant fields of investigation are overlooked when the postulate of maximizing behavior, or global rationality, is accepted: (1) the relation between motivational patterns and behavior, and (2) the change in motivational patterns over time. In contrast, a multiple-motives theory of behavior opens the way for a study of the above points.[16]

A General Model of Individual Behavior

Katona proposed a model of "adaptive behavior" which differs substantially from the model of individual behavior underlying standard economic theory. The principles of adaptive behavior are as follows:

> 1. Human response is a function both of changes in the environment (stimuli) and the "person"....

---

16 Ibid., p. 314.

2. Individuals (and families) function as parts of broader groups. ...
3. Wants are not static. Levels of aspiration are not given once for all time....
4. ... habitual behavior prevails. Careful weighing of alternatives and choosing what appears most appropriate among the perceived alternatives is not an everyday occurence.[17]

This model of adaptive behavior is not based on any a priori principle of rationality or consistent behavior; it is not, therefore, a normative model of behavior. The principles of adaptive behavior, on the contrary, emerged from empirical studies of actual behavior. Although this model has been developed for consumer behavior, it is argued here that the basic principles have a wider application. Most of Katona's work was in relation to short term fluctuation of demand for goods; as a consequence, the questions arise whether the above model of adaptive behavior has been applied to all its potential relevant range, and whether some modifications are needed to tackle other problems than short run cyclical changes in consumers' demand.

Simon proposed a "behavioral model of rational choice" which attempted also to bring more realism to the individual model of behavior used in economics.[18] The approach is, however, substantially different from Katona's. It consists essentially of substituting for the "global rationality" of the classical model an "approximate" rationality which takes into account the limited access to information and computing capacity of individuals. The "approximate" rationality is not considered

---

17 George Katona, "Consumer Behavior: Theory and Findings on Expectations and Aspirations," American Economic Review 58 (May 1968): 20-21.

18 Herbert A. Simon, "A Behavioral Model of Rational Choice," Quarterly Journal of Economics 69 (February 1955): 99-118.

as describing actual behavior, as in Katona's model; it does, however, incorporate some simplifying behavioral features, which are, actually used by individuals in order to overcome their cognitive and computational limitations. Thus, increased realism is achieved without the pretence of actually representing individual behavior.

Although Katona and Simon proceeded from different starting points, their models bear much resemblance and share some features. In the following sub-sections the specific points of these models will be introduced and discussed, especially as they relate to the decision to leave farming.

Habitual Behavior and Genuine Decisions

Katona introduced and stressed the distinction between what he called "routine behavior" and "genuine decisions".[19] Classical economic theory, through the assumption of "global rationality", implies that all economic agents' actions are preceded by a genuine decision. In contrast to this a priori proposition, empirical observation shows that habitual behavior is the rule, at least for consumers. Habitual behavior is a form of behavior where all alternatives for action are not weighed nor even considered before behavior takes place. When circumstances are similar to some previously encountered circumstances, the same behavior is elicited. Unless one adheres to the very strict definition of rationality, habitual behavior

---

[19] See for example George Katona, "Psychological Analysis of Business Decisions and Expectations," American Economic Review 36 (March 1946): 46-51; and Katona, "Rational Behavior and Economic Behavior," pp. 309-313.

cannot be considered as "irrational" since it takes implicitly into account the consequences of previous actions. In many ways, the existence of habitual behavior is consistent with the radical behaviorist theory of "operant conditioning". Habitual behavior, if examined in a context of incomplete and costly information, could possibly be considered to be related to a broader concept of rationality; this aspect will be discussed further in connection with the cognitive aspects of behavior.

Thus, it is asserted that habitual behavior is the rule and that very specific conditions are required to enable genuine decisions to occur. These conditions consist of an unusual situation for which no habitual behavior has been developed; in this case the individual is faced with a problem which he has to solve. Genuine decisions are also made when the expected consequences of the behavior are potentially either very favorable or detrimental, or in game theoretical language, when the "regret" can be large; such situations are rather seldom encountered.

The above distinction between habitual behavior and genuine decision is considered here to be of substantial theoretical as well as empirical interest in the area of occupational mobility. This is discussed below in relation to off-farm movement and migration.

Any active individual must at some early time of his life choose to engage himself in some occupation. Reasons for this choice and circumstances of this choice need not be specified at this time. Hence, some active individuals become farm operators. The activity of farming, of being a farm operator, is considered here as a specific example of an <u>habitual behavior</u>. It is habitual in the sense that consequences of farming and of other alternative occupations are not considered, examined, scrutinized, weighed every day nor even every year. Farm operators do not decide to continue farming, they just continue to farm. Some, in fact many, farm operators, however, do decide to leave farming. It is asserted here that such a behavior results from a genuine decision. Circumstances elliciting such a decision are unusual and conseqences of such a behavior are far-reaching and involve not only the professional life of the operator but, in general, his private and familial life. The decision to leave farming is not taken on the spur of the moment but certainly involves a long and laborious decision making process.

Leaving farming, thus, comes as the outcome of a genuine decision and as a rupture of an habitual behavior. The question immediately arises as to what circumstances may render a genuine decision possible where before only habitual behavior prevailed. Some drastic changes in the environment or some cumulative effect reaching a threshold can be hypothesized. Theoretical reasoning should yield little further insight into what is essentially an empirical question. Thus, the distinction between habitual behavior and genuine decision leads up to a whole new area of empirical research: the investigation of the circumstances which prompt a genuine decision. It is fundamental to grasp that, irrespective of the actual alternatives, a farmer will remain in farming if the prerequirements for a genuine decision do not exist. In this sense behavior depends as much on the subjective readiness to take a decision as on the actual alternatives open for choice.

Level of Aspiration and Satisficing Behavior

Concepts

It is not intended, here, to provide the details of different psychological theories of motivation, but rather to present the common thread which is running through them. According to psychological theories of motivation, action by an individual stems from drives, defined as felt needs; action ceases when the drives, i.e., the needs, are fulfilled. Drives and needs are of an all-or-nothing nature. Drives are fixed at a point in time but may vary through time; this will be examined below.

From the brief foregoing presentation, it is already apparent that psychological theories of motivation differ on a very important point from the model of the economic man: the concept of satiation, which is entirely foreign to standard economic theory based on maximizing behavior,is central to psychological theory of motivation. Psychology considers that an individual acts to fulfill some fixed (and limited) needs where standard economic theory assumes that individuals try to reach as high a point in their satisfaction scale as it is possible. Psychological theories view individual behavior as satisficing, where standard economic theory views it as maximizing.

As mentioned earlier, needs are fixed at any point in time but their level adjusts through time as the individual's level of aspiration adjusts itself. Katona summarized this process as follows:

> 1. Aspirations are not static, they are not established once and for all time.
> 2. Aspirations tend to grow with achievement and decline with failure.

3. Aspirations are influenced by the performance of other members of the group to which one belongs and by that of reference groups.[20]

In a formalized way, goals that are chosen by individuals are "a joint-function of one's estimated ability to reach a goal and the pleasure or satisfaction estimated to result from having achieved it."[21]

Some people have argued that a goal-striving behavior (satisficing behavior), where goals are themselves adjusting upwards with successes or downwards with failures, is equivalent to a maximizing behavior. Such argument would be valid if adjustment in level of aspiration were instantaneous; but, to assume such instantaneous adjustment is to negate the fundamentally dynamic nature of the satisficing model with adjusting level of aspiration. The model, as it stands, includes two different processes, namely, reaching the goals and adjusting goals to past performances. Time is an important variable, and some lag exists between the achievement of goals and the resulting upward adjustment of aspirations. Furthermore, the environment in which individual behavior takes place is constantly evolving: new alternatives for action become possible and other alternatives become impossible; as a result of this changing environment and the time-consuming adjustment of expectations, goals

---

[20] Katona, "Rational Behavior and Economic Behavior," p. 316.

[21] Irving E. Alexander et al., "The Level of Aspiration Model Applied to Occupational Preference," Human Relations 12 (May 1959): 163.

do not coincide with attainable maxima. It must be concluded that a goal-striving (satisficing) model of behavior is not equivalent to a maximizing model of behavior.

It was argued above that the satisficing model of behavior is dynamic is essence; this is seen by Simon as the main reason for the superiority of satisficing models and the strongest argument for prefering them:

> Models of satisficing behavior are richer than models of maximizing behavior, because they treat not only of equilibrium but of the method of reaching it as well. Psychological studies of the formation and change of aspiration levels support propositions of the following kinds.(a) When performance falls short of the level of aspiration, search behavior ... is induced. (b) At the same time, the level of aspiration begins to adjust itself downwards ... (c) If the two mechanisms just listed operate too slowly to adapt aspirations to performance, emotional behavior -- apathy or aggression, for examples -- will replace rational adaptive behavior.[22]

Another important feature of the goal-striving model is that it can accommodate multiple goals whereas a maximizing model, as it was explained earlier, is intrinsically a single-motive model. A goal-striving model allows multiciplity of goals, unequal satisfaction of all these goals and even conflict between goals. Such a model is potentially more able to represent goup decisions when multiple and conflicting goals are the rule.

Implications for Off-Farm Movement

<u>A group decision</u>. It was explained earlier why the decision to leave farming should be considered as a family decision, i.e., a group decision. No attempt is made here to review theories of group-decisions; it suffices to stress that the fundamental difference between

---

[22] Simon, "Theories of Decision Making," p. 263.

models of group-decision and the standard economic model of individual behavior lies in the acknowledgement of conflicting goals which leads to hypotheses about and investigation into the nature of the process by which these conflicts are resolved.

The adoption of a satisficing model of behavior will lead the researcher to investigate goals held by the farm operator, his wife and other members of the family. Invesigation of these goals will likely reveal conflicts: farm operators are usually considered to be primarily concerned with farm-business matters and their wives with family matters. Such a diversity in motivational patterns is highly relevant to the decision to leave farming. Even more important for the decision to leave farming is the way these goals are articulated in the family; this can be investigated empirically. Since goals can change over time, whatever factors might modify one family member's goals and render them consistent with other family members' goals will resolve the conflict. But, whether one family member or the other changes his goals is important to ensuing behavior; for example, if a farmer's wife gives up her goals concerning nonfarm work, this will bear on whether or not the farm operator will remain in farming.

In summary, the identification of goal conflicts and of factors likely to resolve the conflict one way or another is an important element of a study of the pre-retirement decision to leave farming. If these factors influence only one family at a time, the results will allow improved forecast decisions and will be helpful for extension programs dealing with potential exiters; if these factors affect large numbers of farm families at the same time, these results will be useful for aggregate prediction of the number of exiters as well as for policy analysis.

A satisfied farm-family. Let it be assumed that all major goals of a farm family are fulfilled; the family can be said to be satisfied. In such a case, when a satisficing model of behavior is assumed, the family will not decide to leave farming irrespective of what other alternatives are available in other occupations and locations. Going even further, it can be hypothesized that these alternatives will not even be thought of or considered.

The above argument provides the basis for a test since the goals of a family can be identified and their degree of satisfaction can be measured (on a scale): according to a satisficing model of behavior, a satisfied farm-family will not be considering off-farm movement or migration whereas, according to a maximizing model of behavior, a farm family, independently of its level of satisfaction, will weigh costs and benefits of leaving farming. Thus, when maximizing behavior is assumed, the decision to leave farming depends on both the family's preferences (assuming them unambiguously defined) and the differences between benefits derived from available occupational alternatives. When a satisficing behavior is assumed, the level of benefits derived from farming is a crucial conditioning variable: if this level is sufficient, differences in benefits between other alternatives and farming play no role and the operator will stay in farming.

From the foregoing discussion, it appears that a satisficing model of behavior is consistent with the occurence of habitual behavior, whereas maximizing behavior is not.[23] The same habitual behavior (e.g. farming) continues as long as the individual (or the farm-family) is satisfied; it is only when some of his goals cannot be fulfilled that dissatisfaction appears, and that other alternatives come to be considered; this may or may not lead to a change of behavior.

Farm-families' level of aspiration. It was explained above how the level of aspiration of an individual bears on his behavior. It was also stated that the level of aspiration rises with successes and declines with failures, thereby providing an adapting mechanism, and that it is influenced by performances of other individuals belonging to the same group or to reference groups.

The proposition concerning the influence of past experiences on the level of aspiration is particularly relevant to off-farm movement. Among farmers' aspirations are some aspirations concerning income. Low-income farmers, because of their many failures to improve their income position, should have, according to the level of aspiration theory, low income-aspirations. As a consequence, low income would not constitute an incentive for off-farm movement or migration.

---

[23] It is true that a maximizing behavior can be made consistent with any behavior by resorting to unknown adjustment costs, uncertainty regarding success in alternative jobs, psychic costs and benefits for different alternatives, etc. ...; it must be understood that, if this is done, the causes or factors of the action taken are explained by, or inferred from, observed behavior itself which is contrary to the aim of the model to explain behavior by its hypothesized causes. Furthermore, the inferred factors are not easily identified and measured. Thus, a satisficing modeloffers a more operational basis for experimental studies.

Downwards adjustment of the level of aspiration to match the actual level of income allows such low-income farm families to be satisfied and therefore they would not seriously consider other occupational and locational alternatives; despite the availability of an occupation with a higher income, they would continue to stay and farm. Since this downward adjustment of the level of aspiration lags behind performances, only families having received low incomes for a long period of time could be satisfied. A sudden and drastic drop in income, would not be followed immediately by a corresponding downward adjustment of the level of aspiration; the income record of a farm-family thus provides information on the present level of aspiration concerning income. It is seen that a satisficing model of behavior is not inconsistent with income differentials between farm and nonfarm sectors, nor with income differentials between regions.

Level of aspiration depends also on performance of individuals belonging to the same group or reference groups. If it is assumed that low-income farmers consider themselves as belonging to the group made of the farmers of the area, their level of aspiration (concerning income) will be positively affected by higher average income in the area. Thus, low-income farmers' levels of aspiration will differ according to whether they are located in a high-income or a low-income farming area. The same influence can be hypothesized between the average income of the reference group and the level of aspiration. The extent to which farmers take urban communities as their reference groups can be hypothesized to be inversely related to the distance to an urban center or directly related to the number and frequency of exchanges and contacts with an urban center. Given that urban centers are usually characterized by higher monetary income, the level of income aspirations of farmers would be higher the closer they are to an urban center.

Identification of farmers' reference groups and their description in terms of past-performances will shed some light on farmers' levels of aspiration and potential for off-farm movement. Regional analyses may show that farmers in different regions or sub-regions have different reference groups, with different levels of income; this could explain the existence of different levels of aspiration and lack of incentive for off-farm movement and migration. Distance to urban centers can be hypothesized to be negatively related to off-farm movement and migration via its effect on level of aspiration. Consideration of imperfect information can also lead to the same hypothesis on the relationship between off-farm movement and distance to urban centers; this will be examined in the next sub-section.

Conclusion. A satisficing model of behavior displays a number of advantages over a maximizing model of behavior:

1. It is potentially more useful in analyzing how individual goals are articulated in a group situation (e.g. in a family) and how group decisions are taken.

2. When goals are considered to depend on a level of aspiration, itself a function of past-performance of the farmers or of individuals in reference groups, a satisficing model explains the lack of off-farm movement and migration inspite of substantial farm-nonfarm and urban-rural income differentials. This phenomenon has usually been considered as an abnormality in a standard economics context based on a maximizing model of individual behavior.

3. It does not resort to a residual such as "psychic income" to explain behavior, and is therefore more adapted to empirical studies.

Cognitive Aspects of Behavior

The major contention made in this sub-section is that individual behavior is dependent on or elicited by perceived changes in reality; this entails that the study of the process of perception or cognition is of paramount importance to understand and explain individual behavior. This contention can be explained further by stating what it rules out: (1) behavior is not elicited by objective reality or, in other words, a change in objective reality is truly a stimulus only insofar as this change is perceived by the individual, (2) behavior is not considered capricious or random since it is elicited by changes in real conditions (as they are perceived). Thus, both the changes in the environment and the perception of these changes play a role in individual behavior.

Cognition as Limited Knowledge

As it was explained earlier, the standard model of individual behavior ignores all cognitive aspects of behavior. In recent years though, economists have attempted to relax the assumptions of perfect knowledge and foresight which appeared as more and more unrealistic; these assumptions were eliminating from the analysis important economic actions such as job-search, information gathering, speculation and management.

These efforts have resulted in two main approaches: (1) the Bernoullian decision theory whose development is concomitant with a revival of cardinal utility theory and a renewed interest in Bayesian theories of probability.[24] (2) the information-as-a-commodity approach which retains the deterministic framework but simply includes information as an additional commodity; information is then produced and exchanged just as another commodity. This approach may be theoretically interesting but is of very little practical interest.[25] In both cases, economists have tackled the problem of cognition in a very formal way: some axiomatic model of individual behavior (rational or consistent) is assumed without any reference to empirical evidence; such models of individual behavior are, again, normative rather than positive.

These approaches to uncertainty are aptly called: economics with limited knowledge; the word limited expressing that the difference between the objective reality and perceived reality is a matter of a quantitative lack of information; this implies that by securing more of the same kind of information this gap is reducible. Uncertainty and limited knowledge are concepts related to a strictly quantitative conception of information and cognition.[26]

---

[24] For a good everview of Bernoullian decision theory see Dillon, "Review of Bernoullian Decision Theory."

[25] The shape of the production function with respect to information is supposed to be known (with certainty). The problem of measuring information, which is essential for any application of the theory, is evaded.

[26] It should be stressed that Knight's classical distinction between situations of uncertainty and of risk is negated by the Bernoullian theory of decision. Subjective probabilities assigned by the individual to events are a measure of the uncertainty attached to these prospects; no reference to (objective) probabilities of these prospects is needed.

Uncertainty (or limited knowledge) can occur at different levels. First, an individual may not know of all possible courses of action which are available to him, though he may know that more are available to him than he is aware. Second, some consequences of the action chosen may be completely ignored or some mutually exclusive consequences may be known as possible without knowing which one will actually occur. The mere fact that the individual realizes that he has a limited knowledge is also very important, since, in this case, he can decide to look for, buy or produce more information.

The approaches described above, despite their limitations, are successful in making such phenomena as search for nonfarm opportunities, return migration, and information disseminating public programs consistent with the standard economic model of economic behavior.

Search for nonfarm occupation alternatives is possible, because the farm operator knows that some nonfarm occupations are available, and necessary because he has no immediate and complete knowledge of them. The fact that the labor market is not characterized by the exchange of a homogenous product with a well-known price, together with the existence of limited knowledge imposes search of both sellers and buyers of labor.

Search itself is a costly process since it uses resources; one of the main elements of the cost of search is the opportunity cost of time spent in search activities. The assumption of a deterministic relation between information and production (i.e. ultimately returns) is only a way to displace the incidence of uncertainty. Actually, the relationships between search and information and additional information and returns are not deterministic. Individuals take into account this element of uncertainty in their strategy for job-search. The information-as-a-commodity approach fails to account for the important role of strategy in search. The strategy followed by farm operators in searching for nonfarm occupations will heavily depend on their psychological make-up.

Blau and others[27] proposed a conceptual scheme of occupational choice which accounts for uncertainty and which can be adapted to off-farm movement, considered as a change of occupation. Farm operators seriously considering leaving farming have some preferences concerning nonfarm occupations in which they would engage; these preferences may be conceived as being a ranking of these occupations. Whether or not they will actually enter the occupation ranking the highest in their preference does not depend only on their choice but also on someone else's choice. In most cases, actual opportunities of farm operators are determined by other agents who can be aggregated under the label "selector".[28] At every moment farm operators do not know with certainty these alternatives. Some subjective probability

---

[27] Peter M. Blau et al., "Occupational Choice: A Conceptual Framework," in Personality and Social Systems, eds. Neil J. Smelser and William J. Smelser (New York: John Wiley and Sons, 1970), pp. 559-571.

[28] Thus, the term "selector" is used here as a generic name meaning all individuals, practices, agencies and institutions which have some power to choose among job applicants and on which job applicants have no control.

of actually getting a job is attached to each of these occupations. The actual nonfarm occupational choice is a compromise between the farm operators' preferences and these likelihoods of getting a job. These probabilities are certainly revisable on the basis of experience, and this revision is an essential element of job search which bears heavily on the strategy followed. Failures to secure jobs in occupations ranking high in their preferences will force farm operators to look for lower-ranking occupations; as a consequence, the level of aspiration concerning occupational choice will be lowered. Since search is a process, time is also an essential element of job-search: how long will (or should?) a farm operator wait until he adjusts downwards his level of aspiration? All these elements of job-search bear directly on whether the pre-retirement decision to leave farming is made or not.

Engagement of a farm operator in part-time nonfarm work is usually deemed to increase off-farm mobility, because information on nonfarm occupational opportunities is a by-product of this engagement, at little or no cost to the farm operator. Viewed in this way, off-farm work may not facilitate off-farm movement, because, as it was stressed before, the nature of this information on nonfarm occupation may not be conducive to off-farm movement. What can, however, reasonably be assumed is that farmers with off-farm experience are likely to make better informed choices, and that consequently less return movement or migration will occur among them. Return movement and migration are phenomena yet very little investigated, but whose existence has been

been proven.[29] Such return movement is inexplicable in a world of certainty, maximizing behavior and stable preferences, but, it is explicable within the context of Bernoullian decision theory. An individual may make the right choice in term of some criterion (e.g. expected value of utility) and, still, consequences may yield a lower utility, thereby justifying a reversal of the previous decision.

Dissemination in farming areas of information concerning the nonfarm labor market and living conditions in urban areas influences decisions made by farm operators in two ways: (1) it provides them with knowledge of nonfarm job opportunities which they did not have before, and (2) it provides them with more precise expectations concerning the consequences of off-farm movement. Thus, dissemination of information is a policy variable. But too many have assumed the direction of the effects of increased information to farm operators. A priori, better known opportunities and clearer expectations as to the consequences of a pre-retirement decision to leave farming could favor remaining in farming. The effect of such increased information depends on how this information is interpreted by the individual farm operators.

---

[29]See for example: Dale E. Hathaway and Brian B. Perkins, "Occupational Mobility and Migration from Agriculture," In Rural Poverty in the United States, A Report by the President's National Advisory Commission on Rural Poverty (Washington: Government Printing Office, 1968), pp. 185-237; and Eldon D. Smith, "Nonfarm Employment Information for Rural People," Journal of Farm Economics 38 (August 1956): 813-827.

In summary, the approach taken by traditional economists to tackle the cognitive aspects of behavior, has been through the use of a strictly quantitative concept of information. In the following subsection different concepts and conclusions are presented, inspired by a school of psychology, known as Gestalt psychology.

The Gestalt Theory of Perception

Katona summarized the main results of the Gestalt psychology on the importance of perception for human behavior:

> The response is determined by what the stimulus means to the respondent; it changes when the meaning of the stimulus changes. Meanings are not just a matter of subjective interpretation. It is the setting or context of the stimulus, the greater whole of which it forms a part, which determines the meaning of the stimulus. The same stimulus may elicit different responses if it is perceived or understood as the part of the one or other whole.[30]

Katona goes on to explain that, since the individual's reaction depends on the meaning attached to the stimulus, the question of how meaning is acquired will yield the answer to the question of what determines reaction. Meaning can be acquired through mere repetition or through understanding. Understanding implies organizing or re-organizing different parts into a new whole where individual parts acquire a new meaning. Connections, thus, appear between habitual behavior and meaning acquired through repetition, and genuine decisions and meaning acquired through understanding. The distinctive feature of a genuine decision consists in that it is preceded by a cognitive phase where all elements are re-organized into a new whole. This re-organization of elementary stimuli determines how reality or changes in reality are perceived and constitutes the basis of any true decision.

[30] Katona, "Business Decisions and Expectations," pp. 47-48.

The most important insight to be retained from Gestalt psychology is that perception of reality is not solely incomplete i.e. a mixture of known and unknown aspects of reality; perception is the process by which some meaning is conferred to reality through the organization of individual elements. Viewed in this context, cognitive aspects of human behavior, and more specifically of the decision to leave farming, appear more complex than when they are viewed in a context of limited knowledge.

Implications for an empirical study of the decision to leave farming are substantive. Given that farm operators have elements of information on non-farm opportunities and on consequences attached to them, the question arises as to what factors influence the organization of these elements into a whole which renders nonfarm opportunities attractive or unattractive. More important, when such a whole has a meaning which led them to stay in farming, what factors or events will precipitate the re-organization of this whole, leading possibly to the decision to leave farming? Thus, re-appearing at another level, is the already mentioned issue concerning the factors that lead a farm operator (and his family) to a situation where a genuine decision is possible. It is argued here that very little is known on these organization and re-organization processes. This seems a whole new field for investigation, lying largely outside the field of economics stricto-sensu, but whose results would be of paramount importance for the specific decision studied here.

The Time Perspective

Studies in psychology have shown that, while making choices, people take into consideration facts and consequences of their choice belonging to a limited span of time; this span of time is commonly called the "time perspective" of the individual. As Katona wrote: "our time-space ... as of a given moment encompasses some of our past experiences, our perceptions of the present, and our attitudes toward the future."[31]

It was argued earlier that an essential characteristic of off-farm movement or migration lies in their drastic and lasting consequences on all aspects of the farm-family's life. If such contention is justified, the ability to take into account all or only a limited part of the consequences, will bear on the decision taken. Two main types of empirical investigations can be conducted in relation to the time perspective. First, the length of the time perspective may be related to certain socio-economic characteristics of the farm-family. Second, some exogenous factors may affect the length of the time perspective of many farm operators and farm-family at the same time.

One potential factor affecting time perspective could be age: very young adults may tend not to take into account long-term consequences, where mature adults with children may be more concerned with the long-term. Also, to the extent that low-income farmers are always concerned by short-term income inadequacies, they would not develop the habit and skill to weigh long-term consequences of a decision: their time perspective, thus, would be short.

[31] Katona, Psychological Economics, p. 45.

Inasmuch as benefits from off-farm movement or migration are captured on the long run, a short forward time perspective would tend to reduce the mobility of very young adults and low-income farmers.

Unfortunately, very little further help can be provided by theory when this point is reached. Any further progress depends on empirical evidence. As in many fields of psychology, the researcher has to grope for possible relationships using extensively exploratory data analysis.[32]

Social Aspects of Behavior

Search as Social Behavior

Search for nonfarm occupation has already been mentioned, not explicitly as a form of social behavior, but as a necessary consequence of imperfect knowledge. In the process of nonfarm occupational choice, the farm operator is facing what was called a "selector" which restricts the opportunities that are available to him. In economic parlance: the "selector" determines the opportunity set of each individual; uncertainty about alternatives, however, empties the concept of opportunity set of its precise meaning; actual nonfarm opportunities come to be known and available to the farmer as a result of the dynamic process of search and social interaction with the "selector". Empirical study of this process would provide useful insight on how the decision to leave farming is made.

---

[32] For a detailed justification of explanatory data analysis, see chapter 4.

Attachment to the Community

Another major social aspect of the decision to leave farming is the attachment to the local community and, more restrictively, to friends and relatives who leave close to where the farm is. It was mentioned earlier that off-farm movement and migration very often occur concomitantly although they must be conceptually isolated. Attachment to the local community, plays very little role when off-farm movement without migration is considered; it dissuades farm operators and family from migrating but not from looking for nonfarm opportunities which would allow them to retain their home or to live in the same community. What attachment performs is, thus, a lowering of the value attached to remote nonfarm job opportunities, possibly to a point where these may not even be considered. Attachment to the community in this latter case, reduces the opportunity set of the farm operator.

The importance of attachment to the community and familiar surroundings has been recognized for a long time by demographers studying migration, but also by economists who usually integrate it in their study by making it part of a residual quantity called "psychic income." This way of dealing with attachment to community appears as inadequate because it merges different factors which deserves to be considered separately. Thus, recourse to the concept of psychic income detract from incisive empirical analysis.

What needs to be done is to create some measure of this attachment (not necessarily at a ratio-level or interval-level of measurement) and to study empirically how this attachment is related to more tractable socio-economic variables such as age, composition of the family, religion, income, etc. ... Investigation of the extent to which this attachment prevents farm families to even consider migrating, would provide information on whether a genuine decision regarding off-farm movement or migration is possible.

Farming as a Social Activity

It is common among economists, to explain the absence of off-farm movement and migration by some "psychic income" drawn from farming activity. This practice is inadequate for the reasons mentioned in relation to attachment to the community.

It is recognized that the occupation in which one is engaged is an essential element of the individual's position in society. To different occupations correspond certain "social roles" that one has to fulfill. Others have definite expectations concerning the behavior of individuals fulfilling specific social roles. Also, social roles contribute to define the individuals' self-image. The major consequence bearing on the present topic is that an individual's attachment to his occupation, and especially when it is farming, is very real.

The way to incorporate attachment to farming is similar to what was proposed concerning attachment to the community: using a measure of attachment to farming, it is possible to relate this variable to other socio-economic factors and the eventual continuation of farming activities or the decision to leave farming.

Summary and Implications

In this section, the highlights of the theoretical framework developed above are summarized and implications for empirical study are outlined.

Farming is a form of habitual behavior. To leave farming and engage in a nonfarm occupation is the result of a genuine decision. The process by which a farmer evolves from conditions under which habitual behavior takes place to conditions under which a genuine decision concerning occupation is made is ill-known. Such a transition from routine behavior to genuine decision is consistent with a satisficing model of individual behavior and with results of Gestalt psychology concerning perception. This implies that empirical study of this transition and of the factors eliciting and favouring it would be useful.

Individuals should be considered as having a satisficing behavior. They possess a set of goals which they strive to attain; once these goals are attained behavior related to the attainment of these goals stops. These goals are not fixed but flexible upwards or downwards. The level of aspiration defining the level of these elementary goals adjusts downwards or upwards as the individual experiences failures or successes; this level of aspiration is also positively affected by the performances of other individuals belonging to the same group or to reference groups. Such a satisficing and level of aspiration model entails that low income is not, *per se*, an incentive to leave farming because persisting low income entails a corresponding downwards adjustments of the level of aspiration. Empirical study of the income history of each individual and identification of group belonging and reference groups will provide information on the level of aspiration of individuals and consequently on the extent to which their goals are fulfilled.

The relevant decision unit with respect to off-farm movement and migration is the family; therefore, the decision to leave farming is to be considered as a group decision. Most likely, there are some variations in the collegiality of the decisions; study of these variations are needed. Members of a family have different sets of goals which may be partially conflicting. Investigation of these different sets of goals, of their conflicts, and of the way these conflicts are acknowledged and resolved will shed light on the decision making process involved in off-farm movement and migration. Individual behavior is elicited by stimuli which are perceived changes of the environment. Perception is in part limited knowledge, that is, a quantitive lack of information, but is also an interpretation of many elements organized in a whole. Thus, perception is not a mere screen letting pass through only a part of reality, it is a process by which reality is given some meaning. This conception of perception would indicate the necessity to study how farm operators perceive their situation as farmers and other occupational nonfarm alternatives.

Limited knowledge and heterogeneity in the labor market entails the necessity of search, especially for job applicants. Search behavior is a complex form of behavior because strategy plays an important role in it. Search can take place before or after the decision to leave farming has been made. It is thought that most operators would not leave their farm without having found a nonfarm job or having the firm belief they can find one. This entails that some amount of search has been performed before the exit decision is taken. Only job-search performed before the decision is relevant here, as it is the only one which bears on the exit decision. The strategy which is used bears heavily on the outcome of the search and consequently on the eventual decision to leave farming. Detailed descriptive study of actual search behavior of farmers is needed.

Additional information on nonfarm job opportunities and living conditions in urban areas is not necessarily conducive to off-farm movement since it should not be assumed that other alternatives will enable farmers to better fulfill their goals. Additional information, however, is predisposing to better enlightened decision and, thus, reduces the chances of return-movement and migration. Thus, factors such as off-farm work and proximity to urban centers, may not favor off-farm movement and migration but should be expected to reduce return-migration.

Attachment to the community and to farming are elements of the decision to leave farming. The measure of these attachments, by means of scales, and the analysis of relationships between these attachments and other socio-economic variables, as well as the decision concerning off-farm movement would clarify the role they play in off-farm movement and migration.

This chapter has introduced many concepts which are seldom encountered in an applied economic study. Also, some results from positive studies of individual behavior were integrated into the above expanded theoretical framework. The obtained framework is richer, points to many directions where empirical studies would be useful, and provides concepts which are necessary to these studies. On the other hand, this framework is not as well formalized as the benefit-cost model based on maximizing behavior which underlies most economic studies of off-farm movement and migration. The approach taken here can be described as being behavioral: individual behavior is to be observed rather than assumed to be maximizing behavior.

The expanded framework proposed in this chapter, stresses the complexity of the decision to leave farming, the multiplicity of factors of this decision and the importance of the interaction between these factors. This raises the question whether an empirical study based on available secondary data can do justice to this theoretical framework. Because of time-constraints, this empirical study had to rely on secondary data. In the next chapter the data-base used is described and assessed. The adequacy of this data-base is examined further in Chapter VII, which outlines a proposal for further empirical study.

As mentioned earlier, the theoretical framework presented in this chapter draws the attention to many areas of ignorance, e.g., farm operators' perception of nonfarm employment, transition from habitual behavior to genuine decision, etc. This entails that some descriptive and exploratory empirical studies are required. This approach appears to be a departure from the hypothetico-deductive method usually adopted in economics. A methodological issue is raised, it is discussed in the second section of the next chapter.

CHAPTER IV

DATA, DATA ANALYSIS AND HYPOTHESES

This chapter is divided into three main sections. The first section is devoted to a description of the process by which the longitudinal data-base used in this study was obtained, an evaluation of this data-base in relation to the purpose of the study of off-farm movers, and an overview of the variables included in this data-base. In the second section, some methodological problems concerning inference and the adequate statistical methods to use are considered. In the third section, hypotheses, to be tested in the confirmatory part of data analysis, are presented.

The Data Base

The Census of Agriculture

The great majority of the data used in this study originated from the 1966 and 1971 Census of Agriculture. Statistics Canada conducts a Census of Agriculture every five years. This census is, alternatively, in a comprehensive and abbreviated format; thus, the 1966 census was an abbreviated one, and the 1971 census was comprehensive.

The general emphasis in the census, has been on collection of data pertaining to physical aspects of the farm production unit. A few relevant pieces of socio-economic information were, however, collected, thus enhancing the usefulness of census data for a study of the exit of farmers.

Information collected in the 1966 census concerned, generally, the 1966 crop year. Information on part-time work, selected agricultural expenditures, and sales of agricultural products, however, pertained to the twelve month period ending June 1, 1966. Also, sales of agricultural products during year 1965 could be declared if this was considered preferable by the farmer. By the same token, information collected in the 1971 census concerned, generally, the crop year 1971.

Longitudinal data-bases can be obtained by linking (matching) records pertaining to the same entity, but collected at different points in time; such longitudinal data-bases allow the study of changes undergone by these entities. These data-bases are especially valuable for empirical research in the field of migration and occupational mobility.

Prior to the undertaking of the present study, Statistics Canada had completed a match of 1966 and 1971 census records pertaining to the same farm operator.[1] The following section describes

---

[1] This longitudinal data-base will, hereafter, be referred to as the Census of Agriculture Match.

the matching procedure used by Statistics Canada, its shortcomings and the alterations made to it in order that it fits better the requirements of this study.

The Matching Procedure

The objective in creating the Census of Agriculture Match was to link census records for the same operator; in other words the objective was to match people and not holdings, parcels of land or production units. This objective was, however, not clearly formulated and stated; consequently, no strict set of matching rules, consistent with this objective, seems to have been enforced. This Census of Agriculture Match was the first large scale attempt at matching census-records and was, by nature, experimental.

The Census of Agriculture Match was performed in two stages: (1) a computer match based on name and address of operator, and (2) a manual check of computer matches and a manual match of yet unmatched records.

Many difficulties arise in creating longitudinal data-bases, especially when data are not collected with the purpose of producing such bases, as is the case for the Canada Census of Agriculture. Scheuren and Oh stressed that there is no single best matching procedure applicable irrespective of the context:

> In developing procedures for linking records from two or more sources, tradeoffs exist between two types of mistakes: (1) the bringing together of records which are for different entities (mismatches), and (2) the failure to link records which are for the same entity (erroneous non-matches). Whether or not one is able to utilize one's resources in an "optimal" way, it is almost certainly

> going to be true that in most situations of practical interest some mismatching and erroneous non-matching will be unavoidable. How to deal with these problems depends, of course to a great extent on the purposes for which the data linkage is being carried out.[2]

Matches of records obtained through some matching procedure can be of four types: truematches, mismatches, true nonmatches, and erroneous nonmatches. Causes of mismatches and erroneous nonmatches in the Census of Agriculture Match are reviewed below.

A first cause for matching errors was the already mentioned lack of a set of consistent matching rules.

A second cause for errors stemmed from the use of an inadequate list of farms on which to apply the matching procedure. The lists used were, on one hand, the list of operators of census farms in 1971 and, on the other hand, the Central Register of Farms (C.R.F.). The C.R.F. is based on an up-dated list of census farms to which farms (and operators) are added or deleted on the basis of responses to intercensal sample surveys. The important feature of this up-dating is that it is only partial; as a consequence, the up-dated C.R.F. was not a proper list of farm operators neither for 1966 (since it was up-dated) nor for any year between 1966 and 1971 since it was only <u>partially up-dated</u>. The <u>up-dated</u> C.R.F. was used for the Census of Agriculture Match instead of the 1966 list of census operators.

---

[2] Fritz Scheuren and H. Lock Oh, "Fiddling Around with Nonmatches and Mismatches," <u>Proceedings of the Social Statistics Section</u>, American Statistical Association, 1975.

A third cause for matching errors was the restrictions placed on the eligibility of farm operators for matching: the match was performed at the census division level and, as a consequence, only farm operators located in the same census division in 1966 and 1971 could possibly be matched. Thus, an operator who migrated between 1966 and 1971 from one census division to another could not be matched. Also, changes in limits of census divisions between 1966 and 1971 prevented matching some groups of farm operators.

In order to evaluate the quality of this match and to decide on the acceptability of this data-base for this study, a cross-tabulation of age in 1966 by age in 1971 was made, showing the number of matches in each cell. Provided there was no error in recording age, mismatches can be evidenced by the existence of matches with inconsistent age in 1966 or 1971. Such a method does not diagnose, however, mismatches with consistent ages and erroneous nonmatches; it is also affected by incorrect recording or capture of information. The last point is especially relevant, since random assignment of age was performed for census questionnaires where age was not reported; the distribution among age classes of this random assignment is known but its extent is not; thus, it was not possible to account for it. Random assignment of age creates matches with inconsistent ages which show in the age by age cross-tabulation.

It was found that, as far as Saskatchewan was concerned, more than ten percent of the matches in the cross-tabulation displayed inconsistent ages. These results were considered inadequate for the purpose of this study and it was decided to proceed to a thorough manual checking and rematching, using a set of clearly defined rules. This cross-classification also displayed a much higher number of matches with inconsistent ages with an age in 1971 being too low. It was hypothesized that many father-son transitions were missed in the matching procedure.

A new set of rules were decided upon. It was decided to make all farmers in Saskatchewan eligible to be matched; this reduced the chances of erroneous nonmatches due to migration. Special care was taken in devising rules which will limit the number of errors in matching (mismatches) due to father-son transitions; it was considered that by having more stringent matching rules involving first name, initials, and order of given names and initials, this goal would be attained.

The end product of this modified matching procedure was evaluated with the same age in 1966 by age in 1971 cross-classification. Matches with inconsistent ages represented 6.39 percent of the total number of matches. This is still quite high, but represents a substantial improvement on the previous data-base. Furthermore, as mentioned before, the random assignment of age implies the existence of truematches among cases with inconsistent ages.

The above section was devoted to the description of the matching procedure and its appraisal, independently of its use. In the next section factors affecting the suitability of the Census of Agriculture Match for the purpose of the study of the pre-retirement exit of farm operators are examined.

Suitability of the Census of Agriculture Match for a Study of Off-Farm Movement

The validity of the Census of Agriculture Match in relation to a study of off-farm movement can be threatened in two ways: (1) a poor quality of the matching procedure which entails many matching errors, and (2) the under-enumeration of censuses.

Matching Errors

Causes of errors and ways to minimize them have been discussed in general in the previous section. Matching errors can be either mismatches or erroneous nonmatches. A mismatch can involve matching a farmer who left agriculture (exiter) to a new farmer (entrant); then an exit and an entry are suppressed, and information pertaining to an exiter and an entrant are brought together as if they pertained to the same operator. A mismatch can involve matching an exiter to some operator who remained an operator during the period (stayer); this will imply that information pertaining to the stayer in the 1966 census is not brought together with the information of the 1971 census and that unrelated census records are linked; such a mismatch suppresses an exit and alters the information related to a stayer. A mismatch can involve two stayers: then the number of entries, exit and stayer may or may not be altered depending on whether the erroneously unmatched records are matched; in any case, information on stayers is degraded. An erroneous non-match can appear, as it should be clear from the above, as a consequence to a mismatch. It can also arise independently of any mismatch; in this case an exit and an entry are mistakenly created and information related to exiters, entrants and stayers is degraded. Thus it appears that matching errors can alter the number of actual exiters, stayers and entrants and vitiate the data pertaining to these categories in many different ways.

Under-enumeration

It is recognized that, despite all efforts, a census does not enumerate all farms which qualify as a census farm. This discrepancy is called under-enumeration. If under-enumeration is constant (in number) from one census to another, but concerns different sets of farmers, some farm operators cannot be matched even though they stayed in agriculture during the period. This entails the appearance of erroneous entries and exits with its detrimental consequences on estimates of entrants, stayers and exiters, and on information pertaining to these categories. If under-enumeration increases from one census to the next, erroneous exits are created and if under-enumeration decreases, erroneous entries are created.

Statistics Canada performed a Quality Check Survey, immediately after the 1966 and 1971 censuses; these provide an estimate (subject to sampling error) of under-enumerations. Relevant information is shown in Table 4 below. It appears that, as far as the Prairie Provinces are concerned, under-enumeration increased substantially between 1966 and 1971. If this were true, the Census of Agriculture Match would over-estimate the number of exiters, thus entailing a decreased quality in the information pertaining to this class of farm operators.

It is usually considered that, most of the under-enumeration involves small farms. If this were true, some of the distortion in the data could be eliminated by discarding from the analysis the very small holdings; the cut-off point, though, would have to be arbitrary. Another factor seems to have contributed to the increase in under-enumeration in the Prairies: the drastic increase in non-resident farmers. This entailed increased difficulties for the enumerators in interviewing farm operators. The under-enumeration of non-resident farmers may have been especially detrimental to this study, since it was attempted to relate non-residence to off-farm movement.

TABLE 4.--Comparison of Census Totals and Agriculture Quality Check, 1966, 1971, Prairie Provinces

	1966	1971
Census Totals	194,844	174,653
Agriculture Quality Check	197,990	187,900
Net Error		
Total	3,146	13,247
Percentage of Agriculture Quality Check	1.6	7.1
Sampling Error		
Total	733	7,800
Percentage of Agriculture Quality Check	0.4	4.2

SOURCE: Statistics Canada, Census of Agriculture, Catalogue 96-709, Vol. IV, Part 3, and Catalogue 96-609, Vol. V, (5-2).

Conclusion

In spite of the above threats to the validity of the Census of Agriculture Match (matching errors and under-enumeration), it was deemed that the data-base was of potentially fruitful use in an empirical study of factors affecting pre-retirement exit from farming. The exact content of this data-base and its limitations are reviewed in the following subsection.

Content of the Data-Base

Sampling Procedure

The Census of Agriculture Match included all Saskatchewan census-farms; it comprised 61,826 matches (stayers), 23,685 unmatched 1966 records (exiters), and 14,877 unmatched 1971 records (entrants). Records concerning entrants were discarded from the data-base since this study does not deal with the entry phenomenon. Because of the large size of this data-base, it was necessary to take appropriate samples. One sample, with an approximate size of one thousand was taken for every census division in Saskatchewan. One sample, with an approximate size of five thousand, was taken for Saskatchewan as a whole. All samples were stratified according to stayers and exiters, using the same sample proportion for the two groups. Such proportionate stratified samples display greater precision than simple random sampling. Samples were drawn using the SPSS-package, which does not allow the experimenter to control exactly the size of the sample.

Variables in the Data-Base

The complete list of variables included in the data-base is presented in Appendix A, Table A.1. These variables can be considered as falling into the three following categories: census variables, recodings and transformations of census variables, and variables extraneous to the census.

A selection of variables available from the Census of Agriculture Match was made on the basis of their potential relevance for the study of pre-retirement exiters. Variables describing the structure of the farm in 1966 and variables related to the operator himself were retained. Less detailed information was kept from the 1971 census, since the comparison of stayers and exiters was to be based on the common information available for 1966. The main variables are: residence on the farm, age-class of operator, total area of farm, area owned, area rented, value of land and buildings, value of all machinery, value of livestock, total capital value, days worked off-holding, total rent, total value of agricultural products sales, type of farm, and tenure.

Recoding of some variables was performed to comply with the requirement for binary dependent variable by one of the program used. Ratios of census variables were computed to be used as proxies for degree of mechanization, productivity of capital, and productivity of land.[3]

Variables not originating from the census consist of distances to four classes of towns: 2,000 to 4,000; 4,000 to 10,000; 10,000 to 40,000; over 40,000 inhabitants (ie: Saskatoon and Regina). Distances were estimated from the centre of each census subdivision of Saskatchewan to the nearest town of each class. Then, these values were assigned to all farms located in this census subdivision.

---

3 For more details, see Appendix A, Table A.1.

Conclusion

A new longitudinal data-base was assessed with respect to the purpose of this study. It was deemed to be partially inadequate, modified, and improved substantially. The variables included in this data-base were, however, limited in scope. Only distances to nearest urban center could be added to the information on stayers and exiters.

Despite very serious limitations and partly because of the novelty of this data-base, it was thought to be worthwhile to proceed with the empirical analysis.

Inference and Data Analysis

In Chapter III, a theoretical framework was developed and in the preceding section the data-base available for this study was described and assessed. A missing link remains, namely, the articulation of theory and data. An approach to inference and data analysis must be defined before any empirical research is carried out. This is done in this section.

Types of Inference

Zellner listed three types of inference: deductive, inductive and reductive inference.[4]

---

[4] Arnold Zellner, An Introduction to Bayesian Inference in Econometrics (New York: John Wiley & Sons, 1971), pp. 1-12.

Deductive inference consists of logical proof by which statements, or conclusions, are derived from prior statements, or premises, using the rules of logic. Premises and conclusions are consequently logically equivalent; if premises are true, the conclusion is true, or, as Braithwaite stated:

> "In deduction the reasonableness of belief in the premises, as it were, overflows to provide reasonableness for the belief in the conclusion."[5]

Inductive inference involves making inferences from the particular to the general, or, more specifically, making inferences from experiences in the past or in specific settings to predict experiences in the future or in other settings.

Reductive inference, also called "abductive" or "retroductive", consists of the description and the study of facts and the generation of hypotheses explaining these facts. Thus, reductive inference can be considered as the process of idea-formation and is closely linked to the observation of unusual facts. Reductive inference is anterior to inductive inference. Reductive inference is the process by which hypotheses are generated, but it does not include any test or assessment of these hypotheses. Testing hypotheses is in the realm of inductive inference.

---

5 Richard B. Braithwaite, Scientific Explanation (Cambridge: Cambridge University Press, 1953), p. 257.

Inductive Inference and Hypotheses

Falsifiability

The importance of "conceivable falsification", or "falsifiability" of a hypothesis was stressed by Popper. Braithwaite described Popper's position as follows:

> The empirical criterion of rejection for a scientific hypothesis is so fundamental that it is most convenient to treat the meaning of universal sentences expressing empirical generalizations as being determined by the experiences which would refute them.[6]

Thus, a scientific hypothesis is a statement which is potentially falsifiable, and, furthermore, its meaning is determined by the events which would refute it. Such a definition says nothing about how the hypothesis is arrived at; especially, a hypothesis need not be derived logically (i.e., deduced) from higher level hypotheses.[7] The application of the principle of falsification leaves the scientist with a set of not-yet-refuted hypotheses. The concept of falsifiability does not solve, however, the problem of induction. Scientists do distinguish among the not-yet-refuted hypotheses; this led Popper to the definition of the "degree of corroboration"[8] of a hypothesis.

---

6 Ibid., p. 255.

7 For an exposition of the different levels of hypotheses in a deductive scientific system, see Braithwaite, Scientific Explanation, pp. 12-21.

8 Karl R. Popper, Logic of Scientific Discovery (New York: Harper & Row Publishers, 1965), pp. 251-281.

Degree of Corroboration

Theories cannot be verified, ie: stated as true; but they can be corroborated. A theory that yields falsifiable hypotheses can be tested; the number of tests and, more important, the severity of the tests up to which a theory has stood, determines the degree of corroboration. Popper developed his concept of corroboration in opposition to subjective theories of probability which consider probability as representing the degree of confidence that one has in a proposition.[9] According to such logicist-subjective theories of probability, deductive proof and disproof are thus only limiting cases of inductive inferences, cases where the degree of confidence in the inference is 1 or 0; the degree of confidence associated with inductive inference always lies between 0 and 1. "Degree of corroboration" and "degree of confidence" are different concepts; the former is objective in the sense that a quantification of it requires an investigation into the tests up to which a hypothesis as stood, whereas the latter is subjective in the sense that an experimenter assigns a degree of confidence (in a probabilistic form) to a hypothesis and this assigned probability is taken as being the degree of confidence attached to this hypothesis. Quantifications of the degree of corroboration may not be accurate but, in any case, an objective degree of corroboration of a hypothesis exists independently of the experimenter.

---

9 "The Bayesian approach, and Jeffreys' theory in particular, involves a qualification of such phrases as 'probably true' or 'probably false' by utilizing numerical probabilities to represent degrees of confidence or belief that individuals have in propositions." (Zellner, Bayesian Inference, p. 9).

Despite their differences, the concepts of "degree of corroboration" and "degree of confidence" both express the fact that scientists consider different hypotheses differently; and it is this aspect which seems to be of practical interest for applied research.

The Hypothetico-Deductive Framework

Braithwaite considers that the hypothetico-deductive framework is but one species of inductive method. In this framework hypotheses are deduced from higher-level hypotheses, which themselves have been inductively established; these deduced hypotheses are then subjected to tests.[10] Thus, the hypothetico-deductive method allies both deductive inference, by which lower-level hypotheses are derived, and inductive inference. Also, an intermediary or lower-level hypothesis is empirically supported by direct evidence but also by empirical evidence to other hypotheses, of whatever level, belonging to the deductive system. Thus, the degree of corroboration of a hypothesis belonging to a deductive system would depend on the number and severity of tests up to which _this_ hypothesis and _other_ hypotheses of the deductive system have stood. Similarly, the subjective probability or degree of confidence that an experimenter would assign to any hypothesis would depend on the same factors.

---

10 Braithwaite, _Scientific Explanation_, p. 261.

A last word should be said on testing lower-level hypotheses: lower-level hypotheses in most cases, are not derived solely from higher-level hypotheses of a system; some other statements (premises) must be added which, very often, have not been inductively established; thus, by testing lower-level hypotheses, it is the conjunction of these lower-level hypotheses and of the additional premises which is actually tested.

Reductive Inference and Hypotheses

In the preceding sub-section, it was explained that new hypotheses can be deduced from higher-level hypotheses; this is not, however, the only way new hypotheses can arise. Reductive inference was previously described as the process of idea-formation. Reductive inference is a yet little understood type of inference; its importance in the process of scientific inquiry is, nevertheless, being increasingly acknowledged. Reductive inference is thought to be a mixture of description of facts, choice among many possible combination of ideas, and conscious, as well as, subconscious mental activity. In summary, it is not a well formalized form of inference, such as deductive or inductive inference.

Whereas deductive and inductive inference start with premises and hypotheses, reductive inference starts with raw facts. One sees that a debate between the relative importance of reductive inference and inductive inference in scientific inquiry is tantamount to a debate on the role of the "a priori" in scientific inquiry. Bacon rejected the idea that "Anticipations of Nature" (theory) had any role in the advancement of scientific knowledge and argued that the scientist had to lcear his mind of all preconceived ideas and prejudice. Benzécri aptly described the Baconian dream:

> on rêve d'une méthode qui mettrait placidement les idées à l'épreuve des faits; mieux encore: qui distillerait les faits jusqu'à en tirer des idées.[11]

It is suggested, here, that the degree of confidence and degree of corroboration provide a link between hypotheses arrived at through deductive, inductive and reductive inference. Because hypotheses arrived at through reductive inference have not been submitted to many and strict empirical tests, their degree of corroboration is very low; from a subjective point of view, the degree of confidence that a scientist would assign to them would also be very low. On the other hand, hypotheses deduced from higher level hypotheses and belonging to a deductive system are substantiated and supported by the evidence for other hypotheses in the system; thus, their degree of corroboration or their degree of confidence can vary in a wide range depending on the extent to and severity with which this system has been empirically tested. Inductive inference, and more specifically the hypothetico-deductive method, and reductive inference can be viewed as alternative ways of generating scientific hypotheses with different levels of degree of corroboration and degree of confidence.

---

11 Jean-Paul Benzécri, "La Place de l'a priori," in Encyclopaedia Universalis-- Organum (Paris: Encyclopaedia Universalis, 1973), p. 11.

Exploratory and Confirmatory Data Analysis

In statistics, the debate on the role of hypotheses in the initiation of scientific inquiry appears as a rivalry between two schools: the classical school (or mathematical statistics), which relies heavily on strong distributional assumptions and the theory of tests of hypotheses, and the data-analysis school which relies more on the description of data. Tukey expressed his concerns over classical statistics and gave a definition of data-analysis:

> For a long time, I have thought I was a statistician, interested in inferences from the particular to the general. But, as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt.... All in all, I have come to feel that my central interest is in data-analysis.
>
> Large parts of data-analysis are inferential in the sample-to-population sense.... Large parts of data-analysis are incisive, laying bare indications which we could not perceive by simply and direct examination of raw data.... Some parts of data-analysis, ... are allocation, in the sense that they guide us in the distribution of effort.... Data-analysis is a larger and more varied field than inference, or incisive procedures, or allocation.[12]

Data analysis, thus, appears broader than classical statistics and could be considered as covering both reductive inference and inductive inference. Tukey expressed this broader interest:

> To concentrate on confirmation, to the exclusion or submergence of exploration, is an obvious mistake. Where does new knowledge come from? ... There really seems to be no substitute for "looking at the data".[13]

---

[12] John W. Tukey, "The Future of Data Analysis," Annals of Mathematical Statistics 33 (1962): 2.

[13] Idem, "Analyzing Data: Sanctification or Detective Work?" American Psychologist 24 (1969): 83.

In accordance with their wider scope of interests, data-analysts use, and in fact developed, other statistical techniques capable of fulfilling their need for description. Among those, are multivariate techniques such as: factor analysis, principal components, cannonical correlation, discriminant analysis, and cluster analysis.

Implication for this Research

Economic theory constitutes a deductive system of which the highest level hypotheses constitute the model of individual behavior which was critically reviewed in Chapter III. These highest level hypotheses have not been inductively established directly; only deduced lower level hypotheses have been empirically tested. No attempt will be made here to discuss whether realism of these assumptions is necessary or even desirable. But it is argued here that, as was explained before, what has been submitted to tests is the conjunction of the model of individual behavior and additional premises which define the conditions under which this behavior was taking place. Thus, a model of behavior may be valid, for certain simple economic decisions, but may not be valid for the decision to leave farming and migrate. From Chapter III, it should be patent that the degree of confidence in hypotheses deduced from the standard model of individual behavior vary over a wide range.

Data used in this study were described above and their shortcomings were mentioned. Because of these shortcomings, hypotheses, concerning off-farm movement and derived from the standard economic model of individual behaviour, cannot be tested without making many additional assumptions. These assumptions have not been inductively established and as a consequence, the degree of confidence in them, though varying, is generally low. These factors further contribute to the generally low degree of confidence in hypotheses to be tested.

The difficulty in deriving hypotheses with a high degree of confidence from economic theory was an incentive to engage in reductive inference, by means of exploratory data analysis. Reductive inference and corresponding exploratory data analysis techniques are seldom used by economists; this is the reason why such a long methodological discussion was necessary so as to justify the methods used in this study.

Thus, the approach to inference taken in this study included:

1. The use of exploratory data analysis as a potential way to generate hypotheses.
2. The derivation of hypotheses from higher level hypotheses and additional premises, as well as the testing of these hypotheses on the available data.

Exploratory data analysis was performed using contingency tables and discriminant analysis; confirmatory data analysis was performed with techniques, such as the logit model and multivariate logistic model, which impose a constraining structure on the data. These statistical techniques are described in detail in Chapter V.

Hypotheses

In Chapter III, an expanded theoretical framework of the decision to leave farming was presented. This framework treats certain aspects, such as cognition or attachment to the community and to farming, in a different manner. It, furthermore, departs radically from the standard model of individual behavior on other aspects, such as the assumption of a satisficing behavior and the consideration of conflicting goals among the individuals of the family, which is considered to be the decision unit.

In the first section of this chapter, the data-base available for this study was described; it is patent that these data come very short of providing the necessary information to substantiate or establish the various elements of the expanded theoretical framework. This regrettable situation means that only a small portion of the hypotheses suggested in the theoretical framework could be tested; furthermore, in order to perform such tests it was necessary to make further assumptions which, in many cases, were not empirically supported.

The discussion related to hypotheses and their role in scientific inquiry must be kept as a background to this section. It will be seen that the degree of confidence that the experimenter holds for the hypotheses listed below varies over a wide range; for some it is so low that these propositions might have been more aptly called conjectures.

The hypotheses listed below do not correspond to the "null" hypothesis of classical statistical theory but to the "alternative" hypothesis. Actually, what is usually tested is the absence of relationship i.e. the null hypothesis. It must be realized that this is done in part for practical purposes: the distribution of the statistics under the null hypothesis is easier to obtain. From the above discussion on inference it should be clear that stating hypotheses for which a certain minimum degree of corroboration of degree of confidence is the normal process in scientific enquiry. This is the reason why, hypotheses concerning relationships having some (variable) degree of confidence have been stated, even though the absence of relationship will be tested. Confirmatory data analysis, thus, provide <u>indirect</u> evidence to the hypotheses.

<u>Age of Operator</u>

If a maximizing behavior is assumed and if it is further assumed that the nonfarm occupational alternative yields a higher income or satisfaction stream, it follows that older farmers will be less inclined to leave farming to engage in nonfarm occupations. Also, attachment to the community as well as to farming is generally recognized as increasing with age; if so, the cost of movement and migration would increase with age, thereby reducing mobility. If farming is considered as a form of habitual behavior, the further it lasts the deepest it will be entrenched; this would tend to reduce readiness of the farm operator to consider other alternatives and to make a genuine decision regarding occupation. If farm incomes are assumed to be systematically below nonfarm incomes and if other factors are ignored, aspirations with respect to income are more likely to have adjusted downwards toward actual income if the farm operator has been farming for a long time. The foregoing arguments lead us to the first hypothesis:

Hypothesis 1: pre-retirement exit from farming is negatively related to age.

Off-farm Work

If off-farm work is assumed (1) to provide a farm operator with skills which are better rewarded on the nonfarm labor market, (2) to improve the farm operator's knowledge of non-farm opportunities and living conditions in urban areas, (3) to be an indication of a lower level of attachment to farming, and (4) to raise the level of aspiration of the farm operator by putting him in contact with higher standard of living in urban areas, then it follows that off-farm work should be positively related to pre-retirement exit from farming.

Hypothesis 2: pre-retirement exit from farming is positively related to involvement in off-farm work by the farm operator.

Tenure

It has been maintained in the literature that ownership of land and capital acts as an impediment to off-farm movement and migration; ownership can be considered to increase the attachment to the farm and consequently to farming. Since the only information available in the data-base concerns ownership of land, this will be used as a proxy for the degree of ownership of the farm as a whole.

Hypothesis 3: pre-retirement exit from farming is negatively related to the degree the farm family owns the farm land.

Size of the Farm

If the size of the farm is positively related to net income from farming, if benefits from a nonfarm occupation are independent of benefits while farming, and if a maximizing behavior is assumed, then pre-retirement exit from farming should be negatively related to the size of the farm.[14] If a satisficing model is assumed, then the benefits potentially obtained from nonfarm alternatives do not play any direct role; they bear on the decision via the level of aspiration of the farm operator and his family. It can, however, be assumed that the larger the size, the more likely are the aspirations (concerning income and other elements) to be met. Thus, in both cases, pre-retirement exit from farming can be expected to be negatively related to the size of the farm. Since several measures of the size of a farm are possible, Hypothesis 4 will be a composite one.

Hypothesis 4: pre-retirement exit from farming is negatively related to acreage, total capital value, and total sales of agricultural products.

---

[14] Hathaway and Perkins, however, inferred from their results that income from farming and income from non-farm employment which is potentially obtainable by a farm operator are positively related. See chapter 2.

Productivity of Land and Capital

Productivity of land and capital is expected to be positively related to net farm income. With the same assumptions and caveats used in relation to Hypothesis 4, productivity of land and capital can be hypothesized to be negatively related to pre-retirement exit from farming. Given the limitations of the available data, measures of productivity had to be crude: productivity of land was measured by the ratio of total sales of agricultural products to total acreage, and productivity of capital was measured by the ratio of total sales of agricultural products to total capital value of the farm.

Hypothesis 5: pre-retirement exit from farming is negatively related to productivity of land and capital.

Mechanization

Insofar as the degree of mechanization is an indicator of the progressiveness of the farm operator, it can be assumed to be positively related to net farm income and, consequently, to be negatively related, to pre-retirement exit (the same further assumptions and caveats as for Hypothesis 4 apply). Some arguments can be advanced that would tend to support the contrary hypothesis: to the extent that progressive farmers are more business minded, they would be more aware of nonfarm opportunities and more ready to take advantage of them; thus, mechanization, a proxy for progressiveness, would be positively related to pre-retirement exit. The first hypothesis seemed, *prima facie*, more attractive and was retained.

The degree of mechanization was measured by the ratio of the value of machinery and equipment to the total farm acreage.

Hypothesis 6: the degree of mechanization is negatively related to pre-retirement exit from farming.

Distances to Towns

Proximity to urban centers can be considered to reduce the cost of information on nonfarm urban occupational alternatives. If these alternatives yield more benefits, farm operators located close to urban centers would be more likely to leave farming. Also, proximity to urban centers entails more frequent contacts with the urban environment; to the extent that urban residents receive higher incomes, such frequent contacts would tend to raise the income aspirations of farm operators and families and to create dissatisfaction which would provide some incentive to leave farming.

Hypothesis 7: pre-retirement exit from farming is negatively related to distance to towns.

Residence

Residence on the farm is conducive to greater attachment to the farm and farming; on the other hand, non-residence means usually residence in nearest town and thereby entails better access to information on nonfarm occupational alternatives.

Hypothesis 8: pre-retirement exit from farming is positively related to non-residence on the farm.

Conclusion

The foregoing hypotheses are those which could be examined with the limited secondary information available. The main issues raised in Chapter III concerning, especially, the nature of individual behavior (maximizing or satisficing), the nature and role of search behavior, the role of attachment to the community and to farming, could not be investigated. Proposals for further empirical research aimed at these issues are presented in Chapter VII.

Conclusion

In this chapter the data-base available for the study was described. A general approach to inference was proposed, exploratory and confirmatory data analysis were contrasted and their relationships to, respectively, reductive and inductive inference were clarified. Finally, a set of hypotheses were proposed to be tested in the confirmatory part of the data analysis.

Some of the key variables to be analyzed are of a special kind: exiter/stayer, residence, and to a certain extent involvement in off-farm work are qualitative variables. Analysis of qualitative variables, especially when they are considered as being dependent variables, requires special statistical methods. The next chapter is devoted to a review of these methods.

CHAPTER V

UNIVARIATE AND MULTIVARIATE MODELS FOR THE ANALYSIS OF QUALITATIVE DEPENDENT VARIABLES

Economists mostly deal with continuous variables (e.g., prices, quantities, costs, distances) and, consequently, econometricians have concentrated on using and developing statistical models valid for continuous variables. In the last decade some economists and econometricians came to realize that their subject matter comprised many qualitative variables.[1] This is especially true in relation to the study of purchases of durable goods, migration, travel demand and occupational mobility. At the same time it was perceived that statistical methods commonly used were inappropriate for the analysis of qualitative variables; econometricians have thus by necessity, come to the forefront of statistical research concerning the analysis of qualitative variables and also mixtures of qualitative and quantitative variables.

This chapter is a brief survey of the field covering some of the practical situations which an applied econometrician could face. Several of the statistical methods presented here were used in this study of off-farm movement and the results are presented in Chapter VI.

---

[1] Qualitative variables describe alternative states; their values are of no significance, but the probabilities with which they take these values are.

The first section describes statistical methods for one qualitative dependent variable whether it is dichotomous or polytomous.[2] These statistical methods include: the linear model, the logit/probit-type models, discriminant analysis, and the multinomial logit model.

The second section presents statistical methods for the analysis of several qualitative dependent variables. Some general considerations concerning measures of association in contingency tables are stated and several types of measures of association are reviewed. Log-linear models of contingency tables with and without exogenous variables are described. Finally, a linear model for jointly dependent dichotomous variables estimated using the generalized least squares estimator (G.L.S.) is introduced.

One Qualitative Variable

One Dichotomous Variable[3]

Consider a dichotomous dependent variable, y, taking the values zero and one; its expectation is assumed to be a monotonic function of a linear combination of the elements of a vector of explanatory variables, x. This is expressed more formally as follows:

$$E(y_t) = F(x_t\beta),$$

---

[2]A dichotomous variable is a qualitative variable taking two values; another name is binary. A polytomous variable is a qualitative variable taking several (more than two) values.

[3]This section draws on the following sources: Marc Nerlove and S. James Press, Univariate and Multivariate Log-Linear and Logistic Models, Rand Report R-1306-EDA/NIH (Santa Monica: Rand Co., December 1973). A.S. Goldberger, Econometric Theory (New York: John Wiley & Sons, 1964); D.R. Cox, The Analysis of Binary Data (London: Methuen, 1970); A. Zellner and T.H. Lee, "Joint Estimation of Relationships Involving Discrete Random Variables," Econometrica 33 (April 1965): 382-394; N.D. Ashton, The Logit Transformation (New York: Hafner, 1972).

where

y_t = the stochastic variable associated with the t^{th} observation,

x_t = a 1 x (1+k) vector of explanatory variables,

β = a (1+k) x 1 vector of coefficient.

In light of the zero-one values taken by y, the conditional expectation of y, can be interpreted as the conditional probability that y is equal to one.

Depending on the function F( ) that is chosen, several statistical models can be specified. They are presented in the rest of this section.

The Linear Model

This model can be formally expressed as:

$$E(y_t) = x_t\beta,$$

or

$$Y = X\beta + \varepsilon,$$

with $E(\varepsilon) = 0,$

where

Y = a T x 1 vector of observations on the dependent variable,

X = a T x (1+k) matrix of observations on the independent variables,

β = a (1+k) x 1 vector of coefficients,

ε = a T x 1 vector of disturbances.

The linear model cannot be considered as being truly desciptive of the relationship between $E(y_t)$ and the explanatory variables since it does not satisfy the condition that $E(y_t)$ is bounded between 0 and 1 but it is to be considered as an approximation of the true relationship which can lead to close estimates of the parameters β. Figure 1 shows that the approximation of the true functional relationship by the linear probability function can be considered satisfactory in the middle of the range of $X\beta$ but not at the extremes.

When a linear probability function is specified, ordinary least squares (O.L.S.) can be used to estimate the β coefficients. Several problems arise that should make the user cautious. Because y_t is constrained to take only 0 or 1 values, the usual assumption of homoskedasticity of the variance of the disturbance term is not met;[4] the implications are: (1) the O.L.S. estimator is still unbiased but is inefficient, (2) the O.L.S. estimator of the variance of the coefficient estimators is biased and inconsistent, (3) the direction of this bias is unknown, and (4) the usual tests of significance are inapplicable.[5] The binary nature of y_t implies also that R^2 can be equal to 1 only when the values of the explanatory variables are concentrated in two points; thus R^2 will usually be very small and its usual meaning is lost. Besides, it is a known result that, because the dependent variable is not normally distributed, no linear estimation method such as O.L.S. can be efficient.[6]

Another disadvantage of the linear model and the O.LS. method of estimation is that the estimates of the coefficients are highly sensitive to the distribution of the values of the explanatory

---

[4]See Goldberger, Econometric Theory, p. 249.

[5]Implication (4) follows from implications (2) and (3).

[6]See Cox, The Analysis of Binary Data, p. 17.

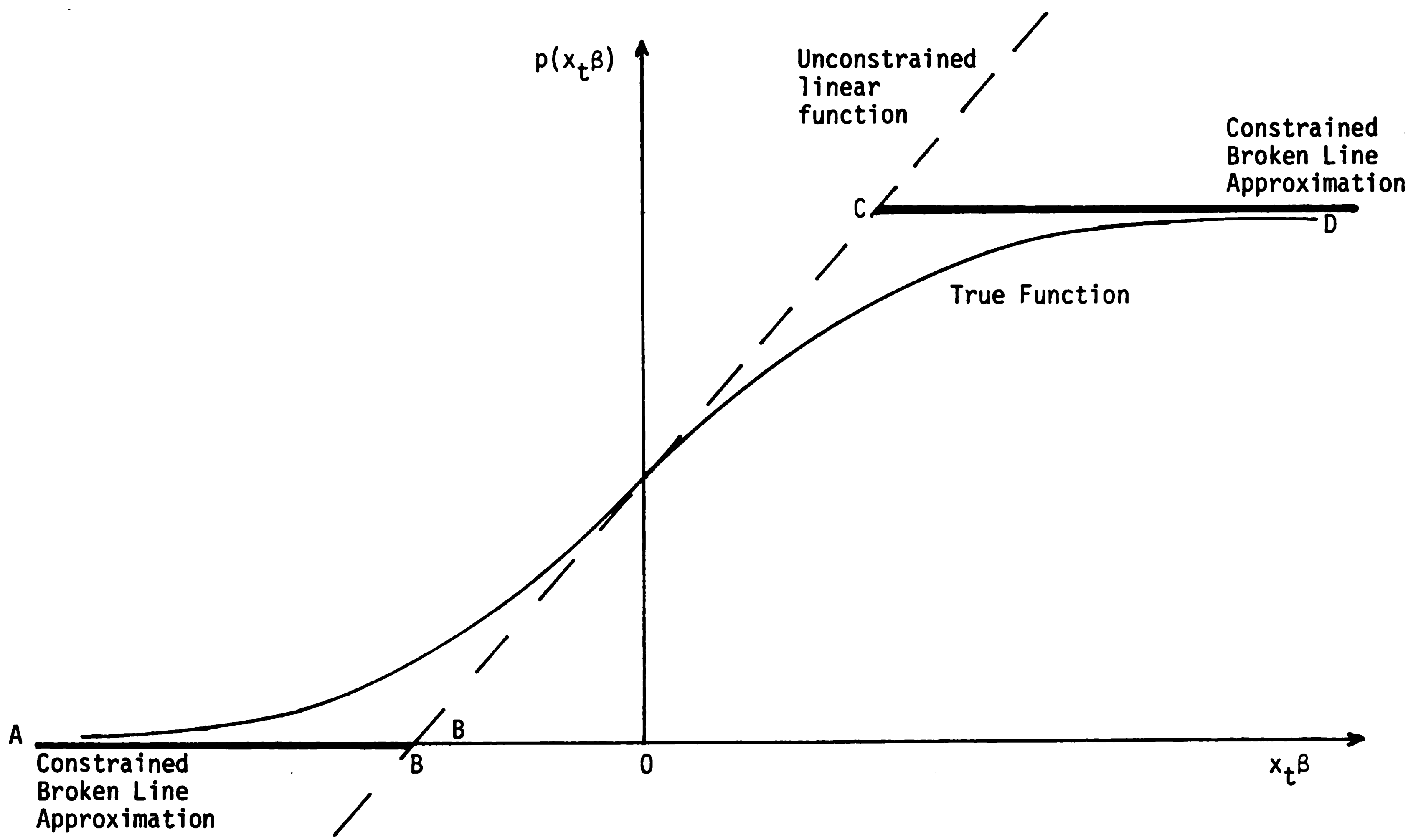

Figure 1. Constrained and unconstrained linear probability function, and true probability function.

Source: Nerlove and Press, Log-Linear and Logistic Models, p. 4.

variables: bunched data on one side will provide very different estimates from the one obtained with data covering regularly the whole span.[7] This last point may not be very relevant to experimental situations since, in such cases, the analyst can control the values taken by the explanatory variables; it may, however, be very relevant to nonexperimental situations where the explanatory variables are not controled.

Because there is heteroskedasticity, the proper estimator is Aitken's generalized least squares estimator (G.L.S.). If the disturbances are assumed not to be autocorrelated, the G.L.S. estimator is a form of weighted least squares where the weights used are the inverse of the variances of the disturbances. These variances are equal to $E(y_t)(1-E(y_t))$, and can be consistently estimated by $\hat{y}_t(1-\hat{y}_t)$, where $\hat{y}_t$ is the calculated value obtained from an O.L.S. estimator; $\hat{y}_t$ may lie outside [0,1] and, consequently, some of the estimated variances may be negative, which is theoretically impossible. In view of this ultimate difficulty the above two-step estimation procedure seems often impractical.

A third possible approach to estimating the linear model, besides O.L.S. and G.L.S., consists of the use of a constrained least squares estimator. Figure 1 shows that the broken line ABCD is a close approximation of the true sigmoid curve except in the two corners B and C. The process of fitting a broken line to a number of points is, however, rather complex; two major shortcomings of this

---

[7] See Nerlove and Press, Log-Linear and Logistic Models, p. 8.

method are: (1) it is reliable only in large samples, and (2) it ignores all the distributional properties of the error terms.

As a consequence of the aforementioned inadequacies of the linear model in the case of a binary dependent variable, some other models were developed and are presented below.

The Logit/Probit-Type Models

Logit and probit models were originally developed in relation to experiments studying the toxicity of drugs in terms of their qualitative effect (e.g., death) on a population of animals or vegetables. In their early development, logit and probit models applied only to the case of a one binary variable dependent on a single contiguous variable. Also, since biologists could experiment on large populations with controls on the independent variable, the data were grouped by value of the independent variable and the dependent variable was considered to be the frequency of occurence of the qualitative effect for each group; thus, the models were truly applied to the study of a <u>continuous limited</u> variable dependent on a continuous independent variable. These models were thereafter extended to deal with several independent variables and ungrouped data.

The common feature of the logit/probit-type models lies in the use of a sigmoid-like function mapping the probability p_t of the event, belonging to [0,1], into a transform taking values on the interval $[-\infty , +\infty]$. The transform of p_t is then expressed as a linear function of the exogenous variables:

$$z_t = F^{-1}(p_t)$$

an

im

wh

Ch

mo

(c.

pr

log

It

and

Thu

on

and

$$z_t = x_t\beta,$$

imply

$$p_t = F(x_t\beta),$$

where

p_t = the conditional probability that $y = 1$ given x_t,

F = a nondecreasing function of $x_t\beta$,

x_t = a 1 x (1+k) vector of independent variables,

β = a (1+k) x 1 vector of coefficients.

Choosing different transform functions of p will yield different models. It should be clear that any cumulative distribution function (c.d.f.) can be chosen. The most widely used, the logit and the probit models are presented below.

The logit model is obtained when the c.d.f. of the standardized logistic distribution is used as transformation; p_t is given by:

$$p_t = 1/(1 + e^{-x_t\beta}).$$

Then, logit (p_t) is defined as follows:

$$\text{logit}(p_t) = \ln [p_t/(1-p_t)].$$

It follows that

$$\text{logit } (p_t) = x_t\beta + \varepsilon_t,$$

and

$$E [\text{logit } (p_t)] = x_t\beta.$$

Thus logit (p_t), which takes values on $[-\infty, +\infty]$, can be regressed on $x_t\beta$ and β can be estimated using O.L.S.

r

wh

Pr

(1)

wei

nei

nume

a s

be

the

The probit model is related to the c.d.f. of the standard normal distribution:

$$P_t = (1/\sqrt{2\pi}) \int_{-\infty}^{v_t} \exp(-1/2u^2)\, du,$$

where

$v_t = x_t\beta$,

x_t = the vector of exogenous variables for the t^{th} observation.

Probit and normit are defined respectively as:

Probit $(p_t) = v_t + 5$,

Normit $(p_t) = v_t$.

The logit and probit models can be estimated in two ways: (1) minimum logit χ^2 (normit χ^2), and (2) maximum likelihood.

The minimum logit χ^2 (normit χ^2) is equivalent to applying weighted least squares on the logits (normits). A solution requiring neither approximation nor iteration can be obtained.

The maximum likelihood function is:

$$L(\beta) = \prod_{t=1}^{T} p_t^{y_t} (1-p_t)^{1-y_t}$$

$$= \prod_{t=1}^{T} [F(x_t\beta)]^{y_t} [1-F(x_t\beta)]^{1-y_t}.$$

This function can be maximized with respect to β, usually by numerical methods since analytical methods generally lead to solving a system of nonlinear equations.

The minimum logit χ^2 (normit χ^2) estimators have been shown to be asymptotically equivalent to the maximum likelihood estimators; they are RBAN (regular best asymptotically normal) and therefore

asy
eit
of
squ
tag
exp
tic
(nc
whe
(2)
grc

rar
con
tha
com
(no
use
is
Gro
mum

Nor
Ame

of

asymptotically efficient.[8] These results imply that, for large samples, either method can be used indifferently. Little is known theoretically of their properties in small samples. Estimators based on the least squares principle (minimum logit χ^2 or normit χ^2) present some advantage over maximum likelihood estimators since only knowledge of the expectation and not of the specific functional form of the distribution is required. On the other hand, (1) the values of the logit χ^2 (normit χ^2) obtained in the minimum logit χ^2 (normit χ^2) are unstable when few observations exist for each group of observations, and (2) the logit χ^2 (normit χ^2) estimation method breaks down when some groups of observations contain only ones or zeros.

This latter point is of special interest to economists who rarely work in an experimental situation and who very often wish to consider many exogenous variables. In such conditions, it is likely that there will be only one observation for each value of the linear combination of the exogenous variables; thus, the minimum logit χ^2 (normit χ^2) method is usually not applicable. Berkson[9] proposed the use of "working values" for cells containing only zeros or ones; this is not possible, however, when all cells contain only zeros or ones. Grouping of observations may therefore be necessary to use the minimum logit χ^2 (normit χ^2) method. Monte Carlo studies[10] show that

---

[8]Ashton, The Logit Transformation, p. 34.

[9]J. Berkson, "Estimate of the Integrated Normal Curve by Minimum Normit χ^2 with Particular Reference to Bio-Assay," Journal of the American Statistical Association 50 (June 1955): 534.

[10]Daniel McFadden, "Quantal Choice Analysis: A Survey," Annals of Economic and Social Measurement 5 (Fall 1976): 373.

minin

larg

maxi

logi

the

a lo

this

take

limi

vari

feas

Pear

χ^2 a

dist

max-

unde

tic

lik

the

of

minimum logit χ^2 estimation with grouping yielded lower variances, larger biases, and comparable mean square errors when compared to maximum likelihood estimation; this would tend to show that minimum logit χ^2, after grouping data and correcting for the bias, is possibly the best method. It should be stressed that grouping of data implies a loss of information which is positively related to the extent of this grouping; consequently, the aforementioned results are to be taken with caution and on the understanding that grouping is of a limited extent. It can be stated that when the number of exogenous variables increases there comes a point where grouping is no longer feasible; maximum likelihood methods should then be used.

The goodness of fit of the model can be assessed through the Pearson χ^2 or logit χ^2 (or normit χ^2).

Tests of hypothesis can proceed in two ways:

1. Using the property that minimum logit χ^2, minimum normit χ^2 and maximum likelihood estimators are asymptotically normally distributed, and estimating asymptotic variances from sample variances.

2. Using the likelihood ratio test obtained by dividing the maximum likelihood under the null hypothesis by the maximum likelihood under the alternative hypothesis; this likelihood ratio is asymptotically distributed as chi-square and easy to obtain when maximum likelihood estimation is performed.

The logit and probit functions are very similar and, usually, the logit model is favored on the basis of the ease of computation of its likelihood function.

Dis

sta
vec
or
tic
cov
bet
the
sev
lin
use
we

dev
app
so
be

lin

whe

Pr

Discriminant Analysis

Discriminant analysis belongs to the family of multivariate statistical techniques related to factor analysis. Suppose that a vector of continuous variables is observed on a set of individuals or objects known to belong to a certain number of different populations; then, discriminant analysis can help the experimenter in discovering whether or not a reduced set of variables discriminates well between these populations. Discriminant analysis can also enable the experimenter to classify an individual or object into one of several populations, on the basis of the individual's score on the linear discriminant function. Thus, discriminant analysis can be used for two purposes: (1) descriptive, or (2) decision-making; here we will concentrate on the former.

Discriminant analysis will be described in detail in the section devoted to the case of one polytomous variable. All the results apply to the special case where individuals belong to two populations, so that a binary variable specifies the population to which they belong. A few specific points deserve to be mentioned here.

Only one discriminant function exists and it is the long-known linear discriminant function of Fisher,[11]

$$F = (N_1 N_2/N^2)\ (y_1-y_2)'\ T^{-1}\ (y_1-y_2)u^1,$$

whose corresponding eigenvalue is:

$$e = (N_1 N_2/N^2)(y_1-y_2)'\ T^{-1}\ (y_1-y_2),$$

---

[11]R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics 7 (September 1936): 179-188.

Th

wh

us

ta

moc

the

The

of

One

den

the

suc

met

cri

Dis

Pro
(Oc

where

N_1, N_2 = the number of individuals in populations 1 and 2,

$N = N_1 + N_2$

y_1, y_2 = the mean vectors of observed variables for populations 1 and 2,

T = the total covariance matrix.

The eigenvalue, e, is the Mahalanobis D^2 premultiplied by $N_1 N_2 / N^2$ which is a measure of the distance between two groups of observations using the metric distance defined by the total covariance matrix.[12]

In the two-class case, linear discriminant analysis is computtationally equivalent to ordinary least squares estimation of a linear model with a binary dependent variable.[13] The regression line and the discriminant function only differ by multiplicative constants. The choice between the two procedures should be based on what kind of distributional assumptions the experimenter is ready to make.

One Polytomous Variable

In the above section, the case of one binary qualitative dependent variable was examined. A qualitative variable can also describe the fact that an individual may belong to more than two categories; such a qualitative variable is called a polytomous variable. Two methods are presented below: a multinominal logit model and discriminant analysis.

---

[12] Jean-Marie Romeder, Méthodes et Programmes d'Analyse Discriminante (Paris: Dunod, 1973), pp. 49-50.

[13] This point is examined in detail in George W. Ladd, "Linear Probability Functions and Discriminant Functions," Econometrica 34 (October 1966): 873-885.

c

d

(

v

ab

se

or

whe

α,β

eac

odd

Int

Multinomial Logit

Theil[14] extended the dichotomous (binomial) logit model to the case of a polytomous variable. He started from the version of the dichotomous logit model, where explanatory variables are of two types: (1) explanatory variables related to the categories of the dependent variable, in which case they enter as a ratio, (2) explanatory variables related to the individual, in which case they enter by themselves. More formally the odd ratio is:[15]

$$(p/1-p) = \exp(\alpha) \prod_{h=1}^{m} x_h^{\beta_h} \prod_{k=1}^{n} (y_{k1}/y_{k2})^{\gamma_k},$$

or in a logarithmic form,

$$\log(p/1-p) = \alpha + \sum_{h=1}^{m} \beta_h \log(x_h) + \sum_{k=1}^{n} \gamma_k \log(y_{k1}/y_{k2}),$$

where

x_h = a variable related to the individual facing a binary choice ($h = 1, . . ., m$),

y_k = a variable taking different values for the different alternatives facing the individual ($k = 1, . . ., n$),

$\alpha, \beta_h, \gamma_k$ = parameters to be estimated.

In the case of a multinomial logit model with P alternatives, each alternative has a probability pi such that $\sum_{i=1}^{P} pi = 1$. The odd ratio can be expressed similarly for each pair of alternatives:

---

[14]H. Theil, "A Multinomial Extension of the Linear Logit Model," International Economic Review 10 (October 1969): 251-259.

[15]Ibid., pp. 251-252.

whe

due

of

par

sid

bil

act

per

pro

var

inf

sul

Dis

met

The

su

po

vi

se

$$p_i/p_j = \exp(\alpha_{ij}) \prod_{h=1}^{m} x_h^{\beta_{hij}} \prod_{k=1}^{n} (y_{ki}/y_{kj})^{\gamma_{kij}},$$

where

$i,j = 1,\ldots,P.$

Many constraints exist on the coefficients of these equalities due to certain symmetries and "circularity" relations; the number of parameters is, thus, only $(m + 1)(P - 1) + n$.[16] This number of parameters may become large when the number of alaternatives considered in the model becomes high. The ratios of pairs of probabilities can be estimated using these equalities; estimation of the actual values of these probabilities can be obtained using the property that they add up to one.

A mathematical expression for the infinitesimal changes in probabilities resulting from infinitesimal variations in explanatory variables can be derived as well as a global measure, based on information theory, of the change in the probability structure resulting from changes in explanatory variables.

Discriminant Analysis

As mentioned previously discriminant analysis is a statistical method which is used to find a set of linear discriminant functions.[17] These discriminant functions allow the analyst to determine if a subset of variables discriminate effectively between the several populations considered and to classify yet unclassified cases (individuals or objects) into one of the original populations.

---

[16] Ibid., p. 253.

[17] Only linear discriminant analysis will be considered in this section.

p

c

s

c

la

ar

ap

fo

Th

to

of

or

nam

nam

qua

Let

The

Ro

Discriminant analysis is quite often used and presented independently of any distributional assumptions; when this is done, discriminant analysis is considered as a descriptive multivariate statistical method which provides a specific insight into what is a complex mass of data. Statistical inference, in the sample to population sense, is then disregarded, and the analyst works in the data analysis framework which was discussed in Chapter IV. Such an approach is taken here in presenting discriminant analysis and, therefore, distributional assumptions will be introduced at the end.[18] This presentation is consistent with the essentially descriptive role to which discriminant analysis was confined in the empirical part of this study.

Discriminant analysis requires the existence of several groups or populations of individuals. Typically, a group is described by a name; for example animals belong to different species, each having a name. The belonging to a certain population can be expressed by a qualitative (usually unordered) variable.

Let: X be the set of N individuals or objects x, on which p variables are measured (x is, thus, a p-vector); X is partitioned into K classes y;

Y be the set of classes y, with N_y individuals each;

$w_x = 1/N$ be the weight assigned to x.

The total weight of the scatter X is

---

[18]The presentation proposed in this section is based on: Romeder, Analyse Discriminate.

The

y w

sca

or

The

cen

whe

the

axis

betw

and

The

or

$$W_X = \Sigma\{w_x | x \varepsilon X\} = 1.$$

The total weight of the scatter y is

$$w_y = \Sigma\{w_x | x \varepsilon y\}.$$

y will also represent the mean vector or center of gravity of the scatter y:

$$y = (1/w_y)\ \Sigma\ \{w_x x | x \varepsilon Y\}$$

or

$$y = (1/N_y)\ \Sigma\ \{x | x \varepsilon Y\}.$$

The total variance of the scatter X, with respect to its mean, or center of gravity is:

$$V_X = \Sigma\{w_x\ (x-\bar{x})^2 | x \varepsilon X\},$$

where $\bar{x}$ is the overall mean vector of x.

Let u be a vector in the R^p space. The variance of u for X is the variance of the orthogonal projection of the scatter X on the axis defined by u. Thus, total variance of u is

$$V_X(u) = \Sigma\{w_x\ u(x-\bar{x})^2 | x \varepsilon X\},$$

between-class variance is

$$V_Y(u) = \Sigma\{w_y\ u(y-\bar{x})^2 | y \varepsilon Y\},$$

and within-class variance is

$$V_y(u) = \Sigma\{w_x\ u(x-y)^2 | x \varepsilon y\}.$$

The theorem of Huygen[19] implies that

$$V_X(u) = \Sigma\{V_y(u) | y \varepsilon Y\} + V_Y(u),$$

or in other words:

total variance = sum of within-class variances
+ between-class variance.

---

[19] Ibid., p. 41.

To

the

Aga

The

as

cova

As

axe

wel

cri

tin

of

fac

rec

or

To these total, within-class, and between-class variances correspond the total, within-class, and between-class covariance matrices:

$$T = (t_{ij}), \quad t_{ij} = (1/N)\Sigma\{(x_i - \bar{x}_i)\ (x_j - \bar{x}_j)|x \varepsilon X\},$$

$$W = (w_{ij}), \quad w_{ij} = (1/N)\{\Sigma\{(x_i - y_i)(x_j - y_j)|x \varepsilon y\}|y \varepsilon Y\},$$

$$B = (b_{ij}), \quad b_{ij} = \Sigma\{(N_y/N)(y_i - \bar{x}_i)(y_j - \bar{x}_j)|y \varepsilon Y\}.$$

Again, the theorem of Huygen implies the following equality:

$$T = W + B.$$

The total, within-class and between-class variances can be expressed as quadratic forms, using the total, within-class, and between-class covariance matrices:

$$V_X(u) = u\ T\ u',$$

$$V_Y(u) = u\ B\ u',$$

$$V_y(u) = u\ W\ u'.$$

As stated above, the objective in discriminant analysis is to identify axes, i.e., linear combination of variables, which "discriminate" well between the different classes of individuals. In linear discriminant analysis, an axis will be considered to be more discriminating the higher the between-class variance will be, as a proportion of the total variance. More formally, the problem of finding the factorial axis which discriminates the best between the classes reduces to the following problem:

$$\max_u\ [u\ B\ u'/u\ W\ u'],$$

or, equivalently,

$$\max_u\ [u\ B\ u'/u\ T\ u'].$$

is

As

is

the

val

fac

to

all

dis

der

the

is

the

i.e

has

and

pe

of

The first discriminanting axis u_1 solution of the above problem, is the eigenvector of $T^{-1}B$ corresponding to the largest eigenvalue λ_1. As with all others eigenvalues of $T^{-1}B$, λ_2 lies between 0 and 1 and is a measure of the discriminating power of this axis.[20]

Similarly, the second discriminating factorial axis u_2 is the eigenvector of $T^{-1}B$ corresponding to the second largest eigenvalue λ_2. This second axis is orthogonal to the first and is the factorial axis which discriminates the best among the axes orthogonal to the first axis.

Similarly, other axes are the eigenvectors corresponding to all the eigenvalues taken in decreasing order. Each is the best discriminating axis among those which are othogonal to the previously derived factorial axes.

The discriminating power of each of these axes, as measured by the corresponding eigenvalue, is zero when the between-class variance is zero, i.e., when the mean vectors of each class project on u at the same point, and is one if the within-class variance is zero, i.e., if all vectors of each class project on u in the same point.

The sum of the eigenvalues corresponding to the factorial axes has no particular meaning, contrary to the case in principal component analysis; more precisely, it cannot be interpreted as being the percentage of the total variance explained by these factorial axes.

The number of discriminant axes is K-1, where K is the number of a priori classes considered, provided that the number of individuals

---

[20]Ibid., pp. 45-47.

(N)

than

ing

assu

vari

accc

for

func

of a

the

are

The

or

wh

Th

e

C
(

(N) is greater than the number of variables (P) which itself is greater than the number of classes.

So far, no distributional assumptions have been made concerning the variables observed on the individuals. If the variables are assumed to follow multivariate normal distributions with equal covariance matrices and if individuals are classified into each class according to a maximum likelihood principle, the partitioning line for classification purposes are hyperplans in R_p and the discriminant functions obtained are the ones previously obtained, independently of any distributional assumptions. This is illustrated below for the two-class case.

The multivariate normal distributions for the two populations are:

$$f_1(x) = 2\pi^{-P/2}\ |\Sigma|^{-1/2}\ \exp[-1/2(x-\mu_1)'\ \Sigma^{-1}(x-\mu_1)],$$

$$f_2(x) = 2\pi^{-P/2}\ |\Sigma|^{-1/2}\ \exp[-1/2(x-\mu_2)'\ \Sigma^{-1}(x-\mu_2)].$$

The classification rule is thus[21]

$$x \varepsilon\ y_i \text{ if } f_1\ (x)/f_2\ (x) > 1,$$

or $$z = \log f_1(x) - \log f_2(x) > 0,$$

where $$z = -\frac{1}{2}\ (x-\mu_1)'\ \Sigma^{-1}(x-\mu_1) + 1/2\ (x-\mu_2)\ \Sigma^{-1}(x-\mu_2),$$

$$= [x - 1/2\ (\mu_1 + \mu_2)]'\ \Sigma^{-1}\ (\mu_1 - \mu_2).$$

Thus, when the variables are distributed as multivariate normals with equal covariance, and when the maximum likelihood principle is adopted

---

[21]Peter A. Lachenbruch, Cheryl Sneeringer, and Lawrence T. Revo, "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality," Communications in Statistics 1 (1973): 41.

the
it
pla
tio
max

ass
dis

of
the
lam
sta

ant
no
fo
ha
we

ar

the discriminant functions are linear. If the parameters are unknown, it is reasonable to replace them by the sample estimates; thus, replacing μ_1 and μ_2 by y_1 and y_2, one finds the same discriminant function as previously derived without normality assumptions, through maximization of the ratio of between-class variance to total variance.

When normality is assumed, but different covariance matrices are assumed for each class, the same method would lead to a quadratic discriminant function.

Classical tests can be performed using the strong assumptions of multivariate normality and equal covariance matrices. A test for the equality of the mean vectors of several classes is based on Wilk's lambda (Λ) which can be approximated either by a F statistic or a χ^2 statistic.[22]

There is some evidence that the robustness of linear discriminant analysis is low and that, consequently, it is badly affected by non-normality; especially, the number of misclassifications increases for some non-normal distribution and not for others. On the other hand, some results showed that linear discriminant analysis performs well for discrete data.[23]

---

[22] Romeder, Analyse Discriminante, pp. 77-81.

[23] Lachenburg, Sneeringer, and Revo, "Robustness of Discriminant Functions," pp. 53-54.

Mea

la

ta

A a

Several Qualitative Variables

Measures of Association in Contingency Tables

A contingency table is a convenient way to cross-classify populations of individuals or objects with respect to two or more qualitative variables (polytomies). In the simple case of two polytomies, A and B, a contingency table has the following format:

		B				Total
		1	2	...	β	
A	1	x_{11}	x_{12}	...	$x_{1\beta}$	$x_{1.}$
	2	x_{21}	x_{22}	...	$x_{2\beta}$	$x_{2.}$
	.	.	.	.	.	
	.	.	.	.	.	
	.	.	.	.	.	
	α	$x_{\alpha 1}$	$x_{\alpha 2}$	...	$x_{\alpha\beta}$	$x_{\alpha .}$
Total		$x_{.1}$	$x_{.2}$	....	$x_{.\beta}$	$x_{..}$

where x_{ij} = the actual count of cell i, j,

$x_{.j}$ = the sum over i of x_{ij},

$x_{i.}$ = the sum over j of x_{ij},

$x_{..}$ = the sum over i and j of x_{ij} and is equal to N the number of observations.

Other representations of this table could be obtained using the same format but replacing x_{ij} by

m_{ij}, the expected cell count,

r_{ij}, the proportion of observations falling in cell i,j, usual

Goodm

vario

Meas

city

tend

Also

tota

of a

tion

pro

Gen

pr

on

sh

C
4
a

p_{ij}, the probability of an observation falling in cell i,j.

After having displayed such a cross-classification, the analyst usually seeks to assess the degree of association between the polytomies. Goodman and Kruskal expressed their concern for a proper use of the various measures of association:

> Our major theme is that the measures of association used by an empirical investigator should not be blindly chosen because of tradition and convention only, although these factors may properly be given some weight, but should be constructed in a manner having operational meaning within the context of the particular problem.[24]

Measures of association will be discussed here, for the sake of simplicity, in the case of two polytomies; many of the results can be extended straightforwardly to the case of three or more polytomies. Also, contingency tables will be considered to cross-classify the total population, and not a sample; the discussion of the measures of association will therefore evade the problem of sample to population inference, and will be in terms of actual cell counts (x_{ij}) or proportions of observations (r_{ij}).

General Considerations

Before the various categories of measures of association are presented, several general comments should be made.

The choice of the adequate measure of association depends heavily on the caracteristics of the polytomies and of the way their relationship is to be considered. The following points are relevant:

---

[24] L.A. Goodman and W.H. Kruskal, "Measures of Association for Cross-Classifications," Journal of American Statistical Association 49 (December 1954): 732-764. This section relies heavily on the above article.

(1)

or

the

the

bri

Thi

be

If

tri

sur

nor

mea

or

it

no

to

or

s

s

g

c

(1) existence or nonexistence of an underlying continuum, (2) existence or nonexistence of an underlying order, (3) symmetry or asymmetry of the way in which polytomies enter the analysis, and (4) definition of the categories of the polytomies. These points are now discussed briefly.

A polytomy may have been dervied from a continuous variable. This process implies a loss of information, and consequently, it may be worthwhile to restore the continuum, even at a substantial cost. If the continuum is not restorable one may want to assume some distribution for the underlying continuous variable and choose a measure of association appropriate for this distribution; a multivariate normal distribution is commonly assumed, in which case the appropriate measure of association should be based on the correlation coefficient.

A polytomy may possess an underlying order. Ignoring this order implies a loss of information; on the other hand the order itself may not be relevant to the question examined. When there is no reason to believe that an underlying order exists, one may want to constrain the measure of association to be independent of the order in which the classes of the polytomies are tabulated. Such considerations dictate whether or not an ordinal measure of association should be chosen.

If a causal relationship is thought to exist on theoretical grounds, the analyst may want to consider each of the polytomies in a different way, thereby introducing an asymmetry in the contingency table; on the other hand if no causal relationship is assumed, polytomies may be treated symmetrically.

ati

pol

hav

lev

act

tic

on

co

an

de

th

re

d

a

o

(

(

Finally, it should be stressed that the measured level of association in a table may depend very much on how the classes of each polytomy are defined; thus, a precise statement of how these classes have been defined should always accompany the reporting of a measured level of association.

The foregoing discussion outlined the importance of the characteristics of the polytomies on the choice of a measure of association; some general remarks on the use of measures of association and on some of their desirable properties are also necessary.

One should not confuse the concept of independence and its corrollary the test of independence, with the concept of association and its corrollary the measure of association. A test of independence should be used to ascertain whether a relationship between the polytomies is likely to exist, and a measure of association is required to assess the type and extent of a relationship.

The concept of association is very imprecise, due to the multidimensional nature of a particular association. More precisely, in a $\alpha x \beta$ two dimensional contingency table there are $\alpha x \beta$-1 total degrees of freedom, α-1 for the rows, and β-1 for the columns, which implies $(\alpha-1)(\beta-1)$ degrees of freedom for the association; this implies that $(\alpha-1)$ $(\beta-1)$ functions are necessary to specify completely the table. This explains why many measures of association have been developed, none of themdescribing completely the association.

The high number of potential candidates as measures of association has led some statisticians to require that a measure of association be invariant under the following transformation:

wher
This
the
sta
and
sid
tio
whe

the
com
wit
att

on
ta
ba
or
pr

T

s

$$r_{ij} \rightarrow t_i \times s_j \times r_{ij}$$

where t_i and s_j are numbers that preserve the equality $\sum_i\sum_j r_{ij} = 1$. This transformation changes the marginal distributions without changing the cross-product ratios in the table. A measure of association which stays invariant under the above transformation is dubbed "margin-free" and one which is not invariant, "margin-sensitive." It can be considered that a margin-sensitive measure of association mixes information on the marginal distributions with information on the association, whereas a margin-free measure does not.

Measures of association are often normalized so that they take the values 0 in case of independence and +1 or -1 in the case of complete association; what complete association means varies, however, with the measure. Normalizing measures of association loses its attractiveness as their interpretability increases.

The foregoing subsection was devoted to general considerations on measures of associations which laid down the way for the presentation of some measures classified into traditional measures, measures based on optimal prediction, measures based on optimal prediction of order, measures of reliability or agreement, and measures based on proportion of explained variance.

Traditional Measures

The traditional measures are mostly based on the Pearson chi-square statistic defined as:

$$\chi^2 = \sum_i \sum_j (x_{ij} - x_{i.}x_{.j}/n)^2/(x_{i.}x_{.j}/n).$$

Yu

an

Ot

p

a

For the special case of a 2 x 2 contingency table the following Yule coefficients are commonly used:

$$Q = (x_{11}x_{22} - x_{12}x_{21})/(x_{11}x_{22} + x_{12}x_{21}),$$

and

$$Y = (\sqrt{x_{11}x_{22}} - \sqrt{x_{12}x_{21}})/(\sqrt{x_{11}x_{22}} + \sqrt{x_{12}x_{21}}).$$

Other measures are:

the mean square contingency,

$$\phi^2 = \chi^2/n;$$

the coefficient of mean square contingency,

$$P = \sqrt{(\chi^2/n)/(1+\chi^2/n)}\ ;$$

Cramer's V,

$$V = \sqrt{\phi^2/\min[(\alpha-1),(\beta-1)]};$$

Tschuprow's T,

$$T = \sqrt{\phi^2/(\alpha-1)(\beta-1)}.$$

The χ^2 statistics is well known to provide a good test of independence; this does not ensure, however, that it is a good measure of association.

The traditional measures of association have the major drawback of not enabling comparison of measured level of association between different contingency tables. This is so because they have no operational meaning.

Measures Based on Optimal Prediction

Consider a two-dimensional table obtained from polytomies A and B without underlying order and continuum. The experimenter's ability to predict the value of the B polytomy may depend on whether th

va

is

kn

pr

me

λ_b

pr

va

me

the value of the A polytomy is known or unknown to him. When the value of A is unknown the experimenter will predict that the B class is the one with largest marginal proportion. When the value of A is known, the best guess is that the B class is the one with largest proportion in the observed A class. This provides the basis for the measure of association defined by:

$$\lambda_b = \frac{(\text{Prob. of error A unknown}) - (\text{Prob. of error A known})}{(\text{Prob. of error A unknown})}$$

λ_b is to be interpreted in terms of the proportion by which errors in predicting B can be expected to be reduced by the knowledge of the value of A for each individual.

Goodman and Kruskal listed the following properties for the λ_b measure of association:

(i) λ_b is indeterminate if and only if the population lies in one column, that is, lies in one B class.

(ii) Otherwise the value of λ_b is between 0 and 1 inclusive.

(iii) λ_b is 0 if and only if knowledge of the A classification is of no help in predicting the B classification. . . .

(iv) λ_b is 1 if and only if knowledge of an individual's A class completely specifies his B class. . . .

(v) In the case of statistical independency λ_b, when determinate, is zero. The converse need not hold: λ_b may be zero without statistical independence holding.

(vi) λ_b is unchanged by permutation of rows and columns.[25]

---

[25]Goodman and Kruskal, "Measures of Association," p. 742.

Symmetrically, λ_a can be defined and interpreted in terms of the increased ability to predict the A polytomy when the B classification is known.

Another measure of association, λ, is defined which is based on the same principle and is interpreted in terms of the increased ability to predict alternatively the A or B polytomy when alternatively the B or A classification is known.

The major shortcoming of λ_a, λ_b and λ is that their value depends on marginal frequencies; in other words two tables displaying the same conditional frequencies but different marginal frequencies will yield different values for λ_a, λ_b and λ, i.e., λ_a, λ_b, and λ are margin sensitive. When one is interested in comparing patterns of conditional frequencies for different tables this can be attained by weighting columns and rows so as to obtain equiprobability of the classes for the different tables.

Another measure of association related to optimal predication is the uncertainty coefficient which is based on information theory.[26]

The expected information, or entropy, or uncertainty of a bivariate distribution is:

$$H(A,B) = \sum_{i=1}^{\alpha} \sum_{j=1}^{\beta} p_{ij} \log (1/p_{ij}).$$

The expected information or uncertainty of the marginal distributions are:

$$H(A) = \sum_{i=1}^{\alpha} p_{i} \log(1/p_{i.}),$$

[26]See Henri Theil, Economics and Information Theory (Chicago: Rand McNally, 1967), pp. 33-35.

and

cor

var

th

"

t

and

$$H(B) = \sum_{j=1}^{\beta} p_{.j} \log(1/p_{.j}).$$

The expected information of the conditional distribution, or conditional uncertainty is:

$$H(B|A) = \sum_{i=1}^{\alpha} \sum_{j=1}^{\beta} p_{ij} \log(p_{i.}/p_{ij}).$$

The asymmetric uncertainty coefficient were B is the dependent variable is defined as:

$$UC_B = [H(B) - H(B|A)]/H(B).$$

Likewise, the a symetric uncertainty coefficient where A is the dependent variable is defined as:

$$UC_A = [H(A) - H(A|B)]/H(A).$$

The asymmetric uncertainty coefficient is the proportion by which "uncertainty" in the dependent variable is reduced when the value of the independent variable is known. UC_A (UC_B) lies between 0 and 1.

A symmetric uncertainty coefficient can be defined as:

$$UC = [H(A) + H(B) - H(A,B)]/[H(A) + H(B)].$$

Measures Based Upon Optimal
Prediction of Order

Suppose that two individuals are picked at random from the population; each falls into a cell of the table. If there is independence between the polytomies A and B, one would expect the order on the A polytomy for the individuals randomly taken is not related to the order on the B polytomy; if a positive association exists these orders should be positively related and conversely, if a negative association exists these orders should be negatively related. On this basis, Goodman and Kruskal proposed the following measure of association:

$$\gamma = (\pi_s - \pi_d)/(1 - \pi_t),$$

where

π_s = the probability of like ordering for two randomly selected individuals,

π_d = the probability of unlike ordering,

π_t = the probability of ties, i.e., of two randomly selected individuals falling in the same cell.

Thus, γ is the difference between the conditional probabilities of like and unlike order, given no ties. The properties of γ are:

(i) γ is indeterminate if the population is concentrated in a single row or column of the cross-classification table.

(ii) γ is 1 if the population is concentrated in an upper-left to lower-right diagonal of the cross-classification table.

(iii) γ is 0 in the case of independence, but the converse need not hold except in 2 x 2 case.[27]

Another measure based on the same principle, but taking into account the occurrences of ties, was proposed:

$$Tau_c = (\pi_s - \pi_d)/[(m - 1)/m]$$

where

m = Min (α,β),

α,β = the dimensions of the contingency table.

---

[27] Goodman and Kruskal, "Measures of Association," p. 749.

Mea

cat

inc

cec

tir

pop

pr

th

w

T

Measures of Agreement

Suppose that a population of individuals can be classified into categories using two procedures; if complete agreement exists, individuals will be classified into the same category by the two procedures. The assignment can be displayed as a two-dimensional contingency table with the same polytomy. No ordering is assumed.

The simplest measure of agreement is the proportion of the population who have been classified identically by the two procedures:

$$p = \sum_i r_{ii}$$

Goodman and Kruskal proposed a measure, λ_r, which can be interpreted as "the relative decrease in error probability as we go from the no information situation to the other-method-known situation."[28]

$$\lambda_r = [\sum_i r_{ii} - 0.5(r_{M.} + r_{.M})]/[1 - 0.5(r_{M.} = r_{.M})],$$

where $r_{M.}$ and $r_{.M}$ are the marginal frequencies of the modal classes. The properties of λ_r are:

1. λ_r ranges between -1 and +1
2. λ= -1 when all frequencies in diagonal cells (r_{ii}) are zero and the modal probability is one
3. λ_r = 1 when the two classification procedures always agree
4. λ is indeterminate only when both methods always classify individuals into one and the same class
5. λ_r takes no particular value when independence exists.

[28]Ibid., p. 757.

Me
of

si

th

of

It

in

ca

or

Co

o

a

g

Measures Based on Proportion of Explained Variance

Bishop et al.[29] propose a measure of association for two-dimensional tables based on the proportion of explained variance; it is, therefore, an analogue for qualitative variables of the coefficient of determination which is commonly used for continuous variables. It is defined as:

$$Tau_b = [\sum_j (1/x_{.j}) \sum_i x_{ij}^2 - (1/n) \sum_i x_{i.}^2]/[n-(1/n) \sum_i x_{i.}^2].$$

Goodman and Kruskal[30] interpretes Tau_b as the relative decrease in the proportion of incorrect predictions when prediction of the row category is based on the conditional proportions $r_{ij}/r_{.j}$ instead of on the marginal proportions only.

Conclusion

Goodman, Kruskal, Costner and others have stressed the importance of choosing a measure of association which is adapted to the purpose at hand and which takes into account the peculiarities of the contingency table. They also urged the use of measures of association which have an operational interpretation and especially of those measures which can be interpreted in terms of the proportional reduction in error of estimation made possible by the relationship.[31]

---

[29] Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland, Discrete Multivariate Analysis: Theory and Practice (Cambridge, Massachusets: The MIT Press, 1975), pp. 390-392.

[30] Goodman and Kruskal, "Measures of Associations," pp. 759-760.

[31] H.L. Costner, "Criteria for Measures of Association," American Sociological Review 30 (June 1965): 341-353.

Log-Linear Models of Contingency Tables

Suppose that a set of P qualitative variables are observed on a certain population of N individuals; each individual can be classified on the basis of the value taken by each qualitative variable for him. We mentioned above that a convenient structure for displaying such a data set is a P- dimensional contingency table. In this section log-linear models for contingency tables are presented, and problems in estimating them as well as in testing hypotheses arediscussed. Some general comments concerning contingency tables and the objectives served in using log-linear models are necessary.

A contingency table is called complete when all cells have a strictly positive probability; in other words these tables display no structural zeros. A complete table may display some cells with zero counts due to sampling; these are called sampling zeros. As it will be explained below, the estimation of incomplete tables is substantially more difficult than the estimation of complete tables, and requires special methods. Estimation of tables with sampling zeros does raise some difficulties, but methods available for tables without sampling zeros can be used once some precautions are taken.

A model describing the underlying data structure of a multidimentional contingency table is a mathematical relationship involving the cell probabilities (p_θ) or the expected count (m_θ).

Bishop et al.[32] list the following four objectives an experimenter may hold when fitting a model to a data-set:

---

[32]Bishop, Fienberg, and Holland, Discrete Multivariable Analysis, p. 311.

1. To describe the data: the model then helps him to understand the relationship between the variables and provides a "smoothing device" by which the most important structural elements of the data are highlighted.

2. To obtain summary statistics of different subtables (configurations) allowing him to decide whether or not the dimensionality of the table can be reduced.

3. To detect outliers, by assessing the goodness of fit of the model for each individual cell.

4. To test whether some variables are associated and to assess the magnitude of their association.

The above definition of models and statement of the purposes they may serve leads us to the actual presentation of log-linear models for contingency tables. The special case of three qualitative variables (three-dimensional table) will be taken as example for expository purposes.

Models for Three-Dimensional Tables

Log-linear models can involve either cell probabilities (p_{ijk}) or expected cell counts (m_{ijk}); the following presentation will use expected cell counts.

The log-linear model is defined by:

$$m_{ijk} = \exp(U_{ijk})/\sum_i \sum_j \sum_k \exp(U_{ijk}),$$

or, in logarithmic form by:

$$\begin{aligned} L_{ijk} &= \log m_{ijk} \\ &= U_{ijk} - \log\left[\sum_i \sum_j \sum_k \exp(U_{ijk})\right], \end{aligned}$$

where U_{ijk} is a parameter specific to each cell. Since the second term is constant over all cells of the table it can be expressed as a single term: u.

By using a parameterization similar to the one used in the analysis of variance (ANOVA) model U_{ijk} can be expressed as follows:

$$U_{ijk} = u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(ij)} + u_{23(jk)} + u_{13(ik)} + u_{123(ijk)}.$$

Thus, another equivalent form of the log-linear model is

$$L_{ijk} = \log (m_{ijk}) = u + u_{1(i)} + u_{2(j)} = u_{3(k)} + u_{12(ij)} + u_{23(jk)} + u_{13(ik)} + u_{123(ijk)},$$

$i = 1, . . ., J;\ j = 1, . . ., I;\ k = 1, . . ., K;$

where the u-terms are parameters on which the following constraints are imposed:

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_k u_{3(k)} = 0,$$

$$\sum_i u_{12(ij)} = \sum_j u_{12(ij)} = \sum_j u_{23(jk)} = \sum_k u_{23(jk)} = \sum_i u_{13(ik)} = \sum_k u_{13(ik)} = 0,$$

$$\sum_i u_{123(ijk)} = \sum_j u_{123(ijk)} = \sum_k u_{123(ijk)} = 0.$$

In words, each u-term must sum to zero over each of the indexing variables. Following the terminology of the ANOVA model the u-terms are interpreted as main and interaction effects:

1. u is the overall mean or overall effect
2. u_1, u_2, u_3 are the main effects of variables 1, 2, 3

3. u_{12}, u_{23}, u_{13} are the two-factor interaction effects

4. u_{123} is the three-factor interaction effect.

A three-factor effect measures the variation between the two factor effects in the two-dimensional tables defined by the values of the third variable. In the same way, a two factor effect measures the variation in the main effect of one variable for the different values of the second variable. Alternatively, a two-factor effect can be interpreted as the average two-factor effect on the same variables for the different values of the third. This implies that if any two-factor effect is constant over the two-way subtables, the three-factor effect is equal to zero. These results, presented for the three-dimensional table case, can easily be extended to higher dimension tables.

The model presented above is only one of the many log-linear models possible for a three-way contingency table. It is characterized by the fact that all u-terms are assumed to be different from zero; such a model is called a saturated model; one in which some u-terms are assumed equal to zero is called an unsaturated model. Whether or not a model is saturated has implications with regard to estimation as well as interpretation of the relationship between variables.

The class of hierarchical models is of special interest and requires definition. First, two u-terms are called relatives when one is subscripted by only a subset of those variables subscripting the second; the u-term with a higher number of subscripts is called a higher relative of the other (e.g., u_{123} is a higher relative of

u_{12} but also of u_{23}, u_{13} and u_1, u_2, u_3). A model is said to be hierarchical if the fact that any u-term is zero implies that all higher relatives are zero.

The importance of such hierarchical models is such that it is worthwhile to give their interpretation in the case of a three-dimensional table:

1. Three-factor effect absent, i.e., $u_{123} = 0$: the two-factor effects for all two-dimensional subtables are constant, and there is "partial association" between each pair of variables

2. Three-factor and one two-factor effect absent, e.g., $u_{123} = u_{12} = 0$: then, $L_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} + u_{13(ik)}$ and variables 1 and 2 are conditionally independent, given the level of variable 3

3. Three-factor and two two-factor effects absent, e.g., $u_{123} = u_{12} = u_{13} = 0$: then, $L_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$ and variable 1 is independent of variables 1 and 2

4. Three-factor and all two-factor effects absent, i.e., $u_{123} = u_{12} = u_{13} = u_{23} - 0$: then, $L_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$ and this is the complete independence model.

5. Noncomprehensive models where one, at least, of the main effects is zero, e.g., $u_{1(i)} = 0$: one variable plays no role and the dimensionality of the table can be reduced by summing up the array over the nonintervening variable; then $L_{ijk} = u + u_{2(j)} + u_{3(k)}$.

Whatever the dimension of a table, the number of independent parameters in a saturated log-linear model is equal to the number of cells in the table.

Models for Higher-Dimension Tables

The model presented above for a three-dimensional table can be extended to higher-dimension tables. A simpler notation can be used, which allows a model of any dimensionality.

$$L_\theta = u + u_1 + u_2 + \ \dots + u_{12} + \dots + u_{12 \dots P}$$

In such notation, θ represents the complete index set; subsets of this index set can be expressed by indexing θ, e.g., θ_1, θ_2. This notation, thus, expresses indifferently probabilities (or their logarithm) of different cells; the indexing is used to indicate in which collapsed subtable (configuration) such cells are considered:

e.g. $\theta_1 = 1,2,$

$$L_{\theta 1} = u_1 + u_2 + u_{12};$$

and $\theta_2 = 1,3,4,5,$

$$L_{\theta 2} = u_1 + u_3 + u_4 + u_5 + u_{13} + u_{14} + u_{15} + u_{34} + u_{35} + u_{45} + u_{134} + u_{145} + u_{345} + u_{1345}.$$

Log-linear models can describe completely the structure of a table; this is shown by the possibility of reconstructing a complete table once the proper number of parameters (I x J) has been specified, i.e., once a saturated log-linear model has been specified.

The main case where hierarchical models cannot adequately describe the data structure of a table is when synergism occurs, i.e., when two factors need to be present together for an effect to appear.

When contingency tables are of high dimensionality, it is obviously interesting to reduce this dimensionality in order to ease the analysis. Reducing the dimensionality of arrays, by summing over certain variables, however, implies a loss of information and may mask some important structural features. It is therefore of important theoretical interest, as well as of practical interest, to know under what conditions arrays can be collapsed.

The following theorem states, in general terms, when variables in a contingency table are collapsible:

> Theorem: Suppose the variables in an s-dimensional array are divided into three mutually exclusive groups. One group is collapsible with respect to u-terms involving a second group, but not with respect to the u-terms involving only the third group, if and only if the first two groups are independent of each other (i.e., the u-terms linking them are 0).[33]

The practical implications of the above theorem are important. If a table is described by a log-linear model where all two-factor effects are different from zero, collapsing the table by summing over any variable will change the u-terms. In other words, the common practice of examining all two-way marginal tables is very misleading in all cases where variables are interdependent. A contrario, an array can be collapsed by summing over any variables that are independent of others; by virtue of the above theorem, the u-terms will not be affected.

Maximum Likelihood Estimation Procedures for Complete Tables

The discrete probability density function for the table depends on the sampling scheme by which the data set was obtained. The three most common sampling schemes are the independent Poisson, simple

---

[33]Ibid., p. 47.

multinomial, and product multinomial sampling schemes, corresponding respectively to the cases where total sample size is not constrained, total sample size is fixed, and the number of observations for certain groups is fixed. Since the kernel of the likelihood function is identical for the three above sampling schemes, the same estimation procedure can be used. In the following, estimation procedures will be examined for these sampling schemes only.

Sufficient statistics are easily obtained; they are configurations of sums (i.e., collapsed tables obtained by summing over certain variables). Simple rules exist, based on what u-terms are included in the model, which allow the determination of which configurations constitute the minimum set of sufficient statistics. The practical implication is that it is possible to derive maximum likelihood estimates of the cell probabilities without estimating the u-terms, and without having to derive the kernel of the likelihood function.

For some models it is possible to write maximum likelihood estimates of the cell probabilities as direct functions of the sufficient statistics; these models are sometimes called direct models. For others, this is not possible and one must resort to iterative methods. Iterative methods yield, in any case, the closed-form estimates when they exist; in practice, estimates are obtained using programs which are based on iterative methods, and therefore, successful whether or not direct estimates exist. Simple rules exist which allow the prior determination of which models have closed-form estimates; it suffices here to mention that no model has such closed-form estimates unless at least one two-factor effect is equal to zero.

If estimates of u-terms are desired (e.g., to assess the extent of relationship between 2 variables), they can be obtained as linear functions of the cell probability estimates subject to the constraints that their sums over each indexed variable are equal to zero.

Goodness of Fit

Several measures of goodness of fit are available. The first is the classical Pearson χ^2 defined as:

$$\chi^2 = \sum_{\theta} [(x_\theta - \hat{m}_\theta)^2 / \hat{m}_\theta],$$

where:

x_θ = the observed cell count,

$\hat{m}_\theta$ = the fitted cell count.

χ^2 is <u>asymptotically</u> distributed as chi-square with the appropriate degrees of freedom (see below).

The second measure of goodness of fit is the likelihood ratio statistic defined as:

$$G^2 = -2 \sum_{\theta} x_\theta \log(\hat{m}_\theta / x_\theta)$$
$$= 2 \sum_{\theta} x_\theta \log(x_\theta / \hat{m}_\theta).$$

G^2 is also asymptotically distributed as chi-square with the appropriate degrees of freedom. The G^2 statistic is less familiar to users and somewhat awkward to compute; it is, however, minimized by maximum likelihood estimation methods, which designates it as the best measure of goodness of fit when such estimation methods are used. Furthermore G^2 can be broken down into parts in two meaningful ways: conditionally and structurally.

Let,

$$L(x) = \sum_\theta x_\theta \log x_\theta,$$

and,

$$L(\hat{m}) = \sum_\theta x_\theta \log \hat{m}_\theta,$$

then,

$$G^2 = -2[L(\hat{m}) - L(x)].$$

Since, for any saturated model, the maximum likelihood estimate of the expected cell count is the actual cell count ($\hat{m}_\theta = x_\theta$), $L(\hat{m}) = L(x)$ and $G^2 = 0$. For any unsaturated model $L(\hat{m}) < L(x)$ and, consequently, G^2 is positive.

In the special case of <u>nested</u> models, G^2 can be decomposed conditionally. Two log-linear models A and B are said to be nested when one (e.g., B) contains only a subset of the u-terms contained by the other (A). Then, the likelihood ratio measure of goodness of fit for model B, $G^2(\beta)$, can be decomposed into two constituents: (1) a measure of the distance between the estimated expected cell counts under model B ($\hat{m}_\theta^B$) and those obtained under model A ($\hat{m}_\theta^A$), and (2) a measure of the distance between the estimated expected cell counts under model A ($\hat{m}_\theta^A$) and the observed cell counts (x). This result is easily derived from the above expression of G^2:

$$\begin{aligned} G^2(B) &= -2[L(\hat{m}^B) - L(x)] \\ &= -2[L(\hat{m}^B) - L(\hat{m}^A) + L(\hat{m}^A) - L(x)] \\ &= -2[L(\hat{m}^B) - L(\hat{m}^A)] -2[L(\hat{m}^A) - L(x)] \\ &= G^2(B|A) + G^2(A). \end{aligned}$$

G^2 $(B|A)$ is the conditional measure of goodness of fit for model B given model A. The following result holds: if $G^2(A)$ and $G^2(B)$ are asymptotically distributed as chi-square with respectively n_A and n_B degrees of freedom, $G^2(A|B)$ is asymptotically distributed as chi-square with $n_B - n_A$ degrees of freedom (n_A and n_B are the degrees of freedom in models A and B). The above decomposition can be repated and, therefore, the results can be extended to any number of nested models. This conditional decomposition is particularly useful since it permits a test for the presence of any subset of u-terms, and especially for a single interaction term.

As mentioned previously, closed form estimates of the expected cell counts, when they exist, are functions of the minimum set of sufficient configurations. In this case, G^2 can be computed directly from these configurations, by adding the G^2 for each sufficient configuration constituting the minimum set; this decomposition of the G^2 statistics in term of the G^2 of these configurations is called the structural decomposition.

Internal Goodness to Fit

X^2 and G^2 statistics are overall measures of goodness of fit; when contingency tables possess many cells it is useful to examine also the internal goodness of fit, i.e., to check each cell for the deviation of the estimated cell count from the actual cell count. This procedure may reveal outliers, which can occur even when the overall fit of the model is good, or specific patterns of positive and negative deviates leading to the choice of a different model.

Testing of Hypotheses

Different hypotheses concerning the structure of the data can be expressed in terms of the main and interaction effects of a log-linear model. For example, direct interaction between two specified variables is equivalent to the two-factor effect involving these variables being different from zero; conversely the hypothesis that these two variables are conditionally independent is equivalent to stating that the same two-factor effect is zero, and other two-factor effects are different from zero. If one wants to test for interaction between two variables, two models can be fitted to the data, one including the two-factor effect, one excluding this term, and two measures of fit are obtained: $G^2(A)$ and $G^2(B)$ (or $\chi^2(A)$ and $\chi^2(B)$). The difference $G^2(A) - G^2(B)$ (or $\chi^2(A) - \chi^2(B)$) is asymptotically distributed as chi-square with $n_A - n_B$ degrees of freedom. The two models fitted are nested since the model corresponding to conditional independence contains one less u-term. If the difference $\Delta G^2 = G^2(B) - G^2(A)$ (or $\Delta\chi^2 = \chi^2(B) - \chi^2(A)$), which measures the increase in goodness of fit when the two-factor effect is added, is significant at the chosen level, the two-factor effect is to be considered as different from zero and the null hypothesis that the two variables are conditionally independent can be rejected.

The advantages of the above method of hypothesis testing are that: (1) the same criterion is used for estimating the model as well as the testing of hypotheses, (2) it is applicable even in the case where the table contains structural zeros, and (3) it is applicable to any level of u-terms as well as any subset of u-terms describing a hypothesis.

Hypothesis testing can also rely on the property that maximum likelihood estimators are asymptotically distributed as normal. Asymptotic t-ratios can be obtained for each coefficient in the model and be used for testing whether this coefficient is different from zero.

Choice of a Model

Since many log-linear models are possible for a specified contingency table, the problem arises of how to choose the "best" model. In this respect the problem does not differ substantially from the choice of a regression model; it is difficult to lay down strict rules which would lead systematically to the best model (the criterion to judge a model is itself not easy to choose and depends greatly upon the objective of the experimenter).

Generally, the choice of a model is reduced to a subset of all possible models, subset defined by the objectives of the study, any *a priori* knowledge concerning the variables, and the sampling scheme. The choice of a model comes out of a search procedure in which models are successively fitted and their goodness of fit assessed.

Several systematic search procedures have been proposed;[34] only one will be presented here.

The first step consists of fitting hierarchical models with u-terms of uniform orders. Thus, for a P-dimensional table, one fits the model with the P-order u-term, then the model with all (P-1)-order u-terms (i.e., without the P-order term), and so on down to the complete independence model with only main effects. These models are nested and embody a set of nested hypotheses. Using a theorem stating

---

[34]Ibid., pp. 311-343.

that likelihood ratio statistics provides a means to perform independent tests of nested hypotheses and using the conditional decomposition of G^2 presented above, it is possible to narrow down the set of adequate models to the intervening models between two models with terms of uniform order. These intervening models are themselves models which are nested into the model with terms of higher uniform order. The same test can be used to choose between these alternative intervening models.

In choosing models one should remember that a simpler model (i.e., including less parameters) is often more informative than a more complex one, because it is more likely to reveal the main structural features of complex data and because it may often yield more stable estimates of the cell probabilities.

A further problem in relation to the search for the best model, which by no means is specific to log-linear models of contingency tables,[35] consists in using the same data for choosing a model and testing hypotheses. Since the model is built on the basis of the information contained in the data, it reflects more and more, with each step of the search procedure, the idiosyncrasies of the sample and less and less the characteristics of the population. One way of dealing with this problem could be to give up some degrees of freedom, but the question as to how many should be given up is unresolved. Another remedy is to split available data into two or more parts and to choose the model using one part and to test it on the other part(s).

---

[35]See for example T. Dudley Wallace, "Pretest Estimation in Regression: A Survey," American Journal of Agricultural Economics 59 (August 1977): 431-443.

Multivariate Log-Linear Models with Exogenous Variables[36]

Nerlove and Press developed a multivariate log-linear model with exogenous variables to analyze relationships between several qualitative variables which are thought to be jointly dependent on a set of exogenous continuous variables. Their model can be described either as an extension of the univariate logit model to the case of several qualitative dependent variables, or as the introduction into the log-linear models for contingency tables of main and interaction effects which are functions of exogenous variables. The second presentation is easier and it will be followed here; many results and discussion included in the section in log-linear models of contingency tables apply to the Nerlove-Press model and consequently only differences will be emphasized in this section.

The Model

The model is similar to the one defined for contingency tables; the definition in terms of cell probabilities and with the simplified notation for any number of dimensions is as follows:

$$p_\theta = \exp(U_\theta)/\sum_\theta \exp(U_\theta).$$

Taking the logarithm:

$$L_\theta = \log p_\theta = U_\theta - \log[\sum_\theta \exp(U_\theta)]$$

---

[36]This section is based on the following works: Marc Nerlove and S. James Press, Univariate and Multivariate Log-Linear and Logistic Models, Rand Report R-1307-EDA/NIH, (Santa Monica: Rand Co., Dec. 1973); Idem, "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data," Center for Statistics and Probability, Discussion Paper No. 1, Northwestern University, Evanston, 1976.

or, $$L_\theta = u + U_\theta,$$

where, $$u = -\log[\sum_\theta \exp(U_\theta)].$$

Using the same parameterization as before it follows that:

$$\begin{aligned} L_\theta = u &+ u_1 + u_2 + \ldots + u_P \\ &+ u_{12} + u_{23} + \ldots \\ &+ \ldots \\ &+ u_{12\,\ldots\,P} \end{aligned}$$

The effect of exogenous variables on the jointly dependent qualitative variables can be introduced into the model by stating that U_θ is a function of these exogenous variables. When restricted to be linear, this function takes the following form:

$$U_\theta = x'U^*_\theta,$$

where x is a vector of exogenous variables and U^*_θ is itself expended in terms of the convenient parameterization described previously:

$$\begin{aligned} u_1 &\equiv x'u^*_1, \\ u_{12} &\equiv x'u^*_{12}, \\ u_{12..P} &\equiv x'u^*_{12...P} \end{aligned}$$

Nerlove and Press present their model as a generalization of the log-linear model of contingency tables. If x is constant, the model reduces to the ANOVA-type model presented before; if x contains true exogenous variables the model can be described as the analogue of the analysis of corariance; it is of the ANACOVA-type.

Maximum Likelihood Estimation

If a multinomial sampling scheme is assumed, the likelihood function (LF) can be derived as follows:

$$p_\theta(n) \equiv \Pr\{\text{nth observation falls in cell } \theta\},$$

$$p_\theta(n) = \exp(U_\theta)/\sum_\theta \exp(U_\theta),$$

then,

$$LF = \prod_{n=1}^{N} \prod_\theta [p_\theta(n)]^{v_\theta^{(n)}} ,$$

where,

$$v_\theta^{(n)} = \begin{cases} 1 \text{ if observation n falls in cell } \theta \\ 0 \text{ otherwise} \end{cases}$$

The u-terms may be assumed to depend on exogenous variables in which case

$$U_\theta = x_n' U_\theta^* .$$

U_θ^* is, like U_θ decomposed into main effect and interaction effects:

$$U_\theta^* = u_1^* + u_2^* + \ldots + u_{12}^* + \ldots + u_{12\ldots P}^* .$$

Estimates of the cell probabilities can be obtained by maximizing the likelihood function either with respect to the u-terms in the case of an ANOVA-type model or with respect to the u*-terms if the model is of the ANACOVA-type.

From the above formulation of the likelihood function, it is clear that the estimation procedure breaks down when the table displays structural zeros. When the table is complete (i.e., no structural zeros exist), but some sampling zeros occur, the fully saturated model cannot be estimated since some p_θ terms vanish from the likelihood by being raised to a zero power; unsaturated models, with the proper number of nonexistent parameters can, however, be fitted.

Conditional Probabilities and Conditional Estimators

Conditional probabilities and estimators will be presented taking the trivariate dichotomy case as an example.

In the case of dichotomies the constraints that every u-term sums to zero over each of the indexing variables implies that every u-term takes only one absolute value, the sign of which depends on the value (0,1) taken by the indexing variables.

Thus, the trivariate dichotomy model is as follows:

$$\log p_{111} = u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23} + u_{123}$$

$$\log p_{110} = u + u_1 - u_2 + u_3 + u_{12} - u_{13} - u_{23} - u_{123}$$

$$\log p_{100} = u + u_1 - u_2 - u_3 - u_{12} - u_{13} + u_{23} + u_{123}$$

$$\log p_{000} = u - u_1 - u_2 - u_3 + u_{12} + u_{13} + u_{23} - u_{123}$$

$$\log p_{001} = u - u_1 - u_2 + u_3 + u_{12} - u_{13} - u_{23} + u_{123}$$

$$\log p_{011} = u - u_1 + u_2 + u_3 - u_{12} - u_{13} + u_{23} - u_{123}$$

$$\log p_{101} = u + u_1 - u_2 + u_3 - u_{12} + u_{13} - u_{23} - u_{123}$$

$$\log p_{010} = u - u_1 + u_2 - u_3 - u_{12} + u_{13} - u_{23} + u_{123}$$

Conditional probabilities can easily be computed from the well-known equality:

$$P(A|B) = P(A)/P(B)$$

Thus

$$P(y_1 = 1|y_2 = 1, y_3 = 1) = p_{111}/(p_{011} + p_{111})$$

$$= \frac{\exp(u+u_1+u_2+u_3+u_{12}+u_{13}+u_{23}u_{123})}{\exp(u-u_1+u_2+u_3-u_{12}-u_{13}+u_{23}-u_{123})+\exp(u+u_1+u_2+u_3+u_{12}+u_{13}+u_{23}+u_{123})}$$

$$= \frac{\exp(u_1+u_{12}+u_{13}+u_{123})}{\exp[-(u_1+u_{12}+u_{13}+u_{123})] + \exp(u_1+u_{12}+u_{23}+u_{123})}$$

$$= \frac{1}{1 + \exp[-2(u_1+u_{12}+u_{13}+u_{123})]}$$

More generally the conditional probability of a cell is given by the formula:[37]

$$p(y_1 = 1|y_2,y_3) = \frac{1}{1 + \exp[-2(u_1+u_{12}v_2 +u_{13}v_3+u_{123}v_2v_3]},$$

where,

$$\begin{cases} v_m = 1 & \text{if } y_m = 1 \\ v_m = 0 & \text{otherwise.} \end{cases}$$

It is seen that conditional probabilities are also of a log-linear type but in u-terms which are different from those present in the original model. In the above formulation, nothing prevents the u-terms to be themselves functions of a set of exogenous variables; whatever function form the dependence of the u-terms on exogenous variables takes, it is preserved in the conditional distribution.

Thus, one can advocate the estimation of the original model from the above conditional probabilities rather than from the joint distribution. Such an estimation procedure would be analogous to the estimation of the equations of a system of linear equations through ordinary least square. In both cases, such a method is not entirely satisfactory since it does not take into account that the dependent variables (whether qualitative or quantitative) are jointly dependent.

---

[37]See Nerlove and Press, Log-Linear and Logistic Models, p. 50.

Such an estimation procedure does present some advantages and consequently deserves some attention.

A conditional likelihood function can be generated from the expression of the conditional probability:

$$LF^*(u_1,u_{12},u_{13},u_{123}) = \prod_{n=1}^{N} \left[1/[1+\exp(-2t_n)]\right]^{y_{1n}} \left[1/[1+\exp(2t_n)]\right]^{1-y_{1n}}$$

where,

$$t_n = u_1 + u_{12}v_{2n} + u_{13}v_{3n} + u_{123}v_{2n}v_{3n},$$

m indexes the dependent variables (m = 1...3)

n indexes the observations (n = 1...N),

and, $v_{mn} = \begin{cases} 1 \text{ if } y_{mn} = 1, \\ \text{o otherwise.} \end{cases}$

This conditional likelihood function can be maximized with respect to the u-terms or, if the u-terms are considered to be functions of exogenous variables, with respect to the corresponding u*-terms. The estimators based on the maximization of the conditional likelihood function are called conditional estimators.

Justification for the use of conditional estimators for jointly dependent qualitative variables can be derived from the fact that it is always possible to infer the joint distribution of qualitative variables from the knowledge of their conditional distribution, whereas this is not generally possible for jointly dependent continuous variables.

Conditional estimators, as defined above, are consistent, asymptotically unbiased and asymptotically normally distributed, but, because the joint dependence of the qualitative variables is ignored, they are generally not efficient. Conditional estimators possess some advantages: (1) they are less costly to derive because they are easier to compute than full maximum likelihood estimators, (2) they allow the use of the many available univariate logit programs, and (3) estimates may be close approximations of the full maximum likelihood estimates. These conditional estimators can, however, only be used for exploratory purposes, since testing hypotheses by means of the likelihood ratio test requires the use of the maximum likelihood method based on the joint distribution of the dependent variables.

Computer Programs

Many univariate logit programs are available and they can be used to obtain conditional estimates.

Nerlove and Press developed a program which can handle up to four dichotomous variables which are jointly dependent on up to sixteen exogenous variables. Main effects only are allowed to be linear functions of exogenous variables, and interaction terms are assumed to be constant. Programs for polytomous variables and with interaction effects dependent on exogenous variables are being developed.

Joint Estimation Using Generalized Least Squares

Zellner and Lee[38] proposed an estimation method of models with jointly dependent dichotomous variables using the generalized least squares estimator (G.L.S.). The case of a system of linear probability

---

[38]Zellner and Lee, "Joint Estimation," pp. 387-392.

functions is chosen for expository purposes, although the method can be generalized with little modification, to logit or probit probability functions.

The system of M linear probability functions can be represented as follows:

$$\begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \cdot & & \cdot\cdot & \cdot \\ \cdot & & \cdot \quad \cdot & \cdot \\ \cdot & & \cdot & \cdot\cdot \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix}$$

where:

p_j = T_j x 1 vector of observed proportions for the j^{th} binary variable,

X_j = a T_j x K_j matrix of observations on K_j exogenous variables,

β_j = a K_j x 1 vector of coefficients,

u_j = a T_j x 1 disturbance vector.

The above system can be represented can be represented more simply as follows:

$$p = X\beta + u,$$

where:

$$p = (p_1', p_2', \ldots, p_M')',$$

$$\beta = (\beta_1', \beta_2', \ldots, \beta_M')',$$

$$u = (u_1', u_2', \ldots, u_M')'.$$

The covariance matrix is:

$$\Sigma = E(uu') = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1M} \\ D_{12} & D_{22} & \cdots & D_{2M} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ D_{M1} & D_{M2} & \cdots & D_{MM} \end{bmatrix}$$

where D_{jk} are diagonal matrices whose coefficients are expressable in terms of the probability that dichotomous variables take the value 1 for each group of observations.[39] The covariance matrix (Σ) can be estimated consistently using single equation procedures. Once such a consistent estimate Σ_e of Σ has been obtained, Aitken's generalized least squares estimator,

$$b = (X' \Sigma_e^{-1} X)^{-1} X' \Sigma_e^{-1} p,$$

can be used to obtain joint estimates.

The major shortcoming of this method is that it requires grouped data, i.e., several observations for the same values of the exogenous variables. When the number of qualitative dependent variables and particularly when the number of continuous exogenous variable is high, the grouping of observations can only be achieved by loosing much information. In such circumstances, maximum likelihood estimation methods can still be used. Also, this method is only valid for dichotomous dependent variables.

---

[39]For the derivation of the exact expression see Zellner and Lee, "Joint Estimation," pp. 388-390.

Conclusion

This chapter has provided an overview of the statistical methods for the analysis of qualitative variables. It only touched the surface of a vast field, trying to demonstrate that standard econometric methods which are familiar to applied econometricians are very inadequate and that adequate techniques do exist or are presently under development.

Some of the statistical methods reviewed in this chapter are more adapted to exploratory data analysis; these are: (1) measures of association in contingency tables (mainly two-dimensional contingency tables), and (2) discriminant analysis (for a dichotomous or polytomous dependent variable).

Other methods, which impose a more constraining statistical structure on the data, are more adapted to confirmatory data analysis, although some exploration is needed to arrive at a model to be tested.

This review showed that several statistical methods are available for the analysis of a single dichotomous variable; computer programs are readily available and, consequently, no major obstacle should hinder the use of these methods by applied econometricians.

In connection with the use of contingency tables, the distinction between tests of independence and measures of association was emphasized. This distinction is often ignored by applied econometricians even though it has some very significant implications for the interpretation of empirical results.[40]

---

[40] This is exemplified in Chapter VI.

Log-linear models of multidimensional contingency tables allow the analysis of several polytomous variables considered to be jointly dependent; computer programs are available and documented.[41] Exogenous variables can be introduced into such log-linear models in the main effects or in the interaction effects. To date, a program for up to four dichotomous variables with exogenous variables entering only into the main effects is available; development of programs for polytomous variables and variable interaction effects are under way. A major shortcoming of the log-linear models of contingency tables, with or without exogenous variables, is that polytomous variables are treated as if there was no underlying order. Consequently, if such ordering does exist, some information is lost.

Several of the above statistical methods were used to analyze the available data on off-farm movement. Discriminant analysis and two-dimensional contingency tables were used in an exploratory mode. A log-linear model of a three-way contingency table with exogenous variables was explored on part of the data and tested on other parts.

---

[41]See, for example, S. J. Haberman, "Log-Linear Fit for Contingency Tables," Applied Statistics 21 (1972): 218-225.

CHAPTER VI

EMPIRICAL FINDINGS

This chapter is divided into three main sections. In the first section, an overview of the extent of total exit and pre-retirement exit from farming in Saskatchewan between 1966 and 1971 is provided. In the second section, exploratory results obtained from cross-classifications and discriminant analyses are presented. In the third section, a log-linear model of pre-retirement exit from farming, considered as jointly dependent with involvement in off-farm work and residence on the farm, is(1) chosen and estimated using only part of the data, and (2) tested on other parts.

Extent of Exit and Pre-retirement Exit from Farming

The Census of Agriculture Match[1] permits a data link between two censuses as exemplified in Table 5. This particular one displays farm operators in Saskatchewan cross-tabulated by age class in 1966 and age class in 1971 for stayers, tabulated by age class in 1966 for exiters, and tabulated by age class in 1971 for entrants. Table 5 was the main instrument in assessing the validity of the Census of Agriculture Match: matches in cells, off the diagonal are potential mismatches.

[1]For a description and an assessment of the Census of Agriculture Match see the section "The Data-Base" in Chapter IV.

Table 5. Number of Stayers, Entrants, and Exiters Classified by Age in 1966 and 1971, Saskatchewan

Age in 1966	Age in 1971								Total (1)-(8)	Exiters	Number of Operators in 1966
	Under 25 (1)	25-34 (2)	35-44 (3)	45-54 (4)	55-59 (5)	60-64 (6)	65-69 (7)	Over 70 (8)			
Under 25 (9)	149	1661	26	29	3	4	1	0	1873	758	2631
25-34 (10)	18	3584	5456	244	47	50	36	28	9463	2374	11837
35-44 (11)	38	135	7720	8916	81	55	34	23	17002	3527	20529
45-54 (12)	57	154	212	9708	8266	293	65	47	18802	5190	23992
55-59 (13)	44	100	82	171	305	5879	179	24	6784	3221	10005
60-64 (14)	26	122	104	133	59	126	3337	96	4003	3491	7494
65-69 (15)	28	73	101	93	38	44	91	1509	1977	2549	4526
Over 70 (16)	11	38	52	67	15	5	11	1245	1444	2973	4417
Total (9)-(16)	371	5867	13753	19361	8814	6456	3754	2972	61348	24083	85431
Entrants	2361	3796	2803	2805	1160	1043	677	710	15355		
Number of Operators in 1971	2732	9663	16556	22166	9974	7499	4431	3682	76703		

Source: 1966-1971 Census of Agriculture Match.

Note: Stayers appear at the intersection of lines (9) to (16) and columns (1) to (8).

Table 5 provides estimates of the gross flows into and out of farming for the 1966-71 period. During this period an estimated 24,083 farm operators left farming while 15,355 entered farming, thereby reducing the total number of farm operators from 85,431 to 76,703. Exiters represented 39.2 percent of 1966 farm operators and entrants represented 20.0 percent of 1971 farm operators. The estimate of the gross rate of exit is considerably larger than the estimate of the net rate of exit. This result strengthens the plea made earlier for the use of data on gross rates of exit and entry, rather than net rates. The Census of Agriculture Match, however, provides an estimate of gross flows which is biased downwards since it masks all entries and exits followed by counter movement occurring within the intercensal period. These results confirm those of Hathaway and Perkins.[2]

The percentage of exiters varies according to the 1966 age class of farm operators as shown below:

Age-Class in 1966	Percentage of Exiters
Under 25	28.8
25 - 34	20.1
35 - 44	17.2
45 - 54	21.6
55 - 59	32.2
60 - 64	46.6
65 - 69	56.3
Over 70	67.3

The percentage of exiters is higher for young farmers under 25 and between 25 and 34 than for mature farm operators between 35 and 44; for age-classes above 45, the percentage of exiters increases with age.

[2]Hathaway and Perkins, "Occupational Mobility," p. 186.

Age of farm operators was used to define and select pre-retirement exiters; 65 was considered to be the normal retirement age and, consequently, all age-classes under 65 in 1966 were considered to represent pre-retirement exiters.

Exploratory Results from Cross-Tabulations and Discriminant Analyses

As explained in Chapter IV, an exploratory data-analysis framework was adopted for this section. Briefly stated, such a framework consists of "looking at the data" in order to observe possible relationships. This look, however, is not free of prior knowledge but corresponds merely to low level of such prior knowledge; also, the structure imposed on the data by the statistical methods is weak.

The exploratory approach was used on a limited subset of the data. More precisely, it was used on data concerning census division 7 of Saskatchewan. Two statistical methods were used: (1) analysis of contingency tables for relationships between decision regarding continuation of farming and categorical or categorized variables, and (2) linear discriminant analysis to identify continuous variables which discriminate well between stayers and exiters.

Contingency Tables

Cross-tabulations of farm operators by decision regarding farming and by residence on the farm, age, off-farm income, sales of agricultural products, days of off-farm work, tenure, total capital value, and acreage of improved land, are in Appendix B. Each table is provided for all operators as well as for farm operators less than 65 years old.

For all tables, the χ^2 statistic takes high values; thus, the test of independence based on the χ^2 statistic leads to the rejection of the independence hypothesis at a level of significance of less than 0.01.

Two types of measures of association are displayed: the asymmetric λ's and the asymmetric uncertainty coefficients (UC).[3] λ_A and UC_A for which the decision regarding farming is considered as dependent are especially relevant. For all tables, λ_A and UC_A (as well as λ_B and UC_B), take very small values as summarized in Table 6. These results imply that knowledge of the value taken by the independent variable entails very little gain in the ability to predict the dependent variable.

The conjunction of low level of significance for the rejection of the independence hypothesis and of low values for the measures of association is not contradictory. As was explained in Chapter V, the χ^2 statistic is not a proper measure of association because it lacks any operational interpretation. The above results clearly exemplify the difference between a test of independence, i.e., the test for the existence of an association between variables, and the measure of an association. In this particular case, it can be concluded that relationships <u>most likely exist</u> between pre-retirement exit from farming and the variables considered, but that these relationships are <u>very</u>

---

[3]As explained in Chapter V, λ_A (or λ_B) can be interpreted as the proportion of improvement in the ability to predict the value of A (or B), considered as a dependent variable, once the value of the other variable, considered as independent, is known; UC_A (or UC_B) can be interpreted as the proportion by which "uncertainty" in the dependent variable is reduced by knowledge of the value of the independent variable.

Table 6. Measures of Association Between Pre-Retirement Exit and Some Selected Variables

	λ_A [a]	UC_A [b]
Residence	0.000	0.021
Age	0.000	0.027
Off-Farm Income	0.000	0.007
Sales of Agricultural Products	0.072	0.070
Days of Off-Farm Work	0.000	0.017
Tenure	0.000	0.029
Total Capital Value	0.053	0.066
Acreage of Improved Land	0.043	0.059

[a] λ_A is known as the asymmetric lambda measure of association; pre-retirement exit is here considered as the dependent variable.

[b] UC_A is the asymmetric uncertainty coefficient; pre-retirement exit is here considered as the dependent variable.

weak. The very low levels of significance for the χ^2-tests are obtained through the availability of a large number of observations.

Discriminant Analyses

Discriminant analysis allows multivariate exploratory data analysis, whereas cross-tabulations are convenient only in the bivariate case. Tables 7 to 12 display results of the discriminant analyses which were performed with the groups of stayers and exiters below 65 years of age.

Table 7 displays standardized coefficients of the discriminant function including a first set of variables describing the general structure of the farm: AVAGE, TOTAREA, LDOWNED, LDRENTED, $LDBLD,

$MACH, $STOCK, TOTCAP, and OFFWORK.[4] The discriminating power of this function is very low, as indicated by an eigenvalue of 0.053 and a Wilk's lambda of 0.949. Variables with highest weight in discriminant function are TOTCAP, $MACH, $LDBLD, and AVAGE; variables with lowest discriminating power are LDRENTED, $GRAIN, and LDOWNED.

Table 8 displays standardized coefficients of a second discriminant function where TOTAREA, TOTCAP, and TOTSALES have been deleted, and ARCROP, ARIMP, ARSF, WAGES, TAXES, TOTRENT, $CATTLE, and $PIG have been added. The discriminating power is also low: the eigenvalue is equal to 0.069 and Wilk's lambda is 0.936. Variables with highest standardized coefficients are $MACH, $STOCK, ARCROP, $WHEAT and those with smallest coefficients are $GRAIN, $PIG, and $CATTLE.

Table 9 displays standardized coefficients of a third discriminant function in which percentage variables have been included: %LDOWNED, %ARCROP, %ARIMP, %ARSF, %ARWOOD, %ARUNIMP, %$WHEAT, %$GRAIN, %$CATTLE, %$PIG, %$LDBLD, and %$MACH. Discriminating power is again very low, with an eigenvalue of 0.047 and a Wilk's lambda of 0.955. Variables with largest discriminating power are %ARUNIMP, %ARCROP, %ARSF, %$LDBLD, TOTCAP, %$WHEAT, %ARIMP and AVAGE. Variables with lowest discriminating power are %LDOWNED, %ARWOOD, TOTSALES, and TOTAREA.

Table 10 displays coefficients of a discriminant function where distances to towns (DIST1, DIST2, DIST3, DIST4), productivity of capital and land (PRODCAP, PRODLD1) are introduced; also, transformations

---

[4]Description of variables can be found in Appendix A, Table A.1.

Table 7. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	-0.442
TOTAREA	-0.301
LDOWNED	-0.159
LDRENTED	-0.008
$LDBLD	-0.740
$MACH	0.767
$STOCK	0.336
TOTCAP	1.149
OFFWORK	-0.315
$WHEAT	-0.368
$GRAIN	-0.083
TOTSALES	-0.207

Note: Eigenvalue = 0.053
Canonical correlation = 0.225
Wilk's Lamda = 0.949
Chi-square = 46.4
DF = 12
Level of significance = 0.0
Percentage of correctly classified cases = 58.6
Proportional chance criterion = 64.6
Score of mean of stayers = 0.123
Score of mean of exiters = -0.411

Table 8. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	-0.301
LDOWNED	-0.126
LDRENTED	-0.241
$LDBLD	-0.175
$MACH	0.700
$STOCK	0.682
ARCROP	0.634
ARIMP	0.135
ARSF	0.144
OFFWORK	-0.239
WAGES	-0.232
TAXES	-0.371
TOTRENT	0.213
$WHEAT	-0.428
$GRAIN	-0.036
$CATTLE	-0.118
$PIG	0.081

Note: Eigenvalue = 0.069
Canonical correlation = 0.254
Wilk's Lambda = 0.936
Chi-square = 59.2
DF = 17
Level of Significance = 0.00
Score of mean of stayers = 0.139
Score of mean of exiters = -0.463

Table 9. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	-0.405
TOTAREA	-0.091
TOTCAP	0.509
OFFWORK	-0.340
TOTRENT	0.272
TOTSALES	-0.085
%LDOWNED	-0.077
%ARCROP	0.686
%ARIMP	0.482
%ARSF	0.603
%ARWOOD	0.082
%ARUNIMP	0.910
%$WHEAT	0.489
%$GRAIN	0.343
%$CATTLE	0.311
%$PIG	0.180
%$LDBLD	-0.577
%$MACH	-0.234

Note: Eigenvalue = 0.047
Canonical correlation = 0.213
Wilk's Lambda = 0.955
Chi-square = 40.68
DF = 18
Level of significance = 0.0
Percentage of correctly classified cases = 62.5
Proportional chance criterion = 64.6
Score of mean of stayers = 0.115
Score of mean of exiters = -0.393

Table 10. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	0.233
OFFWORK	-0.049
TAXES	0.478
TOTRENT	-0.124
%LDOWNED	-0.110
%ARCROP	-0.178
%ARSF	-0.038
%$WHEAT	-0.046
%$LDBLD	0.528
%$MACH	0.239
DIST1	-0.041
DIST2	0.361
DIST3	-0.131
DIST4	-0.282
PRODCAP	0.080
PRODLD1	-0.107
LNTOTAR	-0.407
LNTOTCAP	-0.755
LNTOTSAL	-0.051

Note: Eigenvalue = 0.124
Canonical correlation = 0.332
Wilk's Lambda = 0.890
Chi-square = 102.7
DF = 19
Level of significance = 0.0
Percentage of correctly classified cases = 68.3
Proportional chance criterion = 64.6
Score of mean of stayers = -0.176
Score of mean of exiters = 0.599

of variables are considered: LNTOTAR, LNTOTCAP, and LNTOTSAL. The discriminating power is higher than for previous discriminant functions, but is still low in absolute terms, as indicated by an eigenvalue of 0.124 and a Wilk's lambda of 0.890. Variables with largest standardized coefficients are LNTOTCAP, %$LDBLD, and TAXES; variables with lowest standardized coefficients are %ARSF, DIST1, %$WHEAT, OFFWORK, LNTOTSAL and PRODCAP.

Table 11 displays results from a fifth discriminant analysis where percentage variables have been deleted, distances, productivities, other variables describing the structure of the farm, or transformations of these variables are included. The eigenvalue and Wilk's lambda are respectively equal to 0.127 and 0.887. Variables with highest discriminating power are LNTOTCAP, TOTCAP, LNTOTAR, DIST2 and DIST4. Variables with lowest discriminating power are PRODCAP, OFFWORK, LNTOTREN, and DIST1.

Table 12 provides the standardized coefficients of a last discriminant function with all variables with some discriminating power in previous functions and some others of particular theoretical interest. Discriminating power is somewhat higher than for previous discriminant functions but still low in absolute terms: eigenvalue is 0.144 and Wilk's lambda is 0.874. Variables with largest standardized coefficients are LNTOTCAP, %$LDBLD, LNTOTAR, TOTCAP. Variables with lowest standardized coefficients are PRODCAP, TOTSALES, OFFWORK, LNTOTREN, DIST1 and %LDOWNED.

In Chapter IV a set of hypotheses was stated with respect to the relationships between pre-retirement exit from farming and certain

TABLE 11. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	0.168
TOTAREA	0.168
TOTCAP	0.623
OFFWORK	0.025
TOTRENT	-0.154
TOTSALES	-0.094
DIST1	-0.066
DIST2	0.317
DIST3	-0.155
DIST4	-0.309
MECH1	-0.292
PRODCAP	0.005
PRODLD1	0.175
LNTOTAR	-0.364
LNTOTCAP	-0.954
LNTOTREN	-0.044
LNTOTSAL	-0.132
LNOFFWOR	-0.159

Note: Eigenvalue = 0.127
Canonical correlation = 0.336
Wilk's Lambda = 0.887
Chi-square = 106.7
DF = 18
Level of significance = 0.0
Percentage of correctly classified cases = 71.5
Proportional chance criterion = 64.6
Score of mean of stayers = -0.175
Score of mean of exiters = 0.579

Table 12. Standardized Coefficients of Discriminant Function for Exiters and Stayers, 64 or Less, Census Division 7, Saskatchewan

Variable	Standardized Coefficient
AVAGE	0.170
TOTAREA	0.217
TOTCAP	0.369
OFFWORK	0.034
TAXES	0.233
TOTRENT	-0.160
TOTSALES	-0.024
%LDOWNED	-0.068
%$LDBLD	0.420
%$MACH	0.194
DIST1	-0.057
DIST2	0.279
DIST3	-0.105
DIST4	-0.264
MECH1	-0.273
PRODCAP	0.009
PRODLD1	0.189
LNTOTAR	-0.408
LNTOTCAP	-0.943
LNTOTREN	-0.034
LNTOTSAL	-0.105
LNOFFWOR	-0.180

Note: Eigenvalue = 0.144
Canonical correlation = 0.355
Wilk's Lambda = 0.874
Chi-square = 119.3
DF = 22
Level of significance = 0.0
Percentage of correctly classified cases = 71.1
Proportional chance criterion = 64.6
Score of mean of stayers = -0.188
Score of mean of exiters = 0.623

variables. In the following paragraphs, results from discriminant analyses are used to provide evidence for or counter-evidence to these hypotheses.

Hypothesis 1 states that pre-retirement exit from farming is negatively related to age. AVAGE was not found to be a powerful discriminating variable; furthermore, in all discriminant analyses pre-retirement exit appeared to be positively related to age. Thus, results seem to provide counter-evidence to Hypothesis 1. The fact that age enters linearly in discriminant analysis may explain these unexpected results.

Hypothesis 2 states that pre-retirement exit from farming is positively related to farm operator's involvement in off-farm work. OFFWORK was found to have low discriminating power; it was sometimes positively related and other times negatively related to pre-retirement exit. These results do not support Hypothesis 2.

Hypothesis 3 states that pre-retirement exit from farming is negatively related to the degree the family owns the farm land. LDOWNED and %LDOWNED were found to have little discriminating power. LDOWNED was found to be positively related to pre-retirement exit (but, so was LDRENTED) and %LDOWNED was found to be sometimes positively and other times negatively related to pre-retirement exit from farming. These results do not support Hypothesis 3.

Hypothesis 4 states that pre-retirement exit is negatively related to total acreage, total capital value, and total sales of agricultural products. TOTAREA was found to have little discriminating power, but LNTOTAR showed greater discriminating power; it is disturbing, however, to observe that coefficients of TOTAREA and LNTOTAR take opposite signs when these variables are both included in the discriminant function. TOTCAP and LNTOTCAP were found to have high (in relative terms) discriminating power in all discriminant functions. Again, it is disturbing to observe that coefficients of TOTCAP and LNTOTCAP take opposite signs when these variables are both included in a discriminant function. When only one of TOTCAP and LNTOTCAP is included in the analysis, pre-retirement is found to be negatively related to that variable; this result provides evidence for Hypothesis 4. TOTSALES and LNTOTSALES were found to have very little discriminating power; this does not support Hypothesis 4. In summary, the composite Hypothesis 4 is supported only with respect to the hypothesized negative relationship between pre-retirement exit and total capital value of the farm.

Hypothesis 5 states that pre-retirement exit from farming is negatively related to productivity of land and capital. PRODLD1 was found to have little discriminating power and to be negatively related to pre-retirement exit. PRODCAP was found to have very little discriminating power. These results provide only weak and partial evidence to Hypothesis 5.

Hypothesis 6 states that pre-retirement exit is negatively related to mechanization. MECH1 was found to have little discriminating power and to be negatively related to pre-retirement exit. These results provide some evidence in support of Hypothesis 6.

Hypothesis 7 states that pre-retirement exit from farming is negatively related to distance to towns. DIST1 and DIST3 were found to have very little discriminating power and to be negatively related to pre-retirement exit from farming. DIST2 and DIST4 were found to have some discriminating power and to be respectively positively and negatively related to pre-retirement exit. Empirical evidence seems, therefore, to be ambiguous: on one hand distance to large urban centers (Saskatoon and Regina) is negatively related to pre-retirement exit, but on the other hand distance to small or medium urban centers is either not related or related positively to pre-retirement exit from farming.

Hypothesis 8, concerning relationship between pre-retirement exit from farming and residence, was not investigated because of the theoretical problems involved in including qualitative variables in discriminant analysis.[5]

A Multivariate Log-Linear Model of Pre-retirement Exit from Farming

In the preceding sections statistical analysis of available data was performed, with the decision regarding exit from farming being considered as dependent. Two other decisions made by the farm operator should be considered as jointly dependent with the decision regarding farming: (1) the decision concerning place of residence, and (2) the decision concerning involvement in off-farm work.

In this section a log-linear model of jointly dependent binary variables[6] is used to investigate: (1) the interaction between the

[5]See section on discriminant analysis in Chapter V.

[6]See section entitled "Multivariate Log-Linear Model with Exogenous Variables" in Chapter V.

three above mentioned decisions, and (2) their dependence on a set of exogenous variables.

The statistical analysis proceeded in two main stages: (1) the choice of a model using only part of the data set, and (2) the testing of this model on other parts of the data set.

Choice of a Model

The first step in the search for an adequate model consisted of fitting several saturated[7] log-linear models in which main effects are function of different sets of exogenous variables. This method was necessary because the Nerlove-Press program, which was used, is limited to a maximum of 16 exogeneous variables.

On the basis of the findings using discriminant analysis some variables, e.g., DIST1 and DIST3, were not even considered for inclusion. The objective of this first step is to select the set of exogenous variables to be included in the model to be tested. This exploration can be conducted using conditional estimates; both conditional and full maximum likelihood estimates are, however, displayed in the tables mainly to illustrate the fact that conditional and full likelihood estimates are close.

Table 13 displays conditional and full maximum likelihood estimates of a saturated model including the following exogenous variables: AVAGE, TOTAREA, LDOWNED, TOTCAP, TOTRENT, TOTSALES, %LDOWNED, %$LDBLD, %$STOCK, MECH1, PRODCAP, PRODLD1, LNTOTAR,

---

[7]See Chapter V for detailed explanation: a saturated model is a model where all interaction effects are present.

Table 13. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate and Trivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7, Saskatchewan

Dependent Variable	MAIN EFFECTS																
	CONSTANT	AVAGE	TOTAREA	LDOWNED	TOTCAP	TOTRENT	TOTSALES	%LDOWNED	%$LDBULD	%$STOCK	MECH	PRODCAP	PRODLD	LNTOTAR	LNTOTCAP	LNTOTSAL	TAXES
A Maximum Likelihood Estimates Based on the Joint Probability Function*																	
EXIT	2.23	.00960	.0000649	.000226	.0000616	-.000864	-.0000204	-.00199	.00724	-.00909	-.0598	.0732	.00104	-.263	-.292	-.0611	.00151
	(3.00)	(2.24)	(2.03)	(1.55)	(.32)	(-1.68)	(-.42)	(-1.15)	(1.77)	(-1.51)	(-.62)	(1.08)	(.21)	(-2.54)	(-1.81)	(-.98)	(1.13)
RESID2	-3.14	.00515	.000514	-.000329	-.000469	.00000253	.0000850	.00502	.00789	.0614	.237	.0399	-.0156	-.466	.723	.0820	-.000315
	(-3.65)	(1.14)	(2.13)	(-1,21)	(-1.91)	(.006)	(.61)	(2.24)	(1.67)	(6.68)	(.90)	(.40)	(-1.32)	(-2.62)	(3.79)	(1.16)	(-.17)
OFFWORK2	3.13	-.0216	.0000818	-.000147	.0000203	-.0000399	.000264	.000604	.00281	-.00739	-.0331	.00168	-.00149	-.412	-.0360	.00663	.000754
	(3.27)	(-4.82)	(1.81)	(-.54)	(.69)	(-.80)	(-1.58)	(.30)	(.66)	(-1.07)	(-.29)	(.31)	(-.26)	(-3.32)	(-.18)	(.99)	(.32)
B Maximum Likelihood Estimates Based on the Conditional Probability Function																	
EXIT	2.31	.00981	.0000665	.000227	.0000626	-.000870	-.0000211	-.00201	.00715	-.00912	-.0598	.0753	.000858	-.276	-.277	-.0616	.00147
	(3.08)	(2.29)	(2.07)	(1.56)	(.33)	(-1.69)	(-.44)	(-1.15)	(1.75)	(-1.52)	(-.62)	(1.12)	(.17)	(-2.57)	(-1.72)	(-.99)	(1.11)
RESID2	-3.04	.00485	-.000504	-.000337	-.000476	.0000190	.0000905	.00519	.00801	.0609	.272	.0200	-.0165	-.442	.722	.0857	-.00340
	(-3.25)	(1.07)	(2.00)	(-1.24)	(-1.86)	(.05)	(.63)	(2.32)	(1.61)	(6.55)	(.87)	(.23)	(-1.36)	(-2.19)	(3.56)	(1.18)	(-.18)
OFFWORK2	3.35	-.0210	.0000814	-.000160	.0000499	-.0000289	-.000276	.000521	.00318	-.00767	-.0365	.00168	-.00124	-.415	-.0477	.0123	.000850
	(3.55)	(-4.72)	(1.83)	(-.59)	(.18)	(-.06)	(-1.64)	(.26)	(.74)	(-1.11)	(-.32)	(.03)	(-.22)	(-3.29)	(-.24)	(.18)	(.37)

	BIVARIATE INTERACTION EFFECTS			TRIVARIATE INTERACTION EFFECT
	EXIT	RESID2	OFFWORK2	
EXIT		-.0967	-.0467	-.0442
		(-1.66)	(-.79)	(-.84)
RESID2	-.0967		-.208	-.0442
	(-1.66)		(-3.47)	(-.84)
OFFWORK2	-.0467	-.208		-.0442
	(-.79)	(-3.47)		(-.84)
EXIT		-.0469	.0166	-.115
		(-.77)	(.19)	(-1.04)
RESID2	-.0484		-.156	-.784
	(-.77)		(-2.50)	(-.67)
OFFWORK2	-.0218	-.163		-.0803
	(-.25)	(-2.55)		-.70

Figures in parentheses are asymptotic t-ratios.

*Logarithm of likelihood function = -1194.66

LNTOTCAP, LNTOTSAL, and TAXES. Asymptotic t-ratios can be used to decide which variables should be included in the model. Asymptotic t-ratios associated to coefficients of MECH1, PRODCAP, PRODLD1, LNTOTSAL, and TAXES are low in all three equations and, therefore, these variables need not be included in the final model.

Table 14 displays estimates of a similar model which involves AVAGE, LDOWNED, $MACH, %LDOWNED, %$LDBLD, %$STOCK, DIST2, DIST4, LNTOTAR, LNTOTCAP and LNTOTSAL as exogenous variables. Asymptotic t-ratios associated to coefficients of DIST2, %$LDBLD, %LDOWNED, and $MACH are small and, therefore, these variables need not be retained in the final model.

Table 15 displays estimates of a model which differs from the one displayed in Table 14 only by the deletion of $MACH and DIST2. The asymptotic t-ratio corresponding to DIST4 is small in all three equations and consequently DIST4 need not be retained in the final model.

In Table 16, which displays estimates of another similar model, the asymptotic t-ratios associated to TOTRENT are small, and similarly, TOTRENT will not be retained.

In summary, exogenous variables whose coefficients have large enough t-ratios and which deserve to be retained in the model are AVAGE, TOTAREA, LDOWNED, %LDOWNED, %LDBLD, %$STOCK, LNTOTAR, and LNTOTCAP.

The second step in the search for a suitable model consists of determining what level of interaction among dependent variables should be chosen and, for that level, what interaction terms should be retained.

Table 14. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate and Trivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

Dependent Variable	MAIN EFFECT CONSTANT	AVAGE	LDOWNED	$MACH	%LDOWNED	%$LDBLD	%$STOCK	DIST2	DIST4	LNTOTAR	LNTOTCAP	LNTOTSAL
A Maximum Likelihood Estimates Based on the Joint Probability Function*												
EXIT	1.85 (2.84)	.0104 (2.45)	.000406 (2.79)	-.000377 (-.44)	-.00151 (-.95)	.00680 (1.33)	-.00761 (-1.11)	.00340 (1.44)	-.00400 (-1.50)	-.193 (-2.56)	-.350 (-2.71)	-.0216 (-.47)
RESID2	-2.77 (-3.97)	.00533 (1.18)	-.0000509 (-.29)	-.000518 (-.58)	.00269 (1.46)	.00203 (.37)	.0553 (5.84)	-.00248 (-1.00)	.00435 (1.54)	-2.47 (-2.31)	.561 (3.65)	.0825 (1.54)
OFFWORK2	3.06 (4.03)	-.0217 (-4.85)	-.000221 (-.92)	-.00105 (-.98)	.00108 (.57)	.00109 (.20)	-.00816 (-1.10)	.000358 (.14)	-.00295 (-1.00)	-.238 (-2.94)	-.0772 (-.58)	-.0650 (-1.32)
B Maximum Likelihood Estimates Based on the Conditional Probability Function												
EXIT	1.93 (2.89)	.0105 (2.47)	.000404 (2.79)	-.000360 (-.42)	-.00150 (-.94)	.00676 (1.33)	-.00757 (-1.11)	.00338 (1.42)	-.00392 (-1.47)	-.201 (-2.63)	-.342 (-2.66)	-.0120 (-.45)
RESID2	-2.67 (-3.77)	.00489 (1.08)	-.0000590 (-.34)	-.000641 (-.72)	.00278 (1.51)	.00126 (.23)	.0540 (5.73)	-.00243 (-.98)	.00445 (1.57)	-.249 (-2.33)	.600 (3.87)	.0812 (1.52)
OFFWORK2	3.19 (4.18)	-.0210 (-4.74)	-.000237 (-.98)	-.000976 (-.91)	.00108 (.56)	.00174 (.31)	-.00814 (-1.10)	.000577 (.22)	-.00300 (-1.01)	-.240 (-2.93)	-.0831 (-.63)	-.0616 (-1.24)

	Bivariate Interaction Effects EXIT	RESID2	OFFWORK2	Trivariate Interaction Effect
EXIT		-.0815 (-1.42)	-.0362 (-.62)	-.0228 (-.44)
RESID2	-.0815 (-1.42)		-.183 (-3.10)	-.0228 (-.44)
OFFWORK2	-.0362 (-.62)	-.183 (-3.10)		-.0228 (-.44)
EXIT		-.0475 (-.78)	.0186 (.22)	-.0945 (-.87)
RESID2	-0.389 (-.62)		-.159 (-2.56)	-.0524 (-.46)
OFFWORK2	-.0365 (-.41)	-.159 (-2.54)		-.0368 (-.32)

Figures in parentheses are asymptotic t-ratios

*Logarithm of likelihood function = -1206.59

Table 15. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate and Trivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

Dependent Variable	Main Effects										Bivariate Interaction Effects			Trivariate Interaction Effect
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%LDBLD	%$STOCK	DIST4	LNTOTAR	LNTOTCAP	LNTOTSAL	EXIT	RESID2	OFFWORK2	
A Maximum Likelihood Estimates Based on the Joint Probability Function *														
EXIT	2.09	.00961	.000385	-.00132	.00823	-.00588	-.000971	-.183	-.383	-.0223		-.0848	-.0373	-.0224
	(3.34)	(2.29)	(3.44)	(-.89)	(2.14)	(-1.03)	(-.61)	(-2.58)	(-3.66)	(-.48)		(-1.48)	(-.64)	(-.44)
RESID2	-2.96	.00581	-.000118	.00315	.00422	.0579	.00209	-.233	.508	.0853	.0848		-.184	-.0224
	(-4.39)	(1.31)	(-.81)	(1.85)	(1.07)	(6.99)	(1.20)	(-2.31)	(4.32)	(1.59)	(-1.48)		(-3.11)	(-.44)
OFFWORK2	3.06	-.0220	-.000294	.00160	.00489	-.00432	-.00263	-.220	-.158	-.0652	-.0373	-.184		-.0224
	(4.19)	(-4.97)	(-1.30)	(.87)	(1.20)	(-.68)	(-1.50)	(-2.81)	(-1.45)	(-1.31)	(-.64)	(-3.11)		(-.44)
B Maximum Likelihood Estimates Based on the Conditional Probability Function														
EXIT	2.17	.00976	.000385	-.00132	.00811	-.00596	-.000919	-.192	-.374	-.0218		-.0504	.0178	-.0957
	(3.41)	(2.31)	(3.45)	(-.89)	(2.11)	(-1.05)	(-.57)	(-2.67)	(-3.57)	(-.47)		(-.83)	(.21)	(-.88)
RESID2	-2.86	.00535	-.000138	.00332	.00401	.0572	.00221	-.230	.532	.0844	-.0408		-.158	-.0562
	(-4.14)	(1.19)	(-.95)	(1.96)	(1.01)	(6.93)	(1.27)	(-2.29)	(4.48)	(1.58)	(-.66)		(-2.54)	(-.49)
OFFWORK2	3.21	-.0214	-.000301	.00154	.00523	-.00461	-.00248	-.224	-.157	-.0616	-.0354	-.159		-.0373
	(4.36)	(-4.87)	(-1.32)	(.84)	(1.28)	(-.73)	(-1.40)	(-2.82)	(-1.44)	(-1.24)	(-.40)	(-2.55)		(-.33)

Figures in parentheses are asymptotic t-ratios

* Logarithm of likelihood function: -1208.91

Table 16. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate and Trivariate Effects, Comparison with Conditional Estimates, Census Division 7

Dependent Variable	Main Effects										
	CONSTANT	AVAGE	TOTAREA	LDOWNED	TOTRENT	%LDOWNED	%LDBLD	%$STOCK	LNTOTAR	LNTOTCAP	LNTOTSAL
A Maximum Likelihood Estimates Based on the Join Probability Function*											
EXIT	2.26 (3.64)	.0100 (2.35)	.0000766 (2.98)	.000334 (2.99)	-.000740 (-1.47)	-.00195 (1.17)	.00744 (1.92)	-.00927 (-1.60)	-.230 (-3.15)	-.369 (-3.50)	-.0113 (-2.43)
RESID2	-2.36 (3.33)	.00547 (1.23)	.000333 (1.61)	-000390 (-1.56)	-.000123 (-.30)	.00515 (2.28)	.00468 (1.19)	.0584 (6.94)	-.329 (-2.73)	.471 (3.98)	.107 (1.96)
OFFWORK2	3.11 (419)	-.0215 (-4.83)	.0000614 (1.99)	-.000291 (-1.31)	-.00278 (-.57)	.00122 (.63)	.00392 (.96)	-.00701 (-1.08)	-.267 (-3.22)	-.153 (-1.37)	-.0563 (-1.13)
B Maximum Likelihood Estimates Based on the Conditional Probability Function											
EXIT	2.35 (3.72)	-.0102 (2.40)	.0000777 (3.02)	.000332 (2.99)	-.000740 (-1.47)	-.00197 (-1.17)	.00735 (1.90)	-.00938 (-1.63)	-.238 (-3.20)	-.361 (3.41)	-.0118 (-.25)
RESID2	-2.22 (-3.02)	.00498 (1.10)	.000320 (1.53)	-.000392 (-1.56)	-.000106 (-.256)	.00522 (2.32)	.00448 (1.14)	.0577 (6.87)	-.330 (-2.68)	.499 (4.16)	.105 (1.93)
OFFWORK2	3.26 (4.39)	-.0210 (-4.76)	.0000624 (2.05)	-.000294 (-1.32)	-.000240 (-.51)	-.00117 (.50)	.00435 (1.06)	-.00705 (-1.10)	-.270 (-3.19)	-.153 (-1.37)	-.0526 (-1.06)

	Bivariate Interaction Effects			Trivariate Interaction Effect
	EXIT	RESID2	OFFWORK2	
A Maximum Likelihood Estimates Based on the Joint Probability Function*				
EXIT		-.0968 (-1.68)	-.0484 (-.82)	-.0339 (-.65)
RESID2	-.0968 (-.168)		-.199 (-3.36)	-.339 (-.65)
OFFWORK2	-.0484 (-.82)	-.199 (-3.36		-.0339 (-.65)
B Maximum Likelihood Estimates Based on the Conditional Probability Function				
EXIT		-.0488 -.79	.0166 (.19)	-.111 (-1.01)
RESID2	-.0476 (-.77)		-.162 (-2.61)	(-.50)
OFFWORK2	-.0280 (-.32)	-.160 (-2.57)		.-0623 (-.55)

NOTE: Figures in parentheses are asymptotic t-ratios.
*Logarithm of liklihood function: -1203.30.

Tables 17, 18 and 19, respectively, display estimates of (1) a saturated model, (2) a model with no trivariate interaction effect but all bivariate interaction effects, and (3) a full independence model with no interaction effect; these three models include the same set of exogenous variables.

Tests based on the likelihood ratio can be performed to decide whether to include the trivariate and bivariate terms. These tests must be based on the full maximum likelihood estimates. The test for the trivariate interaction effect is based on:

$$-2 \log \lambda = -2 [-1208.17 + 1207.92] = .50.$$

This statistic is asymptotically distributed as χ^2 with one degree of freedom. Thus the trivariate interaction term is not to be considered different from zero at a level of significance of .15.[8] The statistic for the test for bivariate interaction terms is $-2 \log \lambda = 12.40$ asymptotically distributed as χ^2 with three degrees of freedom; thus, the bivariate interaction terms can be considered to be globally different from zero at a .15 level of significance.

Tests for individual bivariate interaction effects, based on the models displayed in Tables 20, 21, and 22, lead to the following conclusions at a .15 level of significance:

1. The interaction effect between EXIT and OFFWORK2 is not significantly different from zero ($-2 \log \lambda = .89$).

2. The interaction effect between EXIT and RESID2 is significantly different from zero ($-2 \log \lambda = 2.44$).

---

[8]Choice of a high level of significance is justified at this stage by the desire to avoid type II errors, i.e., to avoid deleting interaction terms which are actually different from zero.

Table 17. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate and Trivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

DEPENDENT VARIABLE	MAIN EFFECTS CONSTANT	AVAGE	TOTAREA	LDOWNED	%LDOWNED	% LDBVLD	%$ STOCK	LNTOTAR	LNTOTCAP	BIVARIATE INTERACTION EFFECTS EXIT	RESID2	OFFWORK2	TRIVARIATE INTER-ACTION EFFECT
A Maximum likelihood estimates based on the joint probability function*													
EXIT	2.27 (3.64)	.00976 (2.32)	.0000797 (3.11)	.000335 (3.03)	-.000819 (-.55)	.00817 (2.15)	-.00792 (-1.38)	-.231 (-3.15	-.410 (-4.42)		-.0987 (-1.72)	-.0459 (-.78)	-.0365 (-.71)
RESID2	-2.44 (-3.51)	.00477 (1.08)	.000227 (1.22)	-.000297 (-1.27)	.00496 (2.28)	.00327 (.85)	.0574 (6.97)	-.268 (-2.42)	.558 (5.16)	-.0987 (-1.72)		-.204 (-3.45)	-.0365 (-.71)
OFFWORK2	3.07 (4.18)	-.0212 (-4.81)	.0000656 (2.17)	-.000297 (-1.35)	.00159 (.87)	.00494 (1.22)	-.00637 (-.99)	-.271 (-3.26)	-.222 (-2.25)	-.0459 (-.78)	-.204 (-3.45)		-.0365 (-.71)
B Maximum likelihood estimates based on the conditional probability functions.													
EXIT	2.34 (3.72)	.00993 (2.36)	.0000805 (3.14)	.000333 (3.02)	-.000826 (-.55)	.00807 (2.13)	-.00803 (-1.41)	-.237 (-3.19)	-.402 (-4.34)		-.0499 (-.82)	.0202 (.24)	-.112 (-1.02)
RESID2	-2.28 (-3.17)	.00430 (.97)	.000216 (1.15)	-.000301 (-1.29)	.00503 (2.32)	.00313 (.81)	.0568 (6.91	-.272 (-2.38)	.586 (5.36)	-.0565 (-.92)		-.173 (-2.81)	-.0510 (-.45)
OFFWORK2	3.22 (4.38)	-.0207 (-4.74)	.0000664 (2.22)	-.000302 (-1.37)	.00152 (.83)	.00526 (1.30)	-.00645 (-1.01)	-.273 (-3.23)	-.217 (-2.19)	-.0163 (-.19)	-.161 (-2.58)		-.768 (-.68)

Figures in parentheses are asymptotic t-ratios.

* Logorithm of likelihood function: -1207.92

Table 18. Maximum Lieklihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

DEPENDENT	MAIN EFFECTS									BIVARIATE INTERACTION EFFECTS		
VARIABLE	CONSTANT	AVAGE	TOTAREA	LDOWNED	%LDOWNED	%LDBVLD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
A Maximum likelihood estimates based on the joint probability function*												
EXIT	2.27	.00974	.0000791	.000337	-.000848	.00822	-.00803	-.226	-.415		-.0817	-.0528
	(3.65)	(2.32)	(3.09)	(3.05)	(-.57)	(2.17)	(-1.40)	(-3.11)	(-4.50)		(-1.58)	(-.92)
RESID2	-2.49	.00478	.000215	-.000294	.00494	.00325	.0574	-.256	.557	-.0817		-.186
	(-3.62)	(1.08)	(1.17)	(-1.27)	(2.28)	(.85)	(6.97)	(-2.34)	(5.16)	(-1.58)		(-3.48)
OFFWORK2	3.07	-.0212	.0000650	-.000296	.00158	.00497	-.00645	-.267	-.227	-.0528	-.186	
	(4.19)	(-4.80)	(2.15)	(-1.35)	(.87)	(1.23)	(-1.00)	(-3.23)	(2.30)	(-.92)	(-3.48)	
B Maximum likelihood estimates based on the conditional probability functions												
EXIT	2.37	.00980	.0000787	.000334	-.000821	.00815	-.00785	-.225	-.410		-.0813	-.0454
	(3.77)	(2.33)	(3.08)	(3.04)	(-.55)	(2.16)	(-1.37)	(-3.09)	(-4.47)		(-1.56)	(-.79)
RESID2	-2.32	.00444	.000209	-.000297	.00503	.00317	.0568	-.264	.583	-.0704		-.187
	(-3.25)	(1.00)	(1.12)	(-1.28)	(2.32)	(.82)	(6.92)	(-2.36)	(5.36)	(-1.33)		(-3.48)
OFFWORK2	3.24	-.0207	.0000653	-.000302	.00152	.00537	-.00644	-.269	-.220	-.0613	-.182	
	(4.40)	(-4.74)	(2.18)	(-1.37)	(.83)	(1.32)	(-1.00)	(-3.21)	(-2.23)	(-1.05)	(-3.37)	

Figures in parentheses are asymptotic t-ratios.

* Logorithm of likelihood function: -1208.17

Table 19. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables, Comparison with Conditional Estimates, Census Division 7

DEPENDENT VARIABLE	MAIN EFFECTS CONSTANT	AVAGE	TOTAREA	LDOWNED	%LDOWNED	%$LDBVLD	%$STOCK	LNTOTAR	LNTOTCAP
A Maximum likelihood estimates based on the joint probability function*									
EXIT	2.335	.0101	.0000770	.000344	-.00106	.00785	-.0100	-.210	-.436
	(3.95)	(2.45)	(3.03)	(3.12)	(-.719)	(2.08)	(-1.81)	(-2.94)	(-4.89)
RESID2	-3.103	.00699	.000185	-000302	.00497	.00204	.0586	-.193	.614
	(-4.65)	(1.64)	(1.01)	(-1.30)	(2.31)	(.54)	(7.16)	(-1.81)	(5.80)
OFFWORK2	3.46	-.0220	.0000610	-.000265	.000914	.00420	-.0120	-.251	-.287
	(4.83)	(-5.05)	(2.02)	(-1.22)	(.51)	(1.06)	(-1.93)	(-3.04)	(-3.05)
B Maximum likelihood estimates based on the conditional probability functions									
EXIT									
RESID2				- I D E M -					
OFFWORK2									

Figures in parentheses are asymptotic t-ratios

* Logarithm of likelihood function: -1215.40

Table 20. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

DEPENDENT VARIABLE	MAIN EFFECTS									Bivariate Interaction Effects		
	CONSTANT	AVAGE	TOTAREA	LDOWNED	%LDOWNED	%LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
A Maximum Likelihood Estimates Based on the Joint Probability Function*												
EXIT	2.14 (3.54)	.0105 (2.53)	.0000762 (3.00)	.000333 (3.02)	-.000815 (-.55)	.00802 (2.12)	-.00779 (-1.36)	-.214 (-3.00)	-.405 (-4.43)		-.0753 (-1.47)	
RESID2	-2.51 (-3.65)	.00477 (1.08)	.000215 (1.16)	-.000296 (-1.27)	.00495 (2.28)	.00320 (.84)	.0575 (6.98)	-.255 (-2.33)	.560 (5.19)	-.753 (-1.47)		-.183 (-3.44)
OFFWORK2	2.98 (4.11)	-.0215 (-4.91)	.0000618 (2.06)	-.000309 (-1.41)	.00163 (.90)	.00465 (1.16)	-.00616 (-9.60)	-.257 (-3.14)	-.211 (-2.18)		-1.83 (-3.44)	
B Maximum Likelihood Estimates Based on the Conditional Probability Functions												
EXIT	2.20 (3.71)	.0105 (2.54)	.0000762 (3.01)	.000332 (3.02)	-.000785 (-.53)	.00803 (2.12)	-.00770 (-1.35)	-.213 (-2.99)	-.404 (-4.40)		-.0756 (-1.47)	
RESID2	-2.32 (-3.25)	.00444 (1.00)	.000209 (1.12)	-.000297 (-1.28)	.00502 (2.32)	.00317 (.823)	.0568 (6.92)	-.264 (-2.36)	.583 (5.36)	-.0704 (-1.33)		-.187 (-3.48)
OFFWORK2	3.07 (4.30)	-.0210 (-4.82)	.0000610 (2.04)	-.000316 (-1.44)	.00160 (.88)	.00498 (1.24)	-.00614 (-.96)	-.256 (-3.12)	-.205 (-2.11)		-.177 (-3.30)	

Figures in parentheses are asymptotic t-ratios

*Logarithm of likelihood function: -1208.60

Table 21. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

Dependent Variable	Main Effects									Bivariate Interaction Effects		
	Constant	Age	TOTAREA	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
A Maximum Likelihood Estimates Based on the Joint Probability Function *												
EXIT	2.25 (3.61)	.00987 (2.35)	.0000785 (2.07)	.000337 (3.05)	-.000855 (-.57)	.00817 (2.16)	-.00811 (-1.41)	-.223 (-2.07)	-.414 (-4.48)		-.0754 (1.47)	-.0411 (-.72)
RESID2	-2.98 (-4.43)	.00755 (1.76)	.000194 (1.06)	-.000288 (-1.24)	.00494 (2.29)	.00249 (.66)	.0582 (7.10)	-.207 (-1.93)	.591 (5.53)	-.0754 (-1.47)		
OFFWORK2	3.53 (4.87)	-.0217 (-4.97)	.0000634 (2.08)	-.000254 (-1.16)	.000868 (.48)	.00444 (1.11)	-.0123 (-1.97)	-.258 (-3.10)	-.300 (-3.13)	-.0411 (-.72)		
B Maximum Likelihood Estimates Based on the Conditional Probability Functions												
EXIT	2.37 (3.76)	.00980 (2.33)	.0000786 (3.08)	.000334 (3.04)	-.000821 (-.55)	.00815 (2.16)	-.00785 (-1.37)	-.225 (-3.08)	-.410 (-4.47)		-.0813 (-1.56)	-.0454 (-.79)
RESID2	-2.90 (-4.22)	.00747 (1.74)	.000174 (1.05)	-.000284 (-1.22)	.00487 (2.25)	.00244 (.65)	.0577 (7.07)	-.212 (-1.95)	.597 (5.59)	-.0678 (-1.30)		
OFFWORK2	3.59 (4.86)	-.0218 (-4.98)	.0000642 (2.11)	-.000253 (-1.16)	.000837 (.46)	.00445 (1.11)	-.0123 (-1.98)	-.260 (-3.10)	-.300 (-3.13)	-.0448 (-.78)		

Figures in parentheses are asymptatic t-ratios

* Logarithm of likelihood function: -1214.08

Table 22. Maximum Likelihood Estimates of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Comparison with Conditional Estimates, Census Division 7

Dependent Variable	Main Effects									Bivariate Interaction Effects		
	Constant	Age	TOTAREA	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
A Maximum Likelihood Estimates Based on the Joint Probability Function *												
EXIT	2.45 (4.00)	.00951 (2.27)	.0000793 (3.09)	.000347 (3.15)	-.00111 (-.75)	.00800 (2.12)	-.0103 (-1.86)	-.220 (-3.02)	-.446 (-4.94)			-.411 (-.72)
RESID2	-2.63 (-3.86)	.00421 (.96)	.000205 (1.11)	-.000309 (-1.33)	.00497 (2.30)	.00274 (.72)	.0579 (7.04)	-.240 (-2.22)	.582 (5.45)			-.183 (-3.44)
OFFWORK2	3.06 (4.17)	-.0212 (-4.82)	.0000643 (2.13)	-.000298 (-1.36)	.00159 (.87)	.00489 (1.22)	-.00645 (-1.00)	-.265 (-3.20)	-.224 (-2.28)	-.0411 (-.722)	-.183 (-3.44)	
B Maximum Likelihood Estimates Based on the Conditional Probability Functions												
EXIT	2.46 (3.91)	.00958 (2.28)	.0000789 (3.08)	.000347 (3.15)	-.00111 (-.75)	.00793 (2.10)	-.0103 (-1.85)	-.219 (-3.00)	-.443 (-4.93)			-.0337 (-.59)
RESID2	-2.53 (-3.65)	.00393 (.89)	.000198 (1.07)	-.000315 (-1.36)	.00512 (2.37)	.00275 (.72)	.0577 (7.02)	-.244 (-2.22)	.601 (5.57)			-.186 (-3.47)
OFFWORK2	3.24 (4.40)	-.0207 (-4.74)	.0000653 (2.18)	-.000302 (-1.37)	.00152 (.83)	.00536 (1.32)	-.00644 (-1.01)	-.269 (3.21)	-.220 (-2.23)	-.0613 (-1.05)	-.182 (-3.38)	

Figures in parentheses are asymptatic t-ratios

* Logarithm of likelihood function: -1209.39

3. The interaction effect between RESID2 and OFFWORK2 is significantly different from zero (-2 log λ = 11.82).

Consequently, the bivariate interaction term between EXIT and OFFWORK2 is not retained in the model.

The third step in the search for an adequate model consists of deleting some of the exogenous variables for <u>some</u> of the equations. This is done on the basis of the asymptotic t-ratios corresponding to the coefficients of the exogenous variables. One further problem arises in relation to the presence of both TOTAREA and LNTOTAR in the model; coefficients of both variables are significnatly different from zero but possess opposite signs. On the basis of their asymptotic t-ratios, it was decided to use LNTOTAR only.

Full maximum likelihood estimates of the final model for census division 7 are displayed in Table 23.

Testing the Model

The model was estimated on a number of other census divisions, namely, census divisions 1, 11, 15 and 17; estimates are displayed in Tables 24, 25, 26, and 27. Tests for the coefficients are based on asymptotic t-ratios, with a large number of degrees of freedom and a .05 level of significance.

First, the equation for EXIT is examined. The constant term and the coefficient of LDOWNED are significantly different from zero in none of the census divisions. The coefficient of AVAGE is significantly different from zero in all census divisions. The coefficients of %$LDBLD and RESID2 are significantly different from zero in three census divisions and coefficients of LNTOTAR and LNTOTCAP

Table 23. Maximum Likelihood Estimates* of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Census Division 7

Dependent Variable	Main Effects								Bivariate Interaction Effects		
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
EXIT	1.45 (3.20)	.00957 (2.42)	.000325 (3.33)		.0102 (3.27)		-.155 (-2.35)	-.378 (-4.25)		-.0860 (-1.76)	
RESID2	-2.04 (-4.81)			.00308 (2.49)		.0556 (7.49)	-.228 (-2.55)	.551 (5.36)	-.086 (-1.74)		-.222 (-4.53)
OFFWORK2	3.38 (7.99)	.0226 (-5.50)					-.261 (-3.68)	-.211 (-2.40)		-.222 (-4.53)	

Figures in parentheses are asymptotic t-ratios

* Maximum likelihood estimates are based on the joint probability function. Logarithm of likelihood function: -1220.73

Table 24. Maximum Likelihood Estimates* of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Census Division 1

Dependent Variable	Main Effects								Bivariate Interaction Effects		
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
EXIT	-.0802 (-.15)	.0264 (6.47)	-.0000223 (-.15)		.00287 (.99)		-.225 (-2.40)	-.0716 (-.73)		-.114 (-2.28)	
RESID2	-2.38 (-5.70)			.00293 (2.28)		.0333 (6.91)	-.228 (-2.41)	.636 (6.01)	-.114 (-2.28)		-.271 (-5.58)
OFFWORK2	2.37 (5.65)	-.0197 (-5.06)					-.288 (-3.17)	-.0404 (-.41)		-.271 (-5.58)	

Figures in parentheses are asymptotic t-ratios

* Maximum likelihood estimates are based on the joint probability function. Logarithm of likelihood function: -1272.25

Table 25. Maximum Likelihood Estimates* of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Census Division 11

Dependent Variable	Main Effects								Bivariate Interaction Effects		
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
EXIT	.240 (.44)	.0188 (4.77)	-.000102 (-.76)		.00876 (2.80)		-.200 (-2.67)	-.145 (-1.65)		-.0978 (-2.17)	
RESID2	-2.60 (-6.04)			.00344 (2.91)		.0490 (6.39)	.0264 (.34)	.376 (4.12)	-.0978 (-2.17)		-.324 (-7.43)
OFFWORK2	2.73 (6.30)	-.0252 (-6.69)					-.131 (-1.82)	-.200 (-2.28)		-.324 (-7.43)	

Figures in parentheses are asymptotic t-ratios

* Maximum likelihood estimates are based on the joint probability function. Logarithm of likelihood function:-1370.09

Table 26. Maximum Likelihood Estimates* of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Census Division 15

Dependent Variable	Main Effects								Bivariate Interaction Effects		
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
EXIT	.829 (1.84)	.0153 (3.90)	.000148 (.75)		.00696 (2.36)		-.119 (-1.87)	-.304 (-3.95)		-.183 (-3.77)	
RESID2	-1.54 (-3.65)			.00238 (1.57)		.0835 (8.24)	-.227 (-2.67)	.473 (5.26)	-.183 (-3.77)		-.394 (-8.24)
OFFWORK2	2.46 (6.71)	-.0206 (-5.23)					-.191 (-3.19)	-.131 (-1.79)		-.394 (-8.24)	

Figures in parentheses are asymptotic t-ratios

* Maximum likelihood estimates are based on the joint probability function. Logarithm of likelihood function: -1235.79

Table 27. Maximum Likelihood Estimates* of Main Effects Dependent on Exogenous Variables and Constant Bivariate Interaction Effects, Census Division 17

Dependent Variable	Main Effects								Bivariate Interaction Effects		
	CONSTANT	AVAGE	LDOWNED	%LDOWNED	%$LDBULD	%$STOCK	LNTOTAR	LNTOTCAP	EXIT	RESID2	OFFWORK2
EXIT	.0857 (.18)	.0183 (4.83)	.000103 (.74)		.00890 (3.45)		-.0690 (-.92)	-.272 (-3.76)		-.0928 (-1.64)	
RESID2	-.677 (-1.43)			.00160 (.97)		.0333 (6.01)	-.184 (-1.84)	.362 (4.14)	-.0928 (-1.64)		-.286 (-5.26)
OFFWORK2	2.09 (5.42)	-.0207 (-5.40)					-.113 (-1.62)	-.162 (-2.26)		-.286 (-5.26)	

Figures in parentheses are asymptotic t-ratios

* Maximum likelihood estimates are based on the joint probability function. Logarithm of likelihood function: -1175.04

in two census divisions. EXIT is positively related to AVAGE, a result which contradicts Hypothesis 1 of Chapter IV. EXIT appears unrelated to LDOWNED, a result which contradicts Hypothesis 3. It appears that EXIT is positively related to %LDBLD; such a relationship was not considered in the hypothesis. The hypothesized negative relationship between EXIT and LNTOTAR and LNTOTCAP is still uncertain. EXIT is consistently negatively related to RESID2; this result supports Hypothesis 8 according to which nonresident farm operators are more likely to leave farming before retirement age.

As far as the RESID2 equation is concerned, the main results consist of the strong positive relationship with %$STOCK and the strong negative relationship with OFFWORK2. This can be interpreted as it follows: residence on the farm is positively related to the importance of the livestock enterprise on the farm and is negatively related to the farm operator's involvement in off-farm work.

As far as the OFFWORK2 equation is concerned, the major result (apart from the already mentioned interaction with RESID2) consists of the existence of a negative relationship with AVAGE.

In summary, the statistical analysis of the data using the log-linear model leads to the rejection of all hypotheses stated in Chapter IV except: (1) parts of Hypothesis 4 concerning the relationship between pre-retirement exit and total capital value and total acreage, and (2) Hypothesis 8, concerning the positive relationship between pre-retirement exit from farming and nonresidence on the farm.

Conclusion

The foregoing empirical analysis is disappointing in many respects. Most of the hypotheses presented in Chapter IV are not supported by the statistical analysis of the data.

Although it was recognized at the outset that some of the major variables bearing on the decision to leave farming were not available in the data-base, some relationships between pre-retirement exit and available variables were expected to be identified.

Two main conclusions should be drawn from the foregoing empirical analysis:

1. The decision to leave farming before retirement is not clearly related to a limited set of farm factors; pre-retirement exit is a complex phenomenon which requires detailed analysis based on specially collected information.

2. The failure to identify clear relationships between pre-retirement exit from farming and the considered exogenous variables raises some doubts about the adequacy of the Census of Agriculture Match as a data-base for a study of exiters (or entrants).

From a statistical point of view two important methodological issues were encountered. First, tables of contingencies analyzed in the exploratory phase of the data analysis yielded some highly significant value for the χ^2 statistic, thereby leading to the conclusion that relationshps between variables certainly exist; but in attempting to measure those associations between variables, it was discovered that they were very weak. In summary, the following were exemplified:

1. The χ^2 statistic is a poor <u>measure</u> of association and an

experimenter should rely on other measures such as those which were used in this study;

2. Very significant results concerning the existence of a relationship are consistent with very weak relationships (in the sense that knowledge of one variable is of little help in predicting the value of the other).

Second, in the confirmatory phase of the analysis, the following approach was taken: a model was chosen and estimated using part of the data-set and, then, tested on other parts of this data-set. It followed that several coefficients which apparently were significantly different from zero, on the basis of asymptotic t-ratios obtained using the first part of the data, were shown to be not significantly different from zero, when the model was tested on other parts of the data-set. Thus, it was clearly exemplified that one is likely to arrive at misleading results if a model is chosen, estimated, and tested on the same data. The method used in this study is the proper one, provided enough observations are available.

The general dissatisfaction with the rather sophisticated statistical analysis of available secondary data, prompted a proposal for further research based on "ad hoc" information, which is presented in the following chapter.

CHAPTER VII

PROPOSAL FOR FURTHER EMPIRICAL RESEARCH: A LONGITUDINAL SURVEY

The objective sought in this chapter is twofold: first, to present a critical assessment of the empirical analysis and, second, to propose an additional empirical study to evaluate further the theoretical framework proposed in Chapter III.

Thus, this chapter is divided into two sections. The first section reviews the main propositions of the theoretical framework and discusses the limitations of available secondary data with respect to testing these propostions. In the second section, a proposed longitudinal survey is outlined.

Assessment of the Analysis of Secondary Data

The empirical analysis whose results were presented in Chapter VI will be assessed on the basis of its ability to substantiate the theoretical framework proposed in Chapter III. As a preliminary to this assessment, the main elements of the theoretical framework are inventoried in the following subsection.

Highlights of the Conceptual Framework

The general thrust of the approach in Chapter III is to substitute an empirically relevant model of individual choice for an axiomatic one. In order to develop such a model, theoretical concepts and empirical results have been drawn from several disciplines.

Income maximizing and utility maximizing variants of the axiomatic economic model of choice were described in relation to the decision to leave farming. The behavioral assumptions were briefly and critically reviewed. On the basis of findings in psychology, a general model of behavior and choice was proposed, to be used as a frame on which other elements could be added to fit the specificity of the decision to leave farming. The main features of this general model of adaptive behavior are: (1) human behavior depends on changes in the environment as well as on the person itself, (2) individuals act as parts of larger groups, (3) individual behavior is essentially of a satisficing nature, i.e., individuals try to attain some predetermined goals called wants and aspirations, (4) wants and aspirations are not static, and (5) habitual behavior prevails in most circumstances.

Farming can be considered as a form of habitual behavior whereas exit from farming is the result of a genuine decision. Any habitual behavior takes place with the individual having a certain organized perception of reality which is consistent with the habitual behavior. As long as this perception is unchanged the same habitual behavior will prevail; for a genuine decision to take place, the organized perception needs to change and become inconsistent with the habitual behavior. Very little is known empirically about the transition from habitual behavior to genuine decision and the accompanying restructuration of the perceived environment. Empirical study of this transition and restructuration in the special case of a farmer considering leaving farming would shed some light on the circumstances and factors affecting the decision to leave farming.

Farmers' behavior is to be considered as satisficing: farmers strive to attain goals and aspirations whose level is itself the result of past performances as well as performances of people with whom they associate. Thus, as long as a farmer satisfies his wants and aspirations, there is no reason for him to leave farming.

The decision to leave farming often implies drastic changes in all family members' lives. Very likely, these consequences may be beneficial for some and detrimental to others, according to each individual's goals. Conflicts in goals may arise and the resolution of these conflicts will bear on the action taken. Consequently, the decision to leave farming is to be looked at as a collective decision.

The level of aspiration depends on past performances among which is income remunerating the operator's labor. This implies that not only present income, but also the history (trend and variations) of income bears on the decision to leave farming. Past incomes are factors influencing the decision through the mediation of the level of aspiration.

Any genuine decision, and consequently the decision to leave farming, is based on a perceived economic and social environment. Perception of this economic and social environment plays, therefore, a decisive role in whether the decision to leave farming is made. Perception is both a filtering and a structuring process; in other words, some elements of the environment remain unknown to the farmer (this is limited knowledge) and the elements which are known are organized into a whole which acquires a certain meaning. The extent to which farmers ignore relevant elements is ill-documented. The process by which known elements, especially nonfarm job opportunities and urban conditions of life, are organized appears to be completely unknown.

Imperfect information, with respect to nonfarm occupation alternatives, once it is recognized by the farmer, leads to a process of collection of information usually called job-search. According to the proposed theoretical framework, job-search does not, however, proceed unless the farmer is dissatisfied with his present situation and is considering other alternatives. Job-search and the accompanying information gathering are part of the reorganization of perceived environment. Factors prompting such a job-search, as well as its duration, intensity and other characteristics, are ill-known.[1]

Imperfect information, even after a job-search, may explain the occurrence of decision-making errors and their possible corollary, return migration. The relationship between the amount of knowledge of working conditions in nonfarm occupations and of urban living conditions, and return migration, remains to be investigated.

People, in deciding on an action, take into account benefits and costs occurring only over a limited time span, which is called the time perspective. The action taken is independent of the consequences occurring outside the individual's time perspective. Consequently, the length of this time perspective, because it determines what consequences are or are not accounted for, bears directly on the

---

[1]The so-called job-search literature deduces characteristics of the job-search on the basis of an assumed maximizing behavior and is, therefore, essentially normative. What is proposed here is a positive study of the search process.

individual decision. If the concept of time perspective has any real basis, an inquiry into farmers' time perspective and the factors affecting it is of interest.

The decision to leave farming is affected by the farmer's and the family's attachments to their community as well as to farming as an occupation.

Shortcomings of Available Secondary Data

The data-set which was used in this study and the procedure by which it was obtained were described in Chapter IV. The list of variables which were available for each observation can be found in Appendix A.

A major advantage of the Census of Agriculture Match is that it is exhaustive. Potentially, all stayers, exiters, and entrants between 1966 and 1971 can be identified for all provinces of Canada. Thus, the Census of Agriculture Match allows for good estimates of all gross flows of farmers in and out of agriculture, for each province or for any desired region within a province. In this study only data concerning Saskatchewan were used.

At the time this study is being completed, a 1971-1976 Census of Agriculture Match is being performed, based on an improved methodology in the light of the experience acquired with the 1966-1971 Match. Thus, the same type of analysis could be conducted in the near future for the 1971-1976 period. Furthermore, the 1966, 1971 and 1976 censuses can possibly be linked together to constitute a three-census Match.

The large size of the Census of Agriculture Match renders it quite unwieldy. When multivariate statistical analysis need be performed, samples must be taken.

The main shortcomings of the Census of Agriculture Match in relation to a study of off-farm movement and migration is the limited scope of the information collected in the census. The great majority of the information collected in the census is related to the farm business and, especially, to the physical aspects of agricultural production. Most of this information is a priori irrelevant to the decision to leave farming and, therefore, was left out of the simplified data-base used in this study. As mentioned before, only information describing the overall structure of the farm was retained.

Information concerning the farmer and the family is limited to the age class of the operator, involvement in off-farm work and whether or not the operator resides on the farm. No information concerning net income either from farm or nonfarm sources is available.

Only one piece of information, which was thought to be theoretically relevant, could be added to the data-set: distances from the farm to the nearest communities of four different sizes.

A comparison of the information available from the Census of Agriculture Match to the information which would be necessary to substantiate the previous summarized conceptual framework, clearly shows the inadequacy of the Census of Agriculture Match. For example, the theoretical framework calls for information on present income from farm and nonfarm sources, potential income from nonfarm employment sources for the operator and the family, operator's and spouse's skills and education, family members' goals and ways these goals are acknowledged as valid, attachment to the community and farming as an occupation, knowledge of nonfarm employment opportunities, perception of urban living conditions, etc. The Census of Agriculture provides no such elements of information. In the future more socio-economic information on the operator and the family as well as information on net income are likely to be collected; the census will, however, fall short of the requirements for a detailed study of factors affecting pre-retirement exit from farming.

If such a study is to be performed, ad hoc information has to be collected through a specially designed survey. A brief proposal for such a survey is described in the following section.

A Longitudinal Survey

In this section, the overall objectives of the survey are stated, then the desirable features of a survey designed to reach these objectives are listed, and, finally, an outline of the informational content of such a survey is presented.

Objectives of the Survey

The overall objective of the survey is to substantiate the following elements of the theoretical framework presented in Chapter III:

1. Existence of a satisficing mode of behavior
2. Relevance of community satisfaction and satisfaction with farming as an occupation
3. Discrepancy between perceived and actual socio-economic environment
4. Importance, intensity and duration of the search process

5. Dynamics of the decision process

6. Relevance of operator's and spouse's formal education and professional skills

7. Relevance of the group nature of the decision

8. Influence of the length of the operator's and spouse's time perspecitve and relationship between socio-economic variables and the length of the time perspective

9. Interaction between off-farm movement and involvement in off-farm work.

Desirable Features of the Survey

The pre-retirement exit from farming is a dynamic phenomenon; in classical economic parlance, it is an adjustment to a temporary disequilibrium. In the light of the theoretical framework proposed in this study, the decision to leave farming is a lengthy process, involving the transition from a habitual mode of behavior to an action based on genuine decision; this transition is accompanied by a reorganization of the perceived environment. Given the highly dynamic nature of the decision making process, a longitudinal survey, where information is collected for at least two different points in time, seems to be recommended.

The survey should be as much as possible composed of direct questions calling for simple answers. Given the type of information sought, some nondirect questions should, however, be used. The care with which questionnaires are to be completed together with the difficulty in recruiting and training qualified interviewers call for a survey of limited size. The size of the survey, however, must be sufficient to allow reliable statistical analysis. The major dependent variable is the decision to leave farming, which is a qualitative variable. Other considered variables are qualitative, e.g., community satisfaction, satisfaction with farming, off-farm work, residence on the farm, nonfarm job-search, and migration. Hence, the statistical methods to be used will be those reviewed in Chapter V. The multivariate log-linear model with exogenous variables used in this study would, most likely contribute a major part to the data analysis. Experience with this model shows that, when several jointly dependent variables are used in conjunction with many exogenous variables, the number of observations should not be smaller than 300 to obtain stable estimates. Thus, the survey should include around 400 farm operators.

Such a limited survey cannot pretend to cover a representive sample of a province; it is bound to be exploratory. To avoid, however, being too specific to a certain location, the farmers interviewed should be sampled out of municipalities belonging to several different census divisions.

The survey is to be longitudinal and to consist of two interviews at two or three year intervals. A short interval is necessary so that the same person can be traced to the new residence in case of moving between the first and second interview. A long enough interval is required to ensure the existence, among the sampled farmers, of a sufficient number of pre-retirement exiters, with and without migration. The sample of farmers should include only farmers of age below retirement age since the study of farmers' retirement decisions is not an objective of this survey. All farmers who moved between the two interviews should be traced to their new location whether or not they have left farming.

Informational Content

It is not intended here to draft questionnaires for the proposed longitudinal survey, but merely to outline the type of information to be collected and, sometimes, to indicate the format in which this information would be collected.

In the proposed survey, two interviews are planned implying that two different questionnaires will be prepared. The second interview, however, may cover different ground depending on whether the farmer has stayed in farming or has left farming.

The first interview will aim at collecting the following information:

1. A description of the farm-business (acreage, type of farm, total capital value, total sales, total debts, net farm income, time worked in farm activities, etc.).
2. The socio-economic characteristics of the operator and family (age, level of formal education, vocational training, technical or professional skills, number and ages of children, existence of a potential successor, off-farm work involvement, nonfarm incomes, etc.).
3. Measures of community satisfaction and satisfaction with farming as an occupation.[2]

---

[2]These measures would be based on so-called attitude-scaling methods. Several community satisfaction scales have been developed and could easily be adapted to a survey of Saskatchewan farmers. See, for example, Vernon Davies, "Development of a Scale to Rate Attitude of Community Satisfaction," Rural Sociology 10 (September 1945): 246-255; Clinton Jesser, "Community Satisfaction Patterns of Professionals in Rural Areas," Rural Sociology 32 (March 1963): 56-69; Ronald Johnson and Edward Knop, "Rural-Urban Differentials in Community Satisfaction," Rural Sociology 35 (December 1970): 544-548.

4. The operator's and spouse's expectations concerning nonfarm occupational alternatives (expected income, location, etc.).
5. The operator's forecast as to whether he will still be farming two or three years hence.

In the case of a stayer, the second interview will cover the following points:

1. A description of the farm-business.
2. The socio-economic characteristics of the operator's family that may have changed since the first interview.
3. Measures of community satisfaction and satisfaction with farming as an occupation.
4. A measure of job-search[3] performed in the last year and reason(s) why this search did not result in the farmer leaving farming.
5. The operator's forecast as to whether he will be farming in the near future.

In the case of pre-retirement exiters, the second interview will cover the following points:

1. A description of family members' new occupations (type of work, income, location, etc.).
2. The extent of migration in relation to off-farm movement.
3. Measures of community satisfaction and job satisfaction.[4]

---

[3]Such a measure should be a scale based on a series of questions related to the occurrence of behaviors corresponding to increasing degrees of active involvement in collection of information and job-search.

[4]To be measured by attitude scales as mentioned before.

4. A self-assessment of the change in family members' welfare since off-farm movement.
5. The operator's estimated likelihood of returning to farming.

Data Analysis

Some of the information collected should allow the refining of the conceptual framework presented in this study. No formalized data analysis would be performed on this information which may be expressed in a literary form.

Other information will lenditself to formalized data analysis of both an exploratory and confirmatory type, in accordance to the methodological position described in Chapter IV, and using statistical methods described in Chapter V.

Conclusion

In this chpater we have outlined the main features of a survey designed to make progress along the lines proposed in the theoretical framework. Much work remains in defining both the implementation procedure and the informational content of the survey. Given the multidisciplinary nature of the study, such a survey would benefit greatly from the cooperation of sociologist(s), geographer(s) and psychologist(s), especially in the design of the questionnaire.

CHAPTER VIII

SUMMARY, CONCLUSIONS, AND METHODOLOGICAL CONSIDERATIONS

This chapter is divided into two main sections. The first section summarizes the study and conclusions. In the second section some methodological problems, which arose in conducting this research, are presented and discussed.

Summary and Conclusions

At any point in time, the number of farm operators is equal to the number of farm operators one period earlier, minus the number of farm operators who left farming during that period, plus the number of new farm operators who decided to enter farming during that period; the number of farm operators is, thus, the aggregate and cumulative result of individual entry and exit decisions. Causes which affect these individual entry and exit decisions will, therefore, affect farm demography.

Low incomes in farming, both in relative and in absolute terms, have been noted consistently during recent decades, even when income from off-farm sources has been accounted for. Movement of both hired and self-employed labor out of the farm sector has been and still is considered as the only long term solution.

The concern for Canadian farm operators' welfare has led to the development and implementation of many government programs, both

provincial and federal, aimed at affecting off-farm movement and migration, income from farming, or both. These programs, which purport to bear on individual decisions, have not been designed, however, on the basis of a precise knowledge of the factors affecting these individual decisions.

This research was primarily founded on a belief that the understanding of the variations in the gross movements of farm operators into and out of agriculture is to be based on the understanding of the individual decisions to enter farming, to leave farming before retirement, or to retire from farming. Similarly, the design of appropriate programs aimed at modifying these flows, require a clear understanding of these individual decisions.

Three main objectives were set for this study:

1. To provide a conceptual framework for the analysis of the decision to leave farming before retirement.

2. To appraise the ability of the Canadian Census of Agriculture Match to identify entrants and exiters, and its usefulness as a major data source for analytical studies of entry and exit from farming.

3. To test as many hypotheses, concerning factors affecting off-farm movement, as the available data would permit.

In the process of attaining these objectives the following tasks were performed:

1. A review of the literature on off-farm movement and migration.

2. Development of a theoretical framework for the analysis of

the pre-retirement decision to leave farming, drawing on economics as well as psychology and sociology.

3. Evaluation, both before and after data analysis, of a new longitudinal data base: the Canadian Census of Agriculture Match.

4. Formulation of hypotheses concerning factors influencing the pre-retirement exit of farm operators.

5. A review of statistical methods for the analysis of qualitative dependent variables.

6. An exploratory and confirmatory data analysis of the decisions to leave farming before retirement.

7. A sketch of a proposed longitudinal survey which would further the understanding of pre-retirement exit from farming.

A summary of the aforementioned tasks follows as well as a presentation of conclusions.

A broad review of literature showed that:

1. Some confusion in concepts and words which have been used in the literature (e.g., mobility, movement, migration, off-farm migration, etc.) has been hampering the development of incisive empirical work.

2. The use of aggregate data which provide estimates of net flows in and out of agriculture is inadequate to identify factors of the individual's decision to leave farming.

3. Access to a longitudinal data base is, consequently of the most necessity to yield reliable results on the causes of the decision to leave farming.

4. Empirical studies yielded conflicting results about the

factors influencing the decision to leave farming, especially with respect to the effect of the farm operator's involvement in off-farm work and distance to urban employment centers.

The standard economic model of off-farm movement was briefly presented and its behavioral assumptions were critically reviewed. Standard economic theory postulates a certain type of behavior, namely a utility or profit maximizing behavior. The theoretical framework which was proposed is at variance with this theory in that an attempt was made to draw on positive studies of human behavior in general, and that detailed observation and description of the exit decision-making process was advocated. Briefly stated a behavioral approach was proposed.

The highlights of the proposed theoretical framework for the analysis of the decision to leave farming before retirement are:

1. Farming is considered as a form of habitual behavior and pre-retirement exit from farming is the outcome of a genuine decision.

2. Farm operators' behavior is looked upon as being of a satisficing nature.

3. The decision to leave farming is viewed as a group decision, where the group is the family.

4. Goals of the members of a family are likely to diverge; hence, resolution of these conflicts is of paramount importance to the decision to leave farming.

5. Goals of the farm operator and other family members are constantly adjusted upwards or downwards according to successes or failures in meeting past goals.

6. Any individual or group decision is based on reality as it is perceived and not as it is; perception is characterized by imperfect information and by the organizing of known elements into a meaningful whole.

7. In the process of deciding, individuals only take into account consequences of their choice which occur in a limited period called their "time perspective."

8. The farm operator's and his family's attachment to the local community and to farming as an occupation are important factors related to the decision to leave farming.

9. The decision to leave farming before retirement is contingent on other agents' decisions to provide alternative employment; these agents are given the generic name of "selector."

This study drew upon a newly available longitudinal data base: the Canadian Census of Agriculture Match. This data base was obtained by linking (matching) 1966 and 1971 agricultural census records pertaining to the same farm operator. The matching procedure was based on the surname, first name, and initials of the farm operator. At the onset of this study, the Census of Agriculture Match was assessed on the basis of the age consistency of matched farm operators. The quality was deemed insufficient for the study of off-farm movers and a complete manual match was performed for Saskatchewan, using a more precise and systematic set of matching rules.

A major advantage of the Census of Agriculture Match is that, theoretically, it covers all farm operators in an area. Thus, it is potentially a good basis for providing estimates of gross flows of farmers into and out of agriculture. A major shortcoming of the Census of Agriculture Match in relation to a study of the pre-retirement exit of farm operators is the limited scope of the information which is available. The main available variables considered to be relevant to this study were: residence on the farm, age-class of operator, total area of farm, area owned, area rented, value of land and buildings, value of machinery, value of livestock, total capital value, days worked off the holding, total rent, total value of agricultural products sold, type of farm, and tenure. Some variables thought to be essential for a study of pre-retirement exit of farm operators were not available in the Census of Agriculture Match. These are: level of education, number of dependents, net income from farming and from nonfarm sources, nonfarm occupation alternatives, etc. The only variables which could be added to the Census of Agriculture Match were distances from the farm to four classes of towns.

Only a limited number of hypotheses were formulated: those for which the available data base could provide supporting evidence or counter evidence. Pre-retirement exit from farming was hypothesized to be positively related to the farm operator's involvement in off-farm work and to nonresidence on the farm; it was hypothesized to be negatively related to farm operator's age, degree of ownership of farm land, total acreage, total capital value, total sales of agricultural products, productivity of land and capital, degree of mechanization, and distances to towns.

The pre-retirement decision to leave farming can be expressed for each farm operator by a binary variable whose values are:

0-1 (or stayer - exiter). Such a binary variable is a special case of what are called qualitative variables. Other decisions related to pre-retirement exit, such as residence and involvement in off-farm work can also be expressed by qualitative variables. Thus, the major dependent variables of the empirical study appeared to be qualitative. Consequently, the statistical techniques which are most commonly used by applied econometricians for the analysis of continuous variables were not applicable. A review of statistical methods for the analysis of qualitative dependent variables was necessary before any empirical analysis could proceed.

This review showed that the analysis of contingency tables and discriminant analysis were well suited for exploratory data analysis and that a multivariate log-linear model recently developed by Nerlove and Press was the most adapted for confirmatory data analysis.

The exploratory and confirmatory data analysis performed on the Census of Agriculture Match led to the following conclusions:

1. Pre-retirement exit of farm operators is positively (and not negatively, as it was hypothesized) related to the operator's age.

2. Pre-retirement exit of farm operators is positively related to the value of land and buildings expressed as a percentage of total capital value. Thus, farm operators whose assets consist mainly of land and buildings are more likely to leave farming before retirement.

3. Pre-retirement exit of farm operators is negatively related to residence on the farm. Farmers not living on their farm are more likely to leave farming before retirement. It should be noted however, that underenumeration in the 1971 Census of Agriculture may have involved a high number of nonresident farm operators. This

would explain, in part, the strong relationship found between pre-retirement exit and nonresidence which was evidenced.

4. Results were unconclusive concerning the hypothesized negative relationship between pre-retirement exit, and total acreage and total capital value of the farm.

5. Hypotheses concerning relationships between pre-retirement exit of farm operators on one hand and off-farm work, age, degree of ownership of farm land, total sales of agricultural products, productivity of land and of capital, degree of mechanization and distances to towns, on the other hand, were not supported.

In summary, empirical findings were mostly negative: data provided supporting evidence only for the hypotheses concerning the negative relationships between pre-retirement exit, on the one hand, and total acreage, total capital value, and residence on the farm, on the other hand. This led to two major further conclusions:

1. The decision to leave farming before retirement is very complex, depending on more than farm-business variables.

2. The negative findings raise some doubt about the ability of the present state of the Census of Agriculture Match to identify exiters, entrants, and stayers properly.

The largely negative results of the empirical analysis based on the best available secondary data prompted a proposal for a longitudinal survey of farm operators aimed at collecting information to substantiate the proposed theoretical framework. This survey would cover approximately 400 farm operators at an interval of two or three years. Farm operators who left farming between the two interviews would be traced to their new setting and interviewed. Information collected in these interviews would (1) serve as an input to further developments in the theoretical framework, and (2) be subject to exploratory and confirmatory data analysis.

Methodological Considerations

A Behavioral Approach

The behavioral approach in economics is not new, but it remains of minor importance. It has taken some extension in two areas: (1) the study of large organizations, either firms or public administrations, and (2) the study of consumer behavior to which Katona made major contributions.

Thus, the adoption of a behavioral approach in studying the decision to leave farming before retirement, is an extension of this approach to a new domain. Simple as it may seem, the decision to leave farming is the result of a complex decision making process and, as such, deserves a behavioral study.

A conceptual framework was proposed to assist in empirical analysis; thus, the main concern was to provide a model of behavior, a set of concepts, which were empirically tractable. For example, a utility maximizing income model can be advocated if the concept of "psychic income" derived from farming is introduced; unfortunately, such psychic income is difficult to measure independently and, consequently, leads to an imprecise empirical analysis; on the other hand, measurement of satisfaction from farming as an occupation and attachment to the local community can be measured on an ordinal scale.

In summary, a behavioral approach, by bringing detail into the analysis of the decision-making process was thought to provide concepts with greater empirical relevance.

Statistical Methods and Inference

The approach to inference and data analysis taken in this study is at variance with that taken in the vast majority of empirical economic works.

Three types of inference are usually recognized: deductive inference, inductive inference and reductive inference. Deductive inference consists of logical proof by which statements, or conclusions, are derived from prior statements, or premises, using the rules of logic. Inductive inference involves making inferences from the particular to the general or, more specifically, making inferences from experiences in the past or in specific settings to predict experiences inthe future or in other settings. Reductive inference consists of the description and the study of facts and the generation of hypotheses explaining these facts.

The hypothetico-deductive method, which is predominant in the economic profession, blends deductive and inductive inference. It has been very fruitful but, when strictly applied, it leaves little scope for reductive inference. At the same time as economists have adopted the hypothetico-deductive method they have emphasized the use of ever more sophisticated methods for confirmatory data analysis.

The stand taken in this research is that neither deductive, inductive, nor reductive inference should be neglected; reductive inference, or the process of idea formation, should be given more importance than it is given in the hypothetico-deductive method. As the relative importance of inductive inference is reduced and the relative importance of reductive inference increases, a wider range of statistical methods are used. Inductive inference still requires the confirmatory data analysis of classical statistics, but reductive inference requires exploratory data analysis of a new type. Tukey recently expressed the need for such a broad approach to data analysis:

> The principles and procedures for what we call confirmatory data analysis are both widely used and one of the great intellectual products of our century . . .We can no longer get along without confirmatory data analysis. But we need not start with it. . . . .
>
> Today, exploratory and confirmatory [data analysis] can--and should--proceed side by side.[1]

While the data analysis was being performed, two important technical points, which had been stressed in the chapter on statistical models, were vividly exemplified.

First, the problems of ascertaining the existence of an association between two (or more) variables and of measuring this association are distinct; they consequently require different procedures, namely, tests of independence and measures of association, which rely on different statistics. Thus, very significant results in a test of independence (indicating that variables are most likely dependent) and very weak associations are perfectly consistent; a large number of observations will ensure statistical significance even in the case of weak association. For example, the cross-tabulation[2] of farm

---

[1]John W. Tukey, Exploratory Data Analysis (Reading, Massachusetts: Addison-Wesley Publishing, 1977), pp. vi-vii.

[2]See Table B.1 in Appendix B.

operators, 64 years of age or less of age, by decision regarding farming and tenure in census division 7 of Saskatchewan has a χ^2 statistic of 27.62 which ensures that decisions regarding farming and tenure are dependent at a level of significance of less than 0.01. Nevertheless, the λ_A and UC_A[3] are respectively smaller than 0.001 and equal to 0.029; this means that the ability to predict the decision regarding farming when tenure is known is either unchanged or increased by 2.9 percent, depending on the prediction method. Thus, the association between the two variables is very weak.

Second, it is important to choose the structure of a model on limited part of the data set and to test it on the other part(s). Such procedure is possible only with a large enough data set. In this empirical study the data base was divided in several parts: one for each census division with approximately one thousand observations for each. A log-linear model of the pre-retirement exit from farming was chosen based partially on statistical tests using data pertaining to one census division. This model was then estimated and tested for other census divisions. The stringent procedure showed that some coefficients which appeared to be significantly different from zero following the choice and estimation of the model on part of the data, were not significantly different from zero when the model was tested on data pertaining to other census divisions. This exemplifies the danger of the very common practice consisting of chosing and testing a model on the same data. Such a practice

---

[3] λ_A is the asymmetric λ and UC_A the uncertainty coefficient when the decision regarding farming is considered to be dependent.

misleads the experimenter in failing to reject hypotheses which, on the basis of a more correct statistical procedure, should be rejected.

APPENDICES

APPENDIX A

NAMES AND DESCRIPTIONS OF VARIABLES

APPENDIX A

NAMES AND DESCRIPTIONS OF VARIABLES

Table A.1 Names and Descriptions of Variables

Variable Number	Variable Name	Year	Description
1	EXIT	66	Stayer-Exiter Variable 0.0 Stayer 1.0 Exiter
2		66	Census Division Number
3		66	Census Subdivision Number
4		66	Census-Farm Number
5		66	Crop District Number
6	RESID	66	Residence on the Farm in Previous Year 1.00 9-12 months 2.00 5-8 months 3.00 1-4 months 4.00 non-resident
7	AGECL	66	Age Class of Operator 1.00 under 25 2.00 25-34 3.00 35-44 4.00 45-54 5.00 55-59 6.00 60-65 7.00 65-69 8.00 over 70
8	AGED1	66	Age Class-Dummy 1[a]
9	AGED2	66	Age Class-Dummy 2[a]
10	AGED3	66	Age Class-Dummy 3[a]
11	AVAGE	66	Average of Age Class of Operator
12	TOTAREA	66	Total Area of Farm (Acres)
13	LDOWNED	66	Area-Owned (Acres)
14	LDRENTED	66	Area-Rented (Acres)
15	LDMAN	66	Area-Managed (Acres)
16	$LDBLD	66	Value of Land and Buildings ($100)
17	$MACH	66	Value of All Machinery ($100)

Table A.1 Continued

Variable Number	Variable Name	Year	Description
18	$STOCK	66	Total Livestock Value ($100)
19	TOTCAP	66	Total Capital Value ($100)
20	ARCROP	66	Area-Cropland (Acres)
21	ARIMP	66	Area-Improved Land-Pasture (Acres)
22	ARSF	66	Area-Summer Fallow (Acres)
23	ARWOOD	66	Area-Woodland (Acres)
24	ARUNIMP	66	Area-Other Unimproved Land (Acres)
25	OFFINC	66	Off-Farm Income 1.00 Under $750 2.00 $750-Plus
26	OFFWORK	66	Days Worked Off Holding
27	WORKERS	66	Number Year Round Workers
28	WAGES	66	Cash Wages Paid ($10)
29	TAXES	66	Taxes ($10)
30	RENT$	66	Rent on Cash Basis ($10)
31	RENTSH	66	Rent on Share or Kind Basis ($10)
32	TOTRENT	66	Total Rent = V30 + V31
33	$WHEAT	66	Value-Wheat Sold ($10)
34	$GRAIN	66	Value-Other Grains Sold ($10)
35	$CATTLE	66	Value-Cattle Sold($10)
36	$PIG	66	Value-Pigs Sold ($10)
37	$POULTRY	66	Value-Hens and Chickens Sold ($10)
38	$DAIRY	66	Value-Dairy Products Sold ($10)
39	TOTSALES	66	Value-Total Sales ($10)
40	INST	66	Institutional Farm 1.00 Yes 2.00 No
41		66	Type of Farm 1.00 Dairy 2.00 Cattle-Hogs-Sheep 3.00 Poultry 4.00 Wheat 5.00 Small Grains 6.00 Field Crops

Table A.1 Continued

Variable Number	Variable Name	Year	Description
41			7.00 Fruits and Vegetables
			8.00 Forestry
			9.00 Misc. Specialty
			10.00 Mixed-Livestock
			11.00 Mixed-Field Crops
			12.00 Mixed-Other
42		66	Economic Class
			1.00 $35,000 and over
			2.00 $25,000 - 34,999
			3.00 $15,000 - 24,999
			4.00 $10,000 - 14,999
			5.00 $ 7,500 - 9,999
			6.00 $ 5,000 - 7,499
			7.00 $ 3,750 - 4,999
			8.00 $ 2,500 - 3,749
			11.00 $ 1,200 - 2,499
			12.00 $ 250 - 1,199
			13.00 $ 50 - 249
			21.00 Institutional
43		66	Part Time Work
			1.00 None
			2.00 Less than 7 days
			3.00 7 - 12 days
			4.00 13 - 24 days
			5.00 25 - 48 days
			6.00 49 - 72 days
			7.00 73 - 96 days
			8.00 97 - 126 days
			9.00 127 - 156 days
			10.00 157 - 228 days
			11.00 229 - 365 days
44		66	Tenure
			1.00 Owned
			2.00 Rented
			3.00 Owned-Rented
			4.00 Managed
45		66	Size of Farm
			1.00 1 - 2
			2.00 3 - 9
			3.00 10 - 69
			4.00 70 - 239
			5.00 240 - 399
			6.00 400 - 559
			7.00 560 - 759

Table A.1 Continued

Variable Number	Variable Name	Year	Description
45			8.00 760 - 1,119 9.00 1,120 - 1,599 10.00 1,600 - 2,239 11.00 2,240 - 2,879 12.00 2,880 - Plus
46		66	Total Capital Value 1.00 Under $1,950 2.00 $ 1,950 - 2,949 3.00 $ 2,950 - 3,949 4.00 $ 3,950 - 4,949 5.00 $ 4,950 - 7,449 6.00 $ 7,450 - 9,949 7.00 $ 9,950 - 14,949 8.00 $ 14,950 - 19,949 9.00 $ 19,950 - 24,949 10.00 $ 24,950 - 49,949 11.00 $ 49,950 - 99,949 12.00 $ 99,950 - 149,949 13.00 $149,950 - 199,949 14.00 $199,950 - And Over
47		66	Acreage Improved Land 1.00 0 2.00 1 - 2 3.00 3 - 9 4.00 10 - 69 5.00 70 - 129 6.00 130 - 179 7.00 180 - 239 8.00 240 - 399 9.00 400 - 559 10.00 560 - 759 11.00 760 - 1,119 12.00 1120 - 1,599 13.00 1600 - And over
48		66	Cattle-Sheep 1.00 Yes 2.00 No
49	%LDOWNED	66	Percent-Area-Owned
50	%LDRENT	66	Percent-Area-Rented
51	%LDMAN	66	Percent-Area-Managed
52	%ARCROP	66	Percent-Area-Cropland
53	%ARIMP	66	Percent-Area-Impr. Pasture

Table A.1 Continued

Variable Number	Variable Name	Year	Description
54	%ARSF	66	Percent-Area-Summer Fallow
55	%ARWOOD	66	Percent-Area-Woodland
56	%ARUNIMP	66	Percent-Area-Other Unimproved
57	%$WHEAT	66	Percent-Value-Wheat Sold
58	%$GRAIN	66	Percent-Value-Other Grains Sold
59	%$CATTLE	66	Percent-Value-Cattle Sold
60	%$PIG	66	Percent-Value-Pigs Sold
61	%$POULTRY	66	Percent-Value-Poultry Sold
62	%$DAIRY	66	Percent-Value-Dairy Products Sold
63	%$LDBLD	66	Percent-Value-Land and Buildings
64	%$MACH	66	Percent-Value-All Machinery
65	%$STOCK	66	Percent-Value-Livestock
66	DIST1	66	Distance to Town (2,500-5,000) in miles
67	DIST2	66	Distance to Town (4,000-10,000) " "
68	DIST3	66	Distance to Town (10,000-40,00) " "
69	DIST4	66	Distance to Town (40,000 Plus) " "
70		71	Area-Owned (Acres)
71		71	Area-Rented (Acres)
72		71	Total Area of Farm (Acres)
73		71	Value Land-Building ($100)
74		71	Value Machinery-Equipment ($100)
75		71	Value Livestock ($100)
76		71	Total Capital Value ($100)
77		71	Days Part Time Work
78		71	Value Total Sales ($10)
79		71	Type of Farm
80		71	Economic Class 11.00 50,000 - Plus 12.00 35,000 - 49,999 13.00 25,000 - 34,999 14.00 15,000 - 24,999 15.00 10,000 - 14,999 21.00 7,500 - 9,999

Table A.1 Continued

Variable Number	Variable Name	Year	Description
			22.00 5,000 - 7,499
			31.00 3,750 - 4,999
			32.00 2,500 - 3,749
			41.00 1,200 - 2,499
			42.00 250 - 1,199
			43.00 50 - 249
			51.00 Institutional
81		71	Farm Size (Acres)
			1.00 1 - 2
			2.00 3 - 9
			3.00 10 - 69
			4.00 70 - 239
			5.00 240 - 399
			6.00 400 - 559
			7.00 560 - 759
			8.00 760 - 1,119
			9.00 1,120 - 1,599
			10.00 1,600 - 2,239
			11.00 2,240 - 2,879
			12.00 2,880 - Plus
82		71	Total Capital Value
			1.00 Under $2,950
			2.00 $ 2,950 - 4,949
			3.00 $ 4,950 - 7,449
			4.00 $ 7,450 - 9,949
			5.00 $ 9,950 - 14,949
			6.00 $ 14,950 - 19,949
			7.00 $ 19,950 - 24,949
			8.00 $ 24,950 - 49,949
			9.00 $ 49,950 - 74,949
			10.00 $ 74,950 - 99,949
			11.00 $ 99,950 -149,949
			12.00 $149,950 -199,949
			13.00 $199,950 - Plus
83		71	Part Time Work (Days)
			1.00 None
			2.00 Less than 7
			3.00 7 - 12
			4.00 13 - 24
			5.00 25 - 48
			6.00 49 - 72
			7.00 73 - 96
			8.00 97 - 126
			9.00 127 - 156
			10.0 157 - 228
			11.00 229 - 365

Table A.1 Continued

Variable Number	Variable Name	Year	Description
84		71	Length of Residence in Previous Year 1.00 9 - 12 months 2.00 5 - 8 months 3.00 1 - 4 months 4.00 Non-resident
85	MECH1	66	Mechanization = V17 over V12
86	MECH2	66	Mechanization = V17 over V20 + V21 + V22
87	PRODCAP	66	Product. Capital = V39 over V19
88	PRODLD1	66	Product. Land = V39 over V12
89	PRODLD2	66	Product. Land = V39 over V20 + V21 + V22
90	LNTOTAR	66	LN of Total Area[b]
91	LNTOTCAP	66	LN of Total Capital Value[b]
92	LNTOTREN	66	LN of Total Rent[b]
93	LNTOTSAL	66	LN of Total Sales[b]
94	LNOFFWORK	66	LN of Days of Off-Farm Work[b]
95	RESID1	66	Residence on the Farm in Previous Year 0. Non-resident 1. 1 - 12 months
96	RESID2	66	Residence on the Farm in Previous Year 0. 0 - 4 months 1. 5 - 12 months
97	OFFWORK1	66	Days Off Farm Work 0. None 1. Some
98	OFFWORK2	66	Days Off Farm Work 0. 0 - 24 days 1. 24 - 365 days

[a]AGED1, AGED2, and AGED3 are three binary variables used to express the eight-class qualitative variable AGECL.

[b]LN Stands for "natural logarithm of."

APPENDIX B

CROSS-TABULATIONS

APPENDIX B

CROSS-TABULATIONS

Table B.1 Farm Operators by Decision Regarding Farming and by Length of Resident on Farm, Census Division 7, Saskatchewan

Decision Regarding Farming	Residence on the Farm (Months in Previous Year)				Total
	9-12	5-8	1-4	0	
Stayer	555 (75.5)	37 (60.7)	15 (75.0)	133 (62.1)	740 (71.8)
Exiter	180 (24.5)	24 (39.3)	5 (25.0)	81 (37.9)	290 (28.2)
Total	735	61	20	214	1030

Notes: Figures in parentheses are percentage of column totals.

χ^2 = 18.70, DF = 3, Level of Significance = 0.0.
λ_A = 0.0, λ_B = 0.0,
UC_A = 0.015, UC_B = 0.011;

Table B.2 Farm Operators, 64 or Less, by Decision Regarding Farming and by Length of Residence on Farm, Census Division 7, Saskatchewan

Decision Regarding Farming	Residence on the Farm (Months in Previous Year)				Total
	9-12	5-8	1-4	0	
Stayer	529 (80.6)	29 (65.9)	15 (83.3)	120 (65.6)	693 (76.9)
Exiter	127 (19.4)	15 (34.1)	3 (16.7)	63 (34.4)	208 (23.1)
Total	656	44	18	183	901

Notes: Figures in parentheses are percentage of column totals.

χ^2 = 21.80, DF = 3, Level of Significance = 0.00.
λ_A = 0.0, λ_B = 0.0,
UC_A = 0.021, UC_B = 0.015,

Table B.3 Farm Operators, by Decision Regarding Farming and by Age, County 7, Saskatchewan

Decision Regarding Farming	Age								Total
	Under 25	25-34	35-44	45-54	55-59	60-64	65-69	Over 70	
Stayer	25 (61.0)	111 (81.6)	196 (84.5)	233 (77.4)	85 (70.2)	43 (61.4)	24 (42.1)	23 (31.9)	740 (71.8)
Exiter	16 (39.0)	25 (18.4)	36 (15.5)	68 (22.6)	36 (29.8)	27 (38.6)	33 (57.9)	49 (68.1)	290 (28.2)
Total	41	136	232	301	121	70	57	72	1030

Figures in parentheses are percentage of column totals

χ^2 = 117.24 DF = 7 Level of significance = 0.00

λ_A = 0.121 λ_B = 0.0

UC_A = 0.089 UC_B = 0.028

Statistics for subtable for farm operators less than 65.

χ^2 = 27.56 DF = 5 Level of significance = 0.00

λ_A = 0.0 λ_B = 0.0

UC_A = 0.027 UC_B = 0.009

Table B.4 Farm Operators by Decision Regarding Farming and by Off-Farm Income, Census Division 7, Saskatchewan

Decision Regarding Farming	Off-Farm Income		Total
	Under $750	Over $750	
Stayer	610 (72.6)	130 (68.4)	740 (71.8)
Exiter	230 (27.4)	60 (31.6)	290 (28.2)
Total	840	190	1030

Notes: Figures in parentheses are percentages of column totals.

$\chi^2 = 1.15$, DF = 1, Level of Significance = .28.
$\lambda_A = 0.0$, $\lambda_B = 0.0$,
$UC_A = 0.001$, $UC_B = 0.001$,

Table B.5 Farm Operators, 64 or Less, by Decision Regarding Farming and by Off-Farm Income, Census Division 7, Saskatchewan

Decision Regarding Farming	Off-Farm Income		Total
	Under \$750	Over \$750	
Stayer	564 (78.8)	129 (69.7)	693 (79.5)
Exiter	152 (21.2)	56 (30.3)	208 (20.5)
Total	716	185	901

Notes: Figures in parentheses are percentages of column totals.

$\chi^2 = 6.27$, DF = 1, Level of significance = 0.01.
$\lambda_A = 0.00$, $\lambda_B = 0.0$,
$UC_A = 0.007$, $UC_B = 0.007$,

Table B.6 Farm Operators by Decision Regarding Farming and by Sales of Agricultural Products, Census Division 7, Saskatchewan

Decision Regarding Farming	Sales of Agricultural Products ($)												Total
	35,000 and over	25,000 -34,999	15,000 -24,999	10,000 -14,999	7,500 -9,999	5,000 -7,499	3,750 -4,999	2,500 -3,749	1,200 -2,499	250 1,199	50 249	Inst*	
Stayer	14 (73.7)	33 (82.5)	124 (87.3)	172 (81.1)	107 (69.5)	116 (64.4)	70 (72.9)	48 (62.3)	42 (72.4)	5 (17.5)	7 (35.0)	2 (50.0)	740 (71.8)
Exiter	5 (26.3)	7 (17.5)	18 (12.7)	40 (18.9)	47 (30.5)	64 (35.6)	26 (27.1)	29 (37.7)	16 (27.6)	23 (82.1)	13 (65.0)	2 (50.0)	290 (28.2)
Total	19	40	142	212	154	180	96	77	58	28	20	4	1030

Figures in parentheses are percentages of column totals

χ^2 = 91.65 DF = 11 Level of significance = 0.00

λ_A = 0.083 λ_B = 0.029

UC_A = 0.072 UC_B = 0.020

* Institutional farms, for which sales of agricultural products are not meaningful.

Table B.7 Farm Operators, 64 and Less, by Decision Regarding Farming and Sales of Agricultural Products, Census Division 7, Saskatchewan

Decision Regarding Farming	Sales of Agricultural Products ($)												Total
	35,000 and over	25,000 -34,999	15,000 -24,999	10,000 -14,999	7,500 -7,499	5,000 -7,499	3,750 -4,999	2,500 -3,749	1,200 2,499	250 1,199	50 249	Inst*	
Stayer	14 (73.7)	32 (82.1)	118 (89.4)	165 (85.1)	103 (76.9)	108 (72.0)	64 (77.1)	44 (66.7)	31 (75.6)	5 (23.8)	7 (38.9)	2 (50.0)	693 (76.9)
Exiter	5 (26.3)	7 (17.9)	14 (10.6)	29 (14.9)	31 (23.1)	42 (28.0)	19 (22.9)	22 (33.3)	10 (24.4)	16 (76.2)	11 (61.1)	2 (50.0)	208 (23.1)
Total	19	39	132	194	134	150	83	66	41	21	18	4	901

Figures in parentheses are percentages of column totals

x^2 = 75.13 DF = 11 Level of significance = 0.00

λ_A = 0.072 λ_B = 0.018

UC_A = 0.070 UC_B = 0.017

* Institutional farms, for which sales of agricultural products are not meaningful.

Table B.8 Farm Operators by Decision Regarding Farming and by Days of Off-Farm Work, Census Division 7, Saskatchewan

Decision Regarding Farming	Days of Off-farm Work											Total
	None	1-7	7-12	13-24	25-48	49-72	73-96	97-126	127-156	157-228	229-365	
Stayer	542 (71.5)	21 (84.0)	19 (73.1)	29 (85.3)	16 (69.6)	15 (60.0)	16 (88.9)	15 (71.4)	9 (81.8)	30 (75.0)	28 (57.1)	740 (71.8)
Exiter	216 (28.5)	4 (16.0)	7 (26.9)	5 (14.7)	7 (30.4)	10 (40.0)	2 (11.1)	6 (28.6)	2 (18.2)	10 (25.0)	21 (42.9)	290 (28.2)
Total	758	25	26	34	23	25	18	21	11	40	49	1030

Figures in parentheses are percentages of column totals

$\chi^2 = 15.28$ DF = 10 Level of significance = 0.12

$\lambda_A = 0.0$ $\lambda_B = 0.0$

$UC_A = 0.013$ $UC_B = 0.007$

Table B.9 Farm Operators, 64 or Less, by Decision Regarding Farming and by Days of Off-Farm Work, Census Division 7, Saskatchewan

Decision Regarding Farming	Days of Off-farm Work											Total
	0	1-7	7-12	13-24	25-48	49-72	73-96	97-126	127-156	157-228	229-365	
Stayer	501 (78.0)	20 (87.0)	19 (76.0)	26 (86.7)	15 (68.2)	15 (65.2)	16 (88.9)	15 (75.0)	9 (81.8)	29 (74.4)	28 (58.3)	693 (76.9)
Exiter	141 (22.0)	3 (13.0)	6 (24.0)	4 (13.3)	7 (31.8)	8 (34.8)	2 (11.1)	5 (25.0)	2 (18.2)	10 (25.6	20 (41.7)	208 (23.1)
Total	642	23	25	30	22	23	18	20	11	39	48	901

Figures in parentheses are percentages of column totals

χ^2 = 17.22 DF = 10 Level of significance = 0.07

λ_A = 0.0 λ_B = 0.0

UC_A = 0.017 UC_B = 0.007

Table B.10 Farm Operators by Decision Regarding Farming and by Tenure, Census Division 7, Saskatchewan

Decision Regarding Farming	Tenure				Total
	Owned	Rented	Owned-Rented	Managed	
Stayer	292 (62.8)	64 (68.8)	380 (81.5)	4 (66.7)	740 (71.8)
Exiter	173 (37.2)	29 (31.2)	86 (18.5)	2 (33.3)	290 (28.2)
Total	465	53	466	6	1030

Notes: Figures in parentheses are percentages of column totals.

$\chi^2 = 41.00$, DF = 3, Level of significance = 0.00.
$\lambda_A = 0.0$, $\lambda_B = 0.15$,
$UC_A = 0.034$, $UC_B = 0.021$,

Table B.11 Farm Operators, 64 or Less, by Decision Regarding Farming and by Tenure, Census Division, Saskatchewan

Decision Regarding Farming	Tenure				Total
	Owned	Rented	Owned-Rented	Managed	
Stayer	259 (69.8)	63 (70.0)	367 (84.6)	4 (66.7)	693 (76.9)
Exiter	112 (30.2)	27 (30.0	67 (15.4)	2 (33.3)	208 (23.1)
Total	371	90	434	6	901

Notes: Figures in parentheses are percentages of column totals.

$\chi^2 = 27.62$, DF = 3, Level of Significance = 0.00.
$\lambda_A = 0.0$, $\lambda_B = 0.096$,
$UC_A = 0.029$, $UC_B = 0.016$,

Table B.12 Farm Operators by Decision Regarding Farming and Total Capital Value of Farm, Census Division 7, Saskatchewan

Decision Regarding Farming	Total Capital Value														Total
	Under $1950	$1950 -2949	$2950 -3949	$3950 -4949	$4950 -7449	$7450 -9949	$9950 -14949	$14950 -19949	$19950 -24949	$24950 -49494	$49950 -99949	$99950 -149950	$149950 -199949	$199950 and over	
Stayer	0 (0.0)	1 (33.3)	1 (25.0)	1 (20.0)	3 (23.1)	5 (45.5)	17 (58.6)	21 (56.8)	29 (64.4)	165 (64.0)	317 (76.9)	118 (84.3)	36 (94.7)	26 (78.8)	740 (71.8)
Exiter	2 (100.0)	2 (66.7)	3 (75.0)	4 (80.0)	10 (76.9)	6 (54.5)	12 (41.4)	16 (43.2)	16 (35.6)	93 (36.0)	95 (23.1)	22 (15.7)	2 (5.3)	7 (21.2)	290 (28.2)
Total	2	3	4	5	13	11	29	37	45	258	412	140	38	33	1030

Figures in parentheses are percentages of column totals.

x^2 = 79.82 DF = 13 Level of significance = 0.00

λ_A = 0.055 λ_B = 0.0

UC_A = 0.065 UC_B = 0.022

Table B.13 Farm Operators, 64 or Less, by Decision Regarding Farming and Total Capital Value of Farm, Census Division 7, Saskatchewan

Decision Regarding Farming	Total Capital Value														Total
	Under $1950	$1950 -2949	$2950 -3949	$3950 -4949	$4950 -7449	$7450 -9949	$9950 -14949	$14950 -19949	$19950 -24949	$24950 -49949	$49950 -99949	$99950 -149950	$149950 -199949	$199950 and over	
Stayer	0 (0.0)	0 (0.0)	1 (25.0)	1 (50.0)	2 (25.0)	3 (42.9)	12 (52.2)	18 (60.0)	25 (69.4)	152 (72.7)	302 (80.7)	116 (85.9)	35 (94.6)	26 (81.3)	693 (76.9)
Exiter	2 (100.0)	2 (100.0)	3 (75.0)	1 (50.0)	6 (75.0)	4 (57.1)	11 (47.8)	12 (40.0)	11 (30.6)	57 (27.3)	72 (19.3)	19 (14.1)	2 (5.4)	6 (18.8)	208 (23.1)
Total	2	2	4	2	8	7	23	30	36	209	374	135	37	32	901

Figures in parentheses are percentages of columns totals.

x^2 = 69.01 DF = 13 Level of significance = 0.00

λ_A = 0.053 λ_B = 0.0

UC_A = 0.066 UC_B = 0.021

Table B.14. Farm Operators by Decision Regarding Farming and by Acreage of Improved Land, Census Division 7, Saskatchewan

Decision Regarding Farming	Acreage of Improved Land													Total
	0	1-2	3-9	10-69	70-129	130-179	180-239	240-399	400-559	560-759	760-1119	1120-1599	1600 and over	
Stayer	1 (14.3)	0 (0.0)	4 (44.4)	6 (35.3)	11 (57.9)	28 (54.9)	13 (52.0)	103 (64.4)	124 (71.7)	134 (70.5)	180 (81.8)	99 (88.4)	37 (80.4)	740 (71.8)
Exiter	6 (85.7)	1 (100.0)	5 (55.6)	11 (64.7)	8 (42.1)	23 (45.1)	12 (48.0)	57 (35.6)	49 (28.3)	56 (29.5)	40 (18.2)	13 (11.6)	9 (19.6)	290 (28.2)
Total	7	1	9	17	19	51	25	160	173	190	220	112	46	1030

Figures in parentheses are percentages of column totals

$\chi^2 = 74.75$ DF = 12 Level of significance = 0.00

$\lambda_A = 0.041$ $\lambda_B = 0.021$

$UC_A = 0.061$ $UC_B = 0.017$

Table B.15. Farm Operators, 64 or Less, By Decision Regarding Farming and by Acreage of Improved Land, Census Division 7, Saskatchewan

Decision Regarding Farming	Acreage of Improved Land													Total
	0	1-2	3-9	10-69	70-129	130-179	180-239	240-399	400-559	560-759	760-1119	1120-1599	1600 and over	
Stayer	1 (14.3)	0 (0.0)	4 (57.1)	6 (40.0)	8 (57.1)	24 (60.0)	13 (65.0)	86 (89.9)	117 (78.0)	128 (77.1)	173 (85.2)	96 (88.1)	37 (80.4)	693 (76.9)
Exiter	6 (85.7)	1 (100.0)	3 (42.9)	9 (60.0)	6 (42.9)	16 (40.0)	7 (35.0)	37 (30.1)	33 (22.0)	38 (22.9)	30 (14.8)	13 (11.9)	9 (19.6)	208 (23.1)
Total	7	1	7	15	14	40	20	123	150	166	203	109	46	901

Figures in parentheses are percentages of column totals

x^2 = 62.32 DF = 12 Level of significance = 0.00

λ_A = 0.043 λ_B = 0.011

UC_A = 0.059 UC_B = 0.015

BIBLIOGRAPHY

BIBLIOGRAPHY

Reports and Bulletins

Abrahamson, Jane A. Rural to Urban Adjustment. Ottawa: Department of Regional Economic Expansion, 1968.

Baird, Andrew W., and Bailey, Wilfrid C. Farmers Moving Out of Agriture. Agricultural Experiment Station Bulletin 568, Mississippi State University, 1958.

Bowring, J.R., and Durgin, O.B. Factors Influencing the Attitudes of Farmers Towards Migration Off Farms. Agricultural Experiment Station Bulletin 458, University of New Hampshire, n.d.

Darcovich, W.; Gellner, J.; and Piracha, Z. "Estimates of Low Income in the Farm Sector." Working Paper, Agriculture Canada, Economics Branch, 1977.

Jones, A.R. Policies and Programs for Agriculture--Atlantic Provinces. Publication no. 76/13. Ottawa: Agriculture Canada, 1976.

________. Policies and Programs for Agriculture--Ontario--Quebec. Publication no. 76/14. Ottawa: Agriculture Canada, 1976.

________. Policies and Programs for Agriculture--Western Provinces. Publication no. 76/12. Ottawa: Agriculture Canada, 1976.

Kaldor, Donald R., and Edwards, William M. Occupational Adjustment of Iowa Farm Operators Who Quit Farming in 1959-1961. Agricultural and Home Economics Experiment Station, Special Report 75, Iowa State University, 1975.

McDonald, Stephen L. "Economic Factors in Farm Out-Migration: A Survey and Evaluation of the Literature." Austin, Texas, n.d. (Mimeographed.)

Nerlove, Marc, and Press, S. James. Univariate and Multivariate Log-Linear and Logistic Models. Rand Report R-1306-EDA/NIH. Santa Monica: Rand Co., December 1973.

________. "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data." Center for Statistics and Probability, Discussion Paper No. 1. Evanston, 1976.

Perkins, Brian B., and Hathaway, Dale E. The Movement of Labor Between Farm and Nonfarm Jobs. Agricultural Experiment Station Research Bulletin 13, Michigan State University, 1966.

Articles

Alexander, Irving E., et al. "The Level of Aspiration Model Applied to Occupational Preference." Human Relations 12 (May 1959): 163-170.

Arrow, Kenneth J. "A Difficulty in the Concept of Social Welfare." In Readings in Welfare Economics, pp. 147-168. Edited by K.J. Arrow and T. Scitowsky. Homewood, Illinois: Richard D. Irwin, 1969.

Baumgartner, H.W. "Potential Mobility in Agriculture: Some Reasons for the Existence of a Labor-Transfer Problem." Journal of Economics 47 (February 1965): 74-82.

Bennett, Claude F. "Mobility From Full-time to Part-time Farming." Rural Sociology 32 (June 1967): 154-164.

Benzécri, Jean-Paul. "La Place de l'a priori." In Encyclopaedia Universalis--Organum, pp. 11-24. Paris: Encylopaedia Universalis, 1973.

Berkson, J. "Estimate of the Integrated Normal Curve by Minimum Normit χ^2 with Particular Reference to Bio-Assay." Journal of the American Statistical Association 50 (June 1955): 529-549.

Blau, Peter M. et al. "Occupational Choice: A Conceptual Framework." In Personality and Social Systems, pp. 559-571. Edited by Neil J. Smelser and William J. Smelser. New York: John Wiley and Sons, 1970.

Clawson, Marion. "Aging Farmers and Agricultural Policy." Journal of Farm Economics 45 (February 1963): 13-30.

Colberg, Marshall R., and King, James P. "Theory of Production Abandonment." Rivista Internationale die Scienze Economiche 20 (October 1973): 961-971.

Costner, H.L. "Criteria for Measures of Association." American Sociological Review 30 (June 1965): 341-353.

Davies, Vernon. "Development of a Scale to Rate Attitude of Community Satisfaction." Rural Sociology 10 (September 1945): 246-255.

Diehl, William D. "Farm-Nonfarm Migration in the Southeast: A Costs-Returns Analysis." Journal of Farm Economics 48 (February 1966): 1-11.

Dillon, John D. "An Expository Review of Bernoullian Decision Theory in Agriculture: Is Utility Futility?" Review of Marketing and Agricultural Economics 39 (March 1971): 3-80.

Fisher, R.A. "The Use of Multiple Measurement in Taxonomic Problems." Annals of Eugenics 7 (September 1936): 179-188.

Fox, Karl A. "Guiding Agricultural Adjustments." Journal of Farm Economics 34 (December 1957): 1090-1104.

Garb, Gerald. "Professor Samuelson on Theory and Realism: Comment." American Economic Review 55 (December 1965): 1151-1172.

Goodman, L.A., and Kruskal, W.H. "Measures of Association for Cross-Classifications." Journal of the American Statistical Association 49 (December 1954): 732-764.

Guither, Harold D. "Factors Influencing Decisions to Leave Farming." Journal of Farm Economics 45 (August 1963): 567-576.

Hathaway, Dale E. "Migration from Agriculture: The Historical Record and Its Meaning." American Economic Review 50 (May 1960): 379-391.

________. "Occupational Mobility from the Farm Labor Force." In Farm Labor in the United States, pp. 71-96. Edited by C.E. Bishop. New York: Columbia University Press, 1967.

Hathaway, Dale E., and Perkins, Brian B. "Farm Labor Mobility, Migration and Income Distribution." American Journal of Agricultural Economics 50 (May 1968): 342-53.

________. "Occupational Mobility and Migration from Agriculture." In Rural Poverty in the United States, A Report by the President's National Advisory Commission on Rural Poverty, pp. 185-237. Washington: Government Printing Office, 1968.

Heady, Earl O. "Adjusting the Labor Force to Agriculture." In Agricultural Adjustment Problems in a Growing Economy, pp. 145-159. Edited by Earl O. Heady et al. Ames: Iowa State College Press, 1958.

Hill, Lowell D. "Characteristics of the Farmers Leaving Agriculture in an Iowa County." Journal of Farm Economics 44 (May 1962): 419-426.

Jesser, Clinton. "Community Satisfaction Patterns of Professional in Rural Areas." Rural Sociology 32 (March 1963): 56-69.

Johnson, D. Gale. "Policies to Improve the Labor Transfer Process." American Economic Review 50 (May 1960): 403-418.

Johnson, D. Gale. "Labor Mobility and Agricultural Adjustment." In Agricultural Adjustment Problems in a Growing Economy, pp. 163-174. Edited by Earl O. Heady et al. Ames: Iowa State College Press, 1958.

Johnson, Ronald, and Knop, Edward. "Rural-Urban Differentials in Community Satisfaction." Rural Sociology 35 (December 1970): 544-548.

Johnston, W.E., and Tolley, G.S. "The Supply of Farm Operators." Econometrica 36 (April 1968): 365-382.

Kanel, Don. "Farm Adjustment by Age Groups, North Central States 1950-1959." Journal of Farm Economics 45 (February 1953): 47-60.

_____. "Age Components of Decrease in Numbers of Farmers, North Central States, 1890-1954." Journal of Farm Economics 43 (May 1961): 247-263.

Katona, George. "Psychological Analysis of Business Decisions and Expectations." American Economic Review 36 (March 1946): 44-62.

_____. "Consumer Behavior: Theory and Findings on Expectations and Aspirations." American Economic Review 58 (May 1968): 19-30.

_____. "Rational Behavior and Economic Behavior." Psychological Review 60 (May 1953): 307-318.

Lachenbruch, Peter A.; Sneeringer, Cheryl; and Revo, Lawrence T. "Robustness of the Linear and Quadratic Discriminant Function to Certain Type of Non-normality." Communications in Statistics 1 (1973): 39-56.

Ladd, George W. "Linear Probability Functions and Discriminant Functions." Econometrica 34 (October 1966): 873-885.

McDonald, Stephen L. "Farm Outmigration as an Integrative Adjustment to Economic Growth." Social Forces 34 (December 1955): 119-128.

McFadden, Daniel. "Quantal Choice Analysis: A Survey." Annals of Economic and Social Measurement 5 (Fall 1976): 363-390.

Maddox, J.G. "Private and Social Costs of the Movement of People Out of Agriculture." American Economic Review 50 (May 1960): 392-402.

Machlup, Fritz. "Marginal Analysis and Empirical Research." American Economic Review 36 (September 1946): 519-554.

_____. "Professor Samuelson on Theory and Realism." American Economic Review 54 (September 1964): 733-736.

Machlup, Fritz et al. "Problems of Methodology." American Economic Review 53 (May 1963): 204-236.

Melitz, Jack. "Friedman and Machlup on the Significance of Testing Economic Assumptions." Journal of Political Economy 73 (February 1965): 37-60.

Perkins, Brian B. "Farm Income and Labor Mobility." American Journal of Agricultural Economics 55 (December 1973): 913-920.

Platt, John. "Social Traps." American Psychologist 28 (August 1973): 641-651.

Roy, Prodipto. "Factors Related to Leaving Farming." Journal of Farm Economics 43 (August 1961): 666-674.

Samuelson, Paul A. "Theory and Realism: A Reply." American Economic Review 54 (September 1964): 736-739.

Scheuren, Fritz, and Oh, H. Lock. "Fiddling Around with Nonmatches and Mismatches." 1974 American Statistical Association Proceedings of the Social Statistics Section, 1975, pp. 627-633.

Simon, Herbert A. "Theories of Bounded Rationality." In Decision and Organization, pp. 161-176. Edited by C.B. McGuire and Roy Radner. Amsterdam: North Holland, 1972.

________. "Theories of Decision Making in Economics and Behavioral Science." American Economic Review 49 (June 1959): 253-283.

________. "A Behavioral Model of Rational Choice." Quarterly Journal of Economics 69 (February 1955): 99-118.

Simon, Herbert A., and Stedry, Andrew C. "Psychology and Economics." In The Handbook of Social Psychology, pp. 269-314. Edited by Gardner Lindzey and Elliot Aronson. Reading, Mass: Addison-Wesley, 1969.

Sjaastad, Larry A. "Occupational Structure and Migration Patterns." In Labor Mobility and Population in Agriculture, pp. 8-27. Edited by Iowa State University Center for Agricultural and Economic Adjustment. Ames: Iowa State University Press, 1961.

________. "The Costs and Returns of Human Migration." Journal of Political Economy 70 (October 1962): 80-93.

Smith, Eldon D. "Nonfarm Employment Information for Rural People." Journal of Farm Economics 38 (August 1956): 813-827.

Sofranko, A.J., and Pletcher, W.R. "Factors Influencing Farmers' Expectations for Involvement inAgriculture." Illinois Agricultural Economics 14 (July 1974): 6-12.

Szabo, Michael L. "Depopulation of Farms in Relation to the Economic Conditions of Agriculture on the Canadian Prairies." Geographical Bulletin 7 (Fourth quarter 1965): 187-202.

Theil, Henri. "A Multinomial Extension of the Linear Logit Model." International Economic Review 10 (October 1969): 251-259.

Tolley, George S., and Hjort H.W. "Age-Mobility and Southern Farmer Skill: Looking Ahead for Area Development." Journal of Farm Economics 45 (February 1963): 31-46.

Tukey, John W. "The Future of Data Analysis." Annals of Mathematical Statistics 33 (1962): 1-67.

________. "Analyzing Data: Sanctification or Detective Work?" American Psychologist 24 (1969): 83-91.

Tweeten, Luther G. "Theories Explaining the Persistence of Low Resource Returns in a Growing Economy." American Journal of Agricultural Economics 51 (November 1969): 798-817.

Wallace, T. Dudley. "Pretest Estimation in Regression: A Survey." American Journal of Agricultural Economics 59 (August 1977): 431-443.

Weerdenburg, L.J.M. "Farmers and Occupational Change." Sociologia Ruralis 13 (First quarter 1973): 27-38.

Winkelman, Don. "A Case Study of the Exodus of Labor from Agriculture: Minnesota." Journal of Farm Economics 48 (February 1966): 12-21.

Wong, Stanley. "The F-Twist and the Methodology of Paul Samuelson." American Economic Review 63 (June 1973): 312-325.

Zellner, A., and Lee, T.H. "Joint Estimation of Relationships Involving Discrete Random Variables." Econometrica 33 (April 1965): 382-394.

Books

Ashton, N.D. The Logit Transformation. New York: Hafner, 1972.

Bishop, Y.M.M.; Fienberg, S.E.; and Holland, P.W. Discrete Multivariate Analysis: Theory and Practice. Cambridge: The MIT Press, 1975.

Bogue, Donald J. Principles of Demography. New York: John Wiley and Sons, 1969. Quoted in A.J. Sofranko and W.R. Pletcher, "Factors Influencing Farmers' Expectations for Involvement in Agriculture," Illinois Agricultural Economics 14 (July 1974): 6-12.

Braithwaite, Richard B. Scientific Explanation. Cambridge: Cambridge University Press, 1953.

Cox, D.R. The Analysis of Binary Data. London: Methuen, 1970.

Friedman, Milton. Essays in Positive Economics. Chicago: University of Chicago Press, 1953.

Goldberger, A.S. Econometric Theory. New York: John Wiley & Sons, 1964.

Hicks, J.R. Value and Capital, 2nd ed. Oxford: Clarendon Press, 1946.

Hirschleifer, J. Investment, Interest and Capital. Englewood Cliffs, N.J.: Prentice Hall, 1970.

Katona, George. Psychological Economics. New York: Elsevier, 1975.

Marshall, Alfred. Principles of Economics, 8th ed. London: MacMillan Press, 1920.

Mosteller, Frederick, and Tukey, John W. Data Analysis and Regression. Reading, Massachusetts: Addison-Wesley Publishing, 1977.

Popper, Karl R. Logic of Scientific Discovery. New York: Harper & Row Publishers, 1965.

Romeder, J.M. Méthodes et Programmes d'Analyse Discriminante. Paris: Dunod, 1973.

Savage, L.J. Foundations of Statistics. New York: John Wiley and Sons, 1954.

Theil, Henri. Economics and Information Theory. Chicago: Rand McNally, 1967.

Tukey, John W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Publishing, 1977.

VonNeumann, J., and Morgenstern, O. The Theory of Games and Economic Behavior. Princeton University Press, 1944.

Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York: John Wiley & Sons, 1971.