A LOGIT ANALYSIS OF THE EFFECT
OF RELOCATION ON JOB-QUIT PROBABILITY

Thesis

Submitted by

SEIF E. IBRAHIM TAG-EL-DIN, M.Sc. (Glasgow)

for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics
University of Edinburgh

MARCH, 1980.
I CERTIFY THAT THIS THESIS IS MY OWN COMPOSITION

.................................................................

.................................................................

(Seif E. I. Tag el Din)
A LOGIT ANALYSIS OF THE EFFECT
OF RELOCATION ON JOB-QUIT PROBABILITY

ABSTRACT

Binary logit analysis is adopted to estimate job-quit probability in terms of explanatory variables representing the effect of housing relocation and other characteristics of workers. We utilize a cross-sectional sample of households rehoused by the City Council of Glasgow from condemned properties due to re-development schemes.

The logistic model is adopted for its computational convenience, although a behavioural justification is attempted through the random utility choice theory. The simultaneous logistic model is proposed to allow for the 'within-family' dependence and its properties explored with comparison to a simultaneous linear probability model. Deriving a simple test we support the complementarity hypothesis of husband/wife labour force participation responses.

While exploiting the likelihood estimation techniques we introduce a goodness-of-fit measure which is approximately F-distributed. The most interesting variables turn out to be: the pre-move weekly work hours, change in travel time, change in housing costs, age, skill, and availability of employers in the new area. Policy implications of these findings are outlined.

Using the $S_B$ curve we have established the U-shaped property of the married females' quit probabilities - implying concentration near zero and one probability. We argue that the $S_B$ curve is a useful tool for turnover analysis compared to the survival curve of the demographic approach. On the basis of the $S_B$ curve we develop and test an arc-elasticity formula for average (quit) probability which is easy to compute.

Aggregation bias due to neglect of individuals' heterogeneity in
cross-sectional data is analyzed for the logistic model. Criteria are developed for the variance-elasticity measure of average response probability, and the appropriate formula derived.
I wish to express my gratitude to Dr. Brian Main for the careful and sincere supervision, and the useful advice he has kindly offered me throughout the study. I am equally grateful to Professor J. Wolfe for the genuine concern and encouragement he has kindly expressed towards my work. In addition, I am truly obliged and indebted to Dr. Steve Engelman for giving me access to his data, to extend his previous study. I am thankful to the Edinburgh Regional Computing Centre (E.R.C.C.) for their provision of efficient computing services which have, in fact, been indispensable for the completion of this study.

Moreover, I wish to thank Mr. A.D. McLaren for the useful discussion and advice that he has kindly made possible, and similar thanks are due to Mr. P. Fisk. Regarding the necessary services of correspondence, I thank Mrs. Findlay for her kind readiness and skill; and I thank Mrs. R.W. Chester for typing and organizing this thesis.

I am most obliged to the Administration of Khartoum University for its sponsorship and financial support. And with genuine appreciation, I acknowledge the positive co-operation of my wife, Aysha Abdel-Salam, which she has shown in kind help and faithful support throughout the course of this work.
## CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER I</th>
<th>INTRODUCTION</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>On the Rehousing Problem</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>The Glasgow Labour Market: Some Relevant Observations</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Studies Related to the Relocation Problem</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>The Nature of the Present Study</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>The Nature of the Sample</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER II</th>
<th>MODELS OF QUIT BEHAVIOUR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>28</td>
</tr>
<tr>
<td>2.2</td>
<td>How Voluntary Are Quits</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>The Micro-level Approach for Quits</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>Macro-level Models of Quits</td>
<td>44</td>
</tr>
<tr>
<td>2.4.1</td>
<td>The Role of Institutional Rules and Effect of Unionism</td>
<td>46</td>
</tr>
<tr>
<td>2.4.2</td>
<td>The Role of Specific Human Capital</td>
<td>47</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Job Search Theory and Quits</td>
<td>48</td>
</tr>
<tr>
<td>2.4.4</td>
<td>The Job Hazards Logistic Model of Quit Intentions and Related Comments</td>
<td>51</td>
</tr>
<tr>
<td>2.5</td>
<td>Some Relevant Implications of the Beta Logistic Model</td>
<td>54</td>
</tr>
<tr>
<td>2.6</td>
<td>Job-Satisfaction and Quit Intention</td>
<td>59</td>
</tr>
<tr>
<td>2.7</td>
<td>Individual Preferences, Random Utility and the Logistic Model</td>
<td>66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER III</th>
<th>ASPECTS OF PARAMETER ESTIMATION AND INference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>79</td>
</tr>
<tr>
<td>3.2</td>
<td>Other Probability Models</td>
<td>80</td>
</tr>
<tr>
<td>3.3</td>
<td>Estimation Methods for Non-linear Binary Probability Models</td>
<td>86</td>
</tr>
</tbody>
</table>
# CONTENTS (CONTD.)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Significance Testing and Measurement of Fit</td>
<td>93</td>
</tr>
<tr>
<td>3.5</td>
<td>Other Measures of Goodness of Fit</td>
<td>96</td>
</tr>
<tr>
<td>3.6</td>
<td>The Nature of our Computational Routine</td>
<td>105</td>
</tr>
<tr>
<td>3.7</td>
<td>A Numerical Example</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER IV</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>A SIMULTANEOUS LOGISTIC APPROACH FOR THE ANALYSIS OF</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>QUANTAL CHOICE AMONG HOUSEHOLD MEMBERS</strong></td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>113</td>
</tr>
<tr>
<td>4.2</td>
<td>A Model of Simultaneous Quantal Choice</td>
<td>126</td>
</tr>
<tr>
<td>4.3</td>
<td>The Maximum Likelihood Approach</td>
<td>132</td>
</tr>
<tr>
<td>4.4</td>
<td>Simultaneous Logit and the Log-linear Probability Model</td>
<td>136</td>
</tr>
<tr>
<td>4.5</td>
<td>Significance Testing for $\psi$, the Measure of Interdependence.</td>
<td>141</td>
</tr>
<tr>
<td>4.6</td>
<td>Summary</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER V</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>THE PROBLEM OF AGGREGATE BINARY PREDICTIONS AND</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>ELASTICITIES</strong></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>151</td>
</tr>
<tr>
<td>5.2</td>
<td>Quit Propensity and Policy Control Parameters</td>
<td>155</td>
</tr>
<tr>
<td>5.3</td>
<td>Aggregation Bias and Effect of Heterogeneity in a Binary Logistic Model</td>
<td>162</td>
</tr>
<tr>
<td>5.4</td>
<td>Properties of the Probability Distribution of Quit Propensities</td>
<td>177</td>
</tr>
<tr>
<td>5.5</td>
<td>Point Elasticities of Average Quit Propensity in terms of Employees' Characteristics</td>
<td>196</td>
</tr>
<tr>
<td>5.6</td>
<td>Arc Elasticity and Approximate Prediction Formulae</td>
<td>203</td>
</tr>
<tr>
<td>5.7</td>
<td>Development of the Variance Elasticity Formula</td>
<td>212</td>
</tr>
<tr>
<td>Page No.</td>
<td>Contents</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>219</td>
<td>5.8 One Further Disadvantage of the Linear Probability Model</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>5.9 Summary</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>CHAPTER VI EMPIRICAL MODELS FOR MALES AND FEMALES</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>6.1 Introduction</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>6.2 Definition of Explanatory Variables</td>
<td></td>
</tr>
<tr>
<td>246</td>
<td>6.3 Commentary on the Empirical Models for Males</td>
<td></td>
</tr>
<tr>
<td>265</td>
<td>6.4 Geometrical Representation for the Effect of the Hours Dummy for Males</td>
<td></td>
</tr>
<tr>
<td>267</td>
<td>6.5 Commentary of the Empirical Models for Females</td>
<td></td>
</tr>
<tr>
<td>274</td>
<td>6.6 Geometrical Representation of the Effect of the Hours Variable for the Female Spouse</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>6.6.1 A Model for Female Actual Movers Alone</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>6.7 Summary</td>
<td></td>
</tr>
<tr>
<td>282</td>
<td>CHAPTER VII A GENERAL REVIEW</td>
<td></td>
</tr>
<tr>
<td>282</td>
<td>7.1 Introduction</td>
<td></td>
</tr>
<tr>
<td>283</td>
<td>7.2 Choice of Probability Model</td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>7.3 The Household Effect Problem</td>
<td></td>
</tr>
<tr>
<td>285</td>
<td>7.4 Empirical Finding and Implications</td>
<td></td>
</tr>
<tr>
<td>289</td>
<td>7.5 The Methodology of Turnover Analysis and Its Policy-Orientaton</td>
<td></td>
</tr>
<tr>
<td>291</td>
<td>7.6 The Weekly Work-Hours Dummy and the U-shaped Property for Females' Quit-Probabilities</td>
<td></td>
</tr>
<tr>
<td>292</td>
<td>7.7 Aggregate Bias and the Variance-Elasticity Formula of Binary Response Prediction</td>
<td></td>
</tr>
<tr>
<td>293</td>
<td>7.8 And Finally</td>
<td></td>
</tr>
<tr>
<td>294</td>
<td>APPENDIX (A) Mathematical Notes</td>
<td></td>
</tr>
<tr>
<td>302</td>
<td>APPENDIX (B) Computing Fortran Programmes</td>
<td></td>
</tr>
<tr>
<td>317</td>
<td>BIBLIOGRAPHY</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

This thesis handles the estimation problem of job-quit probability as a function of certain explanatory variables representing the effect of housing relocation and other characters of a sample of workers. While dealing with this specific problem we handle the more general and technical treatment of job-quit behaviour as far as it constitutes an economic problem. This introductory chapter is intended to outline the nature of this thesis of which the data base relates to the experience of Glasgow. As an appropriate background we shall consider in some detail those studies which have immediately started off our thinking about this problem, e.g. Engelman(49) and Mackay(92). The other studies which are of a more general nature are discussed in the next chapter.

Hence, in the first section below we introduce the Glasgow rehousing problem and the labour-market questions that it has stimulated. In section (1.2) we look at some special features of the Glasgow labour market that have emerged out of past experience. Then we move to section (1.3) where we consider the two main empirical studies (Kasper(79) and Engelman(49) which have handled the incidence of relocation on the labour market experience of the Glaswegian workers. Having provided the above background, we then outline the main features and objectives of our study in section (1.4), while briefly describing the data base in the last section.

(1.1) On the Rehousing Problem

The incidence of extensive housing redevelopment schemes on the local labour market experience of the working class families is a problem
that has warranted special attention of some economists in the last decade. As it has been observed by Cullingworth (37, p. 74), these massive schemes, which normally involve the relocation of working class families as currently implemented by the Glasgow Housing Corporation, may run at variance with the social welfare of the relocated families. Yet, in the first place these re-development programmes have been intended to promote the standard of housing and other public utilities in the city. The housing picture of Glasgow, compared to the general Scottish picture, was seen to be characterized by a high proportion of small dwellings of which the greater number were tenements lacking in some essential utilities; e.g. fixed bath, shower or hot water supply. On the whole, it was judged to be below the national standard; see Cullingworth (36). Thus†, between 1956 and 1978 the council has issued 72,888 orders for closing or demolishing houses. The total number of houses condemned over the period 1974-1978 amounted to 10,363— an average of 2,073 houses per year. Moreover, it has been estimated (by the Glasgow Corporation Housing Management) that the program for 1979-1982 would call for the doubling of the average rate to 4,688 per year. However, the criteria for deciding on the suitability or otherwise of a given house is not without differences of opinion. As an extreme example Cullingworth (36) has contrasted, in his rather controversial book, the estimate of 41% given by a housing survey with the estimate of 3.4% given by the official return of the Secretary of State for houses below the standard of decent habitation. On the other hand, it is not only housing standards which have necessitated demolition of houses, but also their placement in the way of a public project (e.g. motor-way, public park, etc.). For these reasons it has been feared that any potential gain in social

† See the Glasgow Corporation, Housing Department, Report 1978.
welfare which these programs may achieve could be partially or totally offset by adverse changes in the social welfare of the rehoused families (Cullingworth(37), Kasper(79), Engelman(49)).

However, the empirical assessment of the welfare effects, which should be rather intricate and subtle as well as costly, is beyond the scope of this study and it appears to be a still remaining potential for future research. One attempt in this direction is a research proposal adopted by Wright(150) in relation to a similar experience in Edinburgh. Alternatively, the more specific question related to the incidence of housing dislocation on one or more aspects of the labour market experience of families has aroused the interests of Kasper(79) and Engelman(49), to whom we shall shortly refer. Relocation naturally involves the departure of a given family† from a residential area or an environment which called for certain past decisions to be made by family members (e.g. as regards employment status, place of work, housing tenure, its size and locality, domestic activities, etc.) to a new environment where some of these past choices have to be revised jointly by the family members. As regards the interdependence of choice of residential location and employment there is some evidence from past experience that it is noticeable especially among working class families. Generally, it is held that the principle governing choice of residential location versus employment location is that individuals tend to minimise the sum of transportation and housing costs; e.g. see Oi(110) and Daniel(38). The latter author has formalized this idea as a simultaneous choice problem for an individual with a given skill level seeking residence and a place of employment. When special income groups have been considered Lowry(91) has given evidence from the U.S.A. that lower income families tend to

† Throughout this study 'family' and 'household' are used interchangeably.
cluster around employment opportunities more closely than higher income families. Similarly Orr\(^{(111)}\) has provided evidence supporting the twin hypotheses that (i) residential location of low income households is more sensitive to employment opportunities than that of higher incomes, and that (ii) residential location of higher income groups is more sensitive to service quality and taxation. Examination of the Glasgow labour market by Mackay et al.\(^{(92)}\) has shown that the journey-to-work distance is indeed an important determinant of the geographical mobility of labour, and that labour mobility is not usually associated with residential mobility.

Hence, on a priori ground it is to be expected that residential relocation policies which do not consider employment choices of working class families could put them into a disequilibrium involving revision of past employment choices and other adjustments in their range of time and budget allocations. But before we consider the nature of the previous studies in the rehousing problem of Glasgow, it is appropriate to give a broad background of the local labour market of Glasgow as examined on the basis of past experience.

(1.2) The Glasgow Labour Market: Some Relevant Observations

The essential features of the local labour market of Glasgow has been examined by Mackay et al.\(^{(92)}\) in comparison with other local labour markets: Birmingham, North Lanarkshire and some new towns. This study has been based on a sample of case-study plants over the period 1959-1966. The latter period has been chosen as it virtually covers two cycles in each labour market area. On this basis it has been possible to contrast behaviour between areas and within areas as labour market conditions changed.

The local labour market of Glasgow has been characterized in this
study as having unemployment rates substantially in excess of the British average throughout the two cycles. The authors have paid special attention to the study of labour turnover analysis in these towns with emphasis on quit-rates. In fact they have argued that the plant quarterly quit rate offers the best measure of turnover which is responsive to market forces, market conditions and personal characteristics of employees, when compared to other measures.

"... we conclude, then, that the crucial test for the labour market theory is the behaviour of voluntary quit"

Mackay et al. (92, p. 143).

In addition the quit rate has been shown to constitute about 70% of all separations for the case-study plants of Glasgow and Birmingham; see details of Table (1.1) below.

However, one consequence of the 'slackness' property of the Glasgow market has been shown by the relatively small degree displayed by its average \(^+\) plant quarterly quit rates as compared to the 'tight' labour market of Birmingham; see Table (1.2). This has been explained on the basis that dispersion of plant quit rates will widen as unemployment falls and narrow as unemployment rises. For Glasgow a negative correlation \((= -0.54)\) has been observed between standard deviations of plant quit rates for each quarter and the quarterly unemployment rates. In all types of market considered the quarterly quit rates of plants were found to be negatively correlated with the level of quarterly unemployment rates, and positively correlated with the number of unfilled vacancies - using a linear model.

\(^+\) The average for each plant is taken by averaging out the quarterly quit rates over the 32 quarters of the 8 year period.
Another interesting property relates to the role of wage-structure and the effect of earning-differentials in slack labour markets as compared to tight labour markets. In the former case, which is typified by Glasgow, the level of wage earnings of plants has been shown to have a systematic and significant negative effect on plants quarterly quit rates, whereas no such evidence has been found in the case of Birmingham. Thus, it has been concluded that in the Glasgow market wage differentials are real - not nullified by non-pecuniary benefits - though they do not explain more than half the observed dispersion in plant quit rates. On the other hand, as the labour market tightens, quits are increasingly composed of employees who change employers frequently and who appear to be little influenced by earnings differentials. Alternatively, for the tight labour market of Birmingham, the recruitment-rate has had a positive significant effect on plant quit-rates whereas no similar evidence was found for Glasgow. It has, then, been concluded that in tight labour markets a high recruitment-rate leads to a high quit-rate and vice versa. Specifically, the observation that plant quit rates are decreasing functions of the level of plant earnings has been taken to imply that the labour market is in a state of disequilibrium.

The contrast of slack and tight labour markets has also been reflected in the effect of skill on quit-rates. If we look at Table (1.3) it can be seen that the average quarterly market quit-rates tend to decrease systematically as the skill level is increased in the case of Birmingham but this pattern is less systematic in the case of Glasgow. This point has been demonstrated by constructing three frequency distribution histograms of plant average quarterly quit-rates as shown below in Figure (1.1) for the skilled, semi-skilled and unskilled workers. It is clearly seen in this figure that the skill level introduces no
significant variation in the distributions of quit-rates in the case of Glasgow, whereas in the case of Birmingham there is marked change in the dispersion of average quarterly quit-rates of plants shown by the histograms, i.e. as the skill level increases the frequency distribution of quit-rates becomes more bunched within the range 1-8 per cent per quarter. However, due to the observation that the market quit-rate in Birmingham has been higher than in Glasgow (see Table (1.1)), we may also notice that the degree of concentration (or bunching) at the lower end of the distribution for the skilled workers' quit-rates is greater in the case of Glasgow than in the case of Birmingham. This point has similarly been explained on the grounds that in tight labour markets employers are more likely to make special efforts to retain skilled employees as quit

† We may also look at Table (1.3) for similar evidence.
incentive is relatively strong and they are difficult to replace.

As regards the effect of job-tenure and age of employees, these variables have been observed to be negatively correlated with quarterly plant quit-rates in Glasgow labour market as well as the other local

Table (1.1)

Market Quarterly Quit Rates, Separation Rates and Quit Percentages

<table>
<thead>
<tr>
<th>area Sex</th>
<th>Glasgow</th>
<th>Birmingham</th>
<th>New Lanarkshire</th>
<th>Small Town</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALES</td>
<td>4.2</td>
<td>5.5</td>
<td>1.9</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.6</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>(70.0)</td>
<td>(72.3)</td>
<td>(70.4)</td>
<td>(60.9)</td>
</tr>
<tr>
<td>FEMALES</td>
<td>5.9</td>
<td>9.0</td>
<td>4.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>8.4</td>
<td>12.9</td>
<td>6.5</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>(70.2)</td>
<td>(69.9)</td>
<td>(64.6)</td>
<td>(36.3)</td>
</tr>
</tbody>
</table>

N.B.: 1st row gives quit-rates and 2nd row gives separation-rates. The parentheses give quits as a percentage of all separations.

Source: arranged from Mackey et al.

labour markets of the case-study plants.

The role of the travel-to-work distance in the geographical mobility of workers is particularly interesting. The evidence which has been drawn from the past experience of the Glasgow labour market, as well as the other markets, supports the general trend that most manual workers seek work close to their homes; e.g. see Orr (111) and Lowry (91). As can be
Table (1.2)

Dispersion of Average Quarterly Plant Quit-Rates

<table>
<thead>
<tr>
<th>Average Plant Quit Rate</th>
<th>Glasgow Males</th>
<th>Birmingham Males</th>
<th>Birmingham Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>8.4</td>
<td>13.9</td>
<td>21.1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.4</td>
<td>3.6</td>
<td>5.0</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>58.2</td>
<td>65.5</td>
<td>45.3</td>
</tr>
</tbody>
</table>

Source: Mackay et al.

Table (1.3)

Average Quarterly Quit-Rates by Level of Skill

(Percentages)

<table>
<thead>
<tr>
<th>Area</th>
<th>Level of Skill</th>
<th>Glasgow</th>
<th>Birmingham</th>
<th>New Lanarkshire</th>
<th>Small Town</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skilled</td>
<td>3.8</td>
<td>3.1</td>
<td>2.4</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Semi-skilled</td>
<td>5.1</td>
<td>6.6</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Unskilled</td>
<td>4.1</td>
<td>8.0</td>
<td>1.7</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Source: Mackay et al.

seen from Table (1.4), most job changes that took place between employers fall within a 5 m. radius (72.1% and 78.8% for Glasgow males and females respectively; 76.5% and 87.2% for Birmingham males and females respectively). If we look at Table (1.5) which relates to distance of travel-to-work journey, we find that the percentages are even bigger of those whose distance of travel-to-work journey is within a radius of five miles. Moreover, as can be seen from Table (1.6) the distance variable is particularly restrictive to the mobility of less skilled people who tend
to choose a shorter distance to work than the more skilled. More generally, it appears that most voluntary job changes were not associated with residential mobility, using the evidence of Table (1.4) below. Hence, the limits of job-search, it is argued, are set by costs associated with long travel-to-work journeys and the high costs of residential mobility. The importance of these factors has led Mackay et al. (92, p.249) to view the conurbations of Glasgow and Birmingham as

"... a series of loosely connected sub-markets with weaker links to the other labour-market areas."

Table (1.4)
Distance of Recruits' Previous Employer from Case-Study Plants

(Percentages)

<table>
<thead>
<tr>
<th>Distance Range (miles)</th>
<th>Birmingham</th>
<th>Glasgow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>Under 2</td>
<td>28.7</td>
<td>55.1</td>
</tr>
<tr>
<td>2 - 5</td>
<td>47.8</td>
<td>32.1</td>
</tr>
<tr>
<td>5 - 20</td>
<td>17.2</td>
<td>10.6</td>
</tr>
<tr>
<td>20+</td>
<td>6.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Total numbers</td>
<td>2,509</td>
<td>577</td>
</tr>
</tbody>
</table>

Source: Mackay et al.
Table (1.5)
Distance of Travel to Work Journeys (Percentages)

<table>
<thead>
<tr>
<th>Distance Range (miles)</th>
<th>Birmingham Males</th>
<th>Birmingham Females</th>
<th>Glasgow Males</th>
<th>Glasgow Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 2</td>
<td>42.1</td>
<td>68.5</td>
<td>42.0</td>
<td>62.0</td>
</tr>
<tr>
<td>2 - 5</td>
<td>42.9</td>
<td>25.6</td>
<td>39.0</td>
<td>26.0</td>
</tr>
<tr>
<td>5 - 20</td>
<td>14.3</td>
<td>5.6</td>
<td>17.7</td>
<td>10.6</td>
</tr>
<tr>
<td>20+</td>
<td>0.7</td>
<td>0.3</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Total numbers</td>
<td>2,306</td>
<td>746</td>
<td>2,663</td>
<td>283</td>
</tr>
</tbody>
</table>

Source: Mackay et al.

Table (1.6)
Distance of Travel-to-work Journeys by Level of Skill

<table>
<thead>
<tr>
<th>Distance Range (miles)</th>
<th>Birmingham Semi-Skilled</th>
<th>Birmingham Skilled</th>
<th>Glasgow Semi-Skilled</th>
<th>Glasgow Skilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 2</td>
<td>33.5</td>
<td>42.6</td>
<td>53.9</td>
<td>37.8</td>
</tr>
<tr>
<td>2 - 5</td>
<td>47.6</td>
<td>44.4</td>
<td>38.0</td>
<td>41.6</td>
</tr>
<tr>
<td>2 - 20</td>
<td>18.4</td>
<td>13.0</td>
<td>7.4</td>
<td>20.5</td>
</tr>
<tr>
<td>20+</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Total numbers</td>
<td>1.078</td>
<td>520</td>
<td>708</td>
<td>1.391</td>
</tr>
</tbody>
</table>

Source: Mackay et al.
Thus, the elasticity of labour supply to a plant will depend to a considerable extent on the existing population within a fairly short travel-to-work journey.

As regards the industrial pattern of recruitment the biggest single source of male labour in both Glasgow and Birmingham is the engineering industry – from which the case-study plants were drawn. Over the past period of investigation the structure of the internal labour market of Glasgow has been characterized as having the least mobility within plants as compared to the other case-study markets. The majority of those recruited as skilled workers have been in that occupational group with their previous plants (more than 80%). The depressed state of the labour market of Glasgow, it is argued, has accounted for the ability of Glasgow plants to meet most of their skilled requirements by recruiting labour already possessing the necessary training and experience.

These are some interesting findings which have emerged from past experience. Now, we may turn to consider the special studies which handled the relocation problem and its incidence on the working class families.

(1.3) Studies Related to the Relocation Problem

Kasper (79) addressed himself to the problem of assessing the employment costs imposed on forced dislocated families as opposed to a control group of voluntary movers. His data base consisted of a group of Glaswegian families who were forced to move at the end of 1970 and another group of voluntary movers. Employment costs have been assumed to reflect in (i) earnings, (ii) employment status, (iii) labour force participation and (iv) the journey-to-work costs. According to Kasper (79) relocation has imposed employment costs for
both forced and voluntary movers but that costs were significantly higher for the former group. His main findings indicate that:

a) The burden of unemployment fell more heavily on forced movers.
b) Relocation raised the journey-to-work costs of the forced relative to the voluntary movers.
c) The labour force participation of forced movers relatively increased.

These findings have been based on the calculation and comparison of simple ratios and percentages intended to evaluate loss of utility due to the disequilibrating effects of the move. One of the limitations of this study is that it has been based on the assumption of a single labour market for the Glasgow market, which, as we have seen in Mackay et al. (92), does not hold for the conurbations of Glasgow due to the high costs of residential mobility and long travel-to-work journeys.

Kasper (79) based his 'single market' hypothesis on a statistic-test which he declared to be rather crude. However, as this restrictive assumption has not been basic for the main findings, the study still provides an interesting insight into the problem.

The second study, and the one which is most directly related to our study, is due to Engelman (49) who has specified a linear probability model to estimate the effect of the move and other related variables on job-quits. The data base of this study is composed of household members who belong to two categories: those who have actually moved to their allocated council houses, called 'movers', and those who have been allocated council houses but have not yet moved, called 'non-movers'. The latter group which possesses similar characteristics to the movers has been used as a control group to isolate the effect of the 'move' per se. As the model related to quits the left-hand variable, Y, is a 0,1 dichotomous variable which equals one if a quit has occurred and
zero otherwise. The set of independent explanatory variables of the model are: age class (A); earnings level (E); skill (S); the after-move family earnings less own wage (F); whether or not the individual is a non-mover (M); whether or not travel-to-work time has decreased (D); direction of change in rent and rates (RR); presence of children below five years (CH) and number of employment opportunities within a radius of 30 minutes from new house for males, or 20 minutes for females (EMP). Then dealing with those who have been employed before the move, and eliminating those who have been sacked or laid off, a linear probability model has been fitted to estimate quit probabilities separately for males and females using OLS method. For each sex group two models are assumed, one for combined movers and non-movers and the other for movers alone. For example, the first model of quit probabilities for the combined males is written as:

\[
( - ) \quad ( - ) \quad (+) \quad ( - ) \quad ( - ) \\
Q = \text{prob}(Y=1) = \beta_0 + \beta_1 A + \beta_2 S + \beta_4 F + \beta_5 M + \beta_6 D \quad (1.1)
\]

where the subscripts refer to the expected signs of the variables' parameter coefficients. All explanatory variables are expressed as 0,1 dichotomous variables except the employment opportunities variable (EMP). Specifically they are expressed as: \( A = 1 \) for age \( > 45 \) years and zero elsewhere; \( E = 1 \) for weekly earnings before the move \( \geq \) £35.50 (the sample median) and zero elsewhere; \( F = 1 \) if earnings of other family members \( \geq 17.50 \) (sample median) and zero elsewhere; \( M = 1 \) for non-movers, and zero elsewhere; \( D = 1 \) for movers who moved closer to their place of work, and zero elsewhere; \( RR = 1 \) if change in rent and rates after the move \( \geq 10.00 \) (sample median), and zero otherwise; \( CH = 1 \) if children under five years of age are present, and zero otherwise. The employment opportunities variable (EMP) and the rent and rates dummy (RR) have been applied only to the model of movers, while the children dummy (CH)
is applied to females alone. The estimated coefficients and their associated t-values appear to support \textbf{a priori} expectations at reasonable significance levels, as can be seen from Table (1.7) above, except the employment opportunities variable (EMP) and the earnings dummy (E) in all models. As for the female models the age dummies ($A_1$ and $A_2$), skill (S), family earnings (F) and the closer-or-further dummy (D) yield rather poor results. It is also noticeable that the employment opportunities variable (EMP) gives negative coefficients which is counter-intuitive, since we expected that greater number of employment opportunities in the new area should be an incentive rather than a disincentive to quit old job. (This point shall be considered among other things in the sixth chapter). The same remarks apply to the positive sign of the closer-or-further dummy (D) of the females models, since it is more plausible \textbf{a priori} that a movement closer to one's job should reduce rather than increase the probability of quitting. However, as both coefficients of (EMP) and (D) are insignificant there is little bother about the sign.

The interesting conclusion of this model is that when the closer-or-further dummy (D) is introduced it implies that those who move closer to their jobs are less likely to quit than non-movers - although the model yields a significant negative coefficient on the quit probability for the non-move dummy (M). Thus, the decisive factor for males has been whether or not the individual has moved closer to his job. However, at this stage there are certain critical points to make with respect to this model.

(1) The use of the linear probability model is subject to some crucial limitations. Basically it cannot guarantee the $|0,1|$ restriction of the probability interval, and that the OLS estimates could be inefficient due to a problem of heteroscedasticity. As we shall
<table>
<thead>
<tr>
<th>Variable</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Number of Observations</th>
<th>194</th>
<th>147</th>
<th>115</th>
<th>96</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient Estimator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>Movers Alone</td>
<td>Combined</td>
<td>Movers Alone</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-0.314</td>
<td>-0.235</td>
<td>A₁</td>
<td>0.018</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.02)</td>
<td>(2.47)</td>
<td></td>
<td>(0.23)</td>
<td>(0.43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.006</td>
<td>-0.043</td>
<td>A₂</td>
<td>-0.111</td>
<td>-0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.56)</td>
<td></td>
<td>(1.01)</td>
<td>(0.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>-0.098</td>
<td>-0.097</td>
<td>E</td>
<td>-0.08</td>
<td>-0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.51)</td>
<td>(1.33)</td>
<td></td>
<td>(0.73)</td>
<td>(0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.208</td>
<td>0.251</td>
<td>S</td>
<td>-0.055</td>
<td>-0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.25)</td>
<td>(3.35)</td>
<td></td>
<td>(0.55)</td>
<td>(0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>-0.192</td>
<td>n.a.</td>
<td>F</td>
<td>0.031</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.31)</td>
<td></td>
<td></td>
<td>(0.36)</td>
<td>(1.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-0.244</td>
<td>-0.259</td>
<td>M</td>
<td>-0.144</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.35)</td>
<td>(3.60)</td>
<td></td>
<td>(1.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP</td>
<td>n.a</td>
<td>-0.061</td>
<td>D</td>
<td>0.031</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.23)</td>
<td></td>
<td>(0.36)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR</td>
<td>n.a</td>
<td>0.111</td>
<td>EMP</td>
<td>n.a</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.361)</td>
<td></td>
<td></td>
<td>(0.51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>n.a</td>
<td>n.a</td>
<td>CH</td>
<td>0.328</td>
<td>0.257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.15)</td>
<td>(2.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers in parentheses represent t-values.
† For females A₁ = 1 for age class (16-24) and zero otherwise; A₂ = 1 for age class (25-44) and zero otherwise.
consider in Chapter III of this study, past experience has shown that even if the linear specification is conceptually valid the estimates are sensitive to prediction bias and specification error. On the other hand, the model is not useful for making aggregate predictions (i.e. average quit probability) since its predictions are totally insensitive to variations in the variance-covariance structure of the explanatory variables (end of Chapter V). Moreover, although the estimation method of the linear probability model is simple and straightforward, yet it does not signal the presence of certain peculiarities in the data (Nerlove and Press (106); also see Chapter VI).

(2) Secondly, the effect of job-tenure has not been allowed for, though it is generally established as an important determinant of quit behaviour; the longer the job-tenure of an employee, the smaller would his quit probability be. As we have mentioned above (section (1.3)) such an effect has been observed by Mackay et al. (92) on the basis of plant quarterly quit rates of Glasgow as well as the other case-study local labour markets. In the next chapter we shall show that job-tenure (or service-length) does in fact occupy an important position in turnover analysis. As job-tenure (see Chapter II, section (2.3)) we shall use a proxy which does have the same negative effect as job tenure.

(3) Thirdly, the data base is a cluster sample with household units as clusters and not a simple random sample of independent individuals. Members of the same household (head, female spouse, son, parent, daughter..) are expected on conceptual grounds to work out their time and budget allocations jointly and interdependently rather than singly. Hence unless there is reason to believe that the "within family" dependence is zero, the treatment of such households'
data as if it were a simple random sample may result in estimation bias. This problem has been considered by Cohen (32), and Altham (5) who analyzed the statistical bias which arises in the treatment of a cluster random sample of families as though it were a simple random sample of independent individuals. They have been particularly interested to solve the problem in the context of multivariate categorical data where the Karl Pearson’s $\chi^2$-statistic is used to test the deviation between observed cell frequencies and expected cell frequencies under different restrictive hypotheses. They have shown that when the within-family dependence is measured by the constant $a$ then the adjusted K. Pearson’s $\chi^2$-statistic is expressible as:

$$\hat{\chi}^2 = \frac{1}{1 + a(K-1)} \chi^2, \quad K = 2, 3, \ldots$$

where $K$ is the family-size. Hence, the unadjusted measure will be biased whenever the within-family dependence is non-zero ($a \neq 0$).

The obvious question to ask is about a similar method of bias-analysis in the case of other statistical models (i.e. linear models, non-linear models). Yet, to our best knowledge this effort has not been paralleled with studies which answer this question when the within-family dependence is non-zero. This could be handled as a future research proposal, but the analogy with Cohen (32) and Altham (5) suggest that such a bias should arise as long as the within-family dependence is not zero.

(4) Finally, although the process of dichotomizing explanatory variables is not uncommon in applied econometrics, yet when the original real-valued variables are available (e.g. age, travel-time, rent and rates) then this process involves sacrifice of otherwise useful statistical information; Nerlove and Press (106, p. 13). It should be more beneficial
to use the whole of the available data (possibly after one-to-one transformations) rather than part of it whenever possible so as to avoid the arbitrariness of grouping and loss of information.

We shall now move to comment on the nature of our study.

(1.4) The Nature of the Present Study

This study has originally been stimulated by the desire to extend and improve on Engelman's (49) — who has kindly offered his data to us — with reference to the above-mentioned four points (i.e. choice of a suitable probability model, allowance for the degree of job commitment and the within-family effect, using the original real-valued data whenever available). Besides allowing for these points and replacing the dummies by original variables (i.e. change in travel time ... etc.) we also aim at testing the effects of other variables like education level, training and housing tenure.

At the same stroke we aim at the more general purpose of improving on the current methods of turnover analysis with special reference to quit behaviour. The main theme of our proposed method is the utilization of a disaggregate model of quit behaviour to arrive at aggregate predictions — or to predict the potential effects on average quit probability due to different possible changes in its explanatory variables. The basic principle has been adopted by Mackay et al. (92, p. 172) when they questioned how the "market quit-rate echoed micro-economic behaviour". They have demonstrated this point in relation to the effect of skill levels on average quit probabilities by comparing ordinary frequency distribution histograms of the plants' average quarterly quit rates under different skill levels; see Figure (1.1) above. By analogy we are asking how average quit probability of a given organization echoes
individuals' behaviour - except that in our case an individual's 'average' quit probability is estimated via a (logistic) model and explained in terms of a set of explanatory variables. Thus, the shape of the frequency histogram of employees' quit probabilities is determined by these explanatory variables. We develop this idea further by fitting to these histograms a smooth curve which is sufficiently flexible and responsive to changes in the explanatory variables, so that it can serve a comparable role as the employee survival curve or stability curve of the demographic-oriented turnover models. The technical details of this method are discussed in Chapter V.

The logistic model has been chosen by us for its computational convenience, although one way in which it can be derived, on the basis of random utility theory of choice, is questioned (Chapter II, section (2.7)). The logistic model automatically satisfies the \([0,1]\) restriction of probability and is not sensitive to the problems of prediction bias, heteroscedasticity or specification error. At the same time, the logistic model is shown to be essential for our technical methodology outlined above.

As regards the job-tenure variable it is usually expressed in number of years, quarters of a year, or months - of work to a given employer. However, in the absence of such data a reasonable proxy can be adopted. For example Stoilov and Raimon(134) used percentage of brief tenure workers in an industry as a proxy for job tenure, while analyzing the inter-industry quit-rates in U.S.A. manufacturing industries. However lack of tenure data in number of years or months compels us to find a proxy which, like job-tenure, can be used to represent the extent of organizational commitment. We define the 'less committed' employees as those who work less than 30 hours per week (normally called part-time workers) and the 'more committed' employees
as those who work more per week. This division gives a special sense to organizational commitment models which are discussed in the second chapter. As we shall see in the last chapter this proxy is, in fact, a very important determinant of quit behaviour. Thus, the less committed employees would be found to be more sensitive to the relocation policy than the more committed.

The possible existence of the "within-family" dependence has been recognized by Engelman (49; p. 164):

"... while we recognized that the most appropriate analytical framework would probably be the family or household such an approach is beyond the scope of this paper!"

The same point has been made by Bowen and Finegan (27, p. 29) in the study of the economics of labour force participation:

"... as desirable as it is on purely conceptual grounds to treat labour force participation decisions of all members of household as being determined simultaneously, it would have required a degree of technical sophistication we do not possess."

As we shall point out in the third chapter, there is a growing literature on the economics of decision making within the family paralleled with econometric formulations. However, we shall argue that the special quantal choice situation, where family members influence each other in making their choices, has not been adequately formulated. In particular we criticize the simultaneous labour force participation model of husband and wife, which has been specified by Ashenfelter and Heckman (9) in terms of a simultaneous linear probability model. Alternatively, we propose the simultaneous logit model to analyze such a
situation, as discussed in Chapter III. But, the effective utilization of this model for the analysis of quit behaviour poses certain difficulties. In the first place, if we are to formulate the responses of all household members in a single model, the problems of quit behaviour and labour force participation become united. The question of labour quit may apply only to those who have had a job, while the labour force participation response applies to those who have not been employed before. The latter individuals may consider entering the labour force in response to the new arising circumstances, or due to job-quit decisions of some family members. For example, the decision of a male head to terminate his job (e.g. due to a very long journey-to-work) and search for a new one could be encouraged by the decision of his wife to accept, e.g. a part-time job, if she has been previously unemployed, or otherwise not to terminate her current job. A further question arises as we require a 'reference member' of the household, who has been previously employed so that other family members may adjust their labour force participation decisions in response to this reference member's quit behaviour. It is reasonable to choose the head of the household - who is always a male in our data - as the reference member. Clearly, the ultimate model specified should depend on the labour force-statuses of other family members. For example, in that part of the sample where both husband and wife have been previously employed, we shall have a simultaneous quit probability model. However, in that part of the data where the wife is not in the labour force we shall have a simultaneous response model where the quit probability applies to husband, and labour force participation probability applies to wife. Thus, we have seen that for a husband/wife size-two family we need to divide the data into two parts depending on the labour force status of the wife - i.e. we partition the data into two cells. Now
if we continue to treat labour force participation as a dichotomous variable and calculate the number of cells needed to partition the data when the family-size increases, we find that we do in fact need a very large sample for different-size families in order to apply the model of job-quit probability on a family basis. For example, for a size-three family we need $2^2 \times 3 = 12$ cells; and more generally we require $2^{m-1} \times m$ cells to partition the data for size-$m$ families. In each cell a special simultaneous response model can be fitted depending on the labour statuses of other household members.

The above discussion reveals that the formulation of our model by using family members in a single model requires a very large amount of data which is beyond our reach. Even if we focus attention on husband/wife size-two families this requires a fairly large sample, since we need sufficient data in each cell for the purpose of model fitting and hypothesis testing. The distribution of families by household size in our data base is shown below for families where the male head has been employed prior to the move:

Table (1.6)

<table>
<thead>
<tr>
<th>Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>19</td>
<td>150</td>
<td>21</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

However, the other alternative - if we cannot fit a simultaneous response model for all household members in a single model - should not be to assume a zero within-family dependence and base our model on the assumption of independent individuals. The hypothesis of zero
within-family dependence itself is testable within the above discussed data set-up of contingency table cross-classification, and again the data limitation does not permit the independence test.

The simplest resolution would, then, be to derive what can be used as a simple random sample out of these family clusters. As suggested by Fisk\(^{(51)}\) we can take male heads of households as one group, and female spouses as another ignoring other members (e.g. son, daughter, parent). These two groups can be treated as simple random samples as long as the family clusters are chosen randomly.

As for the effect of 'within-family' dependence, we may hope to allow for it through the inclusion of a proxy for the change in labour force participation of other family members. A very reasonable proxy of this effect is to compare the total earnings of other family members before the move and after the move, by taking a difference measure. If the difference is negative it implies that one family member's participation has been dropped. If it is positive it implies that one member has taken a paying job. If it is zero it implies that no change has occurred. The effect of changes in their earnings may not be important due to the short period of time over which observation is made. Hence the difference measure should capture more changes in participation rather than changes in earnings.

\[(1.5) \quad \text{The Nature of the Sample}\]

The data we are utilizing in this study has been elicited through direct interviewing by the aid of the Centre for Sample Surveys, Social and Community Planning Research SCPR, which designed the questionnaire and undertook all the interviewing; see Smith\(^{(130)}\). This data has been supplemented by the calculation of employment within given travel
times (using public transportation) on the basis of data from the Greater
Glasgow Transportation Study, 1968, which divided Glasgow into different
zones (742) and travel time from zone to centre were calculated – for
our original data zones have been aggregated to 60 zones.

The data base consists of two broad groups of families: those who
have moved into council housing between July and October 1973, and those
who have received offers from the Council that specific houses have been
allocated to them but have not yet moved. These two groups are labelled
here actual movers, and intending movers in compliance with the current
terminology of housing mobility; e.g. see Cullingworth(36) . The total
number of families in this data are 336 families of different sizes, or
555 individuals (who have not all been previously employed). The sample
of actual movers was chosen randomly from a master list of rehoused
households provided by Glasgow Corporation Housing Department. A team of
trained interviewers contacted the selected households during April and
May 1974, eliciting information about their pre-move and post-move ex¬
perience of (as much as possible) all male members falling within age-
group (16-65) and all female members falling within age group (16-60).
The information included questions about labour force experience, personal,
family and housing characteristics. The intended movers group have been
chosen as a control group to isolate the effect of the "move". The
criterion for a control group has been that they are similar to actual
movers as being eligible for and actually allocated council housing.
Besides, both groups are similar as being residents in the private
sector. This control group has been selected randomly from the list
of current allocations of the Housing Department. They have been asked
similar questions as those asked to actual movers, pertaining to the
period extending from the summer of 1973 until April 1974, which coin¬
cided with the time that has elapsed since actual movers have moved into
council housing.
Out of this data we are utilizing only those who have been employed prior to the move; specifically 184 male heads of households, and 95 female spouses, ignoring other family members so as to avoid any potential bias due to deviation from simple random sampling (see discussion in the above section). It is possible to provide a break-down to the total sample of the previously employed individual by sex, household status (head, female spouse, other), type of mover and the after-move status as given in Table (1.8) below. It can be noticed in this table that when male heads and female spouses are considered alone, the remaining other family members are quite a few cases. Thus, the potential bias which may be avoided by the exclusion of these cases may not be very substantial. Nevertheless, it is in agreement with sound statistical methods that we base our models on a sample which is _a priori_ an independent simple random one, unless we can formulate the model in a way that can capture the possible interdependences, which we proved we cannot. The post-move statuses are labelled as $E_1 =$ for employed with same employer, $E_2 =$ employed with a different employer, $U =$ unemployed and $N =$ not in the labour force. As we shall see in the sixth chapter all previously employed males or females who have left their previous jobs and who have been full-time employees, have found another employer (over the given period of observation), while those who have been part-time may possibly subsequently occupy either of the states $U$ (unemployed), or $N$ (not in the labour force). This will draw our attention to the special difficulties of _less-committed_ workers as we shall comment in the sixth chapter where we provide fuller discussion of our empirical models.
Table (1.8)
Transition Frequencies of all Previously Employed Individuals

<table>
<thead>
<tr>
<th>Post-move Status</th>
<th>Actual Movers</th>
<th></th>
<th>Intending Movers</th>
<th></th>
<th>All Movers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>All</td>
<td>Head</td>
<td>All</td>
<td>Head</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Spouse</td>
<td>All</td>
<td>Spouse</td>
<td>All</td>
<td>Spouse</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td>89</td>
<td>94</td>
<td>44</td>
<td>58</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td>34</td>
<td>38</td>
<td>10</td>
<td>12</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>14</td>
<td>15</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>7</td>
<td>19</td>
<td>19</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Totals</td>
<td>144</td>
<td>154</td>
<td>79</td>
<td>99</td>
<td>40</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>184</td>
<td>201</td>
<td>95</td>
<td>118</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 2

MODELS OF QUIT BEHAVIOUR

2.1 Introduction

It has been observed that there are costs which can be considerable to employers, arising from the voluntary termination of their labour force; e.g. see Lane and Andrew, Holt, Richardson et al., Bartholomew. In the first place, "quitters" disrupt the production process, and delay delivery dates. In anticipation of such disruption, employers may engage in taking on excess labour. But, this would directly raise costs by increasing the number of recruits dealt with in a given year, thereby raising total training costs, personnel department costs, and orientation costs. Secondly, an employer facing a high quit-rate may feel less concerned about his labour force, and less inclined to contribute to their training costs, which will worsen the stability of his labour force even more, and lower their morale. Although some of the leaving employees may have practised on-the-job search, and thus may not have to undergo an intervening period of unemployment (e.g. about 60% in U.S.A.; see Mattila), yet there are others whose expectations of getting a better job may require a longer duration of unemployment than had been initially expected. A long duration of search, it has been asserted by Gronau, leads to a decline in the 'asking wage' of the job-seeker. Hence, although the leaving employee may have rationally calculated the net expected advantage of his decision, yet he may run the risks of income losses and other psychic costs. However, employees may terminate their jobs for reasons other than moving to better offers. For example, married women
may terminate their jobs due to a rise in their 'shadow price of time' (Heckman(68)); or youngsters may leave to re-join schools (Hall and Kastern(66)); or the leaving could be due to health reasons or avoidance of industrial hazard (Viscusi(144)).

Looked upon from the employers' side, the process of voluntary job-termination is treated as wastage of a productive resource. In this context labour turnover is used synonymously with wastage (Bartholomew(15), p. 241). One of the most widely used measures of turnover is the crude rate defined as

\[
LT = \frac{\text{Number of employees who leave in a given interval}}{\text{Average Number employed during the same interval}} \times 100\%
\]

(2.1)

where a high turnover-rate implies low stability and vice versa. There is a wide variety of models for the leaving process which propose different alternative measures in an attempt to overcome the limitations of measure (2.1), as we shall shortly elaborate. Amongst the recent empirical research intended to assess such wastage is that related to the labour force of The British Steel Corporation, by the Department of Employment Gazette(44). A similar study has been carried out by Richardson et al. (122) to examine the factors which caused labour wastage to the London Employers in the public sector, who have been suffering from labour shortages during 1973-74 (specifically, London Transport, British Road Services, the Metropolitan Police, Department of the Environment). On the other hand, the theoretical modelling is intended to describe the process of job leaving, and prescribe means to control turnover. This literature has grown rapidly since the early fifties. Such research seems to have been stimulated by Rice, Hill and Trist (120), who published several empirical frequency
distributions for the duration of employment at the Glacier Metal Company. Ever since a wide variety of Mathematical models have been proposed by different authors to derive measures of wastage on the basis of different specifications for length of service probability distributions; e.g. see Silcock(128), Lane and Andrew(86), Bartholomew(13, 14), Herbst(71), Bowey(28), Bibby(23), Hyman(76) and Clowes(30). We shall refer to this approach relating to particular economic organizations as the micro-level approach when we shortly consider some of its interesting models.

Parallel to the above approach other research goes in the direction of understanding quit behaviour at the national or aggregate level, as a determinant of the general phenomenon of labour mobility; e.g. see Kerr(81), Bluman et al.(26), Stoikov and Raimon(134), Mattila(93), Holt(74, 75), Parson(112, 113), Wickens(149), Burdett(29), Viscusi(144), Medoff(99), Hay and McKenna(72). In this approach the roles of neo-classical economic theory, human capital theory, internal labour market analysis and the theory of information and job research are more noticeable. The main emphasis of such models seems to explain quits in terms of economically meaningful variables. We shall refer to this approach as the macro-level approach while discussing some of its related models.

However, before discussing these points, we may have to define the scope of quit behaviour among other types of separations as described in the section below. In section 2.3 we shall discuss the micro-level models of quits and briefly comment on some relevant points on which we shall elaborate in the fifth chapter. We discuss macro-level models and their economic bases in section 2.4, but we devote a sub-section 2.4.1 to a criticism of Viscusi's(144) logistic model of the effect of industrial injuries on quit intention. In section 2.5 we draw some relevant implications from the Beta logistic model of Heckman and Willis(69)
of the sequential labour force participation of married women. Then, we move to the development of our quit probability theory, which we link up with the random utility theory of choice. We start with section 2.6 where we draw attention to the importance of non-pecuniary benefits of work as a determinant of quit intention. Then, in section 2.7 we propose a simplified theoretical model to describe the latter point together with the role of taste differences among individuals, and thus express the problem within the context of random utility theory.

2.2 How Voluntary Are Quits

Although we understand that quits are voluntary job terminations by individual employees, we need to qualify on what we mean by voluntary. We may adopt the description of labour mobility by Holt (74) and Mortenson (103) in terms of accessions and separations. This description is presented in Figure (2.1) below.

![Mobility in the Labour Force](image-url)
We see that separations are being classified exhaustively into lay-offs, retirements and quits. More precisely the first one refers to action taken by the employer, the second is normally related to the age of the employee, while the third refers to decisions taken by employees. Thus, this view treats quits as employees who are driven to leave their jobs for reasons other than being forced by their employers, nor due to retirement (or expired contracts). In this sense 'quit' is a residual type of separation not directly attributable to the employer's actions or expired contracts.

The same definition seems to be adopted by Heckman\(^{68}\) in his analysis of labour supply, market wage, and shadow price of time, and by Wickens\(^{149}\) in his econometric model of labour turnover in U.K. manufacturing industries. Due to the fact that there are no published time series data on quit rates for manufacturing industries in U.K., the last author estimated quits as an unobserved dependent variable on the basis of the crude data of engagements and disengagements published by the Department of Employment Gazette. Then, Wickens derived an indirect linear equation for quit on the basis that disengagements are the sums of lay-offs and quits at different years. We may get a more direct assessment of the importance of quits at the national level in U.K. by referring to the sample survey for the registered unemployed due to Daniel\(^{39}\), or to similar data published by the General Household Survey, 1971-1977. However, we have to recall that voluntary job leavers at any point of time could either be immediate job changers, or among the registered unemployed or the unregistered unemployed. Hence Daniel's data pertains only to a subset of job-leavers, as it deals basically with the problems of the unemployed. This data is reproduced in Table (2.1) below.
Clearly, the definition which we have adopted above would lead us to consider the 'own accord' leavers as quits. But, Daniel (39, p. 53) disagrees with us, as these leavers have not all been "happy to quit their jobs and confident of finding a new one". Thus, according to this view most of the registered unemployed, who have left of their own accord, were not voluntary leavers. The *24% who have left due to injury and health reasons, and the 15% due to domestic and family reasons are, thereby, excluded. If 'dissatisfaction' were taken as the only criterion for voluntary termination, it leaves 52% of other own accord leavers as a special kind of involuntary disengagements. However, we have shown above, that a broader definition of quits should be based on employee's own decision to terminate his job in response to changes in his own circumstances. The latter may include changes in travel time to work, domestic reasons, or even his health and physical fitness for that kind of job. Those who have left for injury or health reasons are not unemployables as they have not given up looking for alternative jobs, as Daniel has later demonstrated. A particular treatment of the effects of industrial injury and health on quit intention has been examined by Viscusi (144), to which we shall return later. Now, we may examine the General Household Survey data, as compiled in Table (2.2) but we should notice that the percentages do not add to 100%. We see that for 1974 the total percentage for those who have been either sacked or laid off is given as 48% by Daniel (Table (2.1)), but only 39% by Table (2.3). However, if we add the reason of 'temporary jobs' in Table (2.3) to the sacked and laid off, we get 45%, which is quite close to Daniel's figure, 48%. Then, if we subtract the retired percentage we get an 'own accord' percentage of 54% for 1974, using the General Household Survey data, compared to 47% 

* See Table (2.2).
on the basis of Daniel's\(^{(39)}\). Variations during 1971-1977 do not show a specific trend. This can be taken to imply that around 50% of all terminations among the registered unemployed are voluntary terminations.

**Table (2.1) Reasons for leaving last employer for 1974.**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Made redundant</td>
<td>28%</td>
</tr>
<tr>
<td>Dismissed</td>
<td>20%</td>
</tr>
<tr>
<td>Left of own accord</td>
<td>47%</td>
</tr>
<tr>
<td>Retirement</td>
<td>5%</td>
</tr>
</tbody>
</table>

**Table (2.2) Reasons for leaving last job of own accord for 1974.**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury/ill health</td>
<td>24%</td>
</tr>
<tr>
<td>Domestic/family reasons</td>
<td>15%</td>
</tr>
<tr>
<td>Dissatisfaction/dispute with boss/work mates</td>
<td>10%</td>
</tr>
<tr>
<td>Dissatisfaction with working conditions</td>
<td>9%</td>
</tr>
<tr>
<td>Dissatisfaction with nature of work</td>
<td>8%</td>
</tr>
<tr>
<td>Dissatisfaction with hours</td>
<td>5%</td>
</tr>
<tr>
<td>General dissatisfaction</td>
<td>6%</td>
</tr>
<tr>
<td>Inability to cope with work</td>
<td>3%</td>
</tr>
<tr>
<td>Other</td>
<td>10%</td>
</tr>
</tbody>
</table>

100%

Then, to complete the picture we provide the percentages of quits among the registered employed group. This can be shown by Table (2.3) below. The data in this table gives percentage job changes within the last twelve months from the year of observations. Yet, as reasons for job changes are not provided, there is no clear ground to claim that these employees moved "happily and quickly", as they encompass possible periods of search ranging from zero to twelve months. This data is available only for five years which shows an average of 12.9% of job changes per annum for this class of workers in Great Britain.

Unfortunately the analogous situation of reasons for leaving has not yet been examined for the unregistered unemployed. The percentage of females, especially married women, has been shown to be considerably bigger than for males in this class of job-seekers. For example, during the period 1971-1973, the number of unregistered unemployed males varied between 70,000 and 100,000, while the corresponding figures for females were 160,000 and 200,000. This has been explained on the grounds that most of the women were married and had no financial incentive to register, since they were not eligible for unemployment benefit (see the study of the Department of Employment Gazette (43)). It is possible to draw some tentative conclusions from Table (2.5) for the unregistered unemployed. It can be seen that the unregistered unemployment of both sexes shows shorter time spent since leaving the last employer. For this reason it was thought likely that many of these unregistered unemployed "were between jobs and did not bother to claim for a short period, or were not eligible for benefit because they left their previous jobs voluntarily," (Department of Employment Gazette (43, p.1332)).
Table (2.3) Reasons for Leaving Last Job (Registered Unemployed) For Combined Males and Females

<table>
<thead>
<tr>
<th>Year</th>
<th>Made Redundant/sacked</th>
<th>Ill-health</th>
<th>Dissatisfaction</th>
<th>Temporary Job</th>
<th>Retired</th>
<th>Domestic &amp; Other Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>43.1</td>
<td>26.9</td>
<td>14.4</td>
<td>5.0</td>
<td>1.2</td>
<td>15.2</td>
</tr>
<tr>
<td>1972</td>
<td>47.9</td>
<td>22.7</td>
<td>17.3</td>
<td>6.9</td>
<td>1.6</td>
<td>16.0</td>
</tr>
<tr>
<td>1973</td>
<td>38.0</td>
<td>23.0</td>
<td>22.0</td>
<td>7.0</td>
<td>3.0</td>
<td>20.0</td>
</tr>
<tr>
<td>1974</td>
<td>39.0</td>
<td>26.0</td>
<td>12.0</td>
<td>6.0</td>
<td>1.0</td>
<td>21.0</td>
</tr>
<tr>
<td>1975</td>
<td>44.0</td>
<td>11.0</td>
<td>23.0</td>
<td>6.0</td>
<td>2.0</td>
<td>17.0</td>
</tr>
<tr>
<td>1976</td>
<td>43.0</td>
<td>15.0</td>
<td>22.0</td>
<td>6.0</td>
<td>2.0</td>
<td>17.0</td>
</tr>
<tr>
<td>1977</td>
<td>40.0</td>
<td>11.0</td>
<td>22.0</td>
<td>9.0</td>
<td>2.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Note that percentages do not add up to 100% since some people give more than one reason.

Table (2.4) Percentages changed job within last 12 months

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage changed jobs</td>
<td>n.a.</td>
<td>12.7</td>
<td>15.0</td>
<td>15.0</td>
<td>12.0</td>
<td>10.0</td>
<td>n.a.</td>
</tr>
</tbody>
</table>
Table (2.5)
Time since leaving the last job for the unregistered unemployed and those registered

<table>
<thead>
<tr>
<th>Time since leaving last job</th>
<th>Males (%)</th>
<th>Females (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Registered</td>
<td>Unregistered</td>
</tr>
<tr>
<td>Less than 6 months</td>
<td>44</td>
<td>59</td>
</tr>
<tr>
<td>Between 6 and 11 months</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>One year or more</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>Never worked previously</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>


2.3 The Micro-level Approach for Quits

As mentioned earlier (see section 2.1) employers are greatly concerned with the stability of their recruited labour force to guarantee the smoothness of the production process, to minimize total training costs, and to keep up the morale of the labour force. The question which has been subjected to scientific investigation is how to assess the extent of stability at any point of time, and how to advise the employer on controlling turnover at the lowest wastage rate. Earlier surveys carried out during the fifties showed that the proportion of voluntary terminations varied substantially amongst different regions in Great Britain; e.g. see Rice et al.\(^{(120)}\), Silcock\(^{(128)}\). Efforts have been oriented towards explaining such variability within a more general theory of labour turnover. This trend has been stimulated by the description of
Rice, Hill and Trist\(^{(120)}\) to variations in labour turnover as showing regularity which is characteristic of the firm concerned, i.e.

"When .... fluctuations were accounted for there remained a level of turnover which was relatively constant for the institution in which it occurred."

Rice et al. (120, p. 371).

Hence, each firm tends to possess a constant wastage rate, according to this view, depending on the characteristics of the firm concerned (e.g. type of labour contract, the rates of wages offered, opportunities for overtime work, the working conditions and the social relations within the factory.).

This theory has been criticized by Silcock who proposed the life-table mathematical structure to describe the process of labour turnover where 'birth' corresponds to recruit and 'death' corresponds to 'quit'. The general set-up for this approach which has been systematically adopted by subsequent studies is outlined below for reference.

Its main three components are:

(a) The survival function, \( F(t) \); this is defined as the probability that an employee survives for (at least) length of service, \( t \), i.e.

\[
F(t) = \text{Prob}(T \geq t). 
\] (2.2)

(b) The probability density function for the completed length of service is therefore given by:

\[
f(t) = -\frac{d}{dt} F(t).
\]

(c) The force of separation, wastage rate, or turnover rate is defined as:

\[
\text{Prob(employee leaves in } (t,t+\delta t) \mid \text{survives up to } t) \equiv \lambda(t)\delta t.
\]
It is also referred to as the propensity to leave (Bartholomew, p. 183). This function is also called the force of mortality in life-table analysis and given different names in other applications, e.g. the hazard function in life-testing. This function is derived from the relationship that:

\[ f(t) \delta t = \text{Prob(} \text{employee leaves in } (t, t + \delta t) \text{)} \]
\[ = \text{Prob(} \text{survival to } t \text{)} \lambda(t) \delta t \]

so that the wastage rate is expressed as

\[ \lambda(t) = \frac{f(t)}{F(t)} = -\frac{d}{dt} \log F(t) \] (2.3)

In the first place Silcock has criticized the crude turnover rate (2.1) as more dependent on past history than on current status of firms. He then provided the variant which was lacking in the "constant wastage" theory of Rice, Hill and Trist, namely differences in employees' characteristics. Thus, instead of postulating a constant wastage rate, \( \alpha \), for a specific firm, we have to deal with a whole series of different values, \( \alpha_1, \alpha_2, \ldots, \alpha_n \), each one characterizing a small homogeneous group of employees within the firm. He remarks that:

"The constant \( \alpha \) should be regarded not as something determined by the firm alone, but as a result of an interaction between the characteristics of the firm and the characteristics of the individual employees."

(Silcock, (128), p. 434).

The hypothesis of the constant wastage rate can be shown to be consistent with an exponentially distributed length of service:

\[ f(t) = ae^{-\alpha t}, \quad t \geq 0 \]

by noting that \( \lambda(t) = \alpha \) all \( t \) - using definition (2.3). However
Silcock has asserted that there are different wastage rates: \( \alpha_1, \alpha_2, \ldots, \alpha_n \) each one relating to a given group of homogeneous employees. Then, he assumed that these wastage rates follow a gamma probability density:

\[
g(\alpha) = \frac{\lambda^s}{\Gamma(s)} e^{-\lambda \alpha} \alpha^{s-1} \quad \lambda, s > 0
\]

from which he derived a mixed exponential distribution for the completed length of service given by:

\[
f(t) = \int_0^\infty ae^{-\alpha t} g(\alpha) d\alpha = \frac{s}{\lambda} \frac{(\frac{\lambda}{\lambda+t})^s}{\lambda+t}
\]

so that the force of separation is obtained as

\[
\lambda(t) = \frac{s}{\lambda+t}. \quad (2.4)
\]

Thus, he arrived at the interesting conclusion that the wastage rate declines monotonically with the length of service, the longer the latter, the smaller the former. When this mixed exponential model was fitted to the length of service distribution for the Glacier Metal Company, it provided a much better fit than that of the constant rate exponential distribution.

The subsequent works have shown that it is possible to get good fits for different mathematical forms of mixed exponential distributions. For example, Bartholomew\(^{(14)} \), has shown that a simpler mixed exponential distribution for length of service can provide an equally good fit, namely, using only two possible values \( \alpha_1 \) and \( \alpha_2 \) with probabilities \( p \) and \( (1-p) \) respectively. Then the following mixed exponential distribution can be fitted:

\[
f(t) = p \alpha_1 e^{-\alpha_1 t} + (1-p) \alpha_2 e^{-\alpha_2 t} \quad t > 0.
\]

The studies of Herbst\(^{(71)} \), and Clowes\(^{(30)} \) derive similar, though more
extended forms, as we shall shortly consider.

Alongside this approach of describing labour force stability, Lane and Andrew (86) have proposed the use of expected length of service as an alternative to the crude turnover rate (2.1). The former quantity has been derived in analogy with the life-table expectation of life at birth. They have actually compared two departments A and B where A was having a lower crude turnover rate than B. But, it was shown that B was more stable than A in the sense of having higher expected length of service. This anomaly has been explained by the fact that the department B had a relatively high proportion of short service staff and hence had a higher turnover rate, although its men are relatively more 'stable'. However, the use of expected length of services as a measure of stability has been criticized by Benjamin (19) on grounds that vital statisticians in recognition of the inefficient use of crude death rates do not substitute it by the expectation of life as a means of summarizing the life-table. If a single index is necessary, then, they use the standardized mortality rate. However, the difficulties of standardization have already been considered by Silcock (128), p. 432, as it is not easy to suggest a standard set of weights for a large number of firms with different compositions of employees.

As regards the distribution of service length, Lane and Andrew (86) have shown that the log-normal distribution provides a good fit. This finding has led Aitchison (2) and Bartholomew (14), to question the relevance of Kaptynes's law of proportionate effect to the specification of the relationship between duration of employment at job j and a previous job j-1. It has been shown by Bartholomew that although Kaptynes's law does generate the log-normal distribution for the completed length of service, yet it does not explain its relevance in the first job.
Now, before concluding this section it is interesting to consider the organizational commitment model due to Herbst (71), and its modified version due to Clowes (30). Herbst (71) analyzed the process of employees leaving as one involving transitions between five different decisional states of mind as presented in Figure (2.2) below. The parameters $k_1, \ldots, k_s$ represent transition rates for the various states.

\[ N = A \exp(-\alpha t) + B \exp(-\beta t) - C \exp(-\delta t) + N_p \]

where $A, B, C, \alpha, \beta, \delta$ and $N_p$ are various combinations of the constraints $k_1, k_2, \ldots, k_s$ and $N_p$ (the initial number).

Although this model has provided an exceedingly good fit to the data of 40,000 employees over a prolonged period, yet it has been criticized by Clowes (30) as it contains an unnecessarily complex system. It also assigns a zero leaving propensity to the 'permanently committed' which is unrealistic. Besides, it generates five parameters from only two variables which would lead to loss in parameter discrimination.
Clowes\(^{(30)}\) has in fact stressed the last point, as the five parameter system tells too little about how to control the turnover process:

"There are so many parameters derived from so few data that individual parameters do not discriminate very well between high and low turnover rates."

(Clowes\(^{(30)}\); p. 245).

In fact Clowes\(^{(30)}\) has introduced his modified version in order to allow the parameter system of the model to act as policy control parameters. Hence, he proposed a simplified model represented by Figure (2.3) below.

\[
N = \left\{ \frac{N_o}{(k_1 + k_2 - k_3)} \right\} \left( k_1 - k_3 \right) \exp\left( -\left( K_1 + K_2 \right) t \right) + K_2 \exp\left( -k_3 t \right)
\]

Figure (2.3)

Clowes' model of labour turnover

where \(k_1\) is the quit rate of new rates, and \(k_3\) the quit rate for committed employees. Then, he shows that the number remaining is given by

To illustrate the principle of utilizing \(k_1\) and \(k_3\) as policy control parameters, Clowes\(^{(30)}\) examines the effects of modifying these constants on the survival curve. For example, he estimates that for the Glacier Metal Company the savings are about 3.3% of the annual wage.
bill if \( k_1 \), the new recruits quit rate, is halved - assuming that cost
of replacing each man is 20% of the man's salary.

We shall criticize this approach in our fifth chapter on the ground
that \( k_1 \) and \( k_2 \) are too crude to guide the 'screening/training' policy
as Clowes has proposed. We shall adopt a procedure analogous to that
of Silcock of assigning different leaving rates to different groups of
homogeneous employees. On this ground we shall define and utilize a
probability distribution of employee leaving rates whose parameter
system can serve policy-control purposes. However, as our data relates
to employees of different firms, we have no scope to test our model at
the micro-level, though in principle it could be utilized. Our approach
which is fully discussed in Chapter 5 is based on the principle of
utilizing a disaggregate probability model to arrive at aggregate pre-
diction and policy implications. The crux of this principle is to derive
an expression for the expected number of leaving employees directly in
terms of employees vector of characteristics. However, it is not within
our scope to consider the adaptation of the above demographic structural
models to provide similar expressions. We now proceed to consider the
macro-level models.

2.4 Macro-level Models of Quits

The phenomenon of turnover at the national or industrial level is
studied within a more complex set of labour stocks and flows as has been
exemplified by Figure (2.1) above. A massive amount of literature has
grown up rapidly within the last two decades to analyze the complex
process of successful matching of workers and jobs, to explain a diverse
set of phenomena; e.g. accessions and separations, labour force participa-
tion, mobility among the states of employment, unemployment and not in
the labour force, and the Philips curve phenomena. These problems are explained within a unified analytical framework based on the job search theory which has been sparked off by Stigler\(^{(133)}\); see McColl\(^{(94)}\), Hall\(^{(64)}\), Mortenson\(^{(104)}\), Gronau\(^{(60)}\), Seater\(^{(127)}\), Burdett\(^{(29)}\), Wickens\(^{(149)}\) and others. On the other hand human investment theory as pioneered by Becker\(^{(16)}\), Parson\(^{(144)}\), Oi\(^{(109)}\) attempts to explain turnover in terms of the distribution and magnitude of training costs between employers and employees. Besides, internal labour market theory studies the role of institutional rules in determining the extent and nature of mobility in the labour market; see review in Addison and Siebert\(^{(1)}\). However, the general reliance on published statistics and the relative shortage of disaggregate data account for the frequent dependence on aggregate econometric models in testing the implications of the above theories.

The study of quit behaviour has been given special consideration. For example, Berman\(^{(22)}\) finds that the quit rate explains 90% of the variance of wages in U.S.A. for the period (1941-1961). The observation that quits tend to rise with an increase in vacancy/unemployment ratio is sometimes used to explain the negative relationship between unemployment duration and rate of increase in money wage-rate (i.e. the Philips curve phenomenon); see Holt\(^{(74)}\), p. 72. However, at the times of recession, lay-offs provide the opposite force. That is, they rise with a decrease in the vacancy/unemployment ratio, thus, depressing the rate of increase in money wage-rate, and hence the quit rate. Evidence from the U.S.A. for the period (1947-1967) reveals a countercyclical pattern for lay-offs and quits, which implied that total separations vary only slightly over the cycle. Similar evidence as regards the opposite movements of quits and lay-offs has been provided by a group of economists at the I.C.I's Mond Division, U.K.,

\(^{†}\) See Dept. of Employment Gazette\(^{(44)}\)
while explaining the reasons for a low quit rate during the fourth quarter of 1971. They have concluded that the high lay-off rate during that quarter, among other factors, accounted for the drop of quits from the I.C.I. Mond Division. Similar evidence has been given by Stoikov and Remond (134) who obtained a negative regression coefficient for the lay-off rate in their inter-industry quit rates model for U.S.A.

2.4.1 The Role of Institutional Rules and Effect of Unionism

Institutional rules are established by employers, associations, by trade unions and governmental action. These rules tend to create a larger number of sub-markets and make them less interrelated and total mobility tends to be reduced; see Kerr (81). The latter author addressed himself to explain the apparent contradiction observed by other studies for the effect of unionism on voluntary labour mobility. Specifically he considered the claim of Shister (81) that union policies reduce the amount of voluntary mobility on net balance. This finding runs contrary to those of Lispet and Gordon (90) who have shown that union members are more mobile. Recent evidence from U.S.A. by Medoff (99) has shown that unionism tended to reduce voluntary mobility. This contradictory evidence has been attributed by Kerr (81) to the relative dominance of craft rules as opposed to industrial enterprise rules. He comments about the two studies by Shister and by Gordon and Nispet, that

"The two studies come to opposite conclusions because they are based on observations of two contrasting situations."

(Kerr (81); p. 103).

Thus, the latter author has drawn attention to an important question
to be asked in determining the effect of unionism on voluntary mobility: Did the worker belong to a craft or an industrial union? In a simple model he argues that craft workers move horizontally in the craft area, while industrial enterprise workers move vertically in the seniority area, "job rights protect but also confine." Craft unions are believed to reduce inter-occupational movement, movement to unemployment but generally increase movement between employers. On the other hand, industrial enterprise unions reduce movement between employers, or movement to unemployment, but admitting inter-occupational movement.

These findings about the nature of labour mobility, may be formalized below within the theory of human capital, as we shall outline in the next section.

2.4.2 The Role of Specific Human Capital: Different types of movements are sometimes explained by distinguishing between two broad types of investment in human capital as provided on-the-job by employers. On one hand, the 'general training' component tends to be marketable to all potential employers, and may raise productivity of all firms by the same extent. On the other hand, there is the 'firm-specific' training component, which is intended to increase potential productivity of the given firm alone. For example, medical training is a general type, while resources spent by the firm in familiarizing a new recruit with his organization are firm-specific training. The assumption of profit-maximization and loss-minimization involves that the firm may rationally finance the specific-training component, but not the general training; see Becker (16), and Parson (112).

The main implication of this theory is that employees with specific-training have less incentive to quit, and other firms have less incentive to hire them than those with general training. This also has been
shown to imply that quits and lay-offs are inversely related to the amount of specific training. The recognition of the importance of specific training has led Of(109) to treat labour as a quasi-fixed factor of production. Thus explaining that in the short-run variations in output may not induce employers to manipulate or lay off their labour stocks.

Parson(112) has fitted a linear regression model for quit probability on the basis of published data for 24 U.S.A. industries, covering the period 1959-1963. His results confirmed the expectation that quits vary inversely with firm-financed human capital, and directly with employee-financed capital. He constructed certain proxies for these types of training, together with other concomitant variables.

Thus, human capital theory gives an economic meaning to the effect of service length on the propensity to leave, discussed in section (2.3) above. This interpretation of the effect of job-tenure on quits has been pointed out by Stoikov and Raimon(134), p.1289 "The longer the length of service of the employee, the greater his specialization, the fewer the extra-organizational alternatives perceived." The latter authors have analyzed inter-industry quit rates among U.S.A. industries (1963-1966), using length of service as a proxy for specialization, or firm-specific investment. They obtained a negative regression coefficient for this proxy in their linear quit probability model, in agreement with human capital theory.

2.4.3 Job Search Theory and Quits: As we have mentioned above, job search theory attempts to provide a unified analytical framework for the behaviour of workers between jobs and related problems (e.g. see references cited appropriately at page 30 above). This approach is
based on the recognition that information in the labour market is scarce and job-search is a costly process. However, there is a noticeable predominance in this field on analytical relevance and theoretical implication rather than formulation of econometric models. Among the very few econometric models in this area are those due to Parson (113), for the analysis of quit rates over time, and Wickens (149) for the study of quit rates and lay-off rates in U.K. manufacturing industries.

Any worker who believes that search for an alternative job would be profitable is confronted with a choice problem: (i) he may quit as soon as he is convinced that search is rewarding, (ii) he may remain with his current job, searching after work hours, or in spare-time by scanning through appropriate magazines, etc. We shall first comment on Parson's model which defines the critical wage for the quit decision as:

\[ W_c = W_o + TC \]

where \( W_o \) is the current wage, \( TC \) is the cost of transfer and \( W_c \) the critical wage. Let \( f(W_c) \) denote the probability distribution of alternative offers for this worker. It is assumed that wage dispersion exists for people of the same qualification due to the fact that perfect information is uneconomical and job transfer costs retard mobility to higher wage-offers. It is also implicitly assumed that employees aim at maximizing expected income. Units of search are made sequentially where, at each point of time, the job-seeker compares marginal cost for additional units of search with marginal return, or net wage increment, \( NWI \) defined as \( W_A - W_C \), where \( W_A \) is any wage offer. Then the expected value is defined as:

\[ E(NWI) = \int_{W_A} (W_A - W_C)f(W_A)dW_A. \]
Then according to Parson (113), the duration of search is determined by this quantity. (The same concept has been adopted by Hay and McKenna (72) to answer a similar question.) Then, the following quantities are introduced

\[ P = \text{Prob(a job offer)} \]
\[ N = \text{Units of search determined by } E(NWI) \text{ above.} \]

On this basis Parson expressed quit probability as

\[ q = \text{Prob(quit)} = 1 - (1 - p)^N. \quad (2.5) \]

He then proceeds to expand (2.5) as

\[ q = Np - \frac{N(N-1)}{2} p^2 + \text{Higher Order Terms} \]
\[ = Np \quad \text{(see Parson (113); p. 394).} \quad (2.6) \]

However, this approximation can only be defended where \( P \) is very small and \( N \) is not very large, as it is likely to violate the 0, 1 restriction for the probability. Yet, the expectation on the basis of this model is that for the units of search not to be very large, then the offer probability should not be very small! The final model has been estimated by least squares regression analysis by linearizing \( E(NWI) \), the net wage-increment expectation, in terms of the vacancy rate, the wage rate and other real factors like industry demand levels and season of the year. The conclusion has been that job vacancies are a major determinant of monthly variation in industrial quit performance; changes in worker's own wage, and relative wage-rate are seen to be less important than the real factors.

It is worth noting that the linear probability model is commonly adopted for the specification and estimation of turnover equations of quits and layoffs. The models which we considered so far, including Wickens' model (149), all adopt the linear probability specification for
quits. Yet, it has been shown in previous studies that the linear probability model suffers from certain empirical limitations (i.e. violation of the $[0,1]$ restriction, specification and predictive bias, and the problem of heteroscedasticity) though it may, sometimes, give agreeable results with other non-linear probability models. This point will be elaborated on in the next chapter. We may also note that the above models are based on aggregate data, and the dependent variable (quit) is represented by the proportion of leaving employees. Alternatively a disaggregate model can be fitted where the dependent variable is an 0,1 binary random variable representing the individual's quit/stay decision. This will lead us to Viscusi's model(144) for job-hazards and quit rate. This model is also interesting as it postulates a non-linear probability model, namely the logistic function.

2.4.4 The Job Hazards Logistic Model of Quit Intentions and Related Comments

Viscusi (144) developed an econometric model to test the effect of disabling industrial injuries on the quit probability. Specifically, he expressed quit probability in the logistic form as:

$$q(X) = \text{Prob}\{\text{quit}|X\} = \frac{e^{\beta'X}}{1 + e^{\beta'X}} \tag{2.7}$$

where $X$ is a vector of explanatory variables including personal characteristics (e.g. Age, job-tenure, health, hazard perception, schooling), and other measures of general labour conditions (e.g. number of disabling injuries per unit time, percentage of union members, change of employment level during 1969-1976, and rate of new hires). The other interesting variation introduced by this study is that it is based on quit intention rather than actual quits. In particular, two models are
estimated, one for strong quit intention probability and the other for weak quit intention. In both models it has been shown that there is a significant positive sign for the effect of industrial hazard on quit intention, as well as for the effect of employee's health (as expected, e.g. see Tables (2.1) and (2.2) for the health reason of quitting).

However, there has been a pitfall overlooked by Viscusi while estimating the effect of the 0,1 dummy of hazard perception on average quit probability. He has utilized his model (2.7) to compute the difference:

$$\Delta q(X) = q_1(X) - q_0(X) \quad (2.8)$$

where $q_1(X)$ is evaluated at the means of all other explanatory variables with hazard perception present. Similarly $q_0(X)$ is evaluated at the means of all other variables with hazard perception absent. He then gets $\Delta q(X) = 0.11$ to conclude that hazard perception increases average quit intention probability by 100% of strong intention (since $q_0(X) = 0.10$).

A similar approach has been adopted by Medoff in assessing the effect of unionism on average quit and lay-off probabilities on the basis of the logistic model. Yet, there are two criticisms which can be made against the use of relation (2.8) to assess the effect of hazard perception on average quit probability.

In the first place the use of $q(X)$ as an estimate of average quit probability is not justifiable if $X \neq 0$, and the variance-covariance matrix, $E$, for $X$ is not a null matrix. It can be shown* that in the special limiting cases where $X = 0$, or $E = 0$, average probability is always equal to a half and hence the use of the naive formula:

$$q(X) = \frac{e^{\beta X}}{1 + e^{\beta X}} \quad (2.9)$$

* This point will be explored in Chapter 5, section (5.3).
can be justified. However, in the general case where $\bar{X} \neq 0$ and $\Sigma \neq 0$, the use of the naive formula (2.9) is usually biased. It gives an overestimate for the true average choice probability if $\bar{X} > 0$, and an underestimate when $\bar{X} < 0$. The technical details of this point are discussed in the fifth chapter, section (5.3).

In the second place the effect of a variable on average choice probability should be based on the method of deriving aggregate predictions from a disaggregate probability model. This recently developed approach enables us to assess the sensitivity of average quit probability to small or finite changes in the level of hazard perception. It appears that Viscusi has been aiming at such a measure of aggregate elasticity. This idea will be discussed in more detail in Chapter 5, where we consider the application and development of point and arc-elasticity measures to answer similar questions.

In fact this approach has not as yet been adopted in econometric models of labour turnover, and so we are proposing it in this study. In particular, the use of aggregate elasticities which measure the effects of different explanatory variables on average quit probability, provides a useful guide for screening/training policies. In this way the aggregate quit probability model can serve the policy control requirement discussed by Clowes in his organizational commitment model (see section (2.3)). For example we may use the sign and magnitude of the elasticity of average quit probability to certain percentage changes in average travel time of employees. We may utilize our disaggregate model to answer similar questions to illustrate this principle at the macro-level, which should similarly be relevant to micro-level situations. Thus, in principle, individual business firms can utilize disaggregate models which explain quit probability in terms of employees' characteristics.
employer's characteristics, and other labour market conditions. Then, the expected number of leaving employees within a specific period of time can be predicted for different changes in the vector of characteristics. Although this model lacks the components of survival function and distribution of employment duration of the demographic structural models (see section (2.3)), yet it has the advantage of explicitly allowing for other explanatory variables beside job-tenure (or length of service). In fact Silcock (128) has recognized that other variables (i.e. skill, age, sex, distance to work, etc.) are possible determinants of the leaving process, yet length of service has been treated as the sole determinant by the micro-level models. These models are nevertheless quite revealing and genuine efforts, but we believe that the adoption of such a disaggregate probability model may complement them as an aid to policy-making regarding the control of turnover. In the fifth chapter we are elaborating on these points in some detail, but now we briefly discuss some relevant points from the beta-logistic model of Heckman and Willis (69).

2.5 Some Relevant Implications of the Beta Logistic Model

This model has been proposed by Heckman and Willis (69) to study the sequential labour force participation of married women in panel data of a heterogeneous population. In particular they have been interested in the prediction of the participation path of a woman with observed vector of characteristics $X_t$ (at time, $t$) associated with unobserved and chance events in the quantity, $S_t$. The latter quantity has been decomposed into a 'transient component' error $u$ which has zero serial correlation, and a 'permanent component' $v$ which has positive serial correlation over time. Thus, the covariance terms of
$S_t$ are:

$$\text{Cov}(S_t, S_{t'}) = \sigma_v \quad t \neq t$$

$$\sigma_v + \sigma_u \quad t = t \quad \quad (2.10)$$

The fact that the autocorrelation of these $S_t$ over two years is positive \([i.e. \rho = \frac{\sigma_v}{\sigma_v + \sigma_u}]\) means that we have to allow for heterogeneity of observationally homogeneous women, when we want to get the average participation probability of women with the same observed vector, $X_t$. Hence for these women it has been shown that there is a non-degenerate probability distribution $g(\theta)$ for their participation probabilities, $\theta$ in a given year. The structure of this distribution has been shown to depend on the relative sizes of $\sigma_v$ and $\sigma_u$ as defined in (2.10) above. They have argued that the following shapes can be derived for the distribution of participation probabilities at a given point of time for women with the same linear combination, $\beta'X_t$, of observed characteristics. These shapes are defined as in the figure below:

![Figure (2.4)](image)

Distribution of Labour force participation probability for a given year.
Accordingly, Heckman and Willis\(^{69}\) have chosen the Beta Distribution to represent \(g(\theta)\), since its geometrical flexibility permits the generation of these shapes. Thus, they have fitted:

\[
g(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}
\]

where \(B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} \ d\theta\).

They have been particularly interested in the prediction of the following binomial probabilities which represent different sequential paths:

\[
\text{Prob(work j out of n years)} \equiv P(j,n) = \binom{n}{j} \theta^j (1-\theta)^{n-j} \quad (j \leq n)
\]

Thus, they have calculated the average sequential probabilities:

\[
E\left[P(j,n)\right] = \int_0^1 P(j,n) \ g(\theta) \ d\theta
\]

on the basis of the Beta density \(g(\theta)\) above, and compared them with sample participation probabilities over given years. The name beta-logistic has arisen from the fact that

\[
E\left[P(1,1)\right] = \frac{a}{a+b}
\]

Then, by letting \(a = e^{\alpha X}\), and \(b = e^{\beta X}\), they have shown that the average participation probability for any given year is logistic.

The major finding of Heckman and Willis\(^{69}\) (which has been tested on the basis of 1,583 married women from the University of Michigan Panel Study of Income Dynamics) is that \(\hat{a} < \hat{b}\). This result implied that the
distribution of participation probabilities, \( g(θ) \), for married women is U-shaped for any given year. However, according to the authors the beta-logistic model did not perform successfully to predict the sequential paths, \( E[p(j,n)] \). This deficiency has been attributed to the fact that the vector of characteristics, \( \mathbf{x} \), was chosen once and for all for the first year, and all subsequent possible changes in it were ignored. This model also implied that each individual had a given participation probability which was constant over time. The authors show that the beta-logistic model ignores state dependence at the individual level and "overstates the degree of population heterogeneity..."; see the discussion and criticism of this model by Heckman and Willis (69), pp. 48-52).

The idea of sequential labour force participation of women simply means that in each year a woman decides whether to remain in the labour force or stay out of it. This makes it comparable to our leave/stay model. We can also think in terms of a non-degenerate probability distribution of quit probabilities, \( g(θ) \), for married women or male heads of households, at a given point of time. However, heterogeneity in this context is assumed to be reflected in the observed vector of individual's characteristics, \( \mathbf{x} \), rather than unobserved 'permanent components'. The shapes of the resultant \( g(θ) \) should be directly related to the effects of observed characteristics, \( \mathbf{x} \). Thus, we can directly allow for changes in the characteristics vector \( \mathbf{x} \), to reflect in movements in the distribution of quit probabilities, \( g(θ) \). This \( g(θ) \), which we are looking for, should be sufficiently flexible to generate comparable shapes as those of Figure (2.4) above. Yet, we shall not choose the beta function since it does not satisfy certain properties to be outlined in Chapter 5, section (5.4), including flexibility. Instead, we shall argue that Johnson's bounded system, the \( S_B \)
curve provides a better choice for $g(\theta)$.

The U-shaped finding for the distribution of married women’s labour force participation probabilities, is specially interesting. It implies that most women have participation probabilities near zero or near one. It is known* that the Beta distribution possesses a variety of U-shapes over the $0,1$ interval of $\theta$, as long as its two parameters $(a,b)$ satisfy $0 < a \leq b \leq 1$ or $0 < b \leq a \leq 1$. However unless $b = 1$, or $a = 1$ the Beta density will be infinite at the end points $\theta = 1$, or $\theta = 0$; otherwise it will have a finite density equal to $a$ or $b$ whichever of these is equal to one. However, this restrictive property of the Beta U-shaped curve is rather unrealistic as it implies that the closer we approach either of the certainty states ($\theta = 1$, or $\theta = 0$), the more likely we find women with that participation probability. A more realistic picture should allow for a non-zero mode and a non-unity mode while assigning zero probabilities to the unrealistic limits of certainty ($\theta = 1$, or $\theta = 0$), e.g. as in Figure (2.5) below:

![Figure (2.5)](image)

where $g(\theta) = 0$ for $\theta = 1$, or $\theta = 0$. This property is automatically satisfied by the $S_B$ curve. Other technical considerations of this problem are discussed in Chapter 5, where we derive the $S_B$ curve from the logistic model and the multivariate normality assumption of the characteristics vector, $X$. We now move to consider the role of

* See the properties of this distribution in Kendal and Stewart (80, p. 162).
job-satisfaction on quit intention.

2.6 Job-Satisfaction and Quit Intention

The general term 'dissatisfaction' accounts for most leaving intentions by employees; the greater the former, the stronger the latter. This intuitive point underlies the observed pattern of job mobility as described by the annual series of the General Household Survey 1971-1977. In each year a random sample of employees is chosen, and each employee is asked about the level of his satisfaction with his current job. However, since this term is subjective and unmeasurable they have defined five levels of satisfaction: very satisfied, fairly satisfied, neither satisfied nor dissatisfied, fairly dissatisfied, and very dissatisfied. During 1971 employees were also asked to specify their reasons for dissatisfaction (but this was not repeated in subsequent years). Besides, employees were asked whether or not they intended to leave their current jobs. Table (2.6) gives percentages of employees who intend to leave their jobs at different levels of satisfaction for the period 1971-1976. When we calculate correlation coefficients between scores of satisfaction and percentages of quit intentions, we get significant negative correlations as expected (see Table (2.6a)). This implies the simple fact that the greater the state of dissatisfaction, the larger the quit intention probability. However, 'dissatisfaction' with current jobs does not give 100% explanation to quit intention. For example, an average of 40% per annum of 'very dissatisfied' employees do not intend to leave their jobs, while an average of 4% 'very satisfied employees' do intend to leave. This may partly explain the real life fact that employees' decisions to stay or to leave could be attributed to reasons other than mere satisfaction with job.
### Table (2.6)

Percentage Intention to Change Job by Level of Satisfaction

<table>
<thead>
<tr>
<th>Year*</th>
<th>Very Satisfied (2)</th>
<th>Fairly Satisfied (1)</th>
<th>Neither, Nor (0)</th>
<th>Rather Dissatisfied (-1)</th>
<th>Very Dissatisfied (-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>4.7</td>
<td>12.6</td>
<td>30.2</td>
<td>47.8</td>
<td>59.4</td>
</tr>
<tr>
<td>1972</td>
<td>4.8</td>
<td>13.1</td>
<td>31.2</td>
<td>55.6</td>
<td>69.7</td>
</tr>
<tr>
<td>1973</td>
<td>5.0</td>
<td>10.0</td>
<td>27.0</td>
<td>54.0</td>
<td>63.0</td>
</tr>
<tr>
<td>1974</td>
<td>4.0</td>
<td>12.0</td>
<td>28.0</td>
<td>55.0</td>
<td>66.0</td>
</tr>
<tr>
<td>1975</td>
<td>4.0</td>
<td>9.0</td>
<td>21.0</td>
<td>30.0</td>
<td>60.0</td>
</tr>
<tr>
<td>1976</td>
<td>4.0</td>
<td>9.0</td>
<td>18.0</td>
<td>34.0</td>
<td>62.0</td>
</tr>
</tbody>
</table>

* Notice that we are assigning numbers to satisfaction levels to simplify the calculation of correlation coefficients as given in the table below.


### Table (8.6a)

Correlations between job 'satisfaction' scores and quit intentions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>-0.993</td>
<td>-0.947</td>
<td>-0.978</td>
<td>-0.984</td>
<td>-0.920</td>
<td>-0.952</td>
</tr>
</tbody>
</table>
(e.g. family reasons, health reasons, being rehoused, etc.) whether these reasons have already occurred, or are expected to occur. This point has been considered in section (2.2) above.

Now, we may look at Table (2.7) below, which gives reasons for dissatisfaction for a random sample of employees who were not 'very satisfied' with their jobs during 1971. Two objective factors are listed as possible reasons (i.e. pay, and work conditions), while the 'no real reason' factor can be considered as an unexplained taste component. Note that the greater the state of dissatisfaction, the smaller the percentage of this unexplained component, and the milder dissatisfaction, the bigger this percentage. On the other hand, note the reverse pattern for the objective factors, pay and work conditions. It also appears that the general term 'work conditions' explains a greater percentage of dissatisfaction than does pay alone at all levels of dissatisfaction. We may compute the differences between the different percentages assigned to these two factors at the above four levels of dissatisfaction. These differences are given at the bottom of Table (2.7). Notice that the importance of pay relative to work conditions increases as we move from the 'neither satisfied nor dissatisfied' to the 'very dissatisfied', although work conditions explain more than 50% to all the three states of dissatisfaction.

It is also possible to break down the term 'work conditions' into more explicit aspects as in Table (2.8) below. In this table there are 11 aspects of work conditions which may influence an employee's morale within the organization, and therefore his state of 'organizational commitment' (see Figures (2.2) and (2.3)). This point reminds us of the implicit contract theory which asserts the importance and nature of employment terms considered by workers when they accept (or reject) employment offers. Azariadis\(^{(11)}\) has studied the conditions under which implicit contracts may result in full-employment, where the contract
Table (2.7)

Working Persons aged 15 or over who were less than very satisfied with the job by degree of job satisfaction (1971)

<table>
<thead>
<tr>
<th>Reason</th>
<th>Fairly Satisfied</th>
<th>Neither Nor</th>
<th>Rather Dissatisfied</th>
<th>Very Dissatisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay</td>
<td>31.5</td>
<td>34.0</td>
<td>38.0</td>
<td>43.0</td>
</tr>
<tr>
<td>Work conditions</td>
<td>47.9</td>
<td>50.8</td>
<td>56.0</td>
<td>51.8</td>
</tr>
<tr>
<td>No real reason</td>
<td>13.3</td>
<td>8.0</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Other reason</td>
<td>7.3</td>
<td>7.2</td>
<td>5.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Total Percentage</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Differences between 'pay' and 'work conditions' 16.4 16.8 18.0 8.8


is a tacit or open agreement with the firm on certain employment terms "wage-rates, hours worked, employment status, or a combination of all such factors,"; Azariadis (11), p. 1184. This may also explain that job-seekers do not only compare money wage offers, but they value other aspects of work conditions as well. Hence, instead of treating the asking wage as a wage-valued scalar, as is usually done in models of job search, we may define it as a vector containing money-wage offer $W_o$ and other aspects of work conditions, $W_1, W_2, \ldots, W_p$, assuming that they are measurable, i.e.

$$W = (W_o, W_1, W_2, \ldots, W_p) \quad (2.13)$$

Such an approach which allows for non-pecuniary benefits of work in terms
<table>
<thead>
<tr>
<th>Reason for Dissatisfaction with work condition</th>
<th>Fairly Satisfied %</th>
<th>Neither Satisfied nor Dissatisfied %</th>
<th>Rather Dissatisfied %</th>
<th>Very Dissatisfied %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration/organization</td>
<td>16.7</td>
<td>20.1</td>
<td>37.8</td>
<td>39.3</td>
</tr>
<tr>
<td>Did not like the kind of work</td>
<td>15.6</td>
<td>24.0</td>
<td>23.0</td>
<td>25.6</td>
</tr>
<tr>
<td>Physical Working conditions</td>
<td>8.1</td>
<td>11.1</td>
<td>12.2</td>
<td>17.4</td>
</tr>
<tr>
<td>Heavy work</td>
<td>4.7</td>
<td>5.4</td>
<td>4.6</td>
<td>6.9</td>
</tr>
<tr>
<td>Lack of security</td>
<td>6.6</td>
<td>7.9</td>
<td>8.7</td>
<td>9.6</td>
</tr>
<tr>
<td>Long hours</td>
<td>4.0</td>
<td>4.7</td>
<td>5.7</td>
<td>8.7</td>
</tr>
<tr>
<td>Lack of opportunity</td>
<td>5.4</td>
<td>6.3</td>
<td>8.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Shift work</td>
<td>4.5</td>
<td>5.9</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Low status of work</td>
<td>3.4</td>
<td>4.0</td>
<td>7.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Too much travelling (both to work, and with work)</td>
<td>3.6</td>
<td>4.5</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Too much responsibility</td>
<td>0.6</td>
<td>1.1</td>
<td>0.6</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Note: percentages do not add up to those of the second row of Table (2.7) as multiple reasons are permitted.

of aspects of work conditions helps to explain the undeterministic effect of money wage offers in regulating mobility of labour. This limitation of the money wage-rate has already been pointed out by Nickel(108):

"In general there are both people who quit in the certainty of lower wage elsewhere, and people who stay even though there are higher wages elsewhere .... People change jobs for a variety of non-pecuniary reasons."

(p. 193)

Nickel(108) has been concerned with studying the analytical relationship between the personal and socio-economic characteristics of individuals and the life-time wage-structures which they choose. He illustrated his above remark by Figure (2.6) below, though his concern was not with the specification of a quit probability model.

\[ W = w - w_c \]

where \( w_c \) = current employer's offer and

\( w \) = alternative offer net of hiring costs.

Figure (2.6)

Nickel's curve for quit probability.
Notice that the above curve implies:

\[ q(W) = \begin{cases} 
1 & W \geq \bar{W}_c \\
0 & W \leq \bar{W}_c 
\end{cases} \quad \text{for} \quad \bar{W} < W < \bar{W} \tag{2.14} \]

and \( 0 < q(W) < 1 \) for \( \bar{W} < W < \bar{W} \)

This figure can be shaped as an ordinary response curve by putting \( \bar{W} = \infty \) and \( \bar{W} = -\infty \) as in this case it describes a non-linear probability model (e.g. logit or probit). Recalling that \( W = w - w_c \), then the logistic model which reproduces such a sigmoid curve is written as:

\[ q(W) = \frac{e^W}{1 + e^W} \quad -\infty \leq W \leq \infty \tag{2.15} \]

while the probit model that gives a similar description is written as

\[ q(W) = \Phi(W) \quad -\infty \leq W \leq \infty \tag{2.16} \]

where \( \Phi(\cdot) \) is the Normal Distribution Function. See Figure (2.7) below for the sigmoid curve which may suit the logit (2.14), or the probit (2.15) or any other non-linear probability model.

![Figure (2.7)](image URL)

A Sigmoid Response Curve for a Non-linear Probability Model
This representation for the quit probability allows from the non-determinism of the money wage-rate, and the important role of other aspects of work conditions and terms of employment. Thus, at \( W = w - w_c > 0 \), where the net alternative money wage offer \( w \), is greater than current offer \( w_c \), some people may prefer to stay with their current employers. Similarly, where current money wage rate is known to be higher than alternative offers (i.e. \( W < 0 \)) there are people who choose to quit. These two remarks imply in terms of Figure (2.7) that:

\[ \text{Prob}(\text{quit} | W > 0) < 1 \quad \text{and} \quad \text{Prob}(\text{quit} | W < 0) > 0 \quad (2.17) \]

respectively. However, a deterministic wage-quits theory like that adopted by Burdett (29; p.217) asserts that:

\[ \text{Prob}(\text{quit}) = \text{Prob}(W > w_c) \quad (2.18) \]

that is to say:

\[ \text{Prob}(\text{quit} | W > 0) = 1, \quad \text{and} \quad \text{Prob}(W < 0) = 0. \quad (2.19) \]

We may notice other studies adopting different versions of this pure wage-quit theory, e.g. Parsons (113); see equation (2.5); see also Hey and McKenna (72; p.177).

This discussion leads us to the stage where we need to incorporate non-pecuniary benefits in the formulation of a theoretical quit model. Hence, this model should allow for the non-determinism of money wage differentials, and derive the implications of equations (2.17) above. This will lead us to the next section.

2.7 The Relevance of The Random Utility Approach

In the last section we have pointed out the fact that a more realistic definition of the asking wage of a worker should include not
only the money wage, but also other non-pecuniary benefits of work. We have seen on the basis of table (2.5) that the states of different aspects of work explain more than 50% of employees' dissatisfaction with their current jobs. Moreover, as quit intention probability is negatively correlated with the level of 'satisfaction', it means that a model of quit behaviour should allow for these non-pecuniary benefits. However, the main limitation is that these benefits are subjective and not measurable, since it is the individual's perception of them which influences his quit decision. Hence, apart from the measurable characteristics of the employer and those of the employee, there are still unmeasurable components in a quit-behaviour model which include the employee's taste. These elements of the model may be presented in the context of neo-classical utility theory in the manner propounded by McFadden, or Domenich and McFadden, namely within a random utility model. For example, Domenich and McFadden remark that:

"Within a framework of economic rationality and the postulated structure of utility maximization there will be unobserved characteristics such as taste... which vary over the population. These variations may induce variations in observed choice among individuals facing the same measured attributes."

(p. 48)

Random utility theory has in fact been developed to allow for situations where objects of choice are 'lumpy', and where unmeasurable individual tastes and attributes account for a non-degenerate distribution of decision rules across the population. In our present situation this may correspond to the variation of quit responses of employees facing the same vector of attributes. Hence, in this section we shall briefly introduce and consider the relevance of the random utility approach to
the statistical formulation of a model of quit behaviour.

The axioms of random utility theory are presented in contrast with conventional demand analysis where the objects of choice are not lumpy i.e. continuous real-valued. In the latter situation variations due to taste and unobserved attributes of individuals are considered to be uniform across the population. Accordingly the conventional statistical formulation of aggregate demand analysis attributes the disturbance term to measurement error and specification error. Thus, all systematic variations in aggregate demand analysis are interpreted as having been generated by a common variation at the intensive margin of the representative individual's demand (Domenich and McFadden (46), p 50). In this case individuals choose to buy more or less of a commodity, whereas in the case of aggregate qualitative choice the decision is whether to buy or not to buy. In the latter situation systematic variations are all due to shifts at the extensive margin resulting from the distribution of individual's decision rules. Thus, in such models different assumptions are made about the error structure although the basic common assumption of utility maximization is maintained.

The general theory of random utility boils down to the following structure: A randomly selected individual is making a choice from a set of K 'lumpy' alternatives. The jth alternative gives him the utility defined as:

\[ U(x_j, S) = V(x_j, S) + \epsilon(x_j, S) \]

where \( x_j \) is a vector of characteristics of the jth alternative, and \( S \) a vector of individuals' measurable characteristics. \( V(\ldots) \) is a non-stochastic component which reflects the taste of the 'representative' individual with the given vector of attributes for this alternative,
While $\varepsilon(X_j, S)$ is the stochastic ideocyncratic component with mean assumed to be independent of $X_j$. Then using the postulate of utility maximization, it is argued that:

$$\text{Prob(choose } j^{th} \text{ alternative)}$$
$$= \text{Prob} \{ U_j > U_k ; j \neq k = 1, \ldots, K \}$$
$$= \text{Prob} \{ V_j - V_k > \varepsilon_k - \varepsilon_j \} \quad (2.20)$$

Letting $U_i \equiv U(X_i, S)$, $V_i \equiv V(X_i, S)$, and $\varepsilon_i \equiv \varepsilon(X_i, S)$. ($i = 1, \ldots, K$)

Different mathematical models of qualitative choice are then specified depending on the probability distribution structure of the error terms. However, one of the restrictive assumptions which is made to justify the mathematical derivation is that the error terms $\varepsilon_i$ ($i = 1, \ldots, K$) are statistically independent. The derivation of the logistic model can be shown below on the basis of a simple binary choice situation where there are only two alternatives.

In this case we may write:-

$$\text{Prob(choose first alternative)}$$
$$= \text{Prob} \{ V_1 - V_2 > \varepsilon_2 - \varepsilon_1 \}$$
$$\quad (2.21)$$

using the notation of (2.20) above. Then we shall introduce the probability densities $f_1(\varepsilon)$ and $f_2(\varepsilon)$ for the error terms $\varepsilon_1$ and $\varepsilon_2$. The corresponding distribution functions are $F_1(\varepsilon)$ and $F_2(\varepsilon)$ respectively.

Hence, it follows that

$$\text{Prob} \{ V_1 - V_2 > \varepsilon_2 - \varepsilon_1 \} = \int f_1(\varepsilon) f_2(V_2 - V_1 + \varepsilon) \, d\varepsilon \quad (2.22)$$

using the convolution formula.

The random utility theory proceeds to assume particular functional
forms of the distribution functions $F_1(\varepsilon)$ and $F_2(\varepsilon)$, for the error terms $\varepsilon_1$ and $\varepsilon_2$ which are by definition unobservable. One possible choice which justifies the logit is to specify the Weibull Distribution function

$$F_1(\varepsilon) = \exp(-\exp(-c \varepsilon - a_1))$$

and

$$F_2(\varepsilon) = \exp(-\exp(-c \varepsilon + a_2)).$$

Hence,

$$f_1(\varepsilon) = \frac{d}{d\varepsilon} F_1(\varepsilon)$$

$$= e^{-(c \varepsilon + a_1)} \exp(e^{-(c \varepsilon + a_1)})$$

and $f_2(\varepsilon)$ defined by symmetry.

On this basis, expression (2.22) is evaluated as below:

$$\text{Prob}\{V_1 - V_2 > \varepsilon_2 - \varepsilon_1\}^*$$

$$= e^{V_1 - a_1}/(e^{V_2 - a_2 + eV_1 - a_1})$$

then certain simplifications are introduced to express (2.23) in the conventional form of the logistic model. Specifically, the Weibull distribution parameters $a_1$ and $a_2$ are cancelled out by assuming that the error terms $\varepsilon_1$ and $\varepsilon_2$ are identically distributed (Domenich and McFadden (46)). In addition, a linear parameterization is made to the non-stochastic terms $V_1$ and $V_2$ as $V_1 = \beta_1 x_1 + \delta s$ and $V_2 = \beta_2 x_2 + \delta s$, so that it is possible to linearize the difference as

$$V_1 - V_2 = \beta x$$

where the vector $x$ now consists of the differential attributes of the alternatives and the individual's characteristics. This additional simplification implies that we can write the binary choice probability in terms of the conventional logistic model:

*See the derivation in Domenich and McFadden (46).*
Now, we have seen how the above restrictive assumptions have led to the specification of the logit. In particular, the error terms are assumed to be independently identically distributed, with a Weibull functional form. The immediate question is how we can adopt the above theory to derive a quit probability model. In this case, we may think of an employee currently working for employer '2', and who has received an offer from employer '1'. Then (2.24) might be regarded as the employee's quit probability while the 'differential characteristics' in this context stand for differential employment offers, i.e. wage offer and other characteristics of work. However, this approach is subject to certain limitations. First we are not usually in a position to observe alternative employment offers of different employees, so that the differential attributes vector cannot be observed. Secondly, the assumption about the independence and functional form of the error structure is only introduced to simplify the mathematical derivation of the logit and they may not be realistic assumptions. If different functional forms are assumed for the probability distribution of the error structure, we would expect to get different mathematical models of qualitative choice probability. For example, when a Normal probability distribution is assumed for $\varepsilon_1$ and $\varepsilon_2$, it is possible to derive the binary probit model. Hence, the random utility approach justifies the logistic model only under arbitrary restrictive assumptions. Alternatively, we need to base our justification of it on its statistical and computational properties. These considerations are discussed in more detail in the subsequent two chapters and briefly outlined in the next section.
The Logistic Model for Quit Behaviour

The logistic specification of quit probability of an employee can be expressed as a function of the employee's characteristics (personal and family) and his employer's characteristics (e.g. wage offer, racial composition of labour force, etc). If these attributes are all contained in the vector \( x \), then the logistic specification maintains that:

\[
\text{Prob} \{ Y = 1 \} = \frac{e^{\beta' x}}{1 + e^{\beta' x}}
\]

(2.25)

where \( Y \) is a 0, 1 binary random variable which equals one if the decision is to quit, and zero otherwise. The vector \( \beta' = (\beta_0, \ldots, \beta_{p-1}) \) is a vector of parameter coefficients of the explanatory variables vector:

\[ x' = (1, x_1, \ldots, x_{p-1}) \]

The most direct computational advantage of this model is that it reproduces the sigmoid curve of quit probability (figure (2.7)) which stresses the non-determinism of the money wage offer. This curve is always contained within the lower 'zero' asymptote and the upper 'unity' asymptote so that the (0,1) restriction of probability is automatically satisfied. This property is shared by other non-linear
probability models (e.g. the probit) but we shall show in the next chapter that the logistic model has special comparative statistical and computational advantages.

The other property of the logit that makes it appealing to our present study is that it is a basic requirement to our specification of a probability distribution of quit probabilities, which is directly responsive to variations in the explanatory variables. In fact, we shall see in chapter \( \overline{V} \) that the logistic model is a basic component of the \( S_B \) curve to which we have referred via figure (2.5) above. As we have pointed out in section (2.3), p.44, we shall follow Silcock\(^ {128} \) in exploiting a probability distribution of job leaving probabilities, but allow this distribution to be directly translateable from the explanatory variables, \( X \). The derivation of the \( S_B \) curve for the distribution of quit probabilities will be shown in the fifth chapter to depend on the logistic law and the multivariate normality of the explanatory variables, \( X \).
However, the actual data base which we shall utilize for the vector $x$ does not contain employer characteristics - apart from a gross-earnings variable. Thus the analysis only exploits the employees' characteristics variables which we possess in our data base.

2.8 Summary

We have discussed the problem of job-quit behaviour in the light of some selected studies and some data relevant to U.K. We have divided the related models into micro-level and macro-level models. The former models treat the problem of job-quits within the context of a single firm which aims at stabilizing its labour force and reducing wastage. These models rely on the demographic techniques of the life-table. While discussing these models, we have briefly commented on Silcock (128) and Clowes (30) to whom we shall return later in the fifth chapter. On the other hand, the macro-level models treat quit behaviour as a determinant of the general phenomenon of labour mobility. In this approach we notice the roles of neo-classical economic theory, internal labour markets, investment in human capital and job-search theory. However, we have criticized the adoption of the linear probability model, and specifically Parson's (113) linear approximation of quit probability.

We have drawn attention to a pitfall in Viscusi's (144) approach of estimating the effect of hazard perception on average quit intention probability. First, he has been adopting the naive formula of average
probability whose use is very restricted. Secondly, we have argued that such an effect should be measured by the (aggregate) elasticity of average quit probability to hazard perception changes. This alternative approach is based on the principle of deriving aggregate predictions from a dis-aggregate probability model - the main theme of the fifth chapter.

We have questioned the relevance of the random utility approach which is sometimes adopted to justify the logistic model under certain restrictive assumptions. However, we have noted the shortcomings of this theoretical background when applied to quit behaviour. Alternatively, we have supported the logistic model by appealing to its statistical appropriateness and computational convenience as we shall elaborate in the next chapters.

The logistic probability model arises as a special form, although other forms can similarly be derived, as we shall show in the next chapter.
CHAPTER 3

ASPECTS OF PARAMETER ESTIMATION AND INFERENCE

3.1 Introduction

In the last chapter, we have argued that job mobility can be explained in terms of the present circumstances in employment (i.e. employment characteristics), employees' characteristics and their unmeasurable subjective perception of non-pecuniary benefits. We have also considered the relevance of the random utility approach to quit behaviour, the approach which derives a mathematical probability model by postulating the utility maximization axiom and imposing a specific probability distributional structure for the error terms.

We have however criticized this approach as it is not possible to observe alternative job offers, apart from the restrictive assumptions of independence which is made about the error terms. Alternatively, the justification of the logistic model should be sought in terms of its computational convenience, a topic which we shall handle in this chapter in the context of parameter estimation. As regards the special suitability of the logit for the analysis of quit behaviour, this is dealt with in the fifth chapter, where the idea of the distribution of quit probabilities is discussed.

In the last chapter (Section (2.8)) we have expressed job-quit probability, \( \theta \), of an employee as:

\[
\theta = \frac{e^{\hat{\beta}^T x}}{1 + e^{\hat{\beta}^T x}}
\]

where \( x = (1, x_1, \ldots, x_{p-1}) \) is a vector of explanatory variables which contains employees' characteristics (i.e. personal and family), as well as employer's characteristics. However, as we have previously mentioned (section (2.8)), the employers' characteristics variables are not available in our data base - apart from a weekly gross earnings variable. Thus, the analysis and its policy implications are based majorly on the employees' characteristics to which we have access in our data. The parameters coefficient vector \( \hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_{p-1}) \) is assumed unknown and it is the object of this chapter to handle the related estimation problem. The comparative computational advantage of the logit relative to other probability models is considered in the next section. The problems of estimation, significance testing and measurement of goodness-of-fit are discussed in subsequent sections.
A numerical example is supplied at the end of the chapter.

3.2 Other Probability Models

The random utility approach enables us to derive different mathematical models of choice probability depending on the assumed distributional form for the stochastic element of utility. We have been dealing with the binary choice problem whereby two random utilities $U_B$ and $U_C$ are assigned to current employment offer $B$ and alternative offer $C$, respectively, by a random employee with a given set of observed constraints and characteristics. Thus,

$$
\begin{align*}
U_B &= V(X_B, S) + \varepsilon_B \\
U_C &= V(X_C, S) + \varepsilon_C
\end{align*}
$$

(3.2)

where the non-stochastic measurable utilities $V(\cdot)$ depend on employee's observed characteristics, $S$, and characteristics of each of either offer $X_B$ or $X_C$ respectively. The random utility theory of choice asserts that,

$$
\theta \equiv \text{Prob(leave B, accept B)} = \text{Prob}(\varepsilon_C - \varepsilon_B < V_B - V_C) = G(V_B - V_C) = G(\delta'X)
$$

where we let $V_B \equiv V(X_B, S)$, $V_C \equiv V(X_C, S)$ and $\delta'X$ is, as has been defined in (3.1) above, a linear parameterization for the difference $V_B - V_C$. The cumulative distribution function, $G$, is a non-decreasing continuous function of the real scalar, $\delta'X$, and it translates the
latter into the \([0, 1]\) interval. In general the parameters of \(G\) are functions of \(X_B, X_C\) and \(S\). However in order to facilitate empirical analysis, the assumption is made that the parameters of \(G\) are independent of \(X_B, X_C\) and \(S\). Similarly the error terms \(\varepsilon_B\) and \(\varepsilon_C\) are independent of each other and of the vectors \(X_B, X_C\) and \(S\). Apart from the logistic distribution function (3.1), the following forms are sometimes adopted:

\[
G(\beta'X) = \beta'X \quad (3.3)
\]

\[
G(\beta'X) = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}v^2} dv \quad (3.4)
\]

\[
-G(\beta'X) = \frac{1}{\pi} \tan^{-1}(\beta'X) + \frac{1}{2} \quad (3.5)
\]

\[
G(\beta, X) = \beta_0 \beta_1 X_1 \beta_2 X_2 \ldots \beta_{p-1} X_{p-1} \quad (3.6)
\]

Models (3.3), (3.4), (3.5) and (3.6) are called respectively the linear probability model, the probit model, the arctan model and the Gombit model of binary choice; see reviews in Nerlove and Press\(^{106}\), Domenich and McFadden\(^{46}\), Zellner and Lee\(^{151}\). The linear probability model is sometimes adopted for its computational simplicity as it does not require more than the application of the standard least squares results. Let us introduce the 0,1 binary random variable \(Y_i\) which equals one, if the \(i^{th}\) employee decides to leave, and zero otherwise. Then the linear model hypothesis is given as below:

\[
Y_i = \beta' X_i + \varepsilon_i \quad (3.7)
\]

Then if \(E(\varepsilon_i) = 0\) we have \((i = 1, \ldots, n)\)

\[
\begin{align*}
\beta X_i &= E(Y_i) = \text{Prob}(Y_i = 1) = \theta_i
\end{align*}
\]
Moreover, the application of Ordinary Least Square (OLS) requires that

\[ \text{Cov}(Y_i, Y_j) = \begin{cases} 0 & i \neq j \\ \sigma^2 & i = j \end{cases} \quad (i,j = 1, \ldots, n) \]  

Hence the Least Squares Estimator, \( \hat{\beta} \), of the parameter coefficient vector, \( \beta \) is obtained as a solution to the matrix equation below:

\[ (X'X)^{-1} \hat{\beta} = X'Y \]  

where \( X \) is an \( n \times p \) matrix of explanatory variables, \( Y \) is a column vector of \( n \) binary variables. The matrix \( X \) is assumed to have full column rank so that \( (X'X)^{-1} \) does exist. This estimator has been shown in statistical theory to possess certain properties. According to the Gauss-Markov theorem it has the minimum variance among all linear unbiased estimators. In addition, if the error terms \( e_i \) are assumed to be independently identically distributed normal variables, then Rao-Blackwell theorem proves that the least squares estimator is best among all unbiased estimators. In the last case the least squares estimator is also the maximum likelihood estimator.

This is essentially the model which has been adopted by Engelman,\(^{(49)}\), Parson,\(^{(112, 113)}\), Stoikov and Raimon\(^{(134)}\) to describe job-quit probability. It has also been adopted by Bowen and Finegan\(^{(27)}\) in the study of labour force participation rates. However, there are two main reservations which are normally held against the suitability of the linear probability model, namely:

(i) Once the equations system (3.8) has been solved for \( \hat{\beta} \), there is no guarantee that the probability estimates:

\[ \hat{\theta}_i = \hat{\beta} X_i \]  

\[ (3.9) \]
will obey the 0,1 restriction of probability for all \( i \) \((i = 1, \ldots, n)\).

(ii) Since \( Y_i \) is a Bernoulli random variable, then its variance is given by

\[
\text{Var}(Y_i) = E(Y_i^2) - \left[ E(Y_i) \right]^2 = \theta_i (1 - \theta_i) \quad (i = 1, 2, \ldots, n) \tag{3.10}
\]

which defies the constant variance assumption (3.7) above. This implies that \( \hat{\theta} \) of OLS could be inefficient though unbiased. Let us first consider problem (i) above:

The first problem is sometimes resolved by imposing

\[
\hat{\theta}_i = 1 \text{ if } \hat{\theta}_i > 1 \quad \text{and} \quad \hat{\theta}_i = 0 \text{ if } \hat{\theta}_i < 0 \tag{3.11}
\]

The effect of this ad hoc procedure is that we predict the event to occur with certainty for some values of the explanatory variables, while for other values we predict that it is impossible to occur. This limitation is referred to as prediction bias. This problem has been explored by Domenich and McFadden\(^{(46)}\). They have shown that when such extreme ranges of explanatory variables are present and a linear model is fitted, then the magnitude of the parameter estimates are substantially biased below their true values. As a result

"The linear probability model will tend to underestimate the elasticity of the response with respect to explanatory variables for individuals in the intermediate probability range, and overestimate this elasticity in the extreme probability range."

\[(\text{Domenich and McFadden}^{(46)}; \ p. \ 106)\].

Hence, the optimum properties of the least squares method, assuming that
model (3.3) is valid, will be suppressed by prediction bias and specification error if the range of values of explanatory variables violates the 0,1 restriction.

This restriction could be allowed for by solving a problem in quadratic constrained minimization, i.e.

$$\min_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \quad \text{subject to} \quad 0 \leq \beta X_i \leq 1.$$ 

The solution vector $\tilde{\beta}$ of this complex minimization problem is biased but it is more efficient than the OLS estimator $\hat{\beta}$ in the sense of having a distribution which is more concentrated around the true parameter $\beta$. However, this inequality constrained minimization method is not only costly, but it has also been shown to be more sensitive to specification error, and does not eliminate the prediction bias. Moreover, since the inequality constraints apply only to the data base, the 0,1 problem may not be solved if the solution vector $\tilde{\beta}$ is to be applied for prediction on data. For these reasons the use of the inequality constrained model is unrecommendable since the OLS method could even be better.

Apart from the above problems, specification error may occur with the linear probability model if the valid specification is a response curve.

The second problem of heteroscedasticity is sometimes ignored whenever the individual probability estimates, $\hat{\theta}_i$, fall within the range (0.2, 0.8); see Cox$^{(35, \text{p. 167})}$. In this case the variation in individual variances, $\hat{\theta}_i(1-\hat{\theta}_i)$ could be negligible. Note that if this condition is satisfied then the other optimum properties of OLS method could be achieved — there will be no problem of prediction bias and specification error — when the model is valid. However,
if the problem of heteroscedasticity is deemed serious, then weighted least squares could be applied to model (3.12) below as:

$$E \left( \frac{Y_i}{V_i} \right) = \sum_{S=1}^{P} \frac{X_{Si}}{V_i} \beta_S \quad (i = 1, \ldots, n) \quad (3.12)$$

where $\hat{V}_i = \hat{\theta}_i (1 - \hat{\theta}_i)$, and the randomness in the estimates, $\hat{V}_i$, is ignored (Cox(35); p. 31). It is also possible to adopt an iterative estimation procedure for the weights, $\hat{V}_i$. In each iteration the weights are re-estimated until successive estimates $\hat{\beta}_S$ seem to converge; see Tag el-Din(135). However, this approach of weighted least squares has been shown by Goldberg(56) to place excessive weight on extreme observations leading to more unstable estimates than those of the OLS method; moreover the model becomes more sensitive to specification error. Thus past experience has shown that whenever the linear probability specification is assumed, there is no real gain in precision by introducing the weighted least squares version or the above inequality-constrained one.

The problems of prediction bias, specification error and heteroscedasticity are automatically solved by the adoption of a sigmoid-shaped non-linear probability model, e.g. the logistic model (3.1), the probit (3.4), the arctan (3.5) or the Gombit (3.6). These models yield sigmoid curves or ogives bounded within the $[0, 1]$ interval when taken as functions of a single independent variable. There are other ogives which could be adopted in some special applications, such as the models specified below:

$$G(V) = \begin{cases} -(V_o - V) & V < V_o \\ 1 & V \geq V_o \end{cases} \quad (3.13)$$

$$G(V) = \begin{cases} 1 - e^{-(V_o - V)} & V > V_o \\ 0 & V \leq V_o \end{cases} \quad (3.14)$$
\[ G(V) = \frac{(V-V_o)/(V_i-V_o)}{V_o \leq V \leq V_i} \quad (3.15) \]

These truncated forms have been considered by Domenich and McFadden\(^{(46)}\).

Notice that model (3.1) is a cumulative exponential probability distribution giving a one-tailed ogive curve for \( V < V_o \), while model (3.1) gives a one-tailed ogive curve for \( V > V_o \). Model (3.15) is a truncated linear model.

We now turn to the more relevant forms specified by models (3.1), (3.4), (3.5) and (3.6) above. Each of these models could be a rival for the 'true' probability model of binary choice. However, as we shall see in the next section, past experience has given evidence that there is little to choose among these models apart from computational convenience. This consideration mainly accounts for our choice of the logistic model and it leads us to the discussion of the estimation problem.

3.3 Estimation Methods for Non-linear Binary Probability Models

There are two main approaches of estimation for non-linear probability models which are currently adopted. The first approach and the one which is least relevant to our study is due to Berkson\(^{(20)}\). This approach requires sufficiently repeated observations on the \( i^{th} \) vector \( X_i \) of characteristics. These observations could be generated by repeated experimentation under identical conditions described by \( X_i \), as in biological or medical research. It is also possible to a limited extent to adopt it in economics, e.g. in a binary choice model of households' decision on the purchase of consumer durables (Amemiya

\[^*\] In practice identical conditions cannot be precisely achieved.
and Nold (7). An interesting review and discussion of this approach is given by Cox (35) but it is not outlined here. However, it is worthwhile mentioning two limitations of this approach that have been noted in past experience when applied to survey data. First, since some explanatory variables are continuous (e.g. age, income, travel time) we should discreticize them in order to define the cells. As a result, each cell is defined by a unique vector \( X \) which may consist of binary or polychotomous variables which are either naturally discrete, or have been discreticized. However, this process of discreticizing continuous variables involves sacrifice of statistical information as variations are being lumped together. Secondly, the number of cells containing the sampled cases multiply up rapidly as the number of explanatory variables of the vector \( X \) increases, or as finer groupings are made for the continuous variables. For example, a model with five explanatory variables which are all binary will call for \( 2^5 = 32 \) cells, while they rise up to \( 3^5 = 243 \) cells if the variables are all trichotomous. But, in Survey methods the research worker cannot control the \( X_i \)'s to get the desired number of repetitions unless the sample size is very large. Even for a moderate dimension of \( X_i \) to get additional repetitions in every cell, or in some cells, may require increasingly larger samples. This problem is specially difficult when there are certain combinations of \( X \) that are relatively rare to find in survey data, e.g. where \( X \) defines households where the wife does market work, and the husband does home work as it is a theoretical possibility. In addition to these practical problems, it has been shown that Berkson's regression method (which uses cells' means as explanatory variables) introduces an error in variables effect which tends to cause an upward bias in the final estimates. This point has been made and attributed to the process of discreticization by Domenich and McFadden (46) p. 110).
On the other hand, the maximum likelihood approach does not require more than one observation per cell, and it does incorporate all statistical variations as they are. Its only limitation is that the estimates are only calculated numerically as there are no simple analytical solutions for the system of likelihood equations. Let $Y_i$ be the binary choice random variable as defined in model (3.6) above. Similarly, let $\theta_i = \text{Prob}(Y_i = 1)$, then the likelihood function is defined as:

$$L(\theta|Y) = \prod_{i=1}^{n} \theta_i^{Y_i}(1 - \theta_i)^{1-Y_i}$$  \hspace{1cm} (3.16)

where $Y' = (Y_1, ..., Y_n)$ is a set of observed independent binary responses of the randomly sampled individuals. The likelihood is a function of the vector of unknown parameters $\theta' = (\theta_0, \theta_1, ..., \theta_{p-1})$ which determine individual choice probabilities, $\theta_i$ (i = 1, ..., n).

The shape of the likelihood function is, thus, determined by the underlying probability model of the $\theta_i$ and the data on which it is conditioned. The main idea of the maximum likelihood approach is that the estimate of $\theta$ which makes observed data most likely to turn up is the one which is closest to the true value, $\bar{\theta}$. We shall deal with the situation where the $\theta_i$'s follow our logistic model (3.1) although the method can be generalized to other non-linear probability models. Hence, we get the likelihood function:

$$L(\theta|Y) = \frac{\Sigma_{t=0}^{n} \Sigma_{s=0}^{p-1} \beta_s x_{is}^{Y_i}}{\prod_{i=1}^{n} (1 + e^{\Sigma_{s=0}^{p-1} \beta_s x_{is}})$$  \hspace{1cm} (3.17)

where $t_s = \Sigma_{i=1}^{n} x_{is} Y_i$ for $s = 0, ..., p-1$.

Note that the statistics, $\{t_s\}$, are minimal sufficient statistics for unknown parameters $\theta$, using the factorization theorem (see Silvey\(^{129}\), p. 27). Cox\(^{35}\) has observed that both OLS estimates of the linear
probability model and the maximum likelihood estimates of the logistic model are functions of the same sufficient statistics \( \{t_s\} \). This point follows by noting that the right-hand side of equation (3.8) can in fact be written as \( X'Y = t' = (t_0, \ldots, t_{p-1}) \). This point has been extended to the multinomial model analogue of the linear probability model as related to the multinomial logit by Tag el-Din (135).

For computational simplicity it is normally recommendable to work with the log-transformation of model (3.17) above:

\[
\ell(\hat{\beta} | Y) = \log L(\hat{\beta} | Y) = \sum_{s=1}^{p} t_s \hat{\beta}_s - \sum_{i=1}^{n} \log(1 + e^{\sum_{s=1}^{p} \hat{\beta}_s X_{si}})
\]

where \( \ell(\hat{\beta} | Y) \) is the log-likelihood function. The maximum likelihood estimates \( \hat{\beta} \) are then obtained by maximizing \( \ell(\hat{\beta} | Y) \) over the range of \( \hat{\beta} \). The first order condition of the maximum is given by the vector-valued equation below:

\[
\frac{\partial}{\partial \hat{\beta}_s} \ell(\hat{\beta} | Y) = 0
\]

whose solution under certain regularity conditions yields the maximum likelihood estimates, \( \hat{\beta} \). Hence the likelihood equations for the logistic model are given by the system:

\[
\frac{\partial}{\partial \hat{\beta}_s} \ell(\hat{\beta} | Y) = t_s - \sum_{i=1}^{n} X_{si} \frac{\sum_{s=1}^{p} \hat{\beta}_s X_{si}}{1 + \exp(\sum_{s=1}^{p} \hat{\beta}_s X_{si})} = 0 \quad (s = 0, \ldots, p-1)
\]

McFadden (95) has shown that the log-likelihood function (3.18) of
the logistic model is concave in the real parameter vector space, \( \beta \) implying that there is a unique maximum likelihood estimator \( \hat{\beta} \) whenever a maximum exists. The most common numerical procedure adopted for the solution of the system of likelihood equations of the form (3.19) is the Newton-Raphson iterative method as recommended by Silvey\(^{(129)}\). Dixon\(^{(45)}\) discusses the different versions of this procedure which depends on matrix inversion. Basically, this method starts off with an initial guess \( \beta_0 \) which is updated iteratively so that in the \((n+1)\)th iteration we get:

\[
\hat{\beta}^{(n+1)} = \hat{\beta}^{(n)} - [D^{(n)}(\ell)]^{-1} d^{(n)}(\ell)
\]

where

\[
D^{(n)}(\ell) = \frac{\partial^2}{\partial \beta_i \partial \beta_j} \ell(\beta|Y); \ i,j = 0, \ldots, P-1
\]

i.e. the matrix of second order partial derivatives of the log-likelihood function with respect to components of \( \beta \), evaluated at the \( n \)th value, \( \hat{\beta}^{(n)} \). The vector \( d^{(n)}(\ell) \) contains the first order partial derivatives of the log-likelihood, similarly, evaluated at \( \hat{\beta}^{(n)} \): i.e.

\[
d^{(n)}(\ell) = \left\{ \frac{\partial}{\partial \beta_s} \ell(\beta|Y); \ s = 0, 1, \ldots, p \right\}_{\beta = \hat{\beta}^{(n)}}
\]

It is easy to show that

\[
\frac{\partial^2}{\partial \beta_i \partial \beta_j} \ell = - \sum_{k=1}^{p} \frac{\beta_j X_{ik} X_{jk} \sum_{s=1}^{p} \beta_s X_{sk}}{1 + \exp(\sum_{s=1}^{p} \beta_s X_{sk})^2}
\]

while the \( s \)th elements of \( d(\ell) \) has already been given by equations (3.20). The basic Newton Raphson procedure requires at each iteration the inversion of a new matrix \( D^{(n)}(\ell) \) and this has been criticized as
involving excessive computations. Silvey (129) has proposed the replacement of the matrices $D^{(n)}(\xi)$ by their mathematical expectation matrix, $E\left[D^{(n)}(\xi)\right]$ - provided that it is non-singular - and its application at all iterations. This makes the convergence process slower, but substantially reduces the computations. Other modified versions are discussed by Dixon (45). Incidentally, for the logistic model we notice that:

$$E\left[D(\xi)\right] = D(\xi)$$

(3.23)

since $D(\xi)$ is independent of the random-valued vector $Y$. The mathematical expectation of the matrix $D(\xi)$ has another important value. It has been shown that when the log-likelihood is well-approximated by a quadratic function of the parameters at the neighbourhood of its maximum, then the matrix:

$$V(\hat{\beta}) = \left(-D(\xi)\right)^{-1}\bigg|_{\beta = \hat{\beta}}$$

(3.24)

provides a consistent estimate of the variance-covariance matrix of the maximum likelihood estimates, $\hat{\beta}$; e.g. see Cox and Hinkley (34, p.302) Edwards (47, p.195), and Silvey (129, p. 78). When the matrix $D(\xi)$ is negative definite this implies that the log-likelihood is strictly concave and thus the second order condition for the maximum is satisfied, and it should be unique (McFadden (95, p.116). It has also been shown that the maximum is virtually certain in empirical samples of more than twenty observations if the multi-collinearity of the explanatory variables is not serious*. In particular, McFadden (95) has proved the uniqueness

* As Johnston (78, p. 91) has remarked, there is bound to be some multi-collinearity as the hypothesis of orthogonality of the matrix $(X'X)$ is always rejected in economics. Dodonea (41) considered a coefficient of determination $R^2 > 0.7$ as indicative of serious collinearity.
of the maximum on the basis of the logistic model whenever a maximum exists.

In fact the curvature properties of the logistic model makes it computationally more tractable than either of the non-linear probability models mentioned above: the probit (3.4), the arctan (3.5), and the Gombit (3.6). These four ogives are virtually indistinguishable except at arguments yielding extreme probabilities. In a Monte Carlo simulation experiment Domenich and McFadden have shown that the maximum deviation in probability between the probit and logit curves is 0.018, and that between arctan and logit is 0.082. The probit model has been shown to approach extreme values most rapidly, and the arctan model least rapidly, but the deviations in probability between the three curves are always negligible. As regards the Gombit curve, Rogers and Hansen have given empirical evidence that it is virtually indistinguishable from the logit.

Thus, except for computational convenience there is little to choose among these models. In fact, except for a few scattered cases, most related empirical research adopts either the logit or the probit with greater emphasis given to the former model; e.g. see Berkson, Cox, Nerlove and Press, Dodonela, Anderson, Talvitie, Westin, Watson and Westin, Schmidt and Strauss, Ben-Akiva, Richard and Ben-Akiva, Domenich and McFadden, Amemiya, Denton, Akerlof and Main. The computational advantage of the logit over the probit arises from the fact that the logistic distribution function is a closed (or explicit) functional form with convenient curvature properties for numerical maximization. On

* Similar advice has been given to me by Professor Silvey, namely, that as the logistic model is a well-behaved function, the Newton-Raphson procedure should converge to the maximum whenever the latter exists.
the other hand, the probit model is an integral equation which cannot be
expressed in a closed form and so it is comparatively intractable.

3.4 Significance Testing and Measurement of Fit

A typical problem of hypothesis testing is the one of testing the
null hypothesis represented by the system of q-linear constraints on
the p coefficient parameters, \( \beta \):

\[
H_0 : \quad H\beta = 0
\]  

(3.25)

within the more general unrestricted model:

\[
H_1 : \quad H\beta \neq 0
\]

where \( \beta = (\beta_0, \ldots, \beta_{p-1}) \), and \( H \) is a \( q \times p \) matrix of real constants.
The theoretical and practical problems of statistical hypothesis testing
are discussed in various advanced text books, e.g. Cox and Hinkley\(^{(34)}\),
Rao\(^{(118)}\), Silvey\(^{(129)}\). A more common situation arises when the system
of constraints (4.25) is replaced by a single linear constraint on the
parameter coefficients, \( \beta \) i.e.

\[
H_0 : \quad h'\beta = 0
\]  

(3.26)

where \( h \) is a column vector of real constants. The simplest situation
is that where all the components of \( h \) are zero except for one in
the \( s \)th position (\( s = 0, \ldots, p-1 \)). In this case the hypothesis
(3.26) merely indicates that \( \beta_s = 0 \). The general statistical techniques
of significance testing for the binary logit are discussed by Cox\(^{(35)}\),
Nerlove and Press\(^{(106)}\) and others. When the likelihood approach is
adopted, inferences about \( \beta \) are made on the basis of the likelihood
estimates, \( \hat{\beta} \) and the asymptotically efficient covariance matrix \( V(\hat{\beta}) \); see equation (3.24).

Hence, let \( v_{ss} \) be the \( s \)th diagonal element of \( V(\hat{\beta}) \). Then the t-ratio for testing \( \beta_s = 0 \) is defined approximately as,

\[
t_s = \frac{\hat{\beta}_s}{\sqrt{v_{ss}}} \quad (s = 0, \ldots, p - 1).
\] (3.27)

Similarly, if \( 0 < \alpha < 1 \) we may define a \((1 - \alpha)\%\) confidence interval for \( \beta_s \) as

\[
\hat{\beta}_s \pm t(n, \alpha) \sqrt{v_{ss}}
\] (3.28)

where \( t(n, \alpha) \) is the tabulated t value at \( \alpha \)-level of significance and \( n \)-degrees of freedom. More generally, to test the hypothesis that a subset of \( q \) components of the \( p \)-vector \( \hat{\beta} \) are zero (i.e. hypothesis (4.25) with \( H = (I_q', 0) \)) we calculate the log-likelihood ratio statistic:

\[
\ell(\hat{\beta}^*) - \ell(\hat{\beta}) = -2\log\lambda
\] (3.29)

where \( \ell(\hat{\beta}) \) is the unrestricted maximized log-likelihood function and \( \ell(\hat{\beta}^*) \) is the maximized function with \( \hat{\beta} \) restricted to contain only \( q \) non-zero parameters. Then, if this null hypothesis is true, \(-2\log\lambda\) will be asymptotically distributed as chi-squared with \( p - q \) degrees of freedom (Cox \( ^{(35)} \)).

The log-likelihood ratio is also used to provide a summary statistic to assess the explanatory power of the whole vector \( X \), and thus measures the goodness of fit of the model. The calculation of \(-2\log\lambda\) is specially easy when the vector of explanatory variables \( X \) contains a constant term (i.e. \( X_{oi} = 1 \) all \( i \)). In this case the first equation of the system of likelihood equations (3.20) gives:
\[
t_0 = \frac{\sum_{i=1}^{n} X_{oi} \frac{e^{\hat{\beta}'X_i}}{1 + \exp(\hat{\beta}'X_i)}}{\sum_{i=1}^{n} X_{oi} Y_i} = 0 \quad (3.30)
\]

where \( t_0 = \sum_{i=1}^{n} X_{oi} Y_i \equiv \sum_{i=1}^{n} Y_i \). Thus, it follows that

\[
R = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{\theta}_i = n \hat{\theta} \quad (3.31)
\]

and \( \hat{\theta} \) the average of estimated probabilities \( \{\hat{\theta}_i\} \) where \( R \) is the number of 'successes' (i.e. \( Y_i = 1 \)). Hence, when there is a constant term in \( \mathbf{X} \), the method of maximum likelihood makes the sample proportion of successes \( \frac{R}{n} \) always equal to the average of estimated probabilities, \( \hat{\theta} \), where, as before:

\[
\hat{\theta}_i = \frac{e^{\hat{\beta}'X_i}}{1 + e^{\hat{\beta}'X_i}}. \quad (3.32)
\]

However, if the vector \( \mathbf{X} \) does not include a constant term, the relation (3.31) may only hold approximately, with the degree of approximation improving as the components of \( \mathbf{X} \) tend to capture the greatest variability in the data.

Now, suppose the vector \( \mathbf{X} \) includes a constant term, but it is hypothesized that its components explain nothing at all of the variability in observed responses, i.e. \( \beta_s = 0 \) for all \( s = 1, \ldots, p-1 \). In this case all \( \hat{\theta}_i \)'s will be equal to the population parameter \( \theta \), or

\[
\hat{\theta}_i = \frac{e^\theta}{1 + e^\theta} = \hat{\theta} = \frac{R}{n} \quad (i = 1, \ldots, n) \quad (3.33)
\]
Returning to \(-2\log \lambda\) as defined in equation (3.29) we express the restriction that \(\beta_0 \neq 0\) and \(\beta_s = 0\) (s = 1, ..., p-1) as \(\hat{\beta}^* = (\beta_0, 0)\) where \(\hat{\beta}^*\) is the restricted parameter vector under the above null hypothesis. We may also note that

\[ \hat{\beta}_0 = \log \left( \frac{R}{n-R} \right) \]

using equalities (3.33). Hence we can write the log-likelihood function under the above null hypothesis as

\[ \ell(\hat{\beta}^*) = R \log \left( \frac{R}{n-R} \right) - (n - R) \log \left( \frac{n-R}{n} \right) \]  \hspace{1cm} (3.34)

This is the formula which has been adopted by Dodonea\(^{(91)}\). Hence the log-likelihood ratio, \(-2\log \lambda\), can easily be calculated and then compared with the tabulated chi-squared value with p-1 degrees\(^{+}\) of freedom with a given significance level.

3.5 Other Measures of Goodness of Fit

There exists a wide variety of summary statistics proposed to evaluate the general performance of a non-linear probability model; e.g. see Domenich and McFadden\(^{(46)}\), McFadden\(^{(95)}\), Morrison\(^{(102)}\), Neter and Meyenes\(^{(107)}\), Walker and Duncan\(^{(146)}\), and others. Some measures attempt to test the accuracy with which the fitted values of the estimated model approximate with the observed data values. This is done with analogy to linear regression analysis where a coefficient of determination very close to one implies a very good fit. However, the main limitation

\(^{+}\) since in this case \(\hat{\beta}^*\) contains only one non-zero coefficient.
of this approach, as pointed out by Morrison\textsuperscript{(102)}, is that in the case of a disaggregate probability model the true individual probabilities are not observed, and thus, to test how estimated values approximate with observed values in a binary choice model needs special care. One such measure which is based on simple analogy with linear regression analysis is defined as

\[ R^2 = 1 - \frac{S(\hat{\beta})}{S(\beta^*)} \]  

(3.35)

where \( S(\hat{\beta}) = \sum_{i=1}^{n} (Y_i - \hat{\theta}_i(\hat{\beta}))^2 / \hat{\theta}_i(\hat{\beta}) \) and \( S(\beta^*) \) is defined by symmetry - recall that \( \hat{\beta} \) is the vector of maximum likelihood estimates, and \( \beta^* \) is the estimated vector under a given null hypothesis. We shall continue to define \( \hat{\beta}^* = (\beta_0, \Omega) \), i.e. only the constant term, \( \beta_0 \), is assumed to be non-zero under the null hypothesis. McFadden\textsuperscript{(95)} has shown that the values of this index are roughly comparable to the multiple correlation coefficient in ordinary least squares. Another measure which has also been considered by Domenich and McFadden\textsuperscript{(46)} and McFadden\textsuperscript{(15)} as comparable to the multiple correlation coefficient is the likelihood index defined as:

\[ \rho^2 = 1 - \frac{\Lambda(\hat{\beta})}{\Lambda(\beta^*)} \].  

(3.36)

If the sample size is sufficiently large, it has been shown that \((p/p-1)(\rho^2/(1 - \rho^2))\) is distributed approximately as \( F(p-1, p) \) where \( p \) is the dimension of the vector \( \hat{\beta} \) while the dimension of \( \beta^* \) in this case is equal to one as it contains only one non-zero parameter. Domenich and McFadden\textsuperscript{(46)} have shown that the measure \( R^2 \) always tends to give a higher value than the log-likelihood index, \( \rho^2 \).

However, according to Morrison\textsuperscript{(102)}, the two situations of the
general linear model and binary probability model are not directly comparable. In the former situation we do observe both estimated and actual quantities, while in the latter situation we do not observe the actual probabilities. The 0,1 binary responses are clearly not probabilities and so they cannot be directly compared with estimated probabilities - as in the $R^2$ of (3.35) - to measure the closeness of the estimated probabilities to the true probabilities. Even if the model provided a perfect fit in the sense of the true and estimated probabilities being the equal, it would not coincide with the 0,1 outcomes. Morrison\(^{(102)}\) demonstrated this point on the basis of the general definition of the coefficient of determination as the percentage of explained variation ($TV - UV$) to total variation, $TV$ ($UV$ stands for unexplained variation). Thus, the coefficient of determination is defined as:

$$r^2 = \frac{TV - UV}{TV} \quad (3.37)$$

In order to define these quantities, $TV$ and $UV$, Morrison\(^{(102)}\) introduces a probability density function $g(\theta)$ for the individual choice probabilities. Then, the average choice probability or the proportion of successes in the population is given by

$$\theta = \int_0^1 \theta g(\theta) d\theta \quad (3.38)$$

and the proportion of failures is given by $(1 - \theta)$. On this ground Morrison defines total variation as:

$$TV = \theta (1 - \theta ) \quad (3.39)$$

Now, as each individual's response is described by a 0,1 binary random
variable the quadratic prediction errors are defined by Morrison as $(0 - \theta)^2$ if the actual outcome is zero and the true probability prediction is $\theta$. Similarly, if the actual outcome is one, the prediction error is $(1 - \theta)^2$. He then defines the 'mean quadratic error' for the whole population or the unexplained variation as

$$UV = \int_0^1 [(0 - \theta)^2(1 - \theta) + (1 - \theta)^2\theta] g(\theta)d\theta$$

which he simplifies to

$$UV = \int_0^1 \theta^2(1 - \theta)g(\theta)d\theta + \int_0^1 \theta(1 - \theta)^2g(\theta)d\theta . \quad (3.40)$$

Later Morrison recommends the use of the Beta probability density to represent $g(\theta)$:

$$g(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1} \quad 0 < \theta < 1. \quad (3.41)$$

This distribution has been adopted due to its geometrical flexibility, as it is capable of representing unimodal shapes, dish-shapes, or rectangular shapes, depending on its two parameters $\alpha$ and $\beta$. Moreover, its first two moments are simple functions of these two parameters. (Morrison, p. 69). Using this set-up Morrison gets the coefficient of determination $r^2$ as

$$r^2 = \frac{1}{\alpha + \beta + 1} . \quad (3.42)$$

The main analytical conclusion which he has derived on the basis of this coefficient is that the upper bound of $r^2$ is not always one, as in
linear regression analysis. Thus a perfect probability model, whose estimated and true probabilities are equal will only attain its upper bound of $r^2$ which is less than one. This coefficient will attain the maximal upper limit (one) only if $\alpha, \beta \to 0$, but this is the extreme case where true probabilities are exactly equal to the observed 0,1 binary responses. (Note that as $\alpha, \beta \to 0$ then the mass of the Beta distribution will tend to be concentrated at $\theta = 0$ and $\theta = 1$).

This study reveals to us that the corresponding coefficient of determination for any probability model does not necessarily possess the upper bound property of the linear regression model. However, it does not specify a precise means for knowing a priori the upper bound of $r^2$ for any given model. It also relies on the Beta model which is an arbitrary choice. In fact we shall argue in the fifth chapter that the Beta model is not the best flexible distribution to describe binary response probabilities.

We have carried this procedure further to derive the variance ratio:

$$F^* = \frac{\left( \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^2 - \bar{\theta}^2 \right)}{\frac{\sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i)}{p - 1}}$$

which is calculated on the basis of the estimated response probabilities, $\hat{\theta}_i$ while $\bar{\theta}$ has been defined in (3.33). This $F^*$-statistic provides a summary measure for testing the general performance of the model similar to the log-likelihood ratio statistic mentioned above. It is also directly based on estimated probabilities rather than a hypothetical $g(\theta)$ model. However its intuitive appeal depends on the inclusion of a constant term in the vector $X$ of explanatory variables. We have previously seen that the inclusion of the constant term accounts for
equalities:

\[
\sum_{i=1}^{n} \hat{\theta}_i = n \hat{\theta} = R = \sum_{i=1}^{n} Y_i. \tag{3.44}
\]

Recall that the null hypothesis \( \beta = \beta^* = (\beta_0, 0) \) is equivalent to stating \( \theta_i = \hat{\theta} \) \((i = 1, 2, \ldots, n)\). Then, given conditions (3.44) above, it follows that

\[
\sum_{i=1}^{n} \hat{\theta}_i^2 = n \hat{\theta}^2 \text{ if and only if } \hat{\theta}_i = \hat{\theta} \tag{3.45}
\]

\((i = 1, \ldots, n)\)

which implies that

\[
\delta \equiv \sum_{i=1}^{n} \hat{\theta}_i^2 - n \hat{\theta}^2 = 0 \text{ if and only if } \beta' = (\beta_0', 0)
\]

\[
= \beta^*
\]

where \( \beta_0 \neq 0 \) is the constant term and \( 0 \) is a null \((p-1)\)-vector.

Hence the statistic \( \delta \) measures the deviation of the model from the null hypothesis of zero explanatory power of all components of \( \mathbf{X} \).

Since we have \( p-1 \) explanatory variables, \( \delta/(p-1) \) measures 'mean explained variation' per single variable. Of course if a variable's parameter coefficient is very close to zero, then the removal of this variable could raise the quantity \( \delta/(p-1) \) and possibly the \( F^* \) ratio (2.43).

The denominator of this statistic gives the 'mean unexplained variation', or an estimate of the residual variance. It is interesting to note that if the total variation \( n \hat{\theta} (1 - \hat{\theta}) \) is partitioned into explained variation and unexplained variation, we find
\[ n \hat{\theta} (1 - \hat{\theta}) \equiv \sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i) + \delta \]  

(3.46)

where \( \delta = \left( \sum_{i=1}^{n} \hat{\theta}_i^2 \right) n \) measures the variation explained by the set of variables \( X' = (X_1, ..., X_p) \), while \( \sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i) \) measures the residual variation\(^\dagger\). The above identity follows directly by recalling \( \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i \). It is analogous to the Analysis of Variance lemma where the total variation or total sum of squares is partitioned into explained sum of squares and unexplained sum of squares.

For example, in a 'one-way classified' data we may be interested to test the null hypothesis that the means of \( n \) populations are all equal, on the basis of \( n \) sample means; see Rao (118, pp. 244-245).

In this case the deviation from the null hypothesis is tested by the quantity \( \sum_{i=1}^{n} n_i \bar{X}_i^2 - n \bar{X}^2 \), where \( \bar{X}_i \) are the sample means for sample sizes \( n_i \) and \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} \bar{X}_i \) is the grand mean of all data.

In our case we take \( n_i = 1 \) while the \( \hat{\theta}_i 's \) correspond to the sample means, \( \bar{X}_i \) and the \( \hat{\theta} \) corresponds to the grand mean \( \bar{X} \). Thus the way our \( F^* \) ratio is defined makes it comparable to the \( F \)-ratio of the conventional Analysis of Variance table. We may show this analogy by another result due to Cochran\(^{31} \) that if the proportion of successes of a binary response phenomenon in a population is estimated by

\[ P = \frac{1}{n} \sum_{i=1}^{n} Y_i \], where \( Y_i \) are \( 0,1 \) binary variables, then it follows that:

\[ \text{UV} = \int_{0}^{1} \theta (1 - \theta) g(\theta) d\theta = E(\theta (1 - \theta)) \]  

However, our definition \( \sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i) \) is based directly on estimated probabilities.

\(^\dagger\) It is easy to see that the latter quantity is comparable with Morrison's definition of unexplained variation as equation (3.40) can be simplified to give:

\[ \text{UV} = \int_{0}^{1} \theta (1 - \theta) g(\theta) d\theta = E(\theta (1 - \theta)) \]  

However, our definition \( \sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i) \) is based directly on estimated probabilities.
\[ p(1 - p) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - p) \quad \text{(3.47)} \]

Thus, noting that \( \hat{\theta} = \bar{p} \) (see equalities (3.44) we can express our total sum of squares as:

\[ n \hat{\theta} (1 - \hat{\theta}) = \sum_{i=1}^{n} (Y_i - \hat{\theta})^2 . \]

In fact, if we behave as though the individual \( \theta_i \) are sample properties of individual cells \( i = 1, \ldots, n \) and let \( Y_{ij} \) be a 0,1 binary response variable for the \( j \)th individual in the \( i \)th cell, then the probability of falling in the \( i \)th cell is estimated by

\[ \hat{\theta}_i = \frac{1}{m} \sum_{j=1}^{m} Y_{ij} \]

assuming that each \( i \)th (characterized by \( X_i \)) contains \( m \) individuals. Then it is easy to derive the analogue of identity (3.46). In this case

\[ \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij} - \hat{\theta}_i)^2 , \]

\[ \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij} - \hat{\theta}_i), \]

while \( \hat{\theta} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} \). Note that if \( m = 1 \) then this decomposition coincides with out identity (3.46).

Therefore if the dependent variables \( Y_i \) - and thus the parameter estimates, \( \hat{\beta} \) - are normally distributed, our \( F^* \)-test boils down to the conventional \( F(p-1, n-p) \) statistic as in this case the identity (3.46) is nothing different from the ordinary partition of the total sum of squares based on the Analysis of Variance Lemma. However, in practice the \( F \)-test and the \( t \)-statistics are always used without justifying the normality assumption for the sample estimates - e.g. in the linear probability model where the dependent variable is the 0,1 binary variable, \( Y_i \). In our case with the maximum likelihood
estimates $\hat{\beta}$, the Normality assumption$^+$ is used as a large sample property (Cox and Hinkley (34), Silvey (129))

$$\hat{\beta} \rightarrow N(\beta, \left[D(\beta)\right]^{-1})$$

and thus our $F^*$-ratio could be approximated by the $F(p-1, n-p)$ statistic based on Normal theory. In general, the $t$-test, and hence the $F$-test, are described as robust tests which do not rely on strict Normality; see Rao (118, p. 498). However, it may be desirable to carry a rigorous robustness test for our $F^*$ as a problem of future research. Yet, at present, we shall utilize it for its intuitive appeal and compare it with the tabulated $F$-values.

There are other goodness-of-fit measures which are based on the ability of the model to forecast observed responses. The simplest measure is defined as

$$E = \frac{\sum_{i=1}^{n} \left[\delta_i Y_i + (1 - \delta_i)(1 - Y_i)\right]}{n} \quad (3.48)$$

where

$$\delta = \begin{cases} 
1 & \text{if } \hat{\delta}_i > 0.5 \\
0 & \text{if } \hat{\delta}_i \leq 0.5 
\end{cases}$$

It is obvious that if $Y_i = 1$ is always associated with $\delta_i = 1$, and $Y_i = 0$ always associated with $\delta_i = 0$, then $E = 1$. This statistic measures the proportion of correct classifications on the basis of the estimated model. It is sometimes referred to as the diagnostic accuracy of the model (Teather (138)). A similar measure has been defined by

$^+$ In fact this is the basic underlying assumption for the utilization of the approximate $t$-ratios (3.27) and the confidence intervals (3.28).
Morrison (102) as:

\[
E = \int_{0}^{\frac{1}{4}} (1 - \theta)g(\theta)d\theta + \int_{\frac{1}{4}}^{1} \theta g(\theta)d\theta
\]  

(3.49)

which gives fairly close results to (3.48). However this quantity is criticized for its non-allowance for costs of misclassification (Domenich and McFadden (46)). Thus the alternative proposed measures are:

\[
\bar{E} = 1 - \frac{\sum_{i=1}^{n} \left( Y_i (1 - \delta_i) C_1 + (1 - Y_i) \delta_i C_2 \right)}{\sum_{i=1}^{n} \left( Y_i \times \frac{1}{4} C_1 + (1 - Y_i) \times \frac{1}{4} C_2 \right)}
\]  

(3.50)

where \( C_1 \) is the cost of misclassifying an individual to the class \( Y_i = 0 \), and \( C_2 \) is the corresponding cost to class \( Y_i = 1 \). The \( \frac{1}{4} \) which appears in the denominator of \( \bar{E} \) above represents the 'equally likely' hypothesis of classification when the model's information \( \{ \delta_i \} \) is neglected. Domenich and McFadden (46) propose either the adoption of \( C_1 = 1/\hat{\theta} \), and \( C_2 = 1/(1 - \hat{\theta}) \) where \( \hat{\theta} \) is the sample proportion of cases \( Y_i = 1 \), or the equal costs approach: \( C_1 = C_2 = 1 \). The details about this measure can be found from the above cited reference.

3.6 The Nature of Our Computational Routine

The special purpose computer program which we have developed in the course of this study is described more fully in Appendix (B). The definition of the explanatory variables included in the vector \( X \) together with the main findings of this study are given in the sixth chapter.
The empirical analysis of this study is based on the method of maximum likelihood outlined in section (3.3) above. We have adopted the Newton-Raphson method for the numerical solution of the system of likelihood equations (3.20). The problem of starting off with an initial guess $\hat{\theta}^{(0)}$ is considerably reduced by adopting Anderson's proposal that for the binary logit $\hat{\theta} = 0$ always leads to a reasonable speed of convergence. As for the goodness-of-fit measurement we have used the log-likelihood ratio $-2\log \lambda$ and our $F^*$ statistic, together with the diagnostic accuracy (3.48). In the fifth chapter we discuss other computational problems relating to aggregate prediction analysis. These related quantities are also included in our computational routine.

3.7 A Numerical Example

In the way of recapitulation we shall conclude this chapter by a numerical example which demonstrates our computational structure and the subsequent inferences based on the likelihood method. Yet, the discussion and interpretation of the empirical results is not dealt with here. We have devoted the sixth chapter to this purpose, whereas here we merely quantify for comparability and illustration the statistical measures we have defined above. We shall use as our data base the sample of female spouses who were employed before the move into council housing. As we have mentioned in the first chapter, movers are classified into 'actual' and 'intending' movers. The problem is to fit the logistic model for the probability that a randomly selected female spouse would quit her pre-move job. The model will be based on the following set of explanatory variables:
RR ≡ change in rent plus rates on moving.
INC ≡ 0,1 dummy for whether or not the individual has received non-labour income after the move.
SPT ≡ 0,1 dummy for special training.
HRS ≡ 0,1 dummy for hours worked per week before move, i.e. HRS = 1 if hours > 30, and zero otherwise.
MOV ≡ 0,1 dummy for 'actual' versus 'intending' movers.

On the other hand, the dependent variable, $Y_i$, is an 0,1 binary observation which equals one if the individual is observed to have left his job, and zero otherwise. Thus, the problem is to fit the model:

$$\text{Prob}(Y_i = 1) \equiv \theta_i = \frac{e^{\beta'X_i}}{1 + e^{\beta'X_i}} \quad i = 1,2, \ldots, n$$

where $X_i = (RR, INC, SPT, HRS, MOV)$, whereas

$\beta' = (\beta_0, \beta_{RR}, \beta_{INC}, \beta_{SPT}, \beta_{HRS}, \beta_{MOV})$ is the vector of unknown parameter coefficients, containing the constant term $\beta_0$.

The likelihood method involves the numerical maximization of the log-likelihood function:

$$\lambda(\beta) = \sum_{s=0}^{5} t_s \beta_s - 95 \sum_{i=1}^{n} \log(1 + e^{\beta'X_i})$$ (3.51)

where $t_s = \sum_{i=1}^{95} X_{si}Y_i$ (s = 0, ..., 5) and $X_{si}$ is the s$^{th}$ component of $X_i$. We shall adopt the Newton Raphson procedure for the maximization of the log-likelihood function (3.51) as given by the recursive relations (3.21) above. As we have proposed, we start off with the initial value

$$\beta^{(0)} = (0, 0, 0, 0, 0, 0).$$
Then we update it iteratively using the Newton-Raphson procedure. Recall that $d^{(n)}(\ell)$ is the column vector of first order partial derivatives of the log-likelihood function $\ell(\beta^{(n)})$ evaluated at $\beta^{(n)}$ of the nth iteration. Hence, $\beta^{(0)}$ will be updated iteratively until the condition $d(\ell) = 0$ is satisfied, which implies convergence to the vector of maximum likelihood estimates, $\hat{\beta}$. For our particular example we have:

$$
\ell(\beta^{(0)}) = -65.843$$

$$
d^{(0)}(\ell) = \begin{bmatrix}
9.500 \\
-247.685 \\
-5.500 \\
-5.000 \\
-19.500 \\
5.000
\end{bmatrix}
$$

Then, in the seventh iteration we arrive at

$$
\ell(\hat{\beta}) = -34.598$$

$$
d^{(7)}(\ell) = \begin{bmatrix}
-0.00000000 \\
-0.00000000 \\
-0.00000000 \\
0.00000000 \\
0.00000000 \\
0.00000000
\end{bmatrix}
$$

as our convergence condition imposes that

$$
\frac{\partial \ell(\beta)}{\partial \beta_s} \leq 0.05 \times 10^{-6}
$$

for all $s = 0, \ldots, 5$. Thus, we obtain the vector of maximum likelihood estimates:
\[ \hat{\beta}' = (3.412, -0.123, -0.363, -1.476, -3.543, -1.813). \]

The asymptotically efficient variance-covariance matrix for the maximum likelihood estimates, \( \hat{\beta} \), is obtained as:

\[
V(\hat{\beta}) = \begin{pmatrix}
0.702 & -0.023 & -0.293 & -0.340 & -0.448 & -0.345 \\
0.002 & 0.000 & 0.009 & 0.011 & 0.007 & \\
0.631 & 0.138 & 0.193 & 0.208 & \\
0.616 & 0.153 & 0.110 & \\
0.522 & 0.197 & \\
1.026 & 
\end{pmatrix}
\]

Recall that the lower diagonal elements of \( V(\hat{\beta}) \) are exactly the same as the upper diagonal elements since \( V(\hat{\beta}) \) is a symmetric matrix. The square roots of the diagonal elements are the consistent estimates of the standard errors of \( \beta_s \) \((s = 0, \ldots, 5)\). Thus, the standard errors are estimated by the vector

\[
\left( 0.838, 0.040, 0.794, 0.785, 0.722, 1.013 \right).
\]

On the basis of these estimates we compute the t-values corresponding to each maximum likelihood estimate, \( \hat{\beta}_s \); (see formula (3.27)). The absolute t values are given in the vector

\[
\left( 4.072, 3.073, 1.717, 1.880, 4.901, 1.790 \right).
\]

The asymptotic Normality property of \( \hat{\beta} \) implies that we can compare these computed t-values with the tabulated t-values with \((95 - 6 =) 89\) degrees of freedom. The significance levels, corresponding to our computed t-values as read out from the statistical tables are:
which show the probabilities of rejecting the alternative hypothesis \( \theta_s \neq 0 \) when it is true (i.e. the type II error). Besides the t-statistics we may compute other measures of goodness-of-fit. Hence we have computed:

(i) The log-likelihood ratio test:

\[ -2 \log \lambda = 58.676 \] (see definition (3.29)).

This value will then be compared with the tabulated value \( \chi^2(5) = 22.11 \) at 0.0005 significance level. Thus it implies a very good fit for the model.

(ii) Our variance-ratio statistic

\[ F^* = 19.554 \] (see definition (3.43)).

Its computation can be based on the analysis of variance table below:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>d.f.</th>
<th>Mean-square</th>
<th>( F^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between individuals</td>
<td>( \sum \hat{\theta}_i^2 - 95 \times \hat{\theta}^2 )</td>
<td>5</td>
<td>2.387</td>
<td>19.554</td>
</tr>
<tr>
<td>Residual</td>
<td>( \sum \hat{\theta}_i (1-\hat{\theta}_i) )</td>
<td>89</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( 95 \times \hat{\theta}(1-\hat{\theta}) )</td>
<td>94</td>
<td>22.80</td>
<td></td>
</tr>
</tbody>
</table>
(The loss of one degree of freedom to the total variation is due to the estimation of the grand proportion $\hat{\theta}$ on the basis of the sample.)

This value of $F^*$ can then be compared with the tabulated $F(5,94)$ value. We find that $F^* > F(5,94) = 12.1$ at 0.005 significance level which also implies a very good fit for the model. However at the 0.0005 significance level $F(5,94) = 23.8 > F^*$. Thus our $F^*$ gives a more reserved judgement than $-2 \log \lambda$ in this example.

Moreover, if we adopt Morrison's (102) definition of the determination coefficient (see definition (3.37)) but make it on the basis of our analysis of variance format instead of the Beta distribution, we get the estimate

$$r^2 = 0.523$$

which can be compared with the similar measures below:

(iii) The $R^2$ and $\rho^2$ of (3.35) and (3.36) respectively are obtained as

$$R^2 = 0.924 \quad \text{and} \quad \rho^2 = 0.459.$$ 

As expected by Domenich and McFadden (46), $R^2$ is appreciably higher than $\rho^2$. The 'likelihood index' $\rho^2$ is even lower than $r^2$ in this example.

(iv) We may also compute the prediction performance measures $E$ (see (3.48)) and $\overline{E}$ taking the cost of misclassification $C = 1$ (see (3.50)). We get

$$E = 0.8421$$

which implies that 84.21% of the observed responses have been correctly predicted by the model. However, when we adopt $\overline{E}$ we get:

$$\overline{E} = 0.7368$$
which implies that when misclassification involves a constant cost, 73.68% of the observed responses will be correctly predicted by the model. It is interesting to note that we can adapt Morrison's (102) formula (3.49):

\[
E^* = \int_0^1 (1 - \theta) g(\theta) + \int_0^1 \theta g(\theta)
\]

(which he based on the Beta density, \( g(\theta) \)) and write instead

\[
E^* = \frac{1}{95} \sum_{i \in A} (1 - \hat{\theta}_i) + \frac{1}{95} \sum_{i \in B} \hat{\theta}_i
\]

where the set \( A \) contains cases with \( \hat{\theta}_i \leq 0.50 \), and set \( B \) contains cases where \( \hat{\theta}_i > 0.50 \). Then, if we apply the last formula we get:

\[
E^* = 0.8419
\]

which is very close to the value of \( E = 0.8421 \) as expected (see page 105 below).
CHAPTER 4

A SIMULTANEOUS LOGISTIC APPROACH FOR THE ANALYSIS OF
QUANTAL CHOICE AMONG HOUSEHOLD MEMBERS

4.1 Introduction

In the second chapter we have provided a possible behavioural basis for the binary logit to describe the process of job mobility, or leaving. The functional form of the binary logit is written as:

\[ \theta_i = \operatorname{Prob}(Y_i = 1) = \frac{V_i}{1 + e^{-V_i}} \quad (i = 1, ..., n) \]  (4.1)

where \( Y_i \) is a Bernoulli, or binary random variable which takes the value one if the \( i \)th individual decides to leave his current employer, and zero otherwise. The \( V_i \)'s are measurable functions which could be expressed in terms of a set of explanatory variables, \( x_i \) with a generally unknown set of coefficient parameters, \( \beta \), i.e. \( V_i = \beta'x_i \).

However, the derivation of the model (4.1) and the subsequent stage of parameter estimation discussed in the last chapter have been based on the assumption of a simple size-\( n \) random sample. Now, this chapter deals basically with the adaptation of model (4.1) to the case of cluster random sampling, specifically, a sample of households. As we have outlined in the first chapter, the formulation of a quantal (or binary) choice model that allows for interdependent decisions of household members remains to be met. Hence, the main theme of this chapter is to propose such a model which can be of general applicability for quantal choice situations involving interdependent responses of household members.

While discussing the economics of labour force participation, Bowen and Finegan (27, p. 29) have pointed out this problem. Similarly,
Engelman (49, p.164) has expressed the desire for a model which captures interdependence of household members for job quit decisions. However, we have shown in the first chapter that a general model for the analysis of quit decisions on the basis of the household unit entails analysis of labour force participation rates as well, and therefore a considerably larger sample. It is obvious to see that the decision to terminate one's job will be coupled with one of three possibilities (i) moving immediately to another employer, (ii) undergoing a period of unemployment, (iii) dropping out of the labour force. In the context of a multiple adult family any one of these decisions by one member of the household, e.g. the head, may stimulate changes in labour force participation of other members, e.g. the wife. Under the first possibility, if the husband moves to a better job, this may make it less important for the wife to continue in (or to enter) the labour force. However, under the second possibility, where the husband undergoes a period of unemployment, the wife's reaction could be to continue in (or to enter) the labour force - possibly with overtime work. Similarly, if the wife's shadow price of time causes her to drop out of the labour force, the husband may be less inclined to risk leaving his job. These implications could be drawn from the behavioural discussion of household economics which is briefly outlined in the next section on the basis of some selected studies. The dichotomous nature of quit and labour force participation decisions in their simplest form invokes the allowance for such interdependent decisions via a version of model (4.1) as we shall show in section (4.3). As for the general empirical set-up required for fitting a model of job quit allowing for changes in participation pattern of household members, this has been considered in the first chapter and will not be repeated here. We shall discuss the full likelihood for the estimation of our proposed simultaneous logistic model in section (4.4). However, data shortage makes it only
feasible to estimate the interdependence parameter of the husband/wife labour force participation responses. The estimation of this parameter is discussed in section (4.5), while our empirical results are demonstrated in section (4.6). Finally, we give a short summary to the main points of this chapter.

4.2 Models of Household Behaviour

The literature of household economics has grown rapidly within the last two decades; e.g. see Samuelson (124), Becker (16), Gronau (61,62), Cohen et al. (33), Fliesher et al. (52), Wales and Woodland (145), Muth (105), Kniesner (82) Ashenfelter and Heckman (9), Heckman (68) and Lancaster (85) among others.

The idea of a household utility function has been discussed by Samuelson (124), Becker (16), Muth (105) and Lancaster (85). The first three authors dealt with an adaptation of neo-classical consumer theory, while Lancaster discussed the household decision function on the basis of his new theory of consumer behaviour (see Lancaster (85)). However, it appears that most current econometric research in this area is based on the neo-classical household utility model; e.g. see Gronau (61,62), Heckman (68), Ashenfelter and Heckman (9), Wales and Woodland (145) and Kniesner (82).

Samuelson's pioneering contribution has been the theoretical derivation of an optimum condition for dividing a given income, I, among household members. He assumes that the \(i^{th}\) member has a strictly concave utility function, \(u_i(x_i)\), defined on his own consumption vector, \(x_i\). Then, the household utility function is defined as

\[
U = U(u_1(x_1), \ldots, u_m(x_m))
\] (4.2)
for a size-\(m\) household. This function is shown to be strictly quasi-concave on the individual arguments \(u_i(x_i)\). Then, according to Samuelson, the optimal condition for the division of income \(I\) among the \(m\) household members is given by:

\[
\frac{\partial u_i}{\partial x_j} = \frac{3u_i}{3x_j} \quad (i \neq j = 1, \ldots, m),
\]

and then individual members are assumed to optimize on their budgets.

Becker\(^{16}\) and Muth\(^{105}\) defined similar household utility functions while introducing the idea that the household acts as a small factory which produces its own goods and services. According to Muth commodities purchased by household members are used as inputs into the production of goods within the household, and household utility function is, thus, defined directly on the home-made goods. Becker's approach, which has won greater popularity (or criticism), assumes that household members combine market goods and time inputs to produce home goods. The latter quantities, \(Z_i\), are defined as functions of time inputs \(t_i\) and market goods input \(x_i\) via some production function, i.e. \(Z_i = g(t_i, x_i)\) \((i = 1, \ldots, p)\). Hence the household utility function is expressible as:

\[
U = U(Z_1, \ldots, Z_p) = U(x_1, \ldots, x_p; t_1, \ldots, t_p).
\]

In a rather lengthy discussion Becker adopts this model and develops a set of time and budget constraints to derive implications regarding value of time and the role of foregone earnings. In particular, he concludes that household members divide different activities between themselves in order to minimize foregone earnings. It is possible to show that under a certain linearity assumption, this idea can be formulated
as a personnel assignment problem which is a special form of the linear programming transportation problem (see Hadley (63, p.367). Thus, if $t_{ij}$ is the time allocated to the $i^{th}$ activity by the $j^{th}$ member, and if $W_{ij}$ is the amount of foregone earnings per unit time resulting from this allocation, then Becker's principle involves the minimization of total foregone earnings:

$$\text{minimize } F = \sum_{i,j} W_{ij} t_{ij}$$

\hspace{1cm} (4.2)

The constraints require the additional restrictive assumption that for the $i^{th}$ activity there should be (i.e. due to past decisions) fixed allocation of $T_i$ units of time, implying the constraints set

$$\sum_{j=1}^{m} t_{ij} = T_i \quad (i = 1, ..., p)$$

the model (4.2) is completed with the second constraints set:

$$\sum_{i=1}^{p} t_{ij} = T_j^* \quad \text{the time resource for the } j^{th} \text{ household member}$$

$$j = 1, ..., m.$$

Heckman (68) assumes an alternative non-linear programming problem underlying the process of housewives labour force participation. Heckman's econometric model which is based on a neo-classical household utility function, produces a model with a common set of estimable parameters to determine hours of work, together with the probability that the housewife does labour market work. This model is interesting in the sense that it reduces to a single unified model two distinct forms of labour supply functions: the dichotomous labour force participation model, and the conventional model of hours supply. Heckman's principle
is that the participation/non-participation decision is based on com-
parison of the market wage, \( W \), and the shadow price of time \( W^* \). The
former is assumed to depend on skill and labour market experience, while
the latter depends on the husband's wage, the asset income of the house-
hold, price of a vector of consumption goods, time not directly available
for non-market activities, and another set of constraints which represent
past choices (e.g. number of children), and technological state of the
household. Then, we may write:

\[
\text{Prob(wife enters labour force)} = \text{Prob}(W > W^*)
\]

or,

\[
\text{Prob(wife leaves labour force)} = \text{Prob}(W < W^*) .
\]

Now, if \( W > W^* \) as in (4.3) above, Heckman shows that the wife will
enter the labour force and increase her hours of work until the equili-
brium condition:

\[
W - W^* = 0
\]

is reached. Specifically, he considers hours of work to play the role of
a slack variable in non-linear programming to equate market wage and
shadow price of time. This result can similarly be derived on the basis
of the personnel assignment model (4.2) above since the optimal solution
of the transportation problem requires that for the cells where \( t_{ij} > 0 \)
we must have

\[
W_{ij}^* - W_{ij} = 0 \quad (i = 1, \ldots, p; j = 1, \ldots, m)
\]

where \( W_{ij}^* \) is the shadow price of time for the \( j^{th} \) individual doing the
\( i^{th} \) activity (see Hadley\(^{63}\)). However, the linear restriction

\[
\sum_{j=1}^{m} t_{ij} = T_i \quad (i = 1, \ldots, p)
\]

may be subject to criticism since it
presumes fixed time allocations to household activities at a given point
of time.
Gronau (61), using Becker's theory of time allocation, addressed himself to the problem of intra-household allocation of time. His husband/wife model stresses the trichotomy of the housewife's time resource among the states of: work in the market, work at home, and leisure. Using a rather intricate approach he employs this set-up in his model to calculate the housewife's value of time, which he believes to be underrated by the market wage.

Other econometric models dealt with the household interdependent nature via a family labour supply model derived from constrained maximization of a well-behaved neo-classical household utility function which is usually written as:

\[ U = U(L_h, L_w, X) \] (4.4)

where \( L_h \) is the husband's leisure, \( L_w \) = the wife's leisure, and \( X \) = a vector of consumption goods. The constraints are generally expressed in terms of non-labour income, wage-rates for husband and wife and a price vector for consumption goods. Wales and Woodland (145) derived the household supply response functions, for husband and wife, by specifying different functional forms for the utility function (4.4): i.e. Cobb-Douglas and Translog forms. In order to release flexibility of hours worked from institutional restrictions, they have based their model on a sample of self-employed workers. However, their data revealed that the general regularity and curvature conditions for utility maximization are not all satisfied by their estimated functions. They have provided some evidence for interdependence of husband and wife labour supply responses via their respective market wages; namely, an increase in husband's wage does not affect hours worked by wife, but an increase in wife's wage-rate increases hours worked by husband.

However, different results have been obtained by Ashenfetler and
Heckman (9) as regards the last point. They have estimated two labour supply functions for husband and wife respectively based on model (4.4). However, in the subsequent empirical analysis they have adopted the dichotomous labour force participation approach for the labour supply functions. In order to allow for possible autocorrelation in the error terms of the husband equation and wife equation, they have adopted the method of three stage least squares estimation. Their estimated coefficients show that an increase in the husband wage has a significant negative effect on the labour force participation of the wife. The other result which also disagrees with Wales and Woodland (145) is that the wife's wage has no significant effect on the husband's participation.

The interesting feature of the Ashenfelter and Heckman (9) econometric model is that it has been used to test the consistency of a set of neo-classical theoretical restrictions with their data, one of which is the restriction that can be expressed by relation (4.5) below:

\[
\left( \frac{\partial L_h}{\partial W_f} \right)_{u = \text{const.}} = \left( \frac{\partial L_f}{\partial W_h} \right)_{u = \text{const.}}
\]  \hspace{1cm} (4.5)

where \( L_h, L_f \) are leisures of husband and wife respectively, while \( W_h, W_f \) are their respective wage rates. The quantities on either side of relation (4.5) are treated in neo-classical consumer theory as substitution effects evaluated on the basis of a given indifference curve which is a function of \( L_h \) and \( L_f \).

In fact relation (4.5) follows as a special case from the general result that:

"The substitution effect (at constant utility) on the \( i \)th commodity resulting from a change in the \( j \)th price is the same as the substitution effect on the \( j \)th commodity resulting from a change in the \( i \)th price."

Henderson and Quandt (70, p. 36).
Hence, the above result justifies relation (4.5) when \( L_f \) and \( L_h \) are treated as commodities with prices \( W_f \) and \( W_h \) respectively.

Equation (4.5) can be expressed in terms of labour supply responses \( R_h \) and \( R_f \) instead of leisures \( L_h \) and \( L_f \) (as Ashenfelter and Heckman\(^9\) have done\*). The substitution effects

\[
\left( \frac{\partial R_h}{\partial W_f} \right)_u = \text{const.} \quad \text{and} \quad \left( \frac{\partial R_f}{\partial W_h} \right)_u = \text{const.}
\]

will be denoted by \( S_{hf} \) and \( S_{fh} \) respectively for brevity, and condition (4.5) can be written as:

\[
S_{fh} = S_{hf} .
\] (4.6)

The latter condition is usually expressed within the famous Slutsky equation:

\[
\frac{\partial R_i}{\partial W_j} = S_{ij} + R_j \frac{\partial R_i}{\partial \Theta} \quad i = h, f \] (4.7)

where \( R_i \) is the supply response in terms of work hours of husband or wife, \( W_j \) is the wage-rate of either, and \( \Theta \) is non-labour income.

(The derivation of the Slutsky equation (4.7) is given by Henderson and Quandt\(^{70}\).) Ashenfelter and Heckman\(^9\) have based their set of testable restrictions on the last equation (4.7). However, the restriction which concerns us here is that of (4.6) above, and it has been shown to be consistent with Ashenfelter and Heckman's data. This restriction implies that the effect of income-compensated change of the husband's wage on the wife's work effort is equal to the income-compensated change of the wife's wage on the husband's work.

At this point we turn to Kniesner's\(^{62}\) contribution, who has suspected the reliability of data on non-labour income as being usually poor. And since estimates of income compensated wage effects depend on the

\* For simplicity it is assumed that the time budgets are divided exhaustive between market work and leisure.
quantity \( \partial R_i / \partial I \), we have to look for another source to examine the sign of \( S_{ij} \) (see relation (4.6)). Kniesner's approach has been to devise an indirect test for the sign of \( S_{mf} \) (or \( S_{fm} \)) without the need for non-labour income. Then, the resultant sign of this measure can be used to indicate complementarity, substitutability, or independence of the labour supply responses of husband wife, \( R_h \) and \( R_f \) (or equivalently the leisure quantities \( L_h, L_f \)), bearing in mind that

\[
S_{mf} = S_{fm} = \begin{cases} 
> 0 \text{ implies complementarity} \\
= 0 \text{ implies independence} \quad (4.8) \\
< 0 \text{ implies substitutability}
\end{cases}
\]

(see Henderson and Quandt \(^{(70)}\), p. 37).

Kniesner's test depends on the estimation of two labour supply equations: (i) for husband's with working wives, and (ii) for husbands with non-working wives. Let \( \partial R_h / \partial W_h \) and \( \partial R_h / \partial W_h^* \) be the estimated regression coefficients for the gross wage effects in each supply equation respectively (i.e. effect of husband's wage on his own work supply). Then, going through a rather intricate algebraic verification*, Kniesner obtains the following result:

\[
\frac{\partial R_h}{\partial W_h} - \frac{\partial R_h}{\partial W_h^*} \Rightarrow 0 \Leftrightarrow S_{mf} > 0 \quad (4.9)
\]

In other words the difference between gross wage effects in the two equations determines the sign of the net-compensated effect. His empirical results give indirect evidence that for older people \( S_{mf} > 0 \), which he uses to imply complementarity. It is also noteworthy that Ashenfelter and Heckman \(^{(9)}\) obtained a positive sign for the symmetrically defined compensated effects between husband's and wife's

* See proof in Kniesner \(^{(82)}\), pp. 655-659.
supply responses; thus, implying complimentarity of their labour force participation decisions.

This finding of complementarity in the labour force participation of husband and wife has also been confirmed by Daniel (40) on the basis of his sample survey of the unemployed in U.K. Although he has not been specifying any mathematical model for measurement of complementarity, he has shown that:

"where men's wives had some paid employment they themselves were very much likely to be in work"

Daniel (40, p. 24).

He concluded that there was a remarkable positive association between the participation statuses of husband and wife for all ages. In addition, Daniel offered three plausible explanations for this association:

(i) A psychological reason: that there tends to be greater pressure for men with working wives to be in work to maintain their self-esteem and respect.

(ii) A sociological reason: Working wives bring their husbands more in contact with the world of work. This gives them more information about job-opportunities. Thus, they do not suffer the isolation of men where neither spouse went out to work, and "lack of money inhibited other social contact."

(iii) Similarities between married partners: that is, it is more familiar for people to marry people similar to themselves in social and demographic terms. When these similar characteristics are not attractive to employers they leave both partners at a disadvantage. These explanations have not been tested empirically, though they may seem plausible a priori.

Now we mention two other important works in the fields of household models.
Cohen and Stafford(33) developed and simulated a model of household life-cycle behaviour which attempts to integrate several aspects of household choice, allowing for the fact that major areas of household decisions are intertemporally dependent and interrelated with one another (e.g. labour force participation decisions, fertility decisions, spending, allocation of assets and training). Then they have specified their dynamic model as a problem in discrete optimal control, i.e. the family will choose a time path of the various above-mentioned variables with the object of maximizing a time-dependent utility function.

Before concluding this section we may briefly outline Lancaster's theory of household behaviour which we have mentioned above as the alternative approach. Using his "characteristics* consumer theory", Lancaster examined the conditions under which the behaviour of the household might diverge from that of a single individual. The special nature of the household arises from the assumption that it is a small group of closely-knit members. Under this model a consumer who is efficient will, when faced with choices among collections of \( M \) goods possessing between them only \( R \) different characteristics \((R < M)\), need to consume no more than \( R \) different goods. The overall conclusion of this theory is that a household which possesses efficiency properties of a single individual is either dictatorial, or that it has a well-behaved decision function and joint effects in consumption covering at least half the goods characteristics. The existence of joint consumption effects within the household will, thus, reduce the divergence in efficiency properties between the household and the individual, according to Lancaster's theory.

However, although this theory is quite revealing, it it not directly amenable to econometric formulation, e.g. for measuring joint consumption

* The basis of this theory is that consumers place their preferences over the characteristics of the goods rather than their physical units.
effects within households, or the extent of convergence of household beha-
vour to that of a single individual for different choice situations.

In this brief selective review, we notice a situation of binary choice for husband and wife, namely the Ashenfelter and Heckman (9) model of "husband/wife" linear labour force participation equations. The latter equations are in fact linear probability models estimated via the method of 3SLS. Although estimates provide meaningful results, yet as we have seen in the last chapter, linear probability models are subject to prediction bias due to the possible violation of the [0, 1] restriction, apart from the problem of heteroscedasticity. Hence, these regression estimates will be biased at very high or very low values of explanatory variables which result in absorption at either the zero, or the one boundary of probability. Another point relates to Kniesner's indirect measure of complementarity of husband and wife's leisure times. Although Knieser's indirect method (see relation (4.9)) resolves the problem of non-labour income, it does not specify how to test the significance of the indirectly measured income-compensated wage effects. The fact that \( \frac{3R_h}{3W_h} \) and \( \frac{3R_h}{3W_h^*} \) have been highly significant in two different models of husband's labour supply response does not imply that the difference between them is also significant. And, it is the sign of this difference which is taken to determine complementarity or otherwise.

Now, the model which we are proposing derives a measure of inter-
dependence which, like Knieser's procedure, does not rely on availability of non-labour income while its sign and magnitude are sufficient to indicate complementarity, independence or substitutability as in con-
ditions (4.8) above. Moreover, it is analogous to income-compensated wage effect in the sense that it satisfies a similar symmetric restric-
tion condition as that of relation (4.6) above. However, since our
measure is not related to the wage effects, it may not possess the same economic interpretation. As a measure of interdependence it can easily be tested but only within a model of labour supply of the dichotomous type (e.g. Ashenfelter and Heckman's\(^{(9)}\) labour force participation model) or any other model of quantal or binary choice within the household (e.g. the pure quit/stay problem where both spouses are employed). We now turn to our model which will be described in the next section.

4.3 A Model of Simultaneous Quantal Choice

Suppose we are interested in the combined pattern of labour force participation of husband and wife at a given point of time. Let \( Y_i \) be the \( 0,1 \) binary random variable which describes the employed/not employed status of the husband in the \( i^{th} \) household, and let the other random variable \( Z_i \) describe the corresponding status of the wife. Let \( x_i' = (x_{1i}, \ldots, x_{pi}) \) be a vector of characteristics for the husband (e.g. skill, experience, age etc.) and other family characteristics (e.g. wife's wage, non-labour income, etc.). Similarly, define the vector \( s_i' = (s_{1i}, \ldots, s_{qi}) \) as a vector of explanatory variables related to the wife's participation decision.

Then, adopting the simultaneous logit approach which has been developed independently by Schmidt and Strauss\(^{(126)}\) Amemiya\(^{(6)}\) and Nerlove and Press\(^{(106)}\), we can write the following two equations:

\[
\begin{align*}
\log \left( \frac{\text{Prob}(Y_i = 1 | Z_i)}{\text{Prob}(Y_i = 0 | Z_i)} \right) &= \beta' x_i + \delta Z_i \\
\log \left( \frac{\text{Prob}(Z_i = 1 | Y_i)}{\text{Prob}(Z_i = 0 | Y_i)} \right) &= \gamma' s_i + \alpha Y_i 
\end{align*}
\]

where \( \beta' = (\beta_0, \ldots, \beta_p) \) and \( \gamma' = (\gamma_0, \ldots, \gamma_q) \)
are generally unknown parameter coefficient vectors. Note that \( Y_i = 1 \) if husband is employed, and zero otherwise, and that \( Z_i = 1 \) if wife is employed, and zero otherwise. The parameter \( \delta \) measures the effect of labour force participation decision of wife on husband, and the parameter \( \alpha \) measures the effect of husband's labour force participation decision on wife. Then, using model (8.4), it follows that:

\[
\log \frac{\text{Prob}(Y_i = 1 | Z_i = 1)}{\text{Prob}(Y_i = 0 | Z_i = 1)} - \log \frac{\text{Prob}(Y_i = 1 | Z_i = 0)}{\text{Prob}(Y_i = 0 | Z_i = 0)} = \delta \quad (4.11)
\]

and

\[
\log \frac{\text{Prob}(Z_i = 1 | Y_i = 1)}{\text{Prob}(Z_i = 0 | Y_i = 1)} - \log \frac{\text{Prob}(Z_i = 1 | Y_i = 0)}{\text{Prob}(Z_i = 0 | Y_i = 0)} = \alpha \quad (4.12)
\]

Moreover, by expressing conditional probability function as joint probability functions, e.g.

\[
\text{Prob}(Y_i = 1 | Z_i = 1) = \frac{\text{Prob}(Y_i = 1, Z_i = 1)}{\text{Prob}(Z_i = 1)}
\]

we can write equation (4.11) as:

\[
\delta = \log \frac{\text{Prob}(Y_i = 1, Z_i = 1)}{\text{Prob}(Y_i = 0, Z_i = 1)} - \log \frac{\text{Prob}(Y_i = 1, Z_i = 0)}{\text{Prob}(Y_i = 0, Z_i = 0)}
\]

\[
= \log \frac{\text{Prob}(Y_i = 1, Z_i = 1)}{\text{Prob}(Y_i = 0, Z_i = 1)} \cdot \frac{\text{Prob}(Y_i = 1, Z_i = 0)}{\text{Prob}(Y_i = 0, Z_i = 0)} \quad (4.13)
\]

By a similar argument it can be shown that

\[
\alpha = \log \frac{\text{Prob}(Z_i = 1, Y_i = 1)}{\text{Prob}(Z_i = 0, Y_i = 1)} \cdot \frac{\text{Prob}(Z_i = 0, Y_i = 0)}{\text{Prob}(Z_i = 1, Y_i = 0)} \quad (4.14)
\]

Hence, it is obvious from (4.13) and (4.14) that

\[
\delta = \alpha \quad (4.15)
\]
The symmetry restriction (4.15) implies that the effect of the husband's labour force participation status on the wife's decision is equal to the effect of the participation status of the wife on the husband's decision. Notice the analogy between our symmetry restriction (4.15) and that of the neoclassical utility model, i.e. restriction (4.6). However, as we have mentioned above the economic interpretation is not precisely the same, since our measured one is not only defined in terms of the wage or income parameter coefficients. The parameter $\alpha(= \delta')$ represents all potential factors which determine the labour force participation response of the other married partner including his/her wage and non-labour income. Moreover an analogy with conditions (4.8) above for testing direction of interdependence is provided by defining $\psi = \alpha = \delta'$; then it is possible* to show

$$
\psi = \begin{cases} 
> 0 & \text{for complementarity} \\
= 0 & \text{for independence} \\
< 0 & \text{for substitutability.} 
\end{cases} (4.16)
$$

The advantage about this measure $\psi$ is that it can be directly estimated and tested, unlike Kniesner's indirect measure for net-compensated effects whose significance cannot be directly assessed.

Now, let us use model (4.10) to define the individual joint probability functions, $\text{Prob}(Y_i = a, Z_i = b)$ $a, b = 0, 1$. For brevity we use the notation **

$$
\theta_{a,b}^{(i)} = \text{Prob}(Y_i = a, Z_i = b)
$$

and

* See Appendix (A, p.294) for proof.

** This notation will be needed in a later part of this chapter.
\[ \text{Prob}(Y_i = 1 | Z_i = b) = \frac{\theta(i)_{a,b}}{\theta(i)_b} \]

where \( \theta(i)_b = \text{Prob}(Z_i = b) \), averaged over \( a \),

and \( \theta(i)_a = \text{Prob}(Y_i = a) \), averaged over \( b \)

\((i = 1, \ldots, n)\).

Then, using the two equations of model (4.10) we can re-express conditional probability statement as joint probabilities, and write:

\[ \theta(i)_{11} = \theta(i)_{01} e^{\frac{x_i}{\Delta_i}} + \psi \]

using the first equation, together with the symmetry definition of (4.16) above. It follows that:

\[ \theta(i)_{10} = \theta(i)_{00} e^{-\frac{x_i}{\Delta_i}} \]  \hspace{1cm} (4.17)

Similarly, we get for the second equation the following:

\[ \theta(i)_{11} = \theta(i)_{10} e^{y_{i-1} + \psi} \]

\[ \theta(i)_{01} = \theta(i)_{00} e^{-y_{i-1}} \]  \hspace{1cm} (4.19)

Then, if we add the following restriction,

\[ \theta(i)_{11} + \theta(i)_{01} + \theta(i)_{10} + \theta(i)_{00} = 1 \]  \hspace{1cm} (4.20)

it follows from (4.17), (4.18), and (4.20) that:

\[ \theta(i)_{11} = e^{\frac{x_i}{\Delta_i} + y_{i-1} + \psi} / \Delta_i \]

\[ \theta(i)_{10} = e^{\frac{x_i}{\Delta_i}} / \Delta_i \]

\[ \theta(i)_{01} = e^{y_{i-1}} / \Delta_i \]

and \( \theta(i)_{00} = 1 / \Delta_i \)  \hspace{1cm} (4.22)
where \( A_i = (1 + e^{\beta x_i} + e^{\gamma s_i}) \) 
\[ (i = 1, ..., n) \]

or, more shortly:

\[ \theta_{a,b}^{(i)} = \exp(\beta x_i a + \gamma s_i b + \psi_{ab}) / i \]  
\[ (a, b = 0, 1). \]

This model has been utilized by Schmidt and Strauss (126) to describe choice of occupation and of industry as a simultaneous process. Clearly model (4.23) suits two occupational alternatives (e.g. manual, non-manual) and two classes of industry (e.g. agriculture, manufacturing).

In this case choice is made simultaneously and interdependently from two alternative choice sets each containing two states, i.e.

\[ A_{1i} = \{a = 0, 1\} \quad \text{and} \quad A_{2i} = \{b = 0, 1\} \quad (i = 1, ..., n). \]

More generally, we can think of other choice situations which could be represented in terms of more than two interdependent alternative choice sets, each containing more than two states, i.e.

\[ A_{1i} = \{a_{1j} ; j = 1, ..., n_1\} \]
\[ A_{2i} = \{a_{2j} ; j = 1, ..., n_2\} \]
\[ \ldots \]
\[ A_{mi} = \{a_{mj} ; j = 1, ..., n_m\} \]

where \( m \geq 2 \) and \( n_j \geq 2 \) (\( j = 1, ..., m \)). Then, the assumption of interdependence of choice implies that
\[
\varepsilon_k^{(i)} = \text{Prob(choose state } j_k \text{ from set } A_{ki})
\]

where

\[
\varepsilon_k^{(i)} = \frac{m}{k=1} \varepsilon_k^{(i)} a_{kj_n}
\]

The empirical data on which this model can be based may be treated as a cluster size-\(n\) random sample, where the \(i^{th}\) cluster consists of \(m\) points. For other practical examples of such choice situations see Nerlove and Press(106).

However, the variation which we have made in adopting this model is to describe a situation where the choice sets, \(A_{ki}\), may not be distinct, but interdependence occurs only via individuals who make the choice. In other words, if all \(N\) individuals can be grouped into clusters, or households, of fixed size-\(m\), then we have an \(n = \frac{N}{m}\) independent random sample with units as households. Members of the \(i^{th}\) household \((i = 1, \ldots, n)\) will face choice sets \(A_1, \ldots, A_m\) which may not be distinct, but the choice situation satisfies condition (4.24) above. Let us assume for simplicity that \(m = 3\) and \(n_k = 2\) \((k = 1,2,3)\) so that we have three alternative choice sets:

\[
A_1 = \{Y_i = 0,1\}, \quad A_2 = \{Z_i = 0,1\}, \quad A_3 = \{U_i = 0,1\}.
\]

If we define \(\varepsilon_{a,b,c}^{(i)} = \text{Prob}(Y_i = a, Z_i = b, U_i = c)\), then Amemiya(6) has shown that it is possible to express \(\varepsilon_{a,b,c}^{(i)}\) as below:

* Note that we are extending the notations of model (4.23) above.
\[ \theta_{a,b,c}^{(i)} = C_i \exp (\theta \mathbf{x}_1 . a + \gamma \mathbf{s}_1 . b + \lambda \mathbf{v}_1 . c + ab \psi_{ab} + ac \psi_{ac} + bc \psi_{bc} + abc \psi_{abc}) \]  
\( a, b, c = 0, 1, \)

where \( C_i \) is chosen to satisfy the restriction

\[ \sum_{c=0}^{1} \sum_{b=0}^{1} \sum_{a=0}^{1} \theta_{a,b,c}^{(i)} = 1 \quad \text{all } i. \]

Model (4.25) extends model (4.23) to the case of three interdependent dichotomous variables where \( U_i \) is now introduced with corresponding vector \( \mathbf{V}_i \) and parameter coefficients \( \lambda' \) (e.g. a size-3 household: husband, wife, and adult son). The \( \psi \)'s represent the interdependent effects and satisfy the symmetry restrictions:

\[ \psi_{ab} = \psi_{ba}, \quad \psi_{ac} = \psi_{ca}, \quad \psi_{bc} = \psi_{cb}, \]

or more generally the relative position of the \( a, b, c \) subscripts is immaterial.

4.4 The Maximum Likelihood Approach

The parameters' estimation and significance testing on the basis of this model can be done by the method of maximum likelihood as proposed by Nerlove and Press (106), Schmidt and Strauss (126), and Amemiya (6). It has also been observed that there is no identifiability problem for this simultaneous equation model. This has been the case because the process of estimation is performed on the basis of the reduced form of system (4.10) which, in this case, is the set of joint
probability functions (4.23), or (4.25). Let us deal with model (4.23), the "husband/wife" labour force participation model. Then the likelihood function is defined below as:

\[
L = \prod_{i=1}^{n} \prod_{a=0}^{1} \prod_{b=0}^{1} \beta(i)_{a,b} \prod_{i=1}^{n} \frac{1}{\Delta_i} \left( \frac{\beta x_i + \gamma s_i + \psi}{\Delta_i} \right)^{Y_i Z_i} \left( \frac{\gamma s_i}{\Delta_i} \right)^{Y_i (1 - Z_i)} \left( \frac{\beta x_i}{\Delta_i} \right)^{Z_i (1 - Y_i)}
\]

(4.26)

using equations (4.22).

Note that for any value of \( i \) the likelihood function will contain only one factor from the right hand side of the last definition (4.26). To simplify the likelihood function, we may subdivide the \( n \)-sample into (i) \( n_1 \) cases where \( Y_i = 1 \) and \( Z_i = 1 \), (ii) \( n_2 \) cases where \( Y_i = 1 \) and \( Z_i = 0 \), (iii) \( n_3 \) cases where \( Y_i = 0 \), \( Z_i = 1 \) while the remaining number \( n_4 = n - n_1 - n_2 - n_3 \) represents cases where \( Y_i = 0 \) and \( Z_i = 0 \). Then, the likelihood (4.26) can accordingly be factorized as:

\[
L = \prod_{i=1}^{n_1} \left( \frac{\beta x_i + \gamma s_i + \psi}{\Delta_i} \right) \cdot \prod_{i=1}^{n_2} \left( \frac{\beta x_i}{\Delta_i} \right) \cdot \prod_{i=1}^{n_3} \left( \frac{\gamma s_i}{\Delta_i} \right) \cdot \prod_{i=1}^{n_4} \left( \frac{1}{\Delta_i} \right)
\]

(4.27)

Let \( B \) be the set containing \( n_1 \) and \( n_2 \) and \( C \) be the set containing \( n_1 \) and \( n_3 \). Then, we can simplify (4.27) further to get:

\[
L = \exp \left( \sum_{i \in B} \beta x_i + \sum_{i \in C} \gamma s_i + n_4 \psi \right) \prod_{i=1}^{n} \frac{1}{\Delta_i}
\]

while the log likelihood function, which is easier to handle, can be immediately written down as:
\[ \log L = \sum_{i \in B} \beta_i x_i + \sum_{i \in C} \gamma_i s_i + n \psi - \sum_{i=1}^{n} \log \Delta_i. \quad (4.28) \]

The maximum likelihood method of estimation involves the maximization of (4.28) through some numerical procedure, e.g. Newton-Raphson, Powell's algorithm or grid search techniques. Then, the maximum likelihood estimates \( \hat{\beta}, \hat{\gamma}, \) and \( \hat{\psi} \) - assuming they exist - satisfy the likelihood equation:

\[ \frac{\partial}{\partial \beta^*} \log L = 0 \]

where \( \beta^* = (\beta', \gamma', \psi) \) is the compound parameters vector.

The problem of estimation and significance testing is analogous to that described in Chapter III, and will not be discussed here in detail. However, the vector of first-order partial derivatives of \( \log L \) (4.28) with respect to parameters \( \beta, \gamma, \psi \), together with the square matrix of second-order partial derivatives are evaluated at the Appendix (A, p. ). We know that this differential structure is useful for numerical optimization, while the negative of the inverse of the second-derivatives matrix is usually used as asymptotically efficient variance-covariance matrix for the maximum likelihood estimates: \( \hat{\beta}, \hat{\gamma} \) and \( \hat{\psi} \). Hence, this estimation procedure requires that the subsamples \( n_1, n_2, n_3 \) and \( n_4 \) be sufficiently large.

The main advantage of this simultaneous logistic model (4.10) over Ashenfelter and Heckman's linear simultaneous equation model (mentioned in section (4.2)) is that (i) it redresses the limitation

* This approach of specifying two linear probability equations for husband's and wife's labour force participation has also been adopted by Fliesher et al. (52) except that they have added a third equation for the rate of unemployment.
of prediction bias since now the estimates automatically obey the $|0,1|$ restriction of probability. Besides, the problem of heteroscedasticity, to which the linear probability model is susceptible, does not now exist.

(ii) it provides a direct measure of interdependence, $\psi$, which does not rely on non-labour income or wage effects.

However, since our data on $n_1$, $n_2$, $n_3$ and $n_4$ is quite limited, as shown by Table (4.1) below, we cannot provide full testing for this model. We may recall that $n_1$ is the sample number of cases where both spouses are employed, $n_2$ the number of cases where the husband is employed and the wife not employed, $n_3$ the number of cases where the husband is not employed but the wife is employed, and $n_4$ the number of cases where both spouses are not employed. The employment status can be taken after move or before move, but Table (4.1) gives the after-move participation status.

Table (4.1)

<table>
<thead>
<tr>
<th>( Y = 1 )</th>
<th>( Y = 1 )</th>
<th>( Y = 0 )</th>
<th>( Y = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z = 1 )</td>
<td>( Z = 0 )</td>
<td>( Z = 1 )</td>
<td>( Z = 0 )</td>
</tr>
<tr>
<td>( n_1 = 58 )</td>
<td>( n_2 = 84 )</td>
<td>( n_3 = 7 )</td>
<td>( n_4 = 42 )</td>
</tr>
</tbody>
</table>

However, although our data is limited, yet it is possible to provide a reasonable preliminary test for the parameter $\psi$ as we shall examine in the next section.
4.5 Simultaneous Logit and the Log-linear Probability Model

The relationship between the logistic model and the log-linear probability model, which is used as a basis for contingency table analysis, is well documented, e.g. see Bishop, Nerlove and Press. The formulation of the log-linear model of contingency tables due to Birch and developed further by Goodman and others, introduces a convenient re-parameterization of cell probabilities (i.e. \( \theta_{a,b,c} \)) in terms of main effects and different order interaction effects. These interaction effects have been treated by Birch and Goodman as constants. However, Nerlove and Press have shown that it is possible to generalize the analysis by allowing these interaction effects to be functions of exogenous variables.

Now suppose we are still dealing with the three dichotomous variable case, considered previously in model (4.25), where we have defined \( \theta_{a,b,c}^{(i)} \); \( a,b,c = 0,1 \). Then, the log-linear model for this case is expressed as

\[
\log \theta_{a,b,c} = \phi_0 + \phi_a + \phi_b + \phi_c + \phi_{a,b} + \phi_{a,c} + \phi_{b,c} + \phi_{a,b,c}.
\]

(4.29)

The \( \phi \)-parameters of this model represent different main effects and interaction effects: \( \phi_0 \) represents the general effect, while \( \phi_a \), \( \phi_b \), and \( \phi_c \) represent main effects of the three dichotomous variables \( Y, Z \) and \( U \), respectively, at levels \( a, b, c \) (\( a,b,c = 0,1 \)). All possible bivariate interaction effects are denoted by \( \phi_{a,b} \), \( \phi_{a,c} \) and \( \phi_{b,c} \), while the three-order interaction effect is represented by the parameter \( \phi_{a,b,c} \), at levels \( a, b \) and \( c \).

There are certain conditions which are imposed on these parameters for identifiability, namely:
\[
\sum_{a=0}^{1} \phi_a = \sum_{b=0}^{1} \phi_b = \sum_{c=0}^{1} \phi_c = 0
\]
\[
\sum_{a=0}^{1} \phi_{1,2} = \sum_{b=0}^{1} \phi_{a,b} = \sum_{c=0}^{1} \phi_{a,c} = 0
\]
\[
\sum_{a=0}^{1} \phi_{1,2,3} = \sum_{b=0}^{1} \phi_{a,b,c} = 0
\]
\[
\sum_{a=0}^{1} \sum_{b=0}^{1} \phi_{1,2,3} = \sum_{a=0}^{1} \sum_{c=0}^{1} \phi_{a,b,c} = 0
\]

and where

\[
\phi_0 = -\log \sum_{a=0}^{1} \sum_{b=0}^{1} \sum_{c=0}^{1} \exp(\theta_{a,b,c})
\]

At this stage we would like to draw attention to some interesting analogies between the log-linear model (4.29) above and the definition of joint random utility introduced by Ben-Akiva (19) in his travel demand simultaneous choice model. His model, which has been based on travel demand theory, assumes that choices of mode, frequency of trip, and destination of trip are made by any given individual simultaneously from alternative choice sets M, F and D, respectively. The basic theory is that the individual chooses some combination \( m', f', d' \) to maximize the random joint utility function \( U(m, f, d) \) where \( m \in M; f \in F; d \in D \). Ben-Akiva then derives the logistic probability model for the choice probabilities \( \text{Prob}(m, f, d) \) in the same way as we have done in the last chapter, except that his model is the multinomial version defined as:
\[ \theta_{m,f,d} = \text{Prob}(m,f,d) = \frac{e_{m,f,d}}{\sum_{m} \sum_{f} \sum_{d} e_{m,f,d}}. \quad (4.30) \]

The \( V_{m,f,d} \)'s are the non-stochastic components of the random utilities \( U(m,f,d) \). The latter is defined as

\[ U(m,f,d) = V_{m,f,d} + \epsilon_{m,f,d} \quad m \in M; f \in F; d \in D \]

where the stochastic components \( \epsilon_{m,f,d} \) are assumed to have a Weibull distribution.

Then, Ben-Akiva parameterizes the \( V_{m,f,d} \)'s in terms of exogenous variables which would explain the interdependent nature of the choice problem, specifically:

\[ V_{m,f,d} = X_{m} \theta^{(m)} + X_{f} \theta^{(f)} + X_{d} \theta^{(d)} + X_{mf} \theta^{(mf)} + \]
\[ X_{md} \theta^{(md)} + X_{fd} \theta^{(fd)} + X_{mfd} \theta^{(m,f,d)} \quad (4.31) \]

where \( X_{m}, \ldots, X_{m,f,d} \) are vectors of explanatory variables indexed by the aspects to which they pertain. Some variables pertain to particular choices alone (i.e., \( X_{m}, X_{f} \) and \( X_{d} \) for mode, frequency and destination of trips respectively) while other variables relate to different combinations of choices as in (4.31) above. The details of this analysis are given by Ben-Akiva(18), but the immediate observation which we want to make is that of the analogy between (4.31) and the log-linear model (4.29). This analogy is obvious if we see that the \( \theta^{(i)}_{a,b,c} \) can be expressed in the multinomial logistic form (4.30) while using
log $\theta^{(i)}_{a,b,c}$ in place of the $V_{m,f,d}$'s. This gives the impression that Ben-Akiva's measurable utility, $V_{m,f,d}$, is just a special form of the log-linear model (4.29) with effects taken as functions of explanatory variables rather than constants. However, Nerlove and Press (106) have allowed for this possibility while studying more carefully the empirical aspects of the log-linear and logistic models. Yet, they have shown that it is always useful to test interaction effects at the constant level, as introduced by Birch (24) and Goodman (59), before using functions of exogenous variables. It appears, however, that Ben-Akiva has not been aware of the relevance of his model to the log-linear model. In what follows, we are not going to elaborate on the technical details of estimating and testing interaction effects via the log-linear model, but we shall use some established results. For example, it can be shown that the independence hypothesis,

$$\theta^{(i)}_{a,b,c} = \theta^{(i)}_a \cdot \theta^{(i)}_b \cdot \theta^{(i)}_c, \quad a,b,c = 0,1$$

where $\theta^{(i)}_a = \text{Prob}(Y_i = a)$, $\theta^{(i)}_b = \text{Prob}(Z_i = b)$, $\theta^{(i)}_c = \text{Prob}(U_i = c)$ are marginal probabilities, and $\theta^{(i)}_{a,b,c} = \text{Prob}(Y_i = a, Z_i = b, U_i = c)$ is the joint probability function, is equivalent to setting all interaction effects of model (4.29) to zero. The statistic tests for this hypothesis are similarly defined on the basis of model (4.29). Interestingly, it is easy to show that the parameters $\psi_{a,b}$ of model (4.25) are defined in the same form as these statistical tests, except that when the log-linear model is used the statistical tests are based on average cell probabilities $\theta_{a,b,c}$ taken over all $i$. For example the statistic for the zero three-order interaction effect (i.e. $\theta^{1,2,3}_{a,b,c} = 0$) is given by
\[ \psi_{a,b,c}^* = \log_e \left( \frac{\theta_{a,b,c} \theta_{a,e,f} \theta_{g,b,f} \theta_{g,h,c}}{\theta_{g,h,f} \theta_{g,b,c} \theta_{a,h,c} \theta_{a,b,f}} \right) \]  
\hspace{1cm} (4.32)

\[(a,b,c,g,h,f = 0, 1)\]

for a 2 \times 2 \times 2 - contingency table characterized by the three dichotomous variables Y, Z and U. Similarly, other statistical tests for different bivariate interactions can be defined, e.g.

\[ \psi_{a,b}^* = \log \left( \frac{\theta_{a,b,f} \theta_{g,h,f}}{\theta_{a,h,f} \theta_{g,b,f}} \right) \]  
\hspace{1cm} (4.33)

Now, if the \[ \theta_{a,b,c} \]'s are superscripted by \( i \) to indicate the \( i \)th size-3 household, say, with members facing the quit/stay or participation/non-participation decisions, we have seen that this situation can be described by model (4.25) above. Then, it can easily be shown that the \( \psi_{a,b}, \ldots, \psi_{a,b,c} \) parameters of model (4.25) are defined in exactly the same form as measures \( \psi_{a,b}^*, \ldots, \psi_{a,b,c}^* \) exemplified by (4.32) and (4.33) above except for the omission of the \( i \). It is more direct to see that \( \psi \) of model (4.22) can be expressed as

\[ \psi = \log \left( \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} \right) \]  
\hspace{1cm} (4.34)

following from step (4.13) above,

while the corresponding measure in the 2 \times 2 contingency table is

\[ \psi_{a,b}^* = \log \left( \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} \right) \]  
\hspace{1cm} (4.35)

where \( \theta_{a,b} \) are average probabilities aggregated over all individuals falling in cell (a,b) of the contingency table. Notice that it is also possible to write:
\[
\psi = \log \left( \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} \right)
\]

where \( \theta_{a,b} \) are some constants.

Hence, as Nerlove and Press (106) have suggested, we can get preliminary estimates for the parameters \( \psi_{a,b}, \ldots, \psi_{a,b,c} \) of model (4.25), or more specifically for our \( \psi \) parameter, by taking \( \theta_{ab}^* = \theta_{ab} \) and hence applying measure (4.35) in place of (4.34), or generally other contingency table measures like (4.32) or (4.33) in place of \( \psi_{a,b}, \psi_{a,b,c} \).

Thus \( \psi \) is tested through the estimate
\[
\hat{\psi} = \log \left( \frac{\hat{\theta}_{11} \hat{\theta}_{00}}{\hat{\theta}_{10} \hat{\theta}_{01}} \right)
\]

where \( \hat{\theta}_{a,b} = \frac{n_{a,b}}{n} \quad a,b = 0,1 \)
and \( n_{ab} \) = number of cases where \( Y_i = a, \) and \( Z_i = b \)
\( i = 1, \ldots, n \).

Therefore, by treating Table (4.1) as a simple \( 2 \times 2 \) contingency table we can get an estimate for \( \psi \) which appears in the log-likelihood function (4.28).

4.6 Significance Testing for \( \psi \), the Measure of Interdependence

In this section we adopt two statistical procedures for obtaining 95% confidence intervals for \( \psi \), the measure of interdependence for the household spouses vis-à-vis labour force participation. The two procedures which we are about to apply are a Bayesian procedure due to Lindley (90), and a non-Bayesian procedure based on likelihood estimates, proposed by McLaren (98). Beside Table (4.2), which combines actual movers and intending movers, we shall utilize Tables (4.3) for actual movers alone. Since Tables (4.2) and (4.3) refer to the after move participation of husband and wife, we construct corresponding Tables (4.4) and (4.5) for

These \( \{ \hat{\theta}_{a,b} \} \) are in fact the MLE's of the multinomial \( \{ \theta_{a,b} \} \)
the before move participation. Besides, to allow for family size we 
utilize tables for size-two families alone, which are Tables (4.6) and 
(4.7) below.

Before applying the non-Bayesian procedure, it is worth mentioning 
that \( \psi \) can be expressed in two equivalent forms as follows:

\[
\psi = \log \left( \frac{\theta_1(1)}{1 - \theta_1(1)} \right) - \log \left( \frac{\theta_1(0)}{1 - \theta_1(0)} \right) \quad (4.37)
\]

\[
= \log \left( \frac{\theta_1(1)}{1 - \theta_1(1)} \right) - \log \left( \frac{\theta_0(1)}{1 - \theta_0(1)} \right) \quad (4.38)
\]

where

\[
\theta_{a(b)} = \text{Prob}(Y = a | Y = b)
\]

and

\[
\theta_{(a)b} = \text{Prob}(Z = b | Y = a) \quad a, b = 0, 1.
\]

The fact that \( \psi \) can be expressed equivalently in terms of \( \theta_{a(b)} \) 
(conditioning on the column cells of the contingency table) or in terms 
of \( \theta_{(a)b} \) (conditioning on the row cells, is used to describe \( \psi \) as having 
both the column property and the row property for measuring interaction on 
the basis of an \( n \times m \)-contingency table. This measure is also called the 
logistic difference. Thus, this property distinguishes the logistic differ-
ence measure, \( \psi \), from other* measures which do not possess both properties.

The first procedure is based on adopting either of the two equi-
valent forms above, (4.37) or (4.38), and then getting an estimate \( \hat{\sigma}^2 \) 
for the standard error of \( \psi \) where the latter is estimated via formula 
(4.36) above. The resultant formulae for the estimate \( \hat{\sigma}^2 \) are given 
below without proof (see Appendix (A, p.299) for proof. If we adopt 
(4.37) we get:

\[
\hat{\sigma}^2 = 
\]

\[
\]

For example the simple difference measure

\[
\hat{\sigma}^2 = \hat{\theta}_1(1) - \hat{\theta}_1(0).
\]

* For example the simple difference measure
\[ \hat{\sigma}_\psi^2 = \left( \frac{1}{\hat{\theta}_0(0)(1 - \hat{\theta}_0(0))n_0} \right) + \left( \frac{1}{\hat{\theta}_0(1)(1 - \hat{\theta}_0(1))n_1} \right) \] (4.39)

whereas if we use the equivalent definition (4.38) we get

\[ \hat{\sigma}_\psi^2 = \left( \frac{1}{\hat{\theta}_0(0)(1 - \hat{\theta}_0(0))n_0} \right) + \left( \frac{1}{\hat{\theta}_0(1)(1 - \hat{\theta}_0(1))n_1} \right) \] (4.40)

where \( n_{a,b} = \sum_{i=0}^{1} n_{ab} \) and \( n_{a,b} = \sum_{j=0}^{1} n_{ab} \) \( (a,b = 0,1) \)

and \( n_{ab} \) is the number of cases with \( Y = a \) and \( Z = b \).

Then, we evaluate the approximate confidence interval:

\[ \hat{\psi} \pm (1.96) \hat{\sigma}_\psi \] (4.41)

using either of the above formulae for \( \hat{\sigma}_\psi^2 \) which should yield similar results except for a very slight numerical difference.

The second procedure is based on the Bayesian inferential structure of introducing a priori belief about the unknown parameter and then work on the posterior distribution after observing the data, i.e.

\[ \pi(\theta|x) = \pi(\theta)L(x|\theta) \]

where \( \theta \) is a vector of unknown parameters, \( x \) is the observed data, \( \pi(\theta) \) is the prior distribution of \( \theta \), \( L(x|\theta) \) is the likelihood function, and \( \pi(\theta|x) \) is the resultant posterior distribution.

For the case of Bayesian inference about \( \psi \) which is a function of \( \{\theta_{a,b}\} \) we define

\[ L(\theta|n_{a,b} \text{ all } a,b) = \prod_{b=0}^{1} \prod_{a=0}^{1} \theta_{a,b}^{n_{a,b}} . \]
Then Lindlay (89) introduces the prior:

\[ \pi(\theta) = \prod_{b=0}^{1} \prod_{a=0}^{1} \theta_{a,b}^{-1} \]

where \( \{C_{a,b}\} \) \((a,b = 0,1)\) are some constants.

Hence, the posterior distribution on which the analysis is based will be given by

\[ \pi(\theta | n_{a,b} \text{ all } a,b) = \prod_{b=0}^{1} \prod_{a=0}^{1} \theta_{a,b}^{n_{a,b} - 1} \]

However the problem is to get the marginal posterior distribution of \( \psi \), which is defined as

\[ \psi = \log \theta_{11} - \log \theta_{10} - \log \theta_{01} + \log \theta_{00} \]  

and then use the mean of this marginal distribution, and its variance in order to construct an approximate confidence interval for \( \psi \). This procedure has been worked out by Lindlay (89) who used some distributional approximation to normality (see Appendix (A, p.299) for details). The resultant formula for 95\% confidence interval turns out to be

\[ \psi(\pi) \pm (1.96)\sigma_{\psi}^{(\pi)} \]

where \( \psi(\pi) \) is the approximate mean of the marginal posterior distribution, given by:

\[ \psi(\pi) = \log(2n_{11} - 1) - \log(2n_{01} - 1) - \log(2n_{10} - 1) + \log(2n_{00} - 1) \]

and

\[ \sigma_{\psi}^{2(\pi)} = \frac{1}{\Sigma} \frac{1}{\Sigma} \frac{1}{n_{a,b}} \]

is the approximate variance for the estimate.
We may now turn to the results of the calculations which are presented beside the tables to which they correspond (i.e. Tables (4.2) up to (4.7)). This data relates to the Glaswegian households who have been either moved, or notified to move to a new council house. We refer to them as actual movers and intending movers respectively. The combined sample of the two groups consists of 191 households of different sizes, or 168 size two husband/wife families; see Table (4.6). We may also note that the number of intending mover households, where both husband and wife have responded, is too small to stand separate analysis (= 39). For this reason we have given separate treatment only to actual movers, combined sample, and size-two families. The main aim is to assess for each group the sign of the interdependence measure, \( \psi \) and to test its statistical significance. Although we do not expect that \( \psi \) will differ when we compare the before move and after move participation patterns, yet we have allowed for this by looking at the two situations for each group.

### Table (4.2)

<table>
<thead>
<tr>
<th></th>
<th>( n_{00} )</th>
<th>( n_{01} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>42</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>58</td>
</tr>
</tbody>
</table>

\[ \hat{\psi} = 1.481 \]

\( \psi \in (0.554, 2.288) \) non-Bayesian

\( \psi \in (0.114, 0.337) \) Bayesian

### Table (4.3)

<table>
<thead>
<tr>
<th></th>
<th>( n_{00} )</th>
<th>( n_{01} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movers</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>46</td>
</tr>
</tbody>
</table>

\[ \hat{\psi} = 1.458 \]

\( \psi \in (0.511, 2.404) \) non-Bayesian

\( \psi \in (0.100, 0.336) \) Bayesian
### Table (4.4)
Before-move participation of combined sample

<table>
<thead>
<tr>
<th>$n_{00}$</th>
<th>$n_{01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>98</td>
<td>61</td>
</tr>
</tbody>
</table>

$\hat{\psi} = 1.057$

$\psi \in (0.992, 1.936)$ non-Bayesian

$\psi \in (0.021, 0.219)$ Bayesian

### Table (4.5)
Before-move participation of actual movers

<table>
<thead>
<tr>
<th>$\bar{n}_{00}$</th>
<th>$n_{01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>72</td>
<td>51</td>
</tr>
</tbody>
</table>

$\hat{\psi} = 1.016$

$\psi \in (-0.017, 1.926)$ non-Bayesian

$\psi \in (0.012, 0.246)$ Bayesian

### Table (4.6)
Before-move participation of combined sample - size-two families

<table>
<thead>
<tr>
<th>$n_{00}$</th>
<th>$n_{01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>89</td>
<td>52</td>
</tr>
</tbody>
</table>

$\hat{\psi} = 0.944$

$\psi \in (-0.085, 1.974)$ non-Bayesian

$\psi \in (0.006, 0.215)$ Bayesian

### Table (4.7)
After-move participation of combined sample - size-two families

<table>
<thead>
<tr>
<th>$n_{00}$</th>
<th>$n_{01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>76</td>
<td>49</td>
</tr>
</tbody>
</table>

$\hat{\psi} = 1.380$

$\psi \in (0.446, 2.315)$ non-Bayesian

$\psi \in (0.099, 0.338)$ Bayesian
A general look at the results presented in Tables (4.2) to (4.7) reveals the finding that $\psi$, the measure of husband-wife interdependence vis-à-vis labour force participation, tends to be positive at the 95% confidence interval. Hence it appears on the basis of our data that there is a tendency for husband's and wife's participation decision to be complementary. This conclusion agrees with Ashenfetler and Heckman (9) who obtained a positive sign for the symmetrically defined substitution effects of husband and wife labour force participation. Similarly we have seen Kniesner reporting a positive sign for his indirect measure of net-compensated effects (see relation (4.9)) for older people which implied complementarity. However, this analogy should be taken with the awareness that our model is not based on the neoclassical utility model for labour supply where interdependence is explained only through estimated partial wage effects.

If we consider the results individually we note that the Bayesian confidence intervals do not contain zero for any one of the six tables. However, we see that the non-Bayesian confidence interval for the before-move participation pattern of actual movers (Table 4.5)) contains zero, hence implying non-significance of $\psi$ at the 5% level. Obviously since this result pertains to the before move status, it cannot be indicative of a move effect. In addition, we get another non-Bayesian confidence interval containing zero for combined size-two families before move (Table 4.6)). Can we interpret this result as indicative of weaker interdependence before move in size-two families? It is not only that there is no a priori ground to suspect this result, but also that this insignificance is not supported by the second statistical procedure. We may notice that these two insignificant results for Tables (4.5) and (4.6) relate to reduced sample sizes. Thus, recalling that these intervals are based on large sample properties and normality approximation, the latter gets weaker leading to statistical bias, the smaller the sample
size - beside the other possible sources of bias like reporting errors, and non-response. Hence, when these numerical results are considered individually they do not modify our general impression that there is a tendency of husband-wife complementarity. Looking at the individual confidence intervals, we see that although the Bayesian intervals do not contain zero, yet the mean of the posterior distribution around which they vary appears to be always smaller than the sample estimate of $\psi$. Hence, the main reason why the non-Bayesian method has differed in the above two special cases is that it produced relatively bigger standard errors. The latter quantities (i.e. given by (4.10)) are the large sample estimates of the standard errors of the MLE estimates, $\hat{\psi}$. Thus, in small samples they could be biased. The Bayesian confidence intervals are based similarly on large sample properties (see Appendix (A; p.299)), except that both the mean of the posterior distribution and the variance are derived on the basis of large sample theory. These are informal remarks, but a formal comparison between Bayesian and non-Bayesian procedures is not within the scope of this study.

4.7 Summary

In this chapter we have considered the formulation of a quantal choice model which allows for the interdependent response of household members. If large data were available, this model could be used to meet Engelman's requirement that we should allow for interaction of household members in examining their job-quit behaviour. However, when interaction of other household members is considered the problems of job-quits and labour force participation become very closely related. That is, for example, if the male head of household faces the quit/stay decision the reaction of the wife will depend on her previous labour
force participation status.

Ashenfelter and Heckman \(^{(9)}\) have specified a simultaneous-equation linear probability model for the labour force participation of husband and wife. Here, interaction has been defined within a neo-classical labour supply model as income-compensated effects of a change in one partner’s wage on the other partner’s labour force participation. These cross-effects satisfy a symmetry restriction according to neo-classical theory. Ashenfelter and Heckman \(^{(9)}\) tested this restriction within their model, and obtained a positive sign for the income-compensated effect – implying complementarity of husband and wife participation decisions. Although it is an interesting conclusion, we have drawn attention to the limitations of the simultaneous linear probability model as its estimated parameters are susceptible to prediction bias and the consequences of heteroscedasticity. We have also considered Kniesner’s \(^{(82)}\) indirect measure of complementarity in a family labour supply model which does not require data on non-labour income. However the testability of Kniesner’s measure is questionable as it is based on two separate models.

In place of the simultaneous linear probability model we have adopted the simultaneous logit approach for the labour force participation responses of household members. This model meets the inefficiencies of the simultaneous linear probability model of Ashenfelter and Heckman \(^{(9)}\), and it is less sensitive to the identifiability problem. By analogy with the symmetry restriction of the income-compensated effects, the interdependence parameters of our model satisfy a similar symmetry restriction. Yet, these parameters do not only convey the effects of wage changes and non-linear income as the neo-classical model, but the effects of all explanatory variables which influence the labour force participation response of the other partner.
The model has been demonstrated, as in other studies, in the context of husband/wife household.

The rest of the chapter has concentrated on the estimation and statistical significance testing problem of the interdependence parameter. This topic has led us briefly to introduce the log-linear model which is closely related to the simultaneous logistic model. In particular we have adopted two methods of estimation and significance testing: a non-Bayesian and Bayesian procedure. The households of our data base consist of those who have been moved into council housing and those who have been notified to move. The main trend of our results supports the complementarity hypothesis of Ashenfelter and Heckman\(^{(9)}\), Kniesner\(^{(82)}\), and Daniel\(^{(40)}\) for husband and wife's participation responses. Although the sample size is rather small, yet the family size or the move do not seem to affect the general complementarity pattern.
CHAPTER V

THE PROBLEM OF AGGREGATE BINARY PREDICTIONS AND ELASTICITIES

5.1 Introduction

This chapter deals with the econometric problem of utilizing a disaggregate behavioural model for the purpose of making aggregate predictions based on the parameters of the population in question. The relevance of this problem to our study relates to the use which can be made of an estimated behavioural quit probability model to arrive at policy implications, which are normally based on aggregate properties of the population in question, rather than particular individuals. This requirement for a policy-oriented behavioural model has been made clear by Clowes\(^\text{(30)}\) while proposing his micro-level model of labour quits in his remarks:

"Although it is important for any proposed model to provide a good fit to experimental data, it is also important that the model provides a determination of the points at which external control or influence can be exercised over the system. Once the points of control have been determined, it is a relatively easy matter to predict the changes which will occur when the values of the controlling parameters are altered."

Clowes\(^\text{(30)}\), p. 245.

\* We are adopting the terminology of referring to disaggregate econometric models based on individual decision making units as \textit{behavioural} models, while sometimes referring to aggregate models based on groups of different individuals, or population averages, as \textit{empirical} laws (see McFadden and Reed\(^\text{(98)}\)).
The principle of utilizing behavioural models of quantal choice to arrive at aggregate predictions has been originally considered by McFadden and Reed(98), Talvities(136,137), Westin(148) and adopted by other studies, e.g. Watson and Westin(147). The possible aggregation bias resulting from adopting an aggregate model for a quantal choice situation has been discussed by McFadden and Reed(98) on the basis of a probit model and the assumption of a multivariate normal distribution for the vector of explanatory variables, \( X \), on which the model is based. This problem of aggregation bias has also been considered by Talvities(136) on the basis of the logistic model, but without the multivariate normality assumption for \( X \). The idea of deriving aggregate prediction on the basis of a behavioural logistic model has been discussed by Westin(148) who has first proposed Johnson's \( S_B \) curve for making binary choice predictions.

However, such studies as mentioned above appear to be confined to the field of travel demand, and transport model choice where the probits and logits have already gained a rich ground of application. Yet, we see no sign of such efforts being oriented towards similar problems of quantal analysis in labour economics, e.g. turnover analysis, labour force participation analysis, demographic composition and mobility between the states of employment, unemployment, and not in the labour force, or shortly the stocks and flows. In particular, the problem of aggregation bias tends to be ignored by some authors in this field, as we shall shortly note, e.g. Viscusi(144), Schmidt and Strauss(126), and Medoff(100). Thus this chapter attempts to handle the following problems:

* The more familiar analogous problem in the case of a continuous dependent variable (e.g. production of a good, consumption expenditure) has been dealt with by Theil(139) — but it does not concern us here.
(a) How a behavioural quit probability model can be of value to a
given employer or economic organisation to control and influence the
stability of its labour force - assuming that it is sufficiently
large. This will lead us to consider the shortages of some alternative
micro-level models, namely those of Silcock (128) and Clowes (30). On
this basis our behavioural model will assume the additional purpose of assess-
and influencing turnover at the micro-level, beside its direct macro-level
purpose *. However, since our data base for the behavioural models relates to
workers from different employers we shall use one model in a simple simulation
experiment to illustrate its general policy-oriented use. This will lead us to
consider the application and development of certain elasticity measures. Limita-
tions in our data mean that some variables relevant to turnover, particularly
those relating to job characteristics, cannot be investigated empirically in
this study, and the use of our model for policy-making purposes is limited
in this respect.

(b) We shall also explore the importance of aggregation bias with
special reference to the logistic model when the vector of explanatory
variables, \( \mathbf{X} \), is assumed to have a multivariate normal distribution
with mean vector \( \mu \), and variance-covariance matrix, \( \Sigma \). This will
lead us to explore the effect of heterogeneity of individuals on
average choice probability, \( E(\theta) \). Thus, we establish certain general
criteria for the concept of the variance-elasticity of \( E(\theta) \), first
introduced by Westin (148). Accordingly we derive the 'correct
formula' and criticize Westin's formula of variance-elasticity as not
conforming to these criteria.

The plan of the chapter will be as follows: in the next section
we shall discuss the idea of how the employees' characteristics means,
and their associated covariance structure at a given point of time

* We have distinguished between macro and micro level models in Chapter
II. The former type has been defined as relating to workers from
different firms even if the model is disaggregate.
within an organization can serve policy-control purposes. This point is demonstrated by analogy to Silcock (128) on the basis of the probability distribution of quit propensities of the heterogeneous employees. The importance of the explicit allowance for employees' heterogeneity in terms of their variance-covariance structure is explored in section (5.3). As regards the problem of specifying an appropriate mathematical form for the distribution of quit propensities, it is discussed in section (5.4). In this section we argue for the choice of Johnson's bounded system of frequency curves ($S_B$) in the light of two main objectives (i) description of quit structure at a given point of time, and (ii) description of potential policy effects designed to influence the employees' leaving rate. We illustrate this point on the basis of a simple simulation experiment, which shows the sensitivity of the $S_B$ curve to equal percentage changes in certain means of employees' characteristics - assuming that their covariance structure is unaffected. This idea of sensitivity of average quit propensity leads us to discuss the applicability of aggregate point elasticities of average quit propensity to small changes in characteristics means. This is done in section (5.5). In section (5.6) we discuss the problem of arc elasticity, i.e. when the changes in characteristics means are reasonably large. We develop an approximate formula for the computation of arc elasticity which is easy to apply compared to Richard and Ben-Akiva (121) and Westin (148). As regards the variance elasticity of the average quit propensity (i.e. $\sigma$-elasticity) it is discussed in section (5.7). The findings of this section are closely related to those of section (5.3) on the effect of employees' heterogeneity, except that the $\sigma$-elasticity is now worked out more formally. We have found it appropriate in section (5.8) to attract attention to the fact that the linear
probability model is c-inelastic. A brief summary of the main findings of this chapter is given at the end. We may now proceed to the details of these sections:

5.2 Quit Propensity and Policy Control Parameters

We have already proposed the logistic model to express quit probability of a given employee in terms of a set of independent explanatory variables contained in the p-vector $\mathbf{X}$,

$$\mathbf{X} = (X_1, \ldots, X_p)$$

where the $X_i$'s, generally, describe personal, family, and socio-economic characteristics of the employee, his employer's characteristics, and other exogenous labour market conditions. Hence depending on the observed value of $\mathbf{X}$ at a given point of time, each employee in a given economic organization may be assigned a quit propensity $\theta$. We shall assume that:

$$\theta = \frac{e^{\mathbf{\beta}'\mathbf{X}}}{1 + e^{\mathbf{\beta}'\mathbf{X}}} \quad (5.1)$$

where $\mathbf{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ is a $(p+1)$-vector of coefficient parameters.

The problem of estimation and significance testing of $\mathbf{\beta}$ has been discussed in the third chapter, but now we shall assume that either $\mathbf{\beta}$ is known, or that it has been successfully estimated on the basis of past data. Hence, a given employer is able to assign for each one of his employees a $\theta$ value depending on his observed characteristics, $\mathbf{X}$, as we shall assume. Nickel (108) has adopted this principle while studying the relationship between the personal and socio-economic characteristics of individuals and the life-time wage-structures which they choose. He has exploited the notion that
"individuals choose life-time wage structures which vary
in steepness with their own subjective estimation of
their likelihood of quitting. Their choices signal infor-
mation to employers concerning their individual quit
propensities and enable the employers to reward them
appropriately."

(Nickel (108), p. 199)

We have already considered this model among other models of quit
behaviour in the second chapter (see p. 64). Nickel (108) has only used
variability in individual quit propensities to explain wage-structures,
but this idea can be used more formally to understand the process of
labour turnover. Information about individual quit propensities of his
employees, give the employer a general measure of his labour force
stability, or its 'organizational commitment'. Silcock (128) has
adopted this idea to demonstrate that a constant wastage rate, \( \alpha \), of
a given labour force is only compatible with the assumption of per-
fectly homogeneous employees; see Silcock (128), p. 434).

Then, as we have mentioned in Chapter II, Silcock introduced the
hypothesis that: depending on different characteristics of employees
there is a whole series of different values \( \alpha_1, \alpha_2, \ldots, \alpha_n \) of wastage
rates*, or leaving rates, each one characterizing a small homogeneous
group. That is, if the \( i^{th} \) homogeneous group of employees is charac-
terized by the vector of characteristics, \( X_i \), they will possess a
wastage rate, \( \alpha_i \) (\( i = 1, \ldots, n \)). Silcock (128) hypothesizes that
these \( \alpha_i \)'s are described by a gamma probability density:

* Notice that wastage rate and quit propensity measure the same
phenomenon, though on the basis of different models.
\[ g(a) = \frac{\lambda^S}{\Gamma(S)} e^{-\lambda a} \alpha^{-S-1} \quad \beta > 0, S > 0 \tag{5.2} \]

which is completely determined by the parameters: \( S, \lambda \). However, although this model has been proposed to allow for the variability of employees characteristics, yet it does not show how its parameters \( \lambda, S \) are related to the distribution of the characteristics vector, \( \mathbf{X} \). Such an understanding, which is not directly provided by model (5.2) above, may advise the employer on how he might control and influence the stability of his labour force by manipulating the moments of the characteristics vector, \( \mathbf{X} \), either at the screening level or on the job. Yet, Silcock has originally proposed this model to demonstrate the effect of length of service on 'average' wastage rate, using the relation

\[
\phi(t) = \text{Prob\{an employee leaving within } [t, t+\delta t] \text{ his length of service } = t \}
\]

\[
= \left[ \text{Probability Density of } t \right] / \left[ \text{survival function up to } t \right]
\]

\[
= f(t)/F(t)
\]

\[
= - \frac{d}{dt} \log F(t) \quad t = 0, 1, 2, \ldots
\]

where \( \lambda(t) \) is the average wastage rate, or 'force of separation', of employees with completed length of service, \( t \). We have shown in the second chapter how Silcock has used model (5.2) above to yield a monotonically declining wastage rate:

\[
\lambda(t) = \frac{S}{\lambda + t} \quad t = 0, 1, 2, \ldots \tag{5.3}
\]

* Policy at the screening level is based on characteristics of new entrants, while on-job policy is based on provision of training, seniority rights and other work benefits.
Then he considers the effect of small changes in the parameters \( S \) and \( \lambda \) on the force of separation, \( \lambda(t) \), of a given firm's labour force:

"an increase in \( S \) is accompanied by a proportionate increase in wastage rate for all lengths of service. As \( \lambda \) increases, the wastage rate decreases .......

Silcock (128), p. 435

Although these remarks are easy to draw on the basis of (5.3), yet they have no explicit policy implications. Since the parameters \( S \) and \( \lambda \) are not directly connected with the moments of the characteristics vector, \( X \), there is no simple way of influencing \( S \) or \( \lambda \) either at the screening level, or on the job.

Let us now consider the alternative micro-level model due to Clowes\(^{(30)}\) with reference to his quoted remarks in the first page of this chapter. Clowes\(^{(30)}\) has apparently made this point to simplify Herbst's model* of "organizational commitment" which originally provided exceedingly good fit for describing distribution of length of service. But the too many parameters of his model were criticized as not discriminating well between different levels of turnover. Hence, Clowes\(^{(30)}\) adopted a model with only three transitional states,

\[
\begin{array}{c}
\text{new recruit} \\
\text{k}_1 \quad \text{Left} \quad \text{k}_2 \quad \text{Committed employee} \\
\text{k}_3
\end{array}
\]

with only three parameters, \( k_1, k_2 \) and \( k_3 \). The two leaving rate parameters \( k_1 \) and \( k_3 \) are defined as "control parameters" which may be altered through screening/training policies (Clowes\(^{(30)}\), p. 247).

For example, the effect of reducing \( k_1 \) by a given percentage is shown

* See review in the second chapter, section (2.3), p.
numerically and graphically to shift the survival curve upwards as in Figure (5.1) below:

![Graph showing effect of reducing quit rate of new recruits on survival curve.](attachment://figure51.png)

**Effect of reducing quit rate of new recruits on survival curve.**

**Figure (5.1)**

Yet, again these two leaving rate parameters, \( k_1 \) and \( k_2 \), do not really provide a clear guide for different screening/training policies to predict their effects on the survival curve. We may think of various possible policy combinations any one of which may affect a percentage reduction in \( k_1 \). Hence, unless the effect of different policies can be identifiable on the parameter \( k_1 \) its practical use as a control parameter is too limited. But, since policies for controlling turnover are to be spelt out in terms of the components of the characteristics vector \( X \) (e.g. employ less women, more adult males, increase pay, etc.) then, "points at which external control can be exercised" should be defined in terms of the moments of \( X \).

In the remainder of this chapter we shall illustrate how our quit propensity model (5.1) can be used for policy-oriented purposes. Our method is based on the specification of a distribution of quit propensities. However, unlike Silcock we shall allow this distribution to depend directly on the multivariate distribution of the employees.
characteristics vector, $X$. This is a natural consequence of our model (5.1) which explicitly assumes that each group of homogeneous employees characterized by a given $X_i$ have a quit propensity, $\theta_i$ ($i = 1, \ldots, n$). It is a natural consequence because the probability distribution of $\theta$, $g(\theta)$, will be completely determined by the parameters of the distribution of $X$, $f(X)$. Hence, as long as different screening/training policies could be identifiable in terms of changes in the parameters of $f(X)$, then their effects on $g(\theta)$ should be assessable with relative ease. Let us define the two parameters of the multivariate distribution, $f(X)$, as

$$
\begin{align*}
\mathbb{E}(X) &= \mu \\
\text{Var}(X) &= \Sigma
\end{align*}
$$

(5.4)

where $\mu' = (\mu_1, \ldots, \mu_p)$ is the vector of means of the characteristics vector, $X$, across the population. The positive semi-definite variance-covariance matrix $\Sigma$ is defined as $\Sigma = \{\sigma_{ij}\} \quad i, j = 1, \ldots, p$

where

$$
\sigma_{ij} = \text{cov}(X_i, X_j) \quad i \neq j
$$

$$
= \text{Var}(X_i) \quad i = j .
$$

Now, if we make the additional assumption that $f(X)$ is a multivariate normal density, then we know from statistical theory that the two moments $\mu, \Sigma$ characterize the distribution of $X$ completely. The normality assumption in most applications tends to provide a reasonable approximation and it can be justified on the basis of the Central Limit Theorem when the sample size is sufficiently large.

Hence, at a given point of time the employer may characterize his work force in terms of $\mu$ and $\Sigma$, the first two moments of their characteristics distribution. On the other hand, the moments of the
probability density of quit propensity, \( g(\theta) \), should be directly translatable to \( \mu \) and \( \Sigma \). Hence, the average quit propensity, \( E(\theta) \) - or the expected number of employees leaving the organization within a given interval - where

\[
E(\theta) = \int_0^1 \theta g(\theta) \, d\theta
\]  
(5.5)

will be shown to depend directly on the characteristics moments \( \mu \) and \( \Sigma \).

This point will be discussed in more detail later, but it may be useful to draw some analogy with Silcock's model as shown in Table (5.1) below. Note that we have defined wastage as the expected number (or proportion) of employees leaving their jobs within a short interval of time - the latter period being determined by the corresponding time reference of the model. Hence, the employer may influence the expected leaving rate \( E(\theta) \) by manipulating the means of the characteristics vector, \( \mu \), or its dispersion via \( \Sigma \). However, if a policy is expressed in terms of \( \mu \), as normally it is, its effectiveness should be measured by the sensitivity of \( E(\theta) \) to small changes in individual means, \( \mu_i \) (\( i = 1, \ldots, p \)). Hence we may be able to assess the potential effects of different screening/on-job policies on the expected leaving proportion, \( E(\theta) \). We shall elaborate on this point later, but now we note that we have expressed our problem as one of deriving aggregate predictions on the basis of a behavioural model (5.1). The advantage of this approach is that it allows for heterogeneity of employees through the recognition of a non-null variance-covariance matrix \( \Sigma \) of their individual characteristics. In fact, as we shall demonstrate in the next section, aggregation-bias arises as a result of assuming a null variance-covariance matrix, \( \Sigma \).
5.3 **Aggregation Bias and Effect of Heterogeneity in a Binary Logistic Model**

Aggregation bias has been examined by McFadden and Reed [98] in connection with inter-zonal travel demand models via a probit specification. These authors have derived expressions for the biases in aggregate model calibrations resulting from zonal homogeneity assumption in the variables. They have also discussed conditions under which these biases are important. Unlike the latter authors, who have adopted the multivariate normality assumption of the characteristics vector, \( \mathbf{X} \), Talvitie [136] attempted to answer a similar question, without the normality assumption, on the basis of the logistic model. His approach is simpler for a starter, though not sufficiently accurate as he himself has later examined; see Talvitie [137]. It may be described as a means of expressing the bias due to the adoption of the so-called naive formula for average probability of choice:

\[
\hat{p}^* = \frac{e^{\beta' \hat{u}}}{1 + e^{\beta' \hat{u}}}
\]

(5.6)

where \( \hat{u} = \mathbb{E}(\mathbf{X}) \), and \( \beta' \) is the coefficient parameter vector calibrated on the basis of the disaggregate behavioural model (5.1) above. The main limitation of the naive formula (5.6) is that it enforces the homogeneity assumption on the individuals' characteristics. This can be seen by showing that the use of \( \hat{p}^* \) is based on the hypothesis:

\[
\begin{bmatrix}
\Sigma = 0 \\
\text{or,}
\sigma_{ij} = 0
\end{bmatrix}
\]

(5.7)

The naive formula (5.6) reminds us of the approach adopted by Viscusi [144] for assessing the effect of job hazards on average probability of quit intention, as we have mentioned in the second chapter.
### Table (5.1)
Some analogy between Silcock's wastage model and our logistic model

<table>
<thead>
<tr>
<th>Wastage Model</th>
<th>Logistic Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha ) \equiv \text{Wastage-rate (or force of separation) defined by analogy with life-table demographic structure}</td>
<td>( \theta ) \equiv \text{quit (or leaving)probability defined in terms of the logistic model}</td>
</tr>
</tbody>
</table>

1. Each employee possesses a wastage rate: -

\[
0 \leq \alpha_i \leq 1 \quad (i = 1, \ldots, n)
\]

2. Differences in \( \alpha_i \) are due to differences in the characteristics of employees.

3. A given probability distribution \( g(\alpha) \) is assumed for the \( \alpha \)'s which is not explicitly related to the characteristics vector, \( X \).

4. Expected wastage rate

\[
\phi(t) = \frac{S}{\lambda + t} \quad S > 0, \quad \lambda > 0
\]

is derived with the parameters \( S \) and \( \lambda \). It is a decreasing function of length of service, \( t \).

Expected number leaving the establishment (or expected proportion) depends on the two moments of \( \mathbf{X} = \mathbf{\mu} \) and \( \mathbf{E} \), (see (5.4)). Length of service can be allowed for via the characteristics vector \( \mathbf{X} \), but its effect is only assessed empirically.
He has estimated the effect of the job hazard dummy by taking the difference between two estimates of $P^*$, both evaluated at the means of all other explanatory variables, except that one of them contains job hazard at the zero level, and the other has it at the unity level. Using a similar approach Schmidt and Strauss (126, p. 752) have compared the effects of different race-sex combinations on average choice probability of different occupation-industry combinations. They have done so by computing a $P^*$ for each race-sex cell evaluated at means of all other explanatory variables, and then comparing these $P^*$'s. The same approach has been adopted by Medoff (100, p. 387) to assess the effect of unionism on average quit probability, and lay-off probability in the U.S. manufacturing industry. Like Viscusi (144), he takes the difference between two $P^*$'s both evaluated at the means of all other explanatory variables except for difference in the level of unionism. It seems that this naive approach, though inaccurate, yet is not uncommon. It also seems that the approach of measuring effects of certain variables on average choice probability via aggregate elasticities has not gained sufficient popularity among labour economists.

Now, we return to Talvitie's method for correcting the bias in the naive formula $P^*$ of (5.6), when the homogeneity conditions (5.7) are not true. Let us define

$$V_0 = \bar{\theta}'u = E(\theta'X_i) = E(V_i)$$

and similarly let:

$$\sigma^2 = \text{Var}(\theta'X_i) = \theta'\Sigma \theta \quad i = 1, \ldots, n)$$

where $\theta \neq 0$, so that $\sigma^2 = 0$ if and only if $\Sigma = 0$ (since $\Sigma$
is a positive semi-definite matrix).

Hence, the homogeneity conditions (5.7) can be expressed equivalently as \( \sigma^2 = 0 \). For the present we shall assume away the trivial case where \( V_0 = 0 \), but we shall allow for it later.

Then, Talvite's method is to expand the individual choice probabilities, \( \theta_i \), around the naive average probability formula, \( P^* \) of (5.6) such that

\[
\theta_i = e^{V_i/(1 + e^{V_i})}
\]

\[
= P^* + (V_i - V_o) \frac{d\theta_i}{dV_i} \bigg|_{V_i = V_o} + (V_i - V_o)^2 \frac{d^2\theta_i}{dV_i^2} \bigg|_{V_i = V_o} + \text{Higher Order Terms}
\]

where

\[ V_i \equiv \theta' X_i \quad (i = 1, \ldots, n). \]

Then, if the sum of higher order terms is negligible we may apply the mathematical expectation on (5.9) to get:

\[
E(\theta_i) = P^* + \sigma^2 P^* (P^* - 1)(P^* - \frac{1}{4}) \quad (5.10)
\]

Hence, this formula adjusts the naive average probability model by allowing for the dispersion of the \( V_i \)'s across the population. Thus, the naive formula provides a good approximation only under the homogeneity condition, i.e.

\[
E(\theta_i) = P^* \quad \text{only if} \quad \sigma^2 = 0 \quad \text{(given} \ V_o \neq 0). \quad (5.11)
\]

Yet, a stronger form of the result (5.11) can be obtained when the multivariate normality assumption is made for the characteristics vector, \( X \). This situation has been explored by McFadden and Reed\(^{98} \) when the individual choice probabilities, \( \theta_i \)'s, follow the probit model:

\[
\theta_i = \phi(\theta' X_i) \quad (i = 1, \ldots, n)
\]
where \( \phi(\cdot) \) is the normal distribution function. The normality assumption is stated on the basis that individual \( V_i \)'s are independently, identically distributed with mean \( V_o \) and variance \( \sigma^2 \).

Alternatively, using standardized scores:

\[
\frac{V_i - V_o}{\sigma} \sim N(0, 1) \quad i = 1, \ldots, n. \tag{5.12}
\]

Hence, average probability of choice is defined as:

\[
E(\theta) = E\left[ \phi(V) \right] = \int_{-\infty}^{\infty} \phi(V) \phi\left( \frac{V - V_o}{\sigma} \right) \, dV \tag{5.13}
\]

where \( \phi(\cdot) = \frac{d}{dV} \phi(\cdot) \), is the Normal probability density. Then, applying the convolution properties of the normal distribution, McFadden and Reed have shown that it is possible to write:

\[
E(\theta) = \phi\left( \frac{V_o}{1 + \sigma} \right). \tag{5.14}
\]

Note that the equivalent definition for the naive average probability using the probit model is given by:

\[
P^* = \phi(V_o). \tag{5.15}
\]

Hence we may conclude from (5.14) and (5.15) above that given the multivariate normality assumption, (5.12) and the probit specification for binary choice, we can write:

\[
E(\theta) = P^* \text{ if and only if } \sigma^2 = 0. \tag{5.16}
\]

Thus, when the normality assumption is introduced we are able to express condition (5.11) above as an equivalence condition.

We shall now turn to establish this equivalence condition when the individual choice probabilities, \( \theta_i \), follow the logistic model, given
by the first line of relations (5.9) above. Similarly, we assume that \( V_i \)'s follow the normality assumption made at (5.12) above. Then, it follows that:

\[
E(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left( \frac{e^{V \sigma Z}}{1 + e^{V o + \sigma Z}} \right) e^{-Z^2/2} dZ . \tag{5.17}
\]

Now, if we define the standardized score

\[
Z = \frac{V - V_o}{\sigma},
\]

and accordingly transform the integral (5.17) above, we will get:

\[
E(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left( \frac{e^{V o + \sigma Z}}{1 + e^{V o + \sigma Z}} \right) e^{-Z^2/2} dZ . \tag{5.18}
\]

Recalling that the naive formula \( P^* = \frac{e V o}{1 + e o} \), then, the equivalence condition (5.16) follows immediately from the integral equation (5.18) since

\[
E(\theta) = \frac{V o}{1 + e o} \text{ if and only if } \sigma = 0 .
\]

However, the use of the naive formula is justified in the special case where \( V_o = 0 \) (i.e. \( P^* = \frac{1}{2} \)). This point follows directly from equations (5.10), (4.15), and (5.18) when evaluated at \( V_o = 0 \), since in this case it can be shown that:

\[
E(\theta) = \frac{1}{2} \text{ at } V_o = 0, \sigma > 0 \tag{5.19}
\]

The proof of this result on the basis of the logistic model and the normality assumption (i.e. equation (5.18)) can be based on a result due to Johnson (77, p. 174), that it is possible to evaluate integral (5.18) at \( V_o = 0 \) to get:
\[
E(\theta) = \frac{1 + 2 \sum_{n=1}^{\infty} e^{-n^2\sigma^2/2}}{\sqrt{(2\pi)/\sigma} \left[ 1 + 2 \sum_{n=1}^{\infty} e^{-2n^2\pi^2/\sigma^2} \right]}
\]

which has been shown by Johnson to be equal to \( \frac{1}{2} \) whatever the value of \( \sigma \).

Hence we get the implication that at \( V_0 = 0 \), or alternatively at \( E(\theta) = \frac{1}{2} \), any variation in the variance has no effect on average probability of choice.

Thus, the naive formula may still provide a good numerical approximation when \( V_0 = 0 \), and the degree of approximation becoming better as the standard deviation, \( \sigma \), gets smaller approaching the homogeneity condition (5.7). Hence, unless these conditions are guaranteed, the use of the naive formula may result in misleading results.

In order to illustrate this point we are going to utilize one of our empirical models (which will all be discussed in the next chapter) in a simple simulation test. This test involves the drawing of two curves for average quit probability evaluated at a specified range of \( V_0 \) values with \( \sigma^2 \) having a fixed value. The first curve represents the naive formula, \( p^* \) (i.e. \( \sigma = 0 \)), while the second curve represents the integral equation (5.18) above for \( E(\theta) \). Both curves are based on a sample of 157 male heads of households who have been employed prior to the move. The sample proportion of quits is defined as

\[
\hat{\theta} = \frac{\text{Number of individuals who left their jobs over the period of observation}}{\text{Total number of previously employed individuals}} = \frac{157}{157} = 0.25478
\]

where the dichotomous variables \( Y_i \) take the value one if the \( i \)th individual quits, and zero otherwise. Then, using the method of maximum
likelihood, we estimate the model

\[ \hat{\theta}_i = \frac{\hat{\beta}'x_i}{1 + e^{-\hat{\beta}'x_i}} \quad (i = 1, \ldots, 157) \]

where \( \hat{\theta}_i \) are estimated individual quit propensities based on the observed characteristics vector \( X_i \). The components of this vector for any given individual are \( X = (TT, RR, ED, INC, HT, AGE) \)

\[ \begin{align*}
TT &\equiv (0, 1) \text{ Dummy for change in Travel Time to work after move} \\
RR &\equiv \text{Change in housing rent and rates after the move} \\
ED &\equiv \text{Level of education} \\
INC &\equiv \text{Other non-labour income received} \\
HT &\equiv \text{Housing tenure} \\
AGE &\equiv \text{Age in years} \\
\end{align*} \]

The corresponding estimates for the coefficient parameters are expressed in the vector below:

\[ \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_{TT}, \hat{\beta}_{RR}, \hat{\beta}_{ED}, \hat{\beta}_{INC}, \hat{\beta}_{WN}, \hat{\beta}_{AGE}) \]

\[ = (0.069, 1.272, -1.130, -1.130, 0.994, -0.330, -0.060) \]

\[ (0.107) (3.008) (2.526) (2.378) (1.641) (0.655) (2.865) \]

(5.22)

with \(-2 \log \lambda = 33.380, \quad F^* = 5.716\).

The numbers in the parentheses represent t-ratios corresponding to the various components of the MLE, \( \hat{\beta} \). The computed value of the log-likelihood ratio, \(-2 \log \lambda\) can be compared with the tabulated value of the chi-square table at seven degrees of freedom and 1% level of significance (= 6.635), which implies that the model has a very good fit. This result is confirmed at the 5% level of significance by the value of our \( F^* \)-ratio statistic, which we have introduced in the third chapter, section (3.5), as it is greater than \( F(156,7) = 4.85 \). It
is also easy to check that the signs of the coefficient parameters agree with a priori expectations. (This model is only used here for its illustrative purpose, but we have devoted the sixth chapter for more specialized discussion of our empirical models.)

Moreover, we need the sample estimates \( \hat{V}_o \) and \( \hat{\sigma}^2 \) for the parameters \( V_o \) and \( \sigma^2 \) respectively, and we get:

\[
\hat{V}_o = -1.3852 \quad \text{and} \quad \hat{\sigma}^2 = 1.2765 \quad (5.23)
\]

The first result we get is that for the naive formula:

\[
p^* = \frac{V}{1 + e^V}
\]

\[
= 0.20771 \quad \text{at} \quad V = V_o. \quad (5.24)
\]

We notice that \( p^* \) gives an underestimate for the sample average quit probability, \( \hat{\theta} = 0.25478 \) given by result (5.20) above. This is to be expected as long as \( V_o \neq 0 \) and homogeneity condition of individuals' characteristics is rejected by the fact that \( \hat{\sigma}^2 = 1.486 \). Hence, we may proceed with our simulation test to examine the effect of this \( \hat{\sigma}^2 \) at various values of \( V_o \). This will be done by generating two curves defined on the range \( -4.31 \leq V_o \leq 1.84^+ \). The first curve is based on the naive formula (5.24) above, while the second curve is based on the integral equation (5.18) above, which we may rewrite below:

\[\text{+ These end points are respectively the minimum and maximum values of } V_o \text{ in our model. Most cases are concentrated on the negative side as can be seen from the difference between the end points.}\]
\[ E(\theta) = \frac{1}{2} \int_{-\infty}^{\infty} \left( \frac{V_o + \hat{\sigma}Z}{1 + e^{V_o + \hat{\sigma}Z}} \right) e^{-Z^2/2} \, dZ \] (5.25)

\[ 4.31 \leq V_o \leq 1.84. \]

The evaluation of equation (5.25) can be simplified by defining:

\[ v = V_o + \hat{\sigma}Z, \]

and transforming the integral (5.25) accordingly to get

\[ E(\theta) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{v}}{1 + e^{v}} e^{-\frac{(v-V_o)^2}{2\sigma^2}} \, dv \] (5.26)

The above integral could be transformed further by defining

\[ \theta = e^v/(1+e^v), \]

leading to

\[ E(\theta) = \frac{1}{\sigma \sqrt{2\pi}} \int_{0}^{1} \frac{1}{(1-\theta)} e^{-\frac{1}{2\theta^2} \left( \log \left( \frac{\theta}{1-\theta} \right) - V_o \right)^2} \, d\theta \] (5.27)

Hence, the integral (5.25) has now been expressed with finite end points, and may be evaluated by numerical integration over the specified range of \( V_o \).

It is interesting to note that when equation (5.27) is evaluated at the sample estimate of \( \hat{V}_o \) (see results (5.23)), we get average quit probability

\[ E(\theta) = 0.25499 \] (5.28)

which is a very good approximation to the sample proportion of quits, \( \hat{\theta} = 0.25478 \). This agreement is specially interesting because \( E(\theta) \) is only calculated via numerical integration which should naturally involve some sacrifice of precision. We shall return to this point later, but now we adopt the curve related to (5.27) above as the correct one which allows for heterogeneity of employees in terms of \( \hat{\sigma}^2 \), while the curve related to the naive formula (5.24) represents the homogeneity hypothesis, \( \sigma^2 = 0 \). The evaluation of \( P^* \), and \( E(\theta) \) at particular

* Due to lack of a simple closed form for \( E(\theta) \).
points in the interval \(-1.43 \leq V_o \leq 1.84\) has been done by a computer program, while the drawing of the related two curves have been drawn via another program using the graph-plotter (see Appendix (B, p. )). The outcome of the experiment is shown in Figure (5.2) below. The vertical spread between these two curves at different values of \(V_o\) is due to the heterogeneity effect represented by \(\sigma^2 = 1.276\). This spread gets wider and wider as \(V_o\) increases in absolute value, while at very small absolute values of \(V_o\) the spread gets narrower and narrower. As expected the two curves coincide at \(V_o = 0\), where \(E(0) = \hat{P}^* = \frac{1}{2}\). We also see that the 'true' average quit probability (the sample proportion) lies effectively on the \(E(0)\) curve at \(V_o = \hat{V}_o\). The estimate of the naive formula, \(\hat{P}^*\), falls below \(E(0)\).

More specifically we notice that

\[
\begin{align*}
E(P) &< \hat{P}^* \quad \text{where } V_o > 0, \text{ i.e. at } \hat{P}^* > \frac{1}{2} \\
E(P) &= \hat{P}^* \quad \text{where } V_o = 0, \text{ i.e. at } \hat{P}^* = \frac{1}{2} \\
E(P) &> \hat{P}^* \quad \text{where } V_o < 0, \text{ i.e. at } \hat{P}^* < \frac{1}{2}
\end{align*}
\]

(5.29)

In other words, \(\hat{P}^*\) gives an overestimate to \(E(0)\) when the mean linear combination of characteristics, \(V_o\), lies on the positive segment, while \(\hat{P}^*\) gives an underestimate to \(E(0)\) when \(V_o\) lies in the negative segment. These two implications are difficult to work out analytically on the basis of the integral equation (5.27) which we have adopted to run out test. However, they can be derived easily on the basis of Talvitie's approximate formula (5.10) just by putting \(p^* < \frac{1}{2}\).

Of course the extent of the vertical spread between the two curves in Figure (5.2) is accounted for by the magnitude of the variance. In general it is possible to show that the larger the variance, the wider is this extent of vertical spread. This point can easily be illustrated using Talvitie's approximate formula (5.10), but since this is
curve describing $E(0)$

curve describing $P^*$

Figure (5.2)
only a linear approximation the \([0,1]\) restriction for \(E(\theta)\) is not guaranteed at extreme values of \(\sigma\). Alternatively the effect of variance increases can be depicted as in Figure (5.2(a)) below, using the same method of Figure (5.2) above.

![Diagram](attachment:image.png)

**Figure (5.2(a))**

A hypothetical diagram to describe the effect of increases in the standard deviation on average probability

In the above diagram the connected curve (i) represents the naive formula where \(\sigma = 0\). The curve just next to it (ii) represents \(E(\theta)\) when \(\sigma = \sigma_1 > 0\). The effect of \(\sigma_1\) is to stretch the sigmoid curve with the result that its slope is now lower than before at finite \(V_o\) values. When curve (iii) is introduced with \(\sigma_2 > \sigma_1\), the curve gets more stretched, and the slope of the curve lowered even more. We can imagine that successive increases in \(\sigma\) produce curves with lower and lower slopes approaching the zero-slope 'limiting line' at \(E(\theta) = \frac{1}{2}\) for all \(V_o\). Obviously as the slope is lowered (i.e. \(\sigma\) increased)
the extent of vertical spread from the original naive formula curve increases. This "limiting line" argument can be derived rather heuristically from the fact that successive curves (e.g. in Figure (5.2(a)) curves (i), (ii), (iii)) have lower and lower slopes due to increase in the standard deviation, \( \sigma \); i.e.

\[
P^*(1 - P^*) > E(ii)(\theta(1 - \theta)) > E(iii)(\theta(1 - \theta)) > \ldots > 0
\]

where \( P^*(1 - P^*) = \frac{3P^*}{3V_o} \), and, \( \frac{3E(\theta)}{3V_o} = E(\theta(1 - \theta)) \).

The fact that \( 0 \leq P^* \leq 1 \) and \( 0 \leq \theta \leq 1 \) guarantees \( E(\theta(1 - \theta)) \) to be a non-negative number for all values of \( V_o \). Hence as \( \sigma \) increases indefinitely we approach the curve with the lowest possible slope, i.e.

\[
\frac{3E(\theta)}{3V_o} = E(\theta(1 - \theta)) = 0
\]

which implies that:

\[
\lim_{\sigma \to \infty} E(\theta) = \text{const.} \quad -\infty < V_o < \infty
\]

with the boundary condition:

\[
E(\theta) = \frac{1}{2} \quad \text{at} \quad V_o = 0, \ 0 \leq \sigma \leq \infty.
\]

Hence, const. = \( \frac{1}{2} \), or

\[
\lim_{\sigma \to \infty} E(\theta) = \frac{1}{2} \quad -\infty < V_o < \infty.
\]

In other words as the standard deviation \( \sigma \) increases indefinitely the average choice probability approach \( \frac{1}{2} \) for all finite values of \( V_o \) - we are restricting our attention to finite \( V_o \).\(^\dagger\)

\(^\dagger\) The proof of \( \frac{3E(\theta)}{3V_o} = E(\theta(1 - \theta)) \) is given in Section (5.5). See equation (5.55).

\(^\ddagger\) The case where \( |V_o| \to \infty \) is considered by RESULT (5.1), Section (5.6).
We have also noted that the curve with comparatively highest variance is placed at the top of other curves when \( V_o \) is negative, and at their bottom when \( V_o \) is positive. This point bears a special implication for the effect of changes in employees heterogeneity on average quit probability. These implications will be handled more formally in section (5.7) where we shall use integral (5.27) of \( E(\theta) \) to express the above point as

\[
\frac{\partial E(\theta)}{\partial \sigma} = \begin{cases} 
< 0 & \text{for } V_o > 0 \text{ or } P^* > \frac{1}{2} \\
= 0 & \text{for } V_o = 0 \text{ and } P^* = \frac{1}{2} \\
> 0 & \text{for } V_o < 0 \text{ and } P^* < \frac{1}{2}
\end{cases}
\]  
(5.30)

It is very easy to verify these properties on the basis of Talvitie's formula (5.10) where:

\[
\frac{\partial E(\theta)}{\partial \sigma} = P^*(P^* - 1)(P^* - \frac{1}{2}) .
\]  
(5.31)

In other words: the effect of increasing the dispersion of employees, while keeping the linear combination of their characteristics, means constant (i.e. \( V_o = c \)) will depend on the fixed value of \( V_o \). If \( V_o \) is fixed at a negative value, then \textit{ceteris paribus} increasing the dispersion of employees may raise average quit propensity, \( E(\theta) \). Alternatively if \( V_o \) is fixed at a positive value, increasing dispersion may lower \( E(\theta) \). However if \( V_o \) is fixed at zero, then manipulating \( \sigma \) alone may bring no effect. The intuitive appeal of these implications is not trivial but will be illustrated in section (5.7) where a more formal treatment is given to this problem.

The main conclusion of this section is that we have to allow for the heterogeneity of individuals' characteristics explicitly in terms of \( \sigma^2 \) while making aggregate statements on the basis of a behavioural
model. Unless there is reason to suspect that the two curves of Figure (5.2) are very close together, the naive formula approach (adopted by Viscusi (144), Schmidt and Strauss (126), and Medoff (100)) cannot be justified. The method of deriving aggregate prediction on the basis of a behavioural model should therefore be based on an explicit form for the distribution of individual choice probabilities, \( \theta_i \) - or the distribution of quit propensities, \( g(\theta) \), in our case. We shall now turn to elaborate on this point.

5.4 Properties of The Probability Distribution of Quit Propensities

In the last section we have defined average quit probability, \( E(\theta) \), in terms of the probability distribution of the linear combination of characteristics, \( V = \beta'X \), assuming that

\[
V \sim N(V_0, \sigma) .
\]

Thus, we have defined \( E(\theta) \) in terms of the mean of \( V \) (i.e. \( V_0 \)) and its variance \( (\sigma^2) \), as expressed in equation (5.18), or equivalently (5.27). Before we consider the usage of \( V_0 \) or \( \sigma^2 \) as policy control parameters, it is desirable to express \( E(\theta) \) directly in terms of \( g(\theta) \), the distribution of employee quit propensities. We believe that it is desirable to work directly on the basis of \( g(\theta) \) since the different geometrical shapes for \( g(\theta) \) may provide simple descriptions for the state of organizational commitment of a large number of heterogeneous employees. In other words, it is possible to find two economic organizations which experience the same average quit rate per unit of time, but their structures of quit propensities are different due to different compositions of employees. For example, we may encounter the following quit structures:
In Figure (5.3, (a)) it is very unlikely to find an employee with a quit propensity close to one, and a very small proportion of employees have propensities greater than 0.50. In Figure (5.3, (b)) the picture is rather different - although average quit propensity may be equal. In the latter case it is more likely to have employees with propensities very close to one, but again it is more likely to find employees with propensities closer to zero than in case (a). This heterogeneity of quit propensities is even sharper in Figure (5.3, (c)). Here there are two fairly distinct groups of employees as described by the bi-modal curve. This curve suits the description of 'stayers' and 'movers' whose quit propensities are concentrated at very close to zero and
very close to one respectively. This could be the case of a firm which provides "jobs specifically tailored to the requirements of high quit individuals," (see Nickel (108), p. 203). At the macro-level this U-shaped phenomenon has been established by Heckman and Willis (69) for the distribution of labour force participation probabilities of married women. It implied that most women had probabilities of participation close to zero or to one. We have also considered Lane and Andrew (86) where they compared the crude turnover rates of two departments in a branch of U.S. steel industry; by looking at the composition of employees in each department they attributed this difference to the fact that the department with the higher turnover had a predominantly bigger proportion of short-time employees who usually had higher leaving propensities. Hence, looking at the structure of individual quit propensities helps to explain the observed variations in average quit probability.

The second use of \( g(\theta) \) is that it helps to provide a geometrical description for the potential effects of different policies imposed on \( V_o \), the mean linear combination of characteristics, or the variance \( \sigma^2 \). The potential effect of a policy which aims at reducing average quit probability may be assessed not only in terms of changes in \( E(\theta) \), but also in terms of the extent of the shift of \( g(\theta) \) to the left. For example it may be decided to increase average age of employees by recruiting less youngsters (e.g. below 25) since quit propensity declines with age (e.g. see empirical model (5.22). Or the policy may assist employees getting accommodation closer to place of work in order to reduce their average travel time to work, or provide special transport service. However, it is important to compare the expected savings achieved by a given percentage fall in the average quit rate, and the associated costs of policy implementation. Yet, Clowes (30, p. 249)

* We may also recall Figure (1.1) of Chapter I used by Mackay et al. (92) to show the effect of skill on the distribution of plant average quarterly quit-rates.
has overlooked costs of training provision in his estimate for the expected savings to the Glacier Metal Company as a result of halving the quit rate of its new recruits and shifting up the employees survival function. In fact, considering the costs of training and other job benefits, a rational employer may deliberately tolerate a given rate of wastage for his labour force which he may administer by providing special jobs for brief tenure workers, or generally unskilled workers. Hence, in the example just cited, a rational employer may choose one of the two policies, or a combination of them, depending on

(i) Effectiveness: as judged by the sensitivity of average quit propensity, e.g. to a small change in average age, or average travel time to work of employees - or in general other measurable factors.

(ii) Costs: both direct and indirect - of implementing either of the two policies, or a combination of them.

Hence the actual policy adopted will be optimal in the profit maximization/loss minimization sense.

However, we assume that costs are given, and handle the first consideration of effectiveness. We shall represent potential policy effects in terms of shifts in the quit propensity distribution, g(θ), as an alternative to the survival curve approach adopted by Clowes and others. Hence the potential effectiveness of any policy, or combination of policies, can be described by the extent of a shift of the curve towards the left, as this implies greater concentration of quit propensities near zero.

Thus, we shall utilize our curve g(θ), for two illustrative purposes (as outlined above) (i) description of the underlying quit structure, as in Figures (5.3) above, and (ii) description of potential

* For example, see Figure (5.1) above.
policy effects. These purposes require that $g(\theta)$ should possess certain properties as:

(i) **Flexibility:** It should be capable of picturing a wide range of shapes that may occur in practice.

(ii) **Parameterization:** It should possess a parameter system which is easily translatable as policy-control quantities, i.e. the moments $V_0$ and $\sigma^2$ defined above. As Westin\(^{(148)}\) has remarked, this property may also permit the transferability of a behavioural model to predict a new situation where only aggregate data is available in terms of $V_0$ and $\sigma^2$.

(iii) **Preservation:** Since we may use this curve for assessing the effectiveness of different policies aimed at adjusting certain points of its parameter system, then the new curve should belong to the same family of distribution as the original one. In other words, $g(\theta)$ should be preserved under changes of its parameter values.*

Fortunately, the curve, $g(\theta)$, which meets these requirements is readily obtainable from the normality assumption which we have already adopted for the linear combination of employees’ characteristics, $V = \beta'X$, with moments $V_0$ and $\sigma^2$, i.e.

$$V \sim N(V_0, \sigma^2) \quad (5.31)$$

Then, we shall derive $g(\theta)$ using the "method of translation" following the procedure of Johnson\(^{(77)}\). Let us define the standard normal variables:

* For example, we know that if $x$ is normally distributed with parameters $\mu$ and $\sigma$, i.e.,

$$x \sim N(\mu, \sigma)$$

then the normality condition is preserved for all values of $(\mu, \sigma)$, $\sigma > 0$, $-\infty < \mu < \infty$.

However, we may find another distribution, e.g. the Binomial distribution $B(n, p)$, with moments $np$, $np(l - p)$. It is known that if $np = \lambda$, but $p \to 0$ and $n \to \infty$, then we get a Poisson distribution, $P_0(\lambda)$. Hence $B(n, p)$ is not preserved under all parameter values $0 < n < \infty$, $0 < p < 1$.
and introduce the parameters:

\[ \gamma = - \frac{V_o}{\sigma}, \quad \delta = \frac{1}{\sigma} \]  (5.33)

Recalling that, \( \Theta = e^V/(1 + e^V) \), it follows:

\[ \ell(\Theta) = \log(\frac{\Theta}{1-\Theta}) = V \]  (5.34)

Thus, we can rewrite equation (5.32) as:

\[ Z = \gamma + \delta \ell(\Theta) \]  (5.35)

Then, the method of translation can be based on the last equation, using the property that:

\[ g(\Theta) = g(Z) \left| \frac{dZ}{d\Theta} \right| \]  (5.36)

where \( \left| \frac{dZ}{d\Theta} \right| \) is the absolute value of the Jacobian of translation given by

\[ \frac{dZ}{d\Theta} = \frac{\delta}{\Theta(1-\Theta)} \quad 0 < \Theta < 1 \]  (5.37)

and

\[ g(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \]

\[ = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \gamma + \delta \ell(\Theta) \right)^2} \quad -\infty < Z < \infty \]

Hence, we can immediately write:

\[ g(\Theta) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{\Theta(1-\Theta)} e^{-\frac{1}{2} \left( \gamma + \delta \ell(\Theta) \right)^2} \quad 0 < \Theta < 1. \]  (5.38)

The family of distributions, \( g(\Theta) \), described by equation (5.38) has been introduced by Johnson (77) who has called it the bounded system of frequency curves obtained by the translation method, or shortly the \( S_B \) curve.
The flexibility of the $S_B$ curve can be shown for some selected parameter values $\gamma$ and $\delta$ as below.

\[ \gamma = 0; \delta = 1^{1/2} \]

\[ \gamma = 0; \delta = 0.05 \]

\[ \gamma = 0.533; \delta = 0.05 \]

\[ \gamma = 0; \delta = 0.1 \]

Figure (5.4) Different possible shapes of the $S_B$ curve.

The parameterization conditions are satisfied by the fact that $\gamma$ and $\delta$ are simple functions of the characteristic moments $V_o$ and $\sigma^2$; see (5.33). The preservation condition is also satisfied by the fact that any change of the two moments, $V_o$ and $\sigma^2$ which preserves the normality condition, will result in a translated distribution, $g(0)$, which is also a member of the $S_B$ family.

At this stage it is worthwhile recalling the approaches of Morrison (102), and Heckman and Willis (69) who have adopted the Beta distribution for its geometrical flexibility. Particularly, Heckman and Willis (69) adopted the Beta curve to describe the distribution of females sequential labour force participation probabilities, for its capacity of yielding a U-shaped curve such as that of Figure (5.4, ii). The U-shaped curve has been expected to describe the 'movers' and
'stayers' phenomenon of female labour force participation. Similarly Morrison has based his upper bounds theory of correlation coefficients on the assumption that true choice probabilities follow a Beta distribution; see Chapter III, section (3.5). However, the flexibility of the Beta curve is restricted by the fact that it cannot produce a non-zero mode, and a non-unity mode when it is bimodal, as in Figure (5.3, ii), i.e. it gives non-zero points of "high contact". Moreover, it does not meet the parameterization condition since its two parameters* \((a, \beta)\) are not directly related to the moments of the population characteristics. The violation of the preservation condition has been observed by Westin that "a simple additive shift in one of the characteristics determining choice would produce a transformed distribution of probabilities that is not distributed beta".

Thus, these three properties of the \(S_B\) curve advocate it as a convenient representation for our problem. Specifically, it satisfies the two descriptive requirements mentioned** above: for the structure of quit propensities, \(g(\theta)\), and its potential sensitivity to various policy effects. Although our data base relates to workers previously employed by different firms, yet this procedure can equally well suit turnover analysis of a single firm. The numerical illustration which follows is based on our previous empirical model used in section (5.3) above. This time we shall utilize the model to fit the \(S_B\) curve, \(g(\theta)\), and compare it with the actual relative frequency distribution of the estimated quit propensities, \(\hat{\theta}_i\). The six variables of the model are

---

* The general form of the Beta distribution is

\[
f(x) = \frac{1}{B(a, \beta)} x^{a-1} (1-x)^{\beta-1} .
\]

** See bottom of page 180.
defined in (5.21), and the MLE's for coefficient parameters given in (5.22). Then, we shall utilize the fitted $S_B$ curve in a simple simulation experiment to test the potential sensitivity of quit structure to various policy combinations.

The problem of fitting the $S_B$ curve has been considered by Johnson (77), Elderton and Johnson (48), Westin (148), Hill (73), and others (see other references in Chapter IV, Kendal and Stewart (80)). The general form of the transformed sample points used for fitting the $S_B$ curve is given by:

$$l_i(\theta) = \log \left( \frac{\theta_i - \xi}{\xi + \lambda - \theta_i} \right) \quad \xi \leq \theta_i \leq \xi + \lambda$$

where $\xi, \lambda$ are some parameters, while $\theta_i$ are observed data.

There are generally three methods which can be adopted for the curve fitting process which are (i) the method of percentile points, (ii) the method of maximum likelihood and (iii) the method of moments. Johnson (77) has considered the problem of curve fitting for the sample points (5.39) above, under three cases - (a) where both $\xi$ and $\lambda$ are known, (b) where only one of them is unknown, and (c) where both $\lambda$ and $\xi$ are unknown. The first case (a) is considered the easiest since it only involves fitting a normal curve to the transformed observations, $l_i$, and then getting maximum likelihood estimates for $\gamma$ and $\delta$ as

$$\hat{\gamma} = -\frac{\hat{V}_o}{\sigma}, \quad \hat{\delta} = \frac{1}{\sigma}$$

(3.40)

where $\hat{V}_o = \frac{1}{n} \sum_{i=1}^{n} l_i$, and $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} (l_i - \hat{V}_o)^2$.

Clearly, our model suits this case best, since we have

* See Johnson (77, p. 160-162).
\( \lambda = 1, \, \xi = 0.\) However, we differ as regards the treatment of the \( \theta_i \) as observed data since our data base is composed of estimated probabilities, \( \hat{\theta}_i.\) The points of our transformed data are:

\[
\ell_i(\hat{\theta}_i) = \log \left( \frac{\hat{\theta}_i}{1 - \hat{\theta}_i} \right) = \hat{\beta} X_i \quad (i = 1, \ldots, n) \tag{3.41}
\]

where \( \hat{\beta} \) is the MLE of \( \beta \) (see (5.22)).

Hence, our logistic transforms, for which we may fit the normal curve, are not directly observed as required by case (a) above. To get around this problem we shall ignore the randomness associated with \( \hat{\beta} \) and treat the \( \ell_i(\hat{\theta}_i) \) as known sample values. The point of ignoring randomness of estimates is not uncommon, e.g. see Cox(35, p. 32). It has also been applied by Westin(148) for fitting the \( S_B \) curve.

Therefore we shall fit a normal curve to the estimated logistic transforms defined in (5.41) above, using our estimates

\[
\hat{V}_o = \hat{\beta} \hat{U} \quad \text{and} \quad \hat{\sigma}^2 = \hat{\beta} \hat{E} \hat{\beta} = -1.3852 = 1.276 \tag{5.23}
\]

(see numerical results (5.23)).

Then, we fit our \( S_B \) curve, \( g(\hat{\beta}) \), using the two parameter estimates:

\[
\hat{\gamma} = - \frac{\hat{V}_o}{\hat{\sigma}}, \quad \hat{\delta} = \frac{1}{\hat{\sigma}} \tag{5.42}
\]

where

* This term is used by Cox(35) for the transforms of type (3.41).
\[ g(\theta) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{\delta(1-\theta)} \exp\left\{-\frac{1}{2}(\hat{\gamma} + \hat{\delta}\theta)^2\right\} \]
\[ = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\theta(1-\theta)} \exp\left\{-\frac{1}{2\sigma^2} \left(2(\theta) - \hat{\gamma}\right)^2\right\} \]
\[ 0 < \theta < 1. \quad (5.43) \]

We shall subdivide the \([0,1]\) interval for \(\theta\) into 20 small intervals of size 0.05 units each, utilizing the property that

\[ \lim_{\theta \to 0} g(\theta) = 0 = \lim_{\theta \to 1} g(\theta) \]
i.e. the \(S_B\) curve has infinitely 'high contact' at either end of its range. Then, the evaluated ordinates of \(g(\theta)\) at different \(\theta\) points, would be joined by a smooth curve using our program which takes access to the graph-plotter.

The actual frequency histogram is constructed by defining 20 class intervals of size 0.05 each within the \([0,1]\) interval (i.e.,
\((0.00 - 0.05), (0.05 - 0.10), \ldots, (0.95 - 1.00))\). Then our computer program assigns each of the individual estimates \(\hat{\delta}_i\) to one of these class intervals. Next, the relative frequency histogram is obtained by scaling the actual frequencies so as to agree with the condition (5.44) below of unity area under the probability distribution curve:
\[ I = \int_0^1 g(\theta) d\theta = 1. \quad (5.44) \]

Similarly, we draw the histogram utilizing a graph-plotter routine. The outcome of the exercise is shown in Figure (5.5) below. Considering that the sample size is not very large, we see that the fit is reasonably good. We may compare the first two moments of the empirical relative frequency histogram with those of the fitted \(S_B\) curve for quit

* See Johnson (77, p. 158).
propensities. We also want to use the latter's first moment, \( E(\theta) \), as the theoretical average quit propensity. Hence, we evaluate:

\[
E(\theta) = \int_{0}^{1} \theta g(\theta) d\theta \\
= \frac{1}{\sqrt{2\pi}} \int_{0}^{1} \frac{1}{(1-\theta)} \exp \left\{ -\frac{1}{2\sigma^2} \left( \theta(\theta) - \hat{\theta}_0 \right)^2 \right\} . \tag{5.45}
\]

However, the solution of this integral cannot be expressed in a simple closed form, and we should resort to numerical integration. Beside the first moment \( E(\theta) \), we also need to evaluate the integral of the form \( E(\theta(1-\theta)) \) for the second moment:

\[
E(\theta(1-\theta)) = \int_{0}^{1} \theta(1-\theta) g(\theta) d\theta . \tag{5.46}
\]

The use of this term will be discussed shortly. Then these moments \( E(\theta) \), and \( E(\theta(1-\theta)) \) of the \( S_B \) curve will be compared with the corresponding moments of the relative frequency histogram.

\[
\hat{\theta}^* = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i \tag{5.47}
\]

and

\[
\hat{v}^* = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i (1 - \hat{\theta}_i) . \tag{5.48}
\]

As we have previously noticed there is very close agreement between \( E(\theta) \) in (5.45) and \( \hat{\theta}^* \) in (5.47). Similarly \( E(\theta(1-\theta)) \) and \( \hat{v}^* \) are

---

* Note that this last integral (5.42) is identical to that of (5.27) which we have previously developed for the sake of constructing Figure (5.2) below.

** See Appendix (A, p. 95) for the analytical expression of \( E(\theta) \).

*** Previously we have been comparing \( E(\theta) \) and the sample proportion of quits, \( \hat{\theta} \) (see equation (5.20)). But now we are comparing \( E(\theta) \) and \( \hat{\theta}^* \) as defined in equation (5.47). Notice that \( \hat{\theta} = \hat{\theta}^* \) as a consequence of maximum likelihood estimation when the model contains a constant term (see Chapter III, p. 95).
The frequency histogram of estimated individual quit probabilities and the fitted $S_b$ curve.

$N = 157$
in good agreement as shown by Table (5.2) below.

Table (5.2)

Comparison of moments of the $S_B$ curve and those of the relative frequency histogram

<table>
<thead>
<tr>
<th>$S_B$ Curve</th>
<th>Relative Frequency</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\theta) = 0.25499$</td>
<td>$\theta^* = 0.25478$</td>
<td>0.00021</td>
</tr>
<tr>
<td>$E(\theta(1-\theta)) = 0.1507$</td>
<td>$\nu^* = 0.1496$</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

However, this agreement of the two moments should not, naturally, be taken as indicative of goodness of fit for the original logistic model (5.1). Our own experience has shown us that even models of a poor fit (with very low t-ratios, poor log-likelihood ratio) provide similar agreement as shown above. Thus, the picture for the empirical frequency histogram could be very erratic and not clearly related to the corresponding $S_B$ curve, although the first two moments of the two structures are in close agreement. In the latter case the $S_B$ curve may not provide a good fit, but merely gives a simplified picture for the main trend of the irregular frequency histogram which may be useful for comparative reasons when more than one situation is described by such a relative frequency.

Now, having estimated the $S_B$ curve for the structure of quit propensities $g(\theta)$ we move to the next stage of applying a simple simulation experiment to assess the extent of the shifts in $g(\theta)$ as a result of an assumed set of policy combinations. This experiment will also be based on the above model which we have used to construct Figures
Specifically we consider the following policies (while the principle can also be applicable to describe sensitivity of quit structure in a given firm):

(i) Average travel time $\mu_{TT}$ to work is to be reduced by a given percentage.

(ii) Average housing costs $\mu_{RR}$ (Rent plus Rates) are to be increased by another percentage.

(iii) The level of compulsory education $\mu_{ED}$ is to be raised by a given percentage.

For simplicity of exposition we assume that the means are to be changed by constant shifts (e.g. $\mu_{TT} + c_1$ ... etc.) and that they result in 25% equal percentage changes. The constant shifts will guarantee that the characteristics variance, $\sigma^2$, remains constant, while the equal percentage assumption allows comparability of potential effects.

We shall deal with the following policy combinations.

(a) where only policy (i) is applicable.

(b) where policies (i) and (ii) are combined.

(c) where policies (i), (ii) and (iii) are combined.

The hypothesis is that successive introduction of these policies will potentially shift the $S_B$ curve more and more to the left, yielding successively lower and lower average quit probability. Clearly, this effect is due to the signs of the estimated coefficient parameters for travel time, $TT$, housing costs, $RR$, and education $ED$ on the behavioural quit model; (see (5.22)). Recalling that originally we have a (theoretical) average quit probability, $E(0) = 0.25499$, we can now assess the potential effects of the above policy combinations in terms of the expected decline in average quit probability. The results are shown in Table (5.3) below.
Table (5.3)

Potential effects of given policy combinations on $E(\theta) = 0.25499$

<table>
<thead>
<tr>
<th>Policy combination</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\theta)$</td>
<td>0.2365</td>
<td>0.2149</td>
<td>0.1931</td>
</tr>
</tbody>
</table>

As expected we get a declining potential average quit probability as we move from (a) to (c). The geometrical representation of this experiment is shown even clearer in Figures (5.6), (5.7) and (5.8) below. These three figures describe the potential effects of policies (a), (b) and (c) respectively, while the connected curve represents the original quit structure. The spread between this curve and the new disconnected curve represents the potential effect of the policy. As expected, this spread increases, shifting the new curve more and more to the left, as we move from policy (a) to (c), defined above. In other words, quit propensities may get more and more concentrated near zero as a result of these policies, thus restricting workers turnover. Although this conclusion is intuitively plausible, yet this experiment should not be taken too far beyond its illustrative purpose. The implications of our empirical models shall be discussed more fully in the next chapter.

As we have previously mentioned, the extent of the shift of our curve, $g(\theta)$, in response to various policy effects could be used as a measure of their effectiveness, given costs of implementation. Alternatively, the degree of potential sensitivity of the average quit propensity, $E(\theta)$, to small changes in the characteristics moments $\mu$
The effect of policy (a) on the original distribution curve.
Effect of policy combinations (a) and (b) on the original distribution curve.
Effect of policy combinations (a), (b) and (c) on the original distribution

Figure (5.8)
or $\Sigma$, will be indicative of the effectiveness of various policy measures exercised on employees' characteristics means, or their dispersion. In the next two sections we shall pay attention to the case where the screening/on-the-job policies aim at influencing the vector of characteristics means $\mu = (\mu_1, \ldots, \mu_p)$, and not their covariance matrix, $\Sigma$. At the end of the chapter we shall consider the latter problem for the measurement of the sensitivity of $E(\theta)$ to a small change in $\sigma^2$, the dispersion of the characteristics linear combination, $C = \mathbf{g}'\mathbf{x}$.

5.5 Point Elasticities of Average Quit Propensity: In terms of employees' characteristics

There are certain established formulas developed for assessing the elasticity, or degree of responsiveness of average choice probability $E(\theta)$, in terms of its explanatory variables, $\mathbf{x}$; e.g. see Dodonea\(^{(41)}\), Domenich and McFadden\(^{(46)}\), Richard and Ben-Akiva\(^{(121)}\), Westin\(^{(148)}\). In this section we shall utilize some of these results, and introduce certain numerical simplifications which would facilitate their computations. We shall be dealing with aggregate elasticities rather than disaggregate elasticities which relate to individual choice probabilities. The latter concept is defined as

$$
\eta_{ik} = \frac{\partial \theta_i}{\partial x_{ki}} \cdot \frac{\mathbf{x}_{ki}}{\theta_i}
$$

where $\theta_i$ are individual choice probabilities, given by $\theta_i = e^{\beta X_i} / (1 + e^{\beta X_i})$, as usual.

Hence disaggregate elasticities are given by
Aggregate elasticities can similarly be defined as

$$\varepsilon_k = \frac{\beta_k \theta_i (1 - \theta_i)}{\theta_i} \cdot \frac{X_{ki}}{\theta_i}$$

$$= \beta_k (1 - \theta_i) X_{ki} \quad (i = 0, \ldots, n)$$

$$(k = 0, \ldots, p) \quad (5.49)$$

Aggregate elasticities can similarly be defined as

$$\varepsilon_k = \frac{3E(\theta)}{E(\theta)} \cdot \frac{\mu_k}{\mu_k}$$

$$= \frac{E(\theta)}{2\mu_k} \cdot \frac{\mu_k}{E(\theta)} \quad (k = 0, \ldots, p) \quad (5.50)$$

which can be interpreted as percentage change in average choice probability per unit percentage change in the kth population mean. The parameter $E(\theta)$, may not necessarily be based on the distribution of individual choice probabilities, $g(\theta)$, as we have been assuming above. In the case where no distribution is assumed, aggregate elasticities are usually defined by averaging out disaggregate elasticities. For example, Richard and Ben-Akiva (121) adopt the following formula for aggregate elasticity:

$$\hat{\varepsilon}_k = \beta_k (1 - \theta \ast) \hat{\mu}_k$$

$$(5.51)$$

where

$$\theta \ast = \frac{1}{n} \sum_{i=1}^{n} \theta_i$$

as defined in (5.47)

and

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_{ki} \cdot$$

On the other hand, Domenich and McFadden (46) adopt a different formula for aggregate elasticity defined as a weighted average for disaggregate elasticities, $\varepsilon_{ik}$, with weights $\omega_i$ satisfying $\sum_{i=1}^{n} \omega_i = 1$ such that

$$\varepsilon_k = \sum_{i=1}^{n} \omega_i \varepsilon_{ik}$$

where

$$\omega_i = \frac{\theta_i}{\sum_{i=1}^{n} \theta_i} \cdot$$

* See verification in Domenich and McFadden (46, pp. 84-85).
Hence, the Domenich and McFadden formula for aggregate elasticities, is given by

\[ \epsilon_k = \frac{\hat{\beta}_k \sum_{i=1}^{n} \hat{\theta}_i (1-\hat{\theta}_i) X_{ki}}{n \sum_{i=1}^{n} \hat{\theta}_i} \quad (k = 1, \ldots, p) \]  

(5.52)

As for the alternative approach, where average choice probability is based on the probability distribution of individual choice probabilities, \( g(\theta) \), we may directly evaluate aggregate elasticities in terms of expression (5.50) above. Using the \( S_B \) curve specification for \( g(\theta) \) as we have argued in the previous section, we can evaluate expression (5.50) by utilizing the following result, due to Johnson (77, p.173):

\[ \frac{3y_r}{\theta} = -\frac{r}{\sigma} (v_r - v_{r+1}) \]  

(5.53)

where \( v_r = E(\theta^r) \quad r = 1, 2, \ldots \) is the \( r \)th moment about the origin of the \( S_B \) curve. As before \( \gamma = -V_o/\sigma \), and \( \delta = 1/\sigma \), where \( V_o \) and \( \sigma^2 \) are the mean and variance respectively of the linear combination, \( V = \beta'X \), of the explanatory variables, \( X \). It follows that

\[ \frac{\partial y}{\partial V_o} = -\frac{1}{\sigma} = -\delta . \]  

(5.54)

Hence, for \( r = 1 \), we have \( v_1 = E(\theta) \), and using results (5.53) and (5.54), we get:

\[ \frac{3E(\theta)}{3V_o} = \frac{3E(\theta)}{\gamma V_o} \cdot \frac{\gamma V_o}{\partial V_o} = (v_1 - v_2) \]  

\[ = E(\theta(1 - \theta)) . \]  

(5.55)

Then, recalling that \( V_o = \beta_0 + \beta_1 u_1 + \ldots + \beta_p u_p \) we can evaluate aggregate elasticity as

\[ \epsilon_k = \frac{\partial E(\theta)}{\partial \mu_k} \cdot \frac{\mu_k}{E(\theta)} = \frac{E(\theta)}{\partial V_o} \cdot \frac{\partial V_o}{\partial V_o} \cdot \frac{\partial \mu}{\partial V_o} \cdot \frac{\mu_k}{E(\theta)} \]  

\[ = \frac{\beta_k E(\theta(1 - \theta)) \mu_k}{E(\theta)} \quad (k = 1, \ldots, p) . \]  

(5.56)
This is the formula adopted by Westin\(^{(148)}\). Thus, in order to evaluate aggregate elasticities via (5.56) we have to compute the two moments \(E(\theta)\) and \(E(\theta(1 - \theta))\) on the basis of the \(S_B\) curve by numerical integration. However, a simpler version of formula (5.56) can be defined in terms of the corresponding two moments \(\theta^*\) and \(v^*\) of the relative frequency histogram (see equations (5.47) and (5.48)), and Table (5.2)) by using the approximations:

\[
E(\theta) = \frac{1}{n} \sum_{i=1}^{n} \theta_i
\]

\[
E(\theta(1-\theta)) = \frac{1}{n} \sum_{i=1}^{n} \theta_i(1 - \theta_i)
\]

Thus, we can estimate aggregate elasticities by the approximate formula

\[
\hat{\varepsilon}_k = \frac{\hat{\beta}_k \sum_{i=1}^{n} \hat{\theta}_i(1-\hat{\theta}_i) \hat{u}_k}{\sum_{i=1}^{n} \hat{\theta}_i}
\]

The computation of (5.59) is easier since it does not require numerical integration. Moreover, due to the good approximation of (5.57) and (5.58) (see also Table (5.2)), the results of (5.56) and (5.59) are quite close indeed. To get this impression we have used our previous estimated model to calculate aggregate elasticities using formulae (5.56) and (5.59). We also apply the Domenich and McFadden formula (5.52) for comparability. The results are presented in Table 5.4 against the appropriate explanatory variables as defined in (5.24) above: (i) change in travel time (TT), change in housing costs (RR), level of education (ED), housing tenure (HT), non-labour income (INC) and age (AGE).

Thus, as expected we find the results of the first two columns of Table (5.4) in quite good agreement compared to the last columns of
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Using the $S_B$ curve</th>
<th>Using relative frequency</th>
<th>Using Domenich and McFadden's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon_k = \frac{\hat{\beta}_k , E(\theta (1 - \theta)) \mu_k}{E(\theta)}$</td>
<td>$\varepsilon_k = \frac{\hat{\beta}_k , \Sigma \theta_i (1 - \theta_i) \mu_k}{\Sigma \theta_i}$</td>
<td>$\varepsilon_k = k \frac{\Sigma \theta_i (1 - \theta_i) X_{ki}}{\Sigma \theta_i}$</td>
</tr>
<tr>
<td>TT</td>
<td>0.2969</td>
<td>0.2951</td>
<td>0.3733</td>
</tr>
<tr>
<td>RR</td>
<td>-0.3647</td>
<td>-0.3626</td>
<td>-0.3068</td>
</tr>
<tr>
<td>ED</td>
<td>-0.3786</td>
<td>-0.3764</td>
<td>-0.4130</td>
</tr>
<tr>
<td>HT</td>
<td>-0.0469</td>
<td>-0.0466</td>
<td>-0.0414</td>
</tr>
<tr>
<td>INC</td>
<td>0.0674</td>
<td>0.0670</td>
<td>0.0864</td>
</tr>
<tr>
<td>AGE</td>
<td>-1.2239</td>
<td>-1.2166</td>
<td>-1.1004</td>
</tr>
</tbody>
</table>
Table (5.4) in quite good agreement compared to the last column of Domenich and McFadden's formula, which is not related to the distribution of individual choice probabilities.

Thus, we are going to adopt the aggregate elasticity measure which is directly based on the distribution of quit propensities, either formula (5.56), or (5.59). The use of these elasticities gives an economic sense to the estimated parameter coefficients $\hat{\beta}_k$ ($k = 1, \ldots, p$). The latter coefficients, $\hat{\beta}_k$, are not sufficient to describe the economic importance of different variables. In the next chapter we shall encounter a situation where a given $\hat{\beta}_k$ is highly significant and very small in absolute value, but its related elasticity is relatively large.

Similarly, in the current model which we are utilizing we see that the coefficient parameters for housing costs, $\hat{\beta}_{RR^t}$, and for age, $\hat{\beta}_{AGE}$ are almost equal ($= 0.06$) yet the elasticity related to age is noticeably higher (-1.224 compared to -0.365). Let us illustrate this point within the framework of our previous simulation test of the three policy combinations (a), (b), (c); see presentation in pages (191). Then, suppose that average of employees has increased by 25%, same percentage change as other means. If we combine this change with the last combination (c) we get a new combination (d). Then, as before we evaluate average quit probability, $E(8)$, when all four means (of change in travel time, change in housing cost, level of education, and age) have changed. This results in a fourth entry to Table (5.3) above, corresponding to combination (d) with computed value

$$E(8) = 0.1358.$$  

which could be compared with the adjacent number $E(8) = 0.1931$. A clearer description for this point can be provided by constructing a new $S_B$ curve which embodies this fourth combination (d), and then compare
Effect of policy combinations (a), (b), (c) and (d) on the original distribution.

Figure (5.9)
it with its predecessor, Figure (5.6) above. The outcome is given in Figure (5.8) below. As clearly seen by comparing Figures (5.6), (5.7) and (5.8) there is a drastic potential shift in the curve to the left caused by the 25% decrease in employees average age. This implies that average quit propensity is potentially more sensitive to age variation than housing costs, although coefficient parameters may be equal.

5.6 Arc Elasticity and Approximate Prediction Formulae

In the previous section we have been concerned with point elasticities. However, as it is observed in most text books of economics, point elasticities are only relevant to situations of very small changes, while in practice we may be interested to assess the potential effects of larger percentage changes in the explanatory variables. Arc elasticities are defined to meet the latter situation. In practice there are numerous formulae proposed for the computation of arc elasticities given in economics text books, although they end up being approximately the same. However, we shall adopt Samuelson's formula \(^{(124)}, p.382\), whose analogue will be written below. The basic notation which we shall use in this section has repeatedly been defined in previous sections.

Let \( \Delta E(\theta) = E^{(2)}(\theta) - E^{(1)}(\theta) \)

and \( \Delta \mu_k = \mu_k^{(2)} - \mu_k^{(1)} \) \( (5.60) \)

where \( \mu_k^{(1)}, E^{(1)}(\theta) \) are the original values,

while \( \mu_k^{(2)}, E^{(2)}(\theta) \) are respectively the new value of \( \mu_k \) after the change, and the resultant new value of \( E(\theta) \).

Then, define the simple averages of these quantities as

\[ \mu_k = \left( \frac{\mu_k^{(1)} + \mu_k^{(2)}}{2} \right) / 2 \]

and
\begin{equation}
E(\theta) = \left( E^{(1)}(\theta) + E^{(2)}(\theta) \right) / 2 \quad k = 1, \ldots, p.
\end{equation}

Then, Samuelson's formula can be expressed as

\begin{equation}
\rho_k = \frac{\Delta E(\theta)}{E(\theta)} \frac{\Delta u_k}{\mu_k}
\end{equation}

\begin{equation}
= \frac{\Delta E(\theta)}{\Delta u_k} \cdot \frac{\mu_k}{E(\theta)} \quad (k = 1, \ldots, p) \tag{5.61}
\end{equation}

The main problem for the computation of \( \rho_k \) is that, although \( \Delta u_k \) is simple to calculate, yet the resultant potential change in average quit probability, \( \Delta E(\theta) \), cannot be computed on the basis of the original \( S_B \) curve. The original \( S_B \) curve gives \( E^{(1)}(\theta) \), but \( E^{(2)}(\theta) \) is based on a new \( S_B \) curve, and has to be evaluated by numerical integration. This makes computation of arc elasticities rather laborious.

Richard and Ben-Akiva\(^{121}\), whose model has not been based on a distribution of choice probabilities, resorted to the re-estimation of the whole model by imposing the percentage change on the given explanatory variable. Similarly, Westin\(^{148}\) who adopted the \( S_B \) curve for transport model choice probabilities, remarked that

"Although (point) elasticities are useful for predictions of the effects of small changes, an alternative procedure, if the magnitude of the changes being considered is known, is to examine the transformed \( S_B \) curve directly."

Westin (148, p. 10).

The 'alternative' procedure' proposed by Westin is precisely the one which we have been adopting so far for constructing new \( S_B \) curves resulting from specific percentage changes in characteristics means; e.g. see Figures (5.5), (5.6), (5.7), (5.8). Hence, this procedure
involves

(i) the imposition of the specific change on \( \mu_k \), to get \( \mu_k^{(2)} \).

(ii) Allowance for this change in the linear combination of mean, \( V_0^{(1)} \),
to get \( V_0^{(2)} \).

(iii) Thus, we deal with a new \( S_B \) curve with parameters \( \hat{\sigma}^2 \) and \( V_0^{(2)} \).

(iv) Then, evaluate integral (5.45) by numerical methods to get \( E^{(2)}(\theta) \).

However, in this section we provide a plausible approximate method
for the computation of \( E^{(2)}(\theta) \) on the basis of the original \( S_B \) curve
that does not rely on numerical integration. The degree of approximation
is tested by computing \( E^{(2)}(\theta) \) directly on the basis of a new \( S_B \) curve
following the above procedure. When we compare the results of the latter
procedure with our own method we find a very good agreement. Our method
is based on the following mathematical result.

RESULT (5.1)

When the first moment of the \( S_B \) curve is treated as a continuous
function of the linear combination of means, \( V_0 \), it possesses a sigmoid
curve with lower and upper asymptotes at zero and one respectively, i.e.

![Figure (5.9)](image)
This result has already been verified empirically in section (5.3); see Figure (5.2). Its mathematical proof is rather simple to outline if we utilize some of our previous results. The $S_B$ function has been expressed as below

$$g(\theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\theta(1-\theta)} \exp \left\{ -\frac{1}{2\sigma^2} \left( \log \left( \frac{\theta}{1-\theta} \right) - V_o \right)^2 \right\}$$

$$0 < \theta < 1$$

and its first moment, $E(\theta)$ defined as:

$$E(\theta) = \int_0^1 \theta g(\theta) d\theta$$

$$= \frac{1}{\sigma \sqrt{2\pi}} \int_0^1 \frac{1}{(1-\theta)} \exp \left\{ -\frac{1}{2\sigma^2} \left( \log \left( \frac{\theta}{1-\theta} \right) - V_o \right)^2 \right\}$$

We have already shown that this last integral is mathematically equivalent to (see equation (5.25), (5.27)):

$$E(\theta) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \left( \frac{e^{V_o+\sigma Z}}{1+e^{V_o+\sigma Z}} \right) e^{-\frac{Z^2}{2}} dZ.$$

(5.62)

We can rewrite equation (5.62) as:

$$E(\theta) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \left( \frac{1}{1+e^{V_o+\sigma Z}} \right) e^{-\frac{Z^2}{2}} dZ.$$

(5.63)

Now, to establish the upper limit of $E(\theta)$ as $V_o$ approaches $\infty$ we note that

$$\lim_{V_o \to \infty} E(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1 \cdot e^{-\frac{Z^2}{2}} = 1$$

(5.64)

Similarly, the lower asymptote is obtained as
\[
\lim_{V_0 \to -\infty} E(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \, dz = 0.
\] (5.65)

As for the intermediate values \(-\infty < V_0 < \infty\) we utilize the result, that (see equation (5.55))

\[
\frac{\partial E(\theta)}{\partial V_0} = E(\theta(1 - \theta))
\]

where \(0 < E(\theta(1 - \theta)) \leq 1\) (5.66)

implying a positive finite slope for \(E(\theta)\) in terms of \(V_0\) for finite \(V_0\) values.

Therefore, RESULT (5.1) follows from (5.64), (5.65) and (5.66).

This RESULT is very important because it implies that we can get a fairly good linear approximation of \(E(\theta)\) in terms of \(V_0\), as in the region \((a, b)\) of Figure (5.9) above. The linearity approximation implies that

\[
\frac{\Delta E(\theta)}{\Delta \mu_k} = \text{const.} = \frac{\partial E(\theta)}{\partial \mu_k} = \beta_k E(\theta(1 - \theta))
\] (5.67)

which follows from equations (5.50) and (5.56).

Hence, if an arc elasticity has been defined with \(\dot{\mu}_k = \mu_k^{(1)}\), and \(\dot{E} = E^{(1)}(\theta)\), then there will be no considerable difference between point and arc elasticity. Thus, the real problem is not with the ratio \(\frac{\Delta E(\theta)}{\Delta \mu_k}\) rather than the computation of \(E^{(2)}(\theta)\), the mean of the transformed \(S_B\) curve. But this problem is easy to solve on the basis of approximation (5.67), since

\[
\Delta E(\theta) = \beta_k E(\theta(1 - \theta))(\dot{\mu}_k^{(2)} - \dot{\mu}_k^{(1)}),
\]
or alternatively,

$$E^{(2)}(\theta) = E^{(1)}(\theta) + \beta_k E(\theta(1 - \theta))(\mu^*_k - \mu_k)$$

(5.68)

using the definitions (5.57) above.

Therefore, when the percentage change in \(\mu_k\) is not too large, so that linearity is fairly preserved, we can directly use formula (5.68) to predict the potential \(E^{(2)}(\theta)\). The computation of this formula does not require more than \(\mu_k^*\) and the already computed moments of the original \(S_B\) curve, \(E^{(1)}(\theta)\), and \(E(\theta(1-\theta))\). However, a further simplification is given by adopting the corresponding moments of the relative frequency distribution of individual quit propensities, \(\theta^*\), and \(v^*\) (see (5.47) and (5.48) leading to

$$E^{(2)}(\theta) = \left( \frac{\sum_{i=1}^{n} \hat{\theta}_i + \beta_k \sum_{i=1}^{n} (1-\hat{\theta}_i)}{n} \right) \Delta \mu_k$$

(5.69)

This last formula is arithmetically very easy to compute as it does not require any specialized numerical methods.

It should be noticed that we shall get different curves for \(E(\theta)\) in terms of the \(\mu_k\)'s (k = 0, ..., p). The shape of any given curve depends on the sign and value of the associated parameter coefficient, \(\hat{\beta}_k\). Yet linear approximation at any given point \(\mu_k = \hat{\mu}_k\), depends both on \(\hat{\beta}_k\) and the absolute value of \(\hat{\mu}_k\) since the larger the latter, the more likely we will fall on the non-linear segment of the curve, and the smaller, the less likely (see Figure (5.9)). Moreover, the degree of linear approximation also depends on the direction of the percentage change. If \(\hat{\mu}_k\) is fairly large and positive and \(\hat{\beta}\) positive, then a given percentage decrease in \(\mu_k\) may bring it back to the fairly linear region. Hence the extent of linear approximation of \(E(\theta)\) in terms of any \(\mu_k\) at any point \(\hat{\mu}_k\) depends on the value of its point elasticity which combines
both $\hat{\beta}_k$ and $\hat{\mu}_k$ (see definition (5.53)), as well as the direction of change.

Now, it is time to test our approximate formula (5.69) against the exact procedure outlined in page 205 above. At the same stroke we test the above conjectures about the extent of linear approximation of $E(\theta)$ in terms of $\mu_k$. We shall utilize our estimated model which we have repeatedly used in our tests. This time we shall use the three variables a) Change in travel time, b) change in housing cost and c) Age in years. The MLE's of the coefficient parameters are approximately:

$$\hat{\beta}_{TT} = 1.272, \quad \hat{\beta}_{RR} = -0.060 \quad \text{and} \quad \hat{\beta}_{AGE} = 0.060$$

with point elasticities $\varepsilon_{TT} = 0.297, \quad \varepsilon_{RR} = -0.365, \quad \text{and} \quad \varepsilon_{AGE} = -1.224$ respectively.

For each of the means of these three variables we impose certain percentage changes, and evaluate the corresponding potential values of $E(\theta)$. We do this via the exact procedure based on transformed $S_B$ curves, and then via our simple approximate formula (5.69). Then these results can easily be compared as represented in Table (5.6) below. We get the interesting conclusion that our formula gives very good approximations for as large as 50% changes in the characteristics means.

The latter conclusion can easily be justified if we plot $E(\theta)$ against the various new values of the above three means resulting from both positive and negative percentage changes. The corresponding values of $E(\theta)$ are evaluated by the exact procedure based on the $S_B$ curve, and then connected by a curve. The outcome of this test is shown in Figure (5.10) below. Notice that the linearity assumption is very reasonable. Returning to our previous conjectures (see page 208) we see that the slope of $E(\theta)$ in terms of the AGE mean is noticeably
Average Quit Probability against different percentage changes of the variables AGE, TT and RR.

Figure (5.11)
high compared to that of change in housing costs, RR, due to the higher elasticity of the former. Also, see that non-linearity seems to creep in earlier at the positive end of the abcissa for AGE, then at the negative end. This confirms our conjecture that it is not only the value of point elasticity that is important, but also the direction of the change relative to the current value of the $\hat{\mu}_k$. As in the case of AGE which have got a relatively large positive mean, $\hat{\mu}_{AGE} = 34.3$, and $\hat{\beta}_{AGE} < 0$, we would have expected an increase in $\hat{\mu}_{AGE}$ to be closer to non-linearity than a decrease. To sum up, for most practical purposes it appears the linearity assumption performs satisfactorily.

**Table (5.6)**

Comparison of potential average quit propensities at various percentage changes of three characteristics means, using the exact procedure, $S_B$, and our approximate formula (5.66).

<table>
<thead>
<tr>
<th>% Change</th>
<th>$5%$</th>
<th>$10%$</th>
<th>$25%$</th>
<th>$50%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{TT}$</td>
<td>$S_B$</td>
<td>0.2585</td>
<td>0.2623</td>
<td>0.2736</td>
</tr>
<tr>
<td>Approx.</td>
<td>0.2588</td>
<td>0.2626</td>
<td>0.2743</td>
<td>0.2944</td>
</tr>
<tr>
<td>% Change</td>
<td>$-5%$</td>
<td>$-10%$</td>
<td>$-25%$</td>
<td>$-50%$</td>
</tr>
<tr>
<td>$\mu_{RR}$</td>
<td>$S_B$</td>
<td>0.2504</td>
<td>0.2458</td>
<td>0.2324</td>
</tr>
<tr>
<td>Approx.</td>
<td>0.2502</td>
<td>0.2455</td>
<td>0.2317</td>
<td>0.2086</td>
</tr>
<tr>
<td>% Change</td>
<td>$5%$</td>
<td>$10%$</td>
<td>$25%$</td>
<td>$50%$</td>
</tr>
<tr>
<td>$\mu_{AGE}$</td>
<td>$S_B$</td>
<td>0.2397</td>
<td>0.2238</td>
<td>0.1773</td>
</tr>
<tr>
<td>Approx.</td>
<td>0.2393</td>
<td>0.2249</td>
<td>0.1843</td>
<td>0.1284</td>
</tr>
</tbody>
</table>
5.7 Development of the Variance-Elasticity Formula

Up to the present point we have been considering the elasticity of $E(\theta)$ in terms of the characteristics means, assuming that their dispersion remains unchanged. This is only justifiable on the basis that policy makers influence characteristics means by constant shifts* - even though they are expressible in percentage terms. It is also a plausible simplification to the problem if the constant shifts hypothesis is reasonably maintained.

However, we may think of a different situation where, for example, the mean age of employees is not changed but its variance has been changed by recruiting relatively older persons and younger persons. The assessment of the potential effect of such a change (e.g. in the variance of age) on average quit propensity $E(\theta)$ imposes a difficult problem. The analogy between the mean-elasticity of $E(\theta)$ and the latter variance-elasticity of $E(\theta)$ is not straightforward. In the former case a change of a mean need not affect other means, but in the latter case, unless the off-diagonal covariance terms of $E$ are assumed zero, we cannot ignore the effect of a variance change over the off-diagonal elements of the variance-covariance matrix. Hence, it is not possible to isolate individual variance-elasticities for $E(\theta)$, as we have been doing for individual means. Therefore, the variance-elasticity of $E(\theta)$ assesses, not a separate variance effect, but the composite effect due to changes of other covariance terms as well. To express this problem formally, suppose we want to assess the $k^{th}$ variance-elasticity of $E(\theta)$, defined as:

* Recall that if $\text{Var}(X) = \sigma^2$ then $\text{Var}(X+c) = \sigma^2$ where $C$ is fixed independently of $X$. Similarly $\text{Cov}(X,Y) = \text{Cov}(X+c_1, Y+c_2)$ for fixed $c_1$ and $c_2$. 
\[ \phi_k = \frac{\partial E(\theta)}{\partial \sigma_{kk}} \frac{\sigma_{kk}}{E(\theta)} \quad (5.70) \]

where:

\[ \frac{\partial E(\theta)}{\partial \sigma_{kk}} = \frac{\partial E(\theta)}{\partial \sigma} \frac{\partial \sigma}{\partial \sigma_{kk}} , \]

and \( \sigma = \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j \sigma_{ij} \).

The \( \sigma_{ij} \) are the \((i,j)\) elements of the variance-covariance matrix.

Then, noting that \( \sigma_{ij} = \sigma_{ji} \) (all \( i,j \)) we find:

\[ \frac{\partial \sigma}{\partial \sigma_{kk}} = 2 \beta_k \sum_{i=1}^{p} \beta_i \frac{\partial \sigma_{ki}}{\partial \sigma_{kk}} \quad (k = 1, \ldots, p) \]

\[ = 0 \quad \text{if and only if} \quad \sigma_{ki} = 0 . \quad (5.71) \]

However, if \( \sigma_{ik} \neq 0 \) \((k \neq i = 0, \ldots, p)\), then there is no simple way to evaluate derivative \((5.71)\), which in general will not vanish.

Hence, in the case where covariances are non-zero we may only evaluate the composite variance-elasticity, (or \( \sigma \)-elasticity):

\[ \phi = \frac{\partial E(\theta)}{\partial \sigma} \frac{\sigma}{E(\theta)} \quad . \quad (5.72) \]

The evaluation of this formula on the basis of the \( S_B \) curve may be directly based on the previous result \((5.53)\), that

\[ \frac{\partial E(\theta)}{\partial Y} = -\frac{1}{\delta} E(\theta(1 - \theta)) \]

* Note that \( \sigma_{XY} = E(XY) - E(X) \cdot E(Y) \) and \( \sigma_{XX} = E(X^2) - (E(X))^2 \). Suppose \( \sigma_{XX} > 0 \) and \( \sigma_{XY} \neq 0 \). Now, if \( E(X) = K \), but \( X \) becomes more concentrated, then \( E(X^2) \) becomes smaller, and \( E(XY) \) will also be affected. In the limiting case where \( \sigma_{XX} \to 0 \) (i.e. \( E(X^2) \to K^2 \), or \( X \to K \)), then \( E(X \cdot Y) \to KE(Y) \) and \( \sigma_{XY} \to 0 \). Hence, \( \sigma_{XY} \to 0 \) if \( \sigma_X \to 0 \).
where as before \( \gamma = -\frac{V_0}{\sigma} \), and \( \delta = \frac{1}{\sigma} \). We may also recall that 
\( 0 < E(\theta) < 1 \), and \( 0 < E(\theta(1-\theta)) < \infty \). Then

\[
\frac{\partial E(\theta)}{\partial \sigma} = \frac{\partial E(\theta)}{\partial \gamma} \cdot \frac{\partial \gamma}{\partial \sigma} = -\frac{1}{\delta} E(\theta(1-\theta)) \frac{V_0}{\sigma^2} = -\frac{E(\theta(1-\theta))V_0}{\sigma}
\]

Hence*, using definition (5.72) we get the \( \sigma \)-elasticity:

\[
\phi = -\frac{E(\theta(1-\theta))V_0}{E(\theta)}
\]

(5.73)

Notice that \( \phi \) satisfies conditions (5.30) outlined in section (5.3) that it will always have the opposite sign of \( V_0 \):

\[
\phi \begin{cases} 
< 0 & \text{for } V_0 > 0 \\
= 0 & \text{for } V_0 = 0 \\
> 0 & \text{for } V_0 < 0 
\end{cases}
\]

(5.74)

We have seen that these conditions are consistent with the implications of Figure (5.2) above, and that they are obeyed by Talvitie's formula (5.10). However, Westin have adopted the \( S_B \) curve and definition (5.72), as we have done, but the formula he derived and proposed for the \( \sigma \)-elasticity of \( E(\theta) \) is given by

* We have been utilizing a result due to Johnson, formula (5.53) above, which makes the derivation of formulae (5.56) and (5.73) very easy. But we are not sure how Westin proves his formulae, especially (5.75), which is different from ours.
\[
\phi^* = \frac{E\{\delta(1-\delta)(\log \frac{\theta}{1-\theta} - V_0)\}}{E(\theta)}.
\] (5.75)

He does not outline the derivation of (5.75), but it can be shown that \( \phi^* \) is not consistent with the empirical implications of conditions (5.30), or those of Talvitie's formula (5.10). For example, we shall show below that \( \phi^* \) evaluated at \( V_0 \) is not equal to zero, thus violating the property that at \( V_0 = 0 \) the variance is a redundant parameter for \( E(\theta) \). Let us define:

\[
\xi(\theta) = \log \left( \frac{\theta}{1-\theta} \right), \quad \text{and} \quad v(\theta) = \theta(1-\theta).
\]

Then the numerator of Westin's \( \phi^* \) can be expressed

\[
E\left[ v(\theta)\xi(\theta) \right] = \frac{1}{\sigma \sqrt{2\pi}} \int_0^1 \xi(\theta) e^{-\frac{1}{2\sigma^2}(\xi(\theta))^2} d\theta \quad \text{at} \quad V_0 = 0.
\] (5.76)

This integral may not be easy to evaluate but we shall verify that

\[
E[\xi(\theta)] = \frac{1}{\sigma \sqrt{2\pi}} \int_0^1 \frac{\xi(\theta)}{\theta(1-\theta)} e^{-\frac{1}{2\sigma^2}(\xi(\theta))^2} d\theta = 0.
\] (5.77)

Let \( V = \xi(\theta) \), so that \( \frac{dV}{d\theta} = \frac{1}{\theta(1-\theta)} \)

and \( d\theta = \theta(1-\theta)dV \). If we complete the transformation of integral (5.77) we will get:

\[
E(V) = \frac{1}{\sigma \sqrt{2\pi}} \int_0^\infty V e^{-\frac{1}{2\sigma^2} V^2} dV = 0
\] (5.78)

The value of this integral represents the mathematical mean of the random variable \( V \) which is normally distributed with mean zero and
variance $\sigma^2$. Hence, result (5.77) follows. But, as long as $V(\theta)$ is a random variable dependent on $\theta$:

$$E\left\{v(\theta) \lambda(\theta)\right\} \neq E(\lambda(\theta)) = 0.$$ 

Therefore, Westin's formula does not obey the property that at $V_o = 0$, average choice probability is independent of $\sigma^2$, as it is invariably equal to $\frac{1}{2}$ (see equation (5.19)).

More generally at $V_o \neq 0$ we can show that $E(\lambda(\theta)) = V_o$ using the integral transformation which leads to (5.78). However

$$E\{v(\theta)\left[\lambda(\theta) - V_o\right]\} \neq E\left[\lambda(\theta) - V_o\right] = 0.$$ 

Clearly, there is no systematic relationship between Westin's $\phi^*$ and the sign of $V_o$ as should be according to conditions (5.30), or the implications of Figure (5.2a) above. The intuitive appeal of these conditions will be illustrated by a simple example at the end of this section.

Hence, we propose our alternative formula (5.73) for the evaluation of the variance-elasticity of average choice probability, in place of formula (5.75).

It is possible to express the limiting properties of our $\sigma$-elasticity formula rather formally by exploiting RESULT (5.1), recalling that $E(\theta(1 - \theta)) = \frac{\partial E(\theta)}{\partial V_o}$. These properties are expressed by RESULT (5.2) below:

RESULT (5.2): The average quit probability, $E(\theta)$, becomes more and more '\(\sigma\)-inelastic' as $V_o$ approaches zero or infinity in absolute terms, i.e. given $\sigma$-elasticity:

$$\phi = \frac{E(\theta(1-\theta))V_o}{E(\theta)}.$$ 

Then,
This result follows directly from the fact that

\[
\lim_{|V_o| \to \infty} |\phi| = 0 \quad \text{and} \quad \lim_{|V_o| \to 0} |\phi| = 0.
\]

Hence, as implied previously by Figure (5.2a) above, the variance becomes less and less important to \( E(\theta) \) for too small values of \( V_o \) (i.e. \( E(\theta) \neq \frac{1}{2} \)) or too large \( V_o \) values (i.e. \( E(\theta) = 0 \) or \( E(\theta) = 1 \)).

Now, returning to our policy implications we conclude that the effective use of \( \sigma \) as a policy control parameter (while keeping the linear combination of mean characteristics constant at \( V_o \)) depends on the sign and magnitude of \( V_o \). If average quit probability is approximately equal to 50% (i.e. \( V_o = 0 \)), then a policy which influences employees' dispersion alone via \( \sigma \) will not be sufficiently effective in reducing average quit probability. Similarly if average quit probability is already too high (or too low), then the use of \( \sigma \) as policy control parameter may not be effective. In these cases, it is more advisable to influence the mean characteristics via \( V_o \) for policy purposes. Of course special consideration should be taken for the signs and magnitude of separate means-elasticities. However, the use of \( \sigma \) as a policy control parameter when \( V_o \) is neither too small nor too large, depends on the sign of \( V_o \). If \( V_o \) is negative (i.e. \( E(\theta) < \frac{1}{2} \)) it may be more advisable to reduce the variance of employees' characteristics. However, if \( V_o \) is positive (i.e. \( E(\theta) > \frac{1}{2} \)) it may be more advisable to raise the variance of employees' characteristics, in order to reduce average quit probability.

However, we may notice that there is an element of crudeness in the
use of the $\sigma$-elasticity since it conceals the interaction of different variables through the non-zero covariance terms. Hence as a policy control parameter $\sigma$ may not be readily identifiable in terms of the employees' characteristics vector $\mathbf{X}$. We may need to go over rather lengthy and difficult computations in order to assess the ultimate effect of a 5% reduction in $\text{Var}(X_i)$ over $\sigma$, and we may have to impose certain assumptions about the covariance terms. However, despite this limitation the sign and magnitude of the $\sigma$-elasticity may serve an analytical purpose as a guide-line to the direction and magnitude of potential effects of changes in the dispersion structure of employees' characteristics. Such dispersion changes can then be resisted, encouraged or overlooked, depending on their potential aggregate effects as alluded by the sign and magnitude of the $\sigma$-elasticity. It is also possible that a desirable potential effect produced by a percentage increase in a given mean characteristic, $\mu_k$, be reduced by an undesirable effect of an unplanned increase in $\sigma$. In this case it may be more advisable to control $\sigma$ by ensuring a fairly 'constant shift' in $\mu_k$ if the contribution of $\mu_k$ to average quit propensity, $E(\theta)$, is negative. However, if its contribution is positive, making for $E(\theta) > \frac{1}{4}$, we may allow $\sigma$ to increase. The intuitive appeal of this point follows from a simplified example. Suppose the only relevant quantity is average change in travel time, $TT$. Our data base for males shows that $TT$ is positively correlated with $E(\theta)$. Hence, if its mean $\mu_{TT} < 0$ we get $E(\theta) < \frac{1}{4}$, while if $\mu_{TT} > 0$, then $E(\theta) > \frac{1}{4}$. In the former case the lower $\sigma_{TT}$, the lower we expect $E(\theta)$ to be, because when $\sigma_{TT}$ is very small or zero this implies that almost all individual employees have moved closer to their places of work. But if $\sigma_{TT}$ increases this implies that others have been moved further
away, and hence $E(\theta)$ may be pushed up. In this case it may be more advisable to keep $\sigma$ low, or at least to control it. However, if $\mu_{TT} > 0$ a decrease in $\sigma$ may have the opposite effect, because when $\sigma$ is very small it implies that almost all individuals have moved further away from their place of work, and hence given the effect of $TT$ this may push up $E(\theta)$. But, if $\sigma_{TT}$ is increased when $\mu_{TT} > 0$, this implies that some individuals have been moved closer to their places of work, thus making the expectation of a lower average quit propensity. In this case it may be more advisable to allow $\sigma$ to increase.

The idea of $\sigma$-elasticity leads us to pinpoint one limitation of the linear probability model before concluding the chapter.

5.8 One Further Disadvantage of the Linear Probability Model

We have considered the computational disadvantages of the linear probability in the third chapter, as the problems of predictive bias due to the violation of the $[0,1]$ restriction of probability, and the problem of heteroscedasticity. Now after we have utilized the logistic model through the $S_B$ distribution of individual choice probabilities, we become aware of the incapability of the linear probability model to answer the same questions.

In the first place we know that the method of least squares imposes the restriction that

$$
\hat{\Theta} = f(\hat{V}_0)
= \beta_0 + \beta_1 \hat{Y}_1 + \ldots + \beta_p \hat{Y}_p
$$

where

$$
\hat{\Theta} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \quad \hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_{ik}, \quad \hat{\beta}_k \text{ is LSE of } \beta_k
$$

(5.79)
and when the original model is written as

\[ E(Y_i) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi} = \theta_i \]  \hspace{1cm} (5.80)

and \( \text{Var}(Y_i) = \theta_i(1 - \theta_i) \) \( 0 \leq \theta_i \leq 1 \).

However, we have shown in section (5.3) that property (5.79) is not
obeyed by non-linear probability models, e.g. the logit. It is only
obeyed when the sample proportion \( \hat{\theta} \) is equal to \( \frac{1}{2} \), or when the
\( \sigma = \beta' \Sigma \beta = 0 \), where \( \Sigma \) is the covariance matrix of the \( X_i \)'s.

Thus, the disaggregate linear probability model cannot be used to
derive aggregate predictions since average probability based on it
is always \( \sigma \)-inelastic. Formally: Let \( X' \) be multivariate normally
distributed with mean \( \mu \) and covariance matrix, \( \Sigma \). Then individual
probabilities are

\[ \theta_i = \beta X_i \quad 0 \leq \theta_i \leq 1 \]

and average probability is simply:

\[ E(\theta) = \beta' \mu \quad \text{independent of } \Sigma. \]

Hence, if we continue to treat \( \beta \) as a vector of known constant,
and that variations in \( \sigma = \beta' \Sigma \beta \) come about from changes in \( \Sigma \), then
\[ \frac{\partial E(\theta)}{\partial \sigma} = 0 \quad \text{for all } \mu. \]

Nerlove and Press (106) have observed another computational disad-
vantage of the linear probability model as it does not signal information
about certain peculiarities in the data. This point is considered in
the next chapter.
5.9 **Summary**

This chapter centres around the principles and techniques of utilizing a disaggregate probability model of quits to arrive at aggregate predictions related to a given population of heterogeneous employees. We have agreed with Silcock\textsuperscript{(128)} in his turnover model that heterogeneity of employees should be expressed in terms of a non-degenerate distribution of their job-leaving propensities. However, we have argued that the parameters of such a distribution should be directly related to the employees' mean characteristics and the associated variance-covariance matrix. Thus these first and second moments of the explanatory variables which (stochastically) determine leaving behaviour can serve as 'policy-control parameters' - in the sense used by Clowes\textsuperscript{(30)}. Under the multivariate normality assumption for these explanatory variables, the expected proportion of employees leaving their jobs, \( \text{E}(\theta) \), is completely determined by two parameters which are simple functions of these variables' means and their variance-covariance structure respectively. Thus depending on the potential sensitivity of \( \text{E}(\theta) \) to finite changes in the variables means, the latter could be influenced through various screening/on-the-job policies in order to reduce turnover. The measurement of this sensitivity can be done in terms of aggregate point elasticities for small changes in employees' mean characteristics (i.e. means-elasticity), or for small changes in the composite measure of their variance-covariance structure (i.e. \( \sigma \)-elasticity).

The approach which we have adopted is to derive these measurements by specifying a Johnson's \( S_B \) distribution for the employees' leaving propensities. This distribution has arisen as a natural consequence of two basic premises: the adoption of the logistic model and the
multivariate Normality assumption of the explanatory variables. As required the $S_B$ curve is reasonably flexible and its two parameters are simple functions of the means and covariance structure of these explanatory variables. Thus, the $S_B$ curve serves for us two geometrical descriptive purposes: to represent the underlying potential leaving structure of the heterogeneous employees, and the sensitivity of this structure to changes in the explanatory variables — due to policies or otherwise.

In agreement with Westin\(^{(148)}\) we have derived the mean-elasticity formula of $E(\theta)$ on the basis of this $S_B$ curve. This measure gives an economic meaning to the parameters of the logistic model. However, when changes in the characteristics means are reasonably large we require the computation of arc elasticities to assess the sensitivity of $E(\theta)$. The computational procedures adopted by Richard and Ben-Akiva\(^{(121)}\), and Westin\(^{(148)}\) are rather complex and costly. Alternatively, we have derived a computational formula for arc elasticity which is relatively very easy to calculate and gives good approximation to the exact recipe directly based on the $S_B$ curve. This goes as far as mean-elasticities are concerned.

On the other hand, the $\sigma$-elasticity measures of the sensitivity of $E(\theta)$ to small changes in the composite measure, $\sigma$, of the dispersion structure of explanatory variables — when their means are given. We have explored the analytical implications of the $\sigma$-elasticity and the conditions when it is not relevant. These considerations are based on our formula which shows that the $\sigma$-elasticity should be positive when $E(\theta)$ is (fixed) below half, negative when $E(\theta)$ is above half and irrelevant when $E(\theta)$ is exactly equal to one half. Accordingly, we have found that Westin's\(^{(148)}\) formula for $\sigma$-elasticity, which he similarly based on the $S_B$ curve, is mathematically
incorrect as it does not satisfy the above criteria. We have worked out the limiting properties of this measure and given a simple example about its intuitive appeal.

We have concluded the chapter by making the observation that the linear probability model is invariably \(\sigma\)-elastic, which is a serious limitation of its aggregate predictive use.
CHAPTER 6

EMPIRICAL MODELS FOR MALES AND FEMALES

6.1 Introduction

In this chapter we produce and discuss our empirical results concerning the job-quit probabilities of individuals who have been rehoused by the City Council of Glasgow. Our data base consists of 184 male heads of households and 95 married females. These individuals have all been employed before the move into the council housing. However, the rehousing process, whether it be effectively intended (i.e. an offer is received) or has actually occurred, may provoke individuals to revise their past job attachments depending on a certain set of explanatory variables; we have, however, explicitly allowed for this here. It should also be made clear that an element of individual's heterogeneity exists as regards the length of period after the move when quits have been observed. Yet as this data is not available we cannot control for this variation here. We shall define and briefly discuss the relevant set of explanatory variables for the job-quit probability model in the next section (6.2), with special consideration to our data potentials. A more detailed discussion of job leaving econometric models has been given in the second chapter, where we have distinguished between micro-level and macro-level models. Although our disaggregate probability model falls at the macro-level, yet we have argued in the fifth chapter that it may also be applied at the micro-level. This idea falls in line with the remark of McFad

and Reed (98, p. 24) that behavioural (disaggregate) quantal choice models should be regarded as complementary to aggregate models rather than regarding them as perfect substitutes. In the last chapter we have proposed the use of Johnson's $S_B$ curve to describe the distribution of an employee's expected quit probability and outlined its policy implications. We have shown how aggregate predictions can be
made on the basis of a disaggregate model and generalized its use for the micro-level context. These concepts and techniques will be utilized in this chapter. As regards the basic estimation techniques for the logistic model, they have been discussed in the third chapter.

This chapter consists of three main sections; in the section below (6.2), we discuss the explanatory variables and their a priori expected effects on the quit model. In section 6.3 we review the empirical models related to the males. This section also contains a sub-section on the geometrical representation of the effect of the hours variable on the distribution of quit propensities. In section 6.4 we apply the same treatment for females. At the end of the chapter there is a short summary for the main findings.

6.2 Definition of Explanatory Variables

The set of variables which shall be considered here include:

(i) Move-related variables, i.e. those describing the effects of the move into council housing on the measurable aspects of individuals' circumstances.

(ii) Personal, family and employment characteristics of the re-located individuals.

We first consider the move-status variables starting with

(i) Change in Travel Time.

We have seen in the first chapter that previous experience asserts the interaction of residential location choice and choice of work place specially for lower income groups; e.g. Daniel\(^{(38)}\), Orr\(^{(111)}\), Oi\(^{(110)}\), Lowry\(^{(91)}\). Hence, any adverse change in the distance, or travel time to work may have a disequilibrating effect on employees as regards their
current job attachments, _ceteris paribus_. In other words we expect an increased travel time to have a positive coefficient in the probability model of job quits. On the basis of our data we can calculate the change in travel time, \( TT \), as

\[
TT = (\text{Travel time from old house to old job}) - (\text{Travel time from new house to old job})
\]

Hence, a higher value of \( TT \) makes an employee more likely to leave his job.

Of course, there are different ways of expressing change in travel time but the _difference_ measure (6.1) has the merit of being most direct and intuitive. For example, we can define the _ratio_ measure

\[
TT^* = \frac{TT}{(\text{Travel time from new house to old job})}
\]

\[
TT^* = \frac{\text{Travel time from old house to old job}}{\text{Travel time from new house to old job}} - 1 .
\]

It is interesting that the value of \( TT \) (or \( TT^* \)) could be calculated for all individuals whether they are "actual" or 'intending' movers. Hence we are able to test the effect of \( TT \) (or \( TT^* \)) as a simple variable. Alternatively, a multiplicative interaction variable, i.e.

\[
\bar{TT} = TT \times M \quad \text{or} \quad \bar{TT} = TT^* \times M
\]

(6.3)
can be defined, where \( M \) is the 0,1 binary variable for actual versus intending movers. In other words, the variables (6.3) give change of travel time only for actual movers and make it uniformly zero for the others. The way Engelman (49) has allowed for this variable has been to dichotomize the multiplicative term, \( \bar{TT} \), so that it equals one for actual movers who moved closer to their jobs (i.e. \( \bar{TT} < 0 \)) and zero otherwise. He obtained a significant negative coefficient for this dichotomous variable when applied to males but the effect for females
was counter-intuitive (i.e. positive) but with a very low t-value (see Chapter I, Table (1.7)). The significant coefficient for the dichotomous variable is an interesting result; yet this dichotomization does not make the full use of the data variability. It effectively ignores otherwise accessible and useful statistical variability of time unit changes in the travel time to work. Thus, we may adopt either one of the above defined continuous variables.

However, our earlier experimentation with the data has shown us the models with the simple difference measure (6.1) always provide better fits than those with the ratio measure or the other measures (6.3) above. This point can be shown by example in the next sections.

(ii) Change in Housing Costs

Job-leaving usually involves disruption to the continuous flow of earnings, especially if a new job is not immediately offered or received. This risk will be more appreciable the greater the financial obligations that have to be met from a limited household budget. On the other hand, if these financial obligations are released, the risk will be lowered. On this ground, a rise in housing costs may make it less tolerable for an employee to run the risk of leaving his current job, than a fall in housing costs.

In our data the housing cost variable is measurable in money value as the difference between the pre-move and post-move housing rent and rates. However, unlike the travel time variable, TT, the data on changes in rent and rates, RR, relates only to the actual movers. This variable is simply defined as:

\[ RR = (\text{rent and rates of new house} - \text{rent and rates of old house}) \]  

(6.4)
To allow for this variable Engelman\(^{(49)}\) has constructed an 0,1 binary variable which equals one if RR is bigger than the sample median value, and zero otherwise; thus lumping up otherwise useful statistical variability by this dichotomy. Yet, we shall test the effect of RR directly via definition (6.4) above.

(iii) The Actual Move

It may be interesting to assess the partial effect of the actual move per se on the propensity to change the past job attachment. Hence, the 0,1 binary mover variable M used in definitions (6.3) can also be included as a possible explanatory variable. However, the expected sign for the coefficient of the move variable, M, is not determinate a priori. The hypothesis of a positive coefficient for M is based on the belief that the actual move into council housing tends to disturb the pre-move job attachments. This hypothesis can be supported on the grounds that there are certain unobserved components concomitant to the move, apart from the above-mentioned variables, RR and TT, that make the actual mover more likely to leave his previous job than the intending mover. Engelman\(^{(49)}\) obtained a significant negative coefficient for the 'non-move' variable of males, but for females the t-value was too small (see Chapter I, Table (1.7)).

(iv) New Employment Opportunities

One of the basic assumptions in job search theory is that the greater the probability of an employee receiving a 'better' offer, the more he is likely to leave his current job, and the smaller this probability, the less likely. Thus, depending on the spatial distribution of his employment opportunities, the relocated individual may decide to move to
a job in the new locality which may be better than the old job, using certain criteria, e.g. nearness to new house, better wages, better working conditions ... etc. If employment opportunities are randomly distributed in space, then the larger the number of employers in an area, the greater the probability that a random employee finds his appropriate offer. However if the spatial random distribution assumption is invalid for the different employment opportunities, then their crude number, even if large, may not be of interest. In other words, if all or most employers relate to a specialized industry (e.g. baking) in that area, then it is very likely to experience unsuccessful matching of workers and jobs for heterogeneous seeking workers. In addition, the information factor should not be ignored. If the degree of information about job opportunities varies considerably amongst workers, then the effect of the 'crude number' may again be indeterminate. However, if the information level does not vary a lot, while the randomness of the opportunities distribution is approximately maintained, then with heterogeneous moving workers we expect that the larger the number of employment opportunities in the new area, the higher will be the quit rate. In other words, under these conditions the sign of the coefficient parameter of EMP should be positive.

In our data base this variable is defined as the number of employment opportunities within a radius of 30 minutes of the new house for males, and a radius of 20 minutes for females. This data has been gathered with the help of the Department of Employment.

Hence, we can test the effect of this variable, EMP, on the basis of our model. Engelman (49) has also included this variable as a continuous variable but he obtained the opposite sign (negative) which was very insignificant; see Chapter I, Table (1.7).

Let us now move to the other class of variables which relate to personal, family and employment characteristics. We start with the
employment aspect: job-tenure of the pre-move job of an individual.

(v) **Job-tenure**

The importance of this variable has been stressed with special emphasis on what we have called micro-level models (see Chapter II, section (2.3)). The general hypothesis is that the longer the length of service (or tenure) with a given employer of a randomly selected worker, the less likely that the latter will quit. In particular we have seen Silcock's\(^{(128)}\) analytical verification of this point on the basis of a mixed exponential distribution for the completed length of service (see Chapter II, equation (2.4)). The same point has been emphasized by other authors who adopted different distribution models for the completed length of service, e.g. Lane and Andrew\(^{(86)}\), and Bartholomew\(^{(13,14)}\).

The **organizational commitment** model introduced by Herbert\(^{(71)}\), and modified by Clowes\(^{(30)}\) uses job-tenure in a different sense. For example, Clowes' modification distinguishes between two leaving rates: \(k_1\) for short-service employees (or new recruits) and \(k_2\) for long-service employees (or committed employees). Clowes\(^{(30)}\) computed the ratios \(k_1/k_2\) for some sixteen U.S. manufacturing industries and got ratios ranging between 3.21 and 22.95, which is obvious evidence for the significant effect of the degree of organizational commitment.

The use of this variable in what we have called macro-level models gives similar evidence for the negative effect of job-tenure on the intention or the decision to quit; see for example Stoikov and Raimon\(^{(134)}\), and Viscusi\(^{(144)}\), Chapter II, section (2.4). Similar evidence has been obtained by Mackay et al.\(^{(92)}\) for the effect of job-tenure on quit rates in Glasgow (Chapter I). The fact that length of service itself can be explained in terms of another set of explanatory variables (e.g. sex, training, etc.) has been remarked by Bartholomew\(^{(14,15)}\), while
effectively specified in a linear model and explained as a dependent variable by Richardson et al. (122).

However, the breakage of service length, which is often measured in years, into quarterly and monthly units, revealed to Spratling (132) that leaving rates do not always fall with increasing length of service. For example he has shown that the resignation data of bus conductors and porters, appointed by London Transport, rose markedly within the first three months and then it started to fall off. Hence, the estimated effect of tenure on quits may also depend on the unit of measurement (e.g. years, quarters, months, weeks, days per week) although the general effect will be of a monotonically decreasing propensity to leave. The fact that weekly hours per week are seldom used as a measure of organizational commitment in the same way as job-tenure is used, is attributable to the presence of institutional restrictions on work hours - any given employee is not free to make continuous marginal adjustments on his hours of work unless he is self-employed. Nevertheless, it can be argued that longer hours per week with a given employer implies stronger 'organizational commitment' than do shorter hours. The problem of institutional restriction could be released by adopting the following proxy for hours worked per week, HRS:

\[
HRS = 1 \quad \text{if} \quad \text{hours per week} \geq 30 \\
HRS = 0 \quad \text{if} \quad \text{hours per week} < 30.
\]

This convention is used in the annual reports of the General Household Survey to classify female workers into full-time (HRS = 1) and part-time (HRS = 0).

The explanatory capacity for this variable, HRS, should be remarkably strong, as judged by the simple proportions test below. Let Prob(')
denote simple proportion and let:

\[ Y = 1 \text{ if the employee is observed to have left at that point of time} \]
\[ = 0 \text{ otherwise.} \]

Then, we have got the following numbers for male heads of households and female spouses. As for the 175 male heads (deleting those who have been dismissed by the pre-move employer) we find that:

\[
\begin{align*}
\text{Prob}(Y = 1/Male, \text{HRS} = 1) &= 0.254 \\
\text{Prob}(Y = 1/Male, \text{HRS} = 0) &= 1.00
\end{align*}
\]

which could be checked from the table below.

\[
\begin{array}{c|cc}
\text{Y} & 0 & 1 \\
\hline
0 & 0 & 117 \\
1 & 18 & 40 \\
\end{array}
\]

Although we expect result (6.6) to be close to but not exactly equal to unity, it has happened that we have got the extreme result due to the limited sub-sample of males with HRS = 0. (As we shall shortly see this will cause an estimation problem.) The situation for the 95 pre-move employed females is quite similar:

\[
\begin{align*}
\text{Prob}(Y = 1/Female, \text{HRS} = 1) &= 0.16 \\
\text{Prob}(Y = 1/Female, \text{HRS} = 0) &= 0.76
\end{align*}
\]

as based on the table below:
It is particularly interesting to note that all of the male and female quits who worked at least 30 hours/week have got re-employment after the move (40, and 7 respectively). However, none of the 18 male quits who worked less than 30 hours/week - and only 3 out of the 31 female quits - got re-employment. This means that the 'less committed' employees (HRS = 0) in both sexes have had more difficulty in getting re-employed after the move than the 'more committed'. We conclude that the hours variable, HRS - which has been ignored by Engelman (49) - should play an important role in the distribution of quit propensities of both males and females.

(vi) Age

The tendency of leaving rates to decline with age can be seen more clearly by Table (6.1) below of the job-mobility age structure in Great Britain. This table has been compiled from the annual series of the General Household Survey, 1972-1976. It can be seen by visual inspection that mobility, as measured by the proportion of job changes within a 12-month period declines rather systematically with higher age groups. However, it tends to rise after the age of 65 (the retirement age for males) as change of job is very likely due to retirement. Mackay et al. (92) have also observed that age has a negative effect on quitting in the Glasgow local labour market (Chapter I, section (1.3)).

Some authors have noted that the observed negative effect of age on quit probability can be attributed to the negative effect of job-tenure as measured in years (Silcock (128), Burdett (29)). However, both Stoikov and Raimon (134) and Viscusi (144) have introduced the two variables
and obtained significant coefficients:

"quitting bears strong negative relationship to age and tenure"

Viscusi (144, p. 51).

Engelman^9^ did not use age in years but adopted a 0,1 dichotomous variable which is equal to one if the age of the male employee is greater than 45 years, and zero otherwise. He obtained a significantly negative coefficient for males, but not for females for whom he constructed two dichotomous variables of age (see Chapter I, Table (1.7)). However, since our data base contains age in years, we are in a position to utilize it instead of arbitrary grouping.

The negative effect of age on the leaving probability could be explained in terms of human capital theory in the same way as of tenure (see Chapter II, section (2.4)). Generally, the older the employee the more specialized and 'organizationally committed' he would be as he must have developed seniority, pension rights and other ties within the firm - and the less attractive the alternative offers he may contemplate.

(vii) **Skill**

In the light of our brief discussion of human capital theory and internal labour market structures (see Chapter II, section (2.4)), it appears that the partial effect of skill when measured as a single quantity is rather indeterminate a priori. If skill is acquired by 'general training' that may increase an employee's productivity at all potential firms by the same quantity, then its partial effect on the leaving propensity can be positive. While, if skill is gained by a firm's 'specific training' which may only increase the employee's productivity for that particular firm, then its partial effect can be negative. We have seen how Parsons^12^
## TABLE (6.1)

Job-Mobility by Age Classes, U.K.

<table>
<thead>
<tr>
<th>Age Classes</th>
<th>Males</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16-17</td>
<td>0.38</td>
<td>0.32</td>
<td>0.50</td>
<td>0.32</td>
<td>0.34</td>
<td>0.38</td>
<td>0.44</td>
<td>0.41</td>
<td>0.42</td>
<td>0.30</td>
</tr>
<tr>
<td>18-24</td>
<td>0.39</td>
<td>0.33</td>
<td>0.50</td>
<td>0.35</td>
<td>0.27</td>
<td>0.35</td>
<td>0.37</td>
<td>0.41</td>
<td>0.38</td>
<td>0.23</td>
</tr>
<tr>
<td>25-34</td>
<td>0.23</td>
<td>0.32</td>
<td>0.30</td>
<td>0.22</td>
<td>0.15</td>
<td>0.29</td>
<td>0.26</td>
<td>0.26</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>35-44</td>
<td>0.14</td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
<td>0.09</td>
<td>0.15</td>
<td>0.19</td>
<td>0.15</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>45-54</td>
<td>0.10</td>
<td>0.12</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>55-64</td>
<td>0.03</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
<td>0.05</td>
<td>0.06</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>65+</td>
<td>n.a.</td>
<td>0.13</td>
<td>0.07</td>
<td>0.10</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

has demonstrated this point by adopting two proxies: one for "general" and another for 'specific' training, where he obtained positive and negative coefficients respectively. As for Viscusi\(^{(144)}\) he has defined skill as a dichotomous variable which is equal to one if the employee attends a training program to improve his skill, and zero otherwise. He obtained positive coefficients for this proxy in his two "strong" quit intention models but with very poor t-values (i.e. 0.15 and 0.18 respectively), although his two "week" intention models provided better t-values (i.e. 1.43 and 1.46 respectively). In this case the opposite effects of the two types of training (general, and specific) have been obscured in a single variable, though it seems that the effect of general training is slightly predominant. Alternatively, Engelman's\(^{(49)}\) proxy for skill is another 0,1 dichotomous variable which equals one if the social economic group (SEG) of the employee is described by either of the codes 2, 5, 8 or 9, and equal zero for the codes 6, 7, 10 or 11. The definitions of this coding system of the SEG of employees can be found out from any one of the General Household Survey series, 1971-1977. Engelman\(^{(49)}\) obtained negative coefficients for the effect of this variable (as shown in Chapter I, Table (1.1)) with very poor t-values for females (i.e. 0.55, 0.07) and better t-values for males (i.e. 1.55, 1.33). Similarly the effect of this SEG dichotomy does not allow for the different types of training.

We are not in a better position to represent the effect of skill in a way compatible to the human capital theory. Yet, we have adopted two different proxies of skill which our earlier data experimentation has shown as reasonable: i.e.

(i) whether or not the employee received special training for his usual job before the move. Of course this is again representable by a 0,1 dichotomy, denoted as SPT.

(ii) The level of education can be used as another proxy for skill.
"Structureless" labour markets are associated with lower skill workers - who are usually of lower education. Such markets have been characterized by Kerr (81) as accounting for the greater mobility in the labour markets. However among the same class of educated people the nature of human capital would again account for the possible indeterminacy of the education effect on job-mobility. This variable is represented here as a 0,1 dichotomy (for lack of more refined data) which is equal to one if the level of education is not below secondary school level and zero otherwise, denoted ED.

(viii) Change in Family Earnings

We have already discussed the relevance of the economic decision of one household member on the other members. In particular, we have outlined the relevance of changes in the labour force participation within the family to the job quit behaviour of other family members. If a previously employed household member (wife, son, parent) is now unemployed, this may make it less tolerable for the head to quit his current job and undergo a possible spell of unemployment. On the other hand, if a previously unemployed member has now been employed, this may make it more tolerable for the head to leave. Thus, we may define:

\[ FE = E_1 - E_0 \]

where \( E_0 \) = sum of earnings of other household earnings before the move; and \( E_1 \) = sum of their earnings after the move. A rise in \( FE \) implies (i) either that a previously unemployed member has now been employed, (ii) or some other family member(s) has had an increment (e.g. due to bonus or promotion). A fall in \( FE \) can be defined by symmetry. However, since the time difference between the pre-move and post-move is relatively small (seven to ten months), it is more likely that a rise (or fall) in
FE accrues as a result of the first reason (i) above.

The hypothesis we want to test is that this FE variable (6.7) will have a negative effect on the job leaving probability. The effect of this variable has not been allowed for by Engelman (49). Alternatively, he dichotomized $E_1$ (see (6.7) above) so that it is equal to one if $E_1$ is above the sample median and zero elsewhere. However, this dichotomous variable (for which Engelman (49) obtained a significant positive coefficient; see Chapter I, Table (1.1)) does not capture the effect of changes in labour force participation within the family as it ignores $E_0$. It only relates to the post-move dispersion of the "between families" earnings. Thus, it embodies the effect of family size since for multiple adult families we expect to find $E_1 > 0$, while for single adult families it is always true that $E_1 = 0$. This bias due to the family size effect does not relate to our measure FE as it is capable of assuming negative zero or positive values. Hence, we shall incorporate our variable FE, for the effect of changes in family earnings on the leaving probability for both males and females.

(ix) Other Variables

There are other explanatory variables which may be used to specify the leaving probability model:

(a) Unearned income received before move (e.g. family supplement or allowance, child benefit, bonus, etc. (INC)).

(b) Wage-rate offered by the pre-move employer (W).

(c) Previous housing tenure (i.e. owned or rented) (HT)

(d) Number of dependent children in the family (NCH)

We shall consider these variables below, in turn:

(a) The data we have about INC is simply a dichotomous variable which
is equal to one if such unearned income has been received and zero otherwise. However, the type of income received, its source and quantity are not known. Hence, the \textit{a priori} expectation for the effect of this variable cannot be made, but it may be tested empirically.

(b) The housing tenure variable, HT, is also representable by a dichotomous variable where $HT = 1$ if previous house was owned, and zero elsewhere. We may expect this variable to have a negative effect as housing ownership is more indicative of housing stability, and therefore employment stability. However, this hypothesis has to be qualified by the length of stay at previous house since $HT = 1$ need not be associated necessarily with a longer staying duration at past address - and it is the latter quantity which is indicative of housing stability. Hence, the empirical effect of the housing tenure variable, HT, will reflect the degree of concentration of the stay duration distribution among owners and non-owners, e.g. a significant negative coefficient may indicate higher concentration of long duration of stay at past address among owners than among non-owners.

(c) The incorporation of the number of children as an explanatory variable allows for the effect of dependents in the family. The expectation is that the greater the number of dependents, the smaller is the job leaving probability. However, the use of 'children' is subject to certain reservations as it is important to allow for their ages. For a female spouse the presence of children below five years may induce her to stay at home to provide child care, while if all children are at school, she may find it more rational to spend the day time in paid work. Engelman (49) has tested the effect of the presence of children below five years on the basis of his linear probability model of female quits (see Chapter I, Table (1.7)). It is also questionable whether all children as defined in our sample can be regarded as dependents since some of them do
paid work - specially those above fifteen years. To the extent that this is the case, the hypothesis of a negative effect of the number of children on the head's quit propensity is questionable. Similarly it can be argued that family size or number of adult numbers in the household may have self-cancelling effect on the model.

(d) The effect of wage earnings is another possible source of variation of quit decisions. Given the level of non-pecuniary benefits and transfer costs, the higher the wage received by an employee the less likely he would be to move out. However, the empirical sign of the wage variable should depend on the level of non-pecuniary rewards of work or what we have called the subjective rate of marginal substitution between pecuniary and non-pecuniary benefits; see Chapter II, section (2.7). Moreover, the relevant quantity should be the relative wage, i.e. current wage-rate of an employee relative to what he can get elsewhere (the opportunity wage). Otherwise comparison of absolute wage-earnings on a cross-sectional basis has rather vague implications. The relative wage measure has been used by Richardson et al. (122) who obtained, as expected, a significant negative coefficient on their leaving probability model. However, other authors used the absolute wage measure and similarly obtained significant coefficients, e.g. Viscusi (144) and Stoikov and Raimon (134). In the latter case it appears that the range of observations do not deviate much from the relative wage model, i.e. lower absolute wage earnings tend to lie below their opportunity levels while higher absolute wage earnings tend to lie at or above their opportunity levels. On the other hand, Engelman (49) constructed a 0,1 dichotomous variable from the gross earnings of the previously employed worker for his linear quit model but he obtained rather poor t-values (see Chapter I, section (1.1)). The absence of net wage-earnings is another limitation of our data and there is not much to expect from this crude variable, as we shall shortly examine.
The sample correlations matrix for these variables is given in Table (6.2a) for males and Table (6.2b) for females. The corresponding sample means are given in Table (6.3) below. If we look at Table (6.2a) we find that the highest correlation for males exists between the two variables for children (i.e. number of children, and presence of children below five years). The positive correlation 0.55 agrees with intuition. The next highest correlation -0.52 is a negative one between AGE and presence of children below five years (CH < 5) and again this agrees with common-sense, although the negative correlation with the number of children (NCH) is quite low (-0.15). The same impressions can be obtained by looking at the females matrix of Table (6.2b). However, we find a rather high correlation 0.69 existing between non-labour income received (INC) and the number of children (NCH). This result is rather interesting. It throws light on the expectation that this income is related to the number of dependents in the family, e.g. child benefit, family income supplement, etc.). The corresponding correlation coefficient for males is also positive but relatively low, 0.21. However, for females the highest correlation, 0.760, exists between AGE and the move variable, M. It implies that female actual movers are on average older than intending movers. However as $R^2 = 0.58 < 0.70$, this may not cause a serious collinearity problem*. The negative correlation - 0.38 between the hours of work dummy variable, HRS and the presence of children below five years (CH < 5) is also plausible. This appears as the biggest correlation for females between HRS and the rest of the variables. We may also note that the change in rent and rates variable is positively correlated with age in the case of both males and females; 0.24 for males and 0.41 for females. It is also positively correlated with their size of gross earnings, GE, as expected -

* See comments in footnote of Chapter III, p. 91.
<table>
<thead>
<tr>
<th></th>
<th>RR</th>
<th>TT</th>
<th>FE</th>
<th>AGE</th>
<th>ED</th>
<th>SPT</th>
<th>INC</th>
<th>HT</th>
<th>MOV</th>
<th>NCH</th>
<th>CH &lt;5</th>
<th>GE</th>
<th>HRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>1.0</td>
<td>-0.026</td>
<td>-0.005</td>
<td>0.415</td>
<td>-0.163</td>
<td>-0.130</td>
<td>0.206</td>
<td>0.104</td>
<td>0.157</td>
<td>0.126</td>
<td>-0.249</td>
<td>0.198</td>
<td>0.142</td>
</tr>
<tr>
<td>TT</td>
<td>1.0</td>
<td>-0.293</td>
<td>0.014</td>
<td>-0.034</td>
<td>-0.081</td>
<td>0.091</td>
<td>-0.091</td>
<td>-0.267</td>
<td>0.175</td>
<td>0.087</td>
<td>0.028</td>
<td>-0.076</td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>1.0</td>
<td>-0.013</td>
<td>-0.031</td>
<td>0.001</td>
<td>0.128</td>
<td>-0.123</td>
<td>0.071</td>
<td>0.051</td>
<td>-0.028</td>
<td>0.081</td>
<td>0.170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>1.0</td>
<td>-0.056</td>
<td>-0.136</td>
<td>0.043</td>
<td>0.043</td>
<td>0.76</td>
<td>-0.147</td>
<td>-0.477</td>
<td>0.278</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>1.0</td>
<td>0.214</td>
<td>-0.051</td>
<td>0.082</td>
<td>-0.117</td>
<td>0.033</td>
<td>0.088</td>
<td>0.146</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>1.0</td>
<td>-0.182</td>
<td>-0.081</td>
<td>-0.003</td>
<td>-0.129</td>
<td>-0.017</td>
<td>0.105</td>
<td>0.079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>1.0</td>
<td>0.033</td>
<td>-0.013</td>
<td>0.694</td>
<td>0.175</td>
<td>0.045</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>1.0</td>
<td>0.010</td>
<td>0.104</td>
<td>0.048</td>
<td>0.140</td>
<td>-0.082</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOV</td>
<td>1.0</td>
<td>-0.100</td>
<td>-0.115</td>
<td>0.176</td>
<td>0.117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCH</td>
<td>1.0</td>
<td>0.437</td>
<td>-0.201</td>
<td>-0.193</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH &lt;5</td>
<td>1.0</td>
<td>-0.262</td>
<td>-0.330</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRS</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE (6.2b)
Correlation Matrix for Females
**TABLE (6.2a)**

**Correlation Matrix for Males**

<table>
<thead>
<tr>
<th></th>
<th>RR</th>
<th>TT</th>
<th>FE</th>
<th>AGE</th>
<th>ED</th>
<th>SPT</th>
<th>INC</th>
<th>HT</th>
<th>MOV</th>
<th>NCH</th>
<th>CH &lt;5</th>
<th>GE</th>
<th>HRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>1.00</td>
<td>0.102</td>
<td>0.066</td>
<td>0.236</td>
<td>0.171</td>
<td>-0.072</td>
<td>-0.017</td>
<td>0.128</td>
<td>0.179</td>
<td>0.064</td>
<td>-0.131</td>
<td>0.220</td>
<td>0.053</td>
</tr>
<tr>
<td>TT</td>
<td>1.00</td>
<td>-0.088</td>
<td>-0.023</td>
<td>0.090</td>
<td>-0.074</td>
<td>-0.027</td>
<td>-0.076</td>
<td>0.178</td>
<td>0.112</td>
<td>0.075</td>
<td>-0.082</td>
<td>-0.201</td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>1.00</td>
<td>-0.048</td>
<td>-0.006</td>
<td>-0.135</td>
<td>-0.077</td>
<td>0.057</td>
<td>-0.002</td>
<td>-0.150</td>
<td>0.013</td>
<td>0.088</td>
<td>0.318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>1.00</td>
<td>-0.205</td>
<td>0.122</td>
<td>-0.001</td>
<td>0.053</td>
<td>-0.025</td>
<td>-0.147</td>
<td>-0.518</td>
<td>0.362</td>
<td>0.099</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>1.00</td>
<td>0.186</td>
<td>0.120</td>
<td>0.042</td>
<td>-0.042</td>
<td>0.028</td>
<td>0.150</td>
<td>-0.166</td>
<td>-0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>1.00</td>
<td>0.153</td>
<td>-0.055</td>
<td>0.028</td>
<td>0.077</td>
<td>0.080</td>
<td>-0.073</td>
<td>0.130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>1.00</td>
<td>0.051</td>
<td>-0.032</td>
<td>0.207</td>
<td>0.093</td>
<td>-0.120</td>
<td>-0.099</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>1.00</td>
<td>-0.008</td>
<td>-0.004</td>
<td>-0.008</td>
<td>-0.119</td>
<td>-0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOV</td>
<td>1.00</td>
<td>-0.004</td>
<td>-0.008</td>
<td>-0.119</td>
<td>-0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCH</td>
<td>1.00</td>
<td>0.549</td>
<td>-0.226</td>
<td>-0.181</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH &lt;5</td>
<td>1.00</td>
<td>-0.323</td>
<td>-0.111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>1.00</td>
<td>0.140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRS</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
though the correlations are rather low (0.22 and 0.20 respectively). The size of variables correlations in general are rather low. But we have to recall that these are simple correlations.

The above remarks are the most general features of the sample correlation matrices. We may also look at the sample means in Table (6.3). We note that average change in travel time TT for males (21.7 minutes) is more than twice as big as the average change for females (i.e. 8.7 minutes). The proportion of those whose education level is not below senior secondary school ED is also higher for males (0.43) than for females (0.25). This judgment also applies as regards attainment of special training, SPT.

(Notice that ED and SPT are positively correlated as expected for both males and females, although the correlations are not very big: 0.19 for males, and 0.21 for females). It is also noticeable that earning of 'other household members' has decreased on average for both sexes as a result of the move - as FE is negative - implying reduced labour participation in the average family after the move. At this stage it might be observed that certain family characteristics are not having the same means for males and females (e.g. housing tenure, HT, children below five years, CH<5). This is attributable to the fact that these two samples of male heads and female spouses do not match strictly as husbands and wives. Apart from the problem of non-response of certain wives and certain husbands, there is the observation that 14 of the male heads say they are not married, i.e., single, widowed, divorced or separated. In fact, if we match up wife and husband responses, who were employed before the move, we get only 61 cases - as it appears from the 2 x 2 contingency Table (3.2) of Chapter III. Partly, this non-response problem of both spouses in the household has accounted for our inability to adopt a full simultaneous logit model analysis. Hence, comparison of male heads' empirical models and those of female spouses in the following single
### TABLE (6.3)

Sample Means of the Explanatory Variables

#### (A) Males

<table>
<thead>
<tr>
<th>Variable</th>
<th>RR</th>
<th>TT</th>
<th>FE</th>
<th>AGE</th>
<th>ED</th>
<th>SPT</th>
<th>INC</th>
<th>HT</th>
<th>MOV</th>
<th>CHD</th>
<th>CHD&lt;5</th>
<th>GE</th>
<th>HRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>10.018</td>
<td>21.665</td>
<td>-4.116</td>
<td>33.838</td>
<td>0.428</td>
<td>0.526</td>
<td>0.127</td>
<td>0.220</td>
<td>1.468</td>
<td>0.561</td>
<td>111.010</td>
<td>0.902</td>
<td></td>
</tr>
</tbody>
</table>

#### (B) Females

<table>
<thead>
<tr>
<th>Variable</th>
<th>RR</th>
<th>TT</th>
<th>FE</th>
<th>AGE</th>
<th>ED</th>
<th>SPT</th>
<th>INC</th>
<th>HT</th>
<th>MOV</th>
<th>CHD</th>
<th>CHD&lt;5</th>
<th>GE</th>
<th>HRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>9.789</td>
<td>8.400</td>
<td>-6.502</td>
<td>34.431</td>
<td>0.442</td>
<td>0.253</td>
<td>0.263</td>
<td>0.179</td>
<td>0.168</td>
<td>1.021</td>
<td>0.305</td>
<td>33.336</td>
<td>0.557</td>
</tr>
</tbody>
</table>
equation models should be made with some special care.

Let us now start with the male heads' models as we shall do in the next section.

6.3 Commentary on the Empirical Models for Males

The number of male heads who were employed before the move into the council housing consists of 184 cases, which reduce to 175 after removal of those who have either been dismissed or made redundant. Specifically these individuals had been doing paid work before their move into the council houses between July and October, 1973. The information as regards subsequent changes in their employment attachments had been elicited later between April and May, 1974, i.e. a period between seven and ten months has elapsed between the pre-move and post-move data. The number of actual movers among these 175 cases is 127, which is quite big compared to the 48 individuals who had also been allocated council housing but had not yet moved, i.e. 'intending' movers; the latter people have been chosen as a control group to isolate the effect of the move (see Chapter I, for the data description). As these remarks apply to males and females alike they are not to be repeated later for females.

The estimates which we are about to review emerge from the numerical maximization of the likelihood equation based on the logistic probability model

\[
\text{Prob}(Y_i = 1) = \frac{e^{\beta'X_i}}{1 + e^{\beta'X_i}}
\]

where \( Y_i = 1 \) if the \( i^{th} \) individual quits at any time over the period of observation

\( = 0 \) otherwise.
$X_i$ is the observed column vector of explanatory variables for the $i^{th}$ individual; $X' = (X_{1i}, \ldots, X_{pi})$ and $\beta' = (\beta_0, \ldots, \beta_p)$ is the associated unknown parameter coefficient vector containing a constant term, $\beta_0$.

The techniques of estimation and significance testing adopted in this chapter have already been discussed in Chapter III (e.g. see section (3.6)). However, the first difficulty that we have met with, the numerical estimation of model (6.8) above, is that when the hours dummy, HRS, is included in the vector $X_i$ together with the constant term, $\beta_0$, we never get a convergent solution. This problem is attributable to the fact that

$$\text{Prob}(Y_i = 1 | \text{HRS} = 0) = 1.0;$$

in other words, all previously employed males with HRS = 0 (i.e. working for at most 30 hours per week) have quit over the period of observation. This property makes it impossible in practice to get the MLE of the parameter coefficient of HRS. This point has been made more clearly by Nerlove and Press (106) who observed that

"...if the data tells us that the probability must be one for the explanatory variable to equal one, the estimated coefficient must be $+\infty$. A computer cannot represent a number that large; the attempt to estimate such a coefficient results in a failure to converge ... for maximizing the likelihood function, or occasionally an apparent convergence but with a very large number of iterations and peculiar values for the affected coefficient and the constant term ..."

Nerlove and Press (106, p. 69).

In the case of our model the probability must be one for the explanatory
variable HRS to equal zero.

It is noticeable that this problem is peculiar to the maximum likelihood approach but can not be felt when applying the linear probability model. Yet, according to Nerlove and Press\(^{(106)}\) lack of such a signal should be regarded as a shortcoming of the linear probability model in the same way as failure of a numerical matrix inversion routine to signal singularity is considered a weakness. There are two possible ways out of this difficulty.

(i) We could ignore the constant term altogether; but this would entail undesirable limitations, as we shall shortly see.

(ii) Alternatively, we may run the maximum likelihood procedure separately for cases where HRS = 1 (i.e. 157 cases) and make another run on the combined sample ignoring the offending variable, HRS. The advantage of this compromise is that it enables us to test the effect of this variable HRS indirectly by comparing the empirical or the fitted \(S_B\) curves of the distribution of estimated quit propensities under these two models.

As regards the alternative of dropping the constant term \(\beta_0\), it is undesirable a priori. This will adversely affect the slope of the probability plane in terms of the other explanatory variables as illustrated by the single-variable model below. In fact we have tried this alternative possibility and obtained uninteresting and perverse coefficient values (e.g. a positive coefficient for the AGE variable) and so we have given up this alternative. Thus, for the rest of this section we shall be adopting the alternative procedure (ii) outlined above. We first review the empirical results pertaining to the combined group (with both HRS = 1, and HRS = 0), and later we give separate treatment to the cases where HRS = 1.
(i) Curve fitted with a non-zero constant term.

(ii) Curve fitted with a zero constant term.

The tabular representation of results is done column-wise in terms of the vector of maximum likelihood estimates, $\hat{\theta}$ and the associated point elasticities. These are given in the same column, while the corresponding $t$-values are produced in the adjacent column together with their significance levels of the null hypotheses $\delta_j = 0; S = 0, \ldots, p$. Let us first consider the left hand column of Table (6.4) which includes the variables: RR for change in rent plus rates; TT, change in travel time after the move; FE, change in family earnings after the move; AGE, as given in years; ED, for level of education; SPT, indicating attainment or otherwise of special training; INC, for receipt of non-labour income since the move; MOV, indicating whether the individual is an actual mover or not; HT is the housing tenure dummy indicating whether the pre-move house was owned or otherwise; while 'con' stands for the constant term. Notice that the parameter coefficient estimates for the variables RR, TT, FE, AGE and ED are specially interesting as they possess the expected signs and are highly significant. AGE seems to have

Note that significance levels are written to the nearest tabulated value of $t$ which is lower than the computed $t$-value. As we are restricted by the tabulated values the significance levels are only approximate.
TABLE (6.4)
Results for Combined Sample of Males

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \hat{\beta} ) (elast.)</th>
<th>t (sig.)</th>
<th>( \hat{\beta} ) (elast.)</th>
<th>t (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td>2.120</td>
<td>2.687 (0.01)</td>
<td>1.965</td>
<td>2.500 (0.02)</td>
</tr>
<tr>
<td>RR</td>
<td>-0.078</td>
<td>3.107 (0.01)</td>
<td>-0.082</td>
<td>3.193 (0.01)</td>
</tr>
<tr>
<td></td>
<td>(-0.359)</td>
<td></td>
<td>(-0.383)</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>0.037</td>
<td>2.878 (0.01)</td>
<td>0.037</td>
<td>2.831 (0.01)</td>
</tr>
<tr>
<td>FE</td>
<td>-0.044</td>
<td>2.423 (0.02)</td>
<td>-0.044</td>
<td>2.470 (0.02)</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td></td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>AGE,</td>
<td>-0.058</td>
<td>2.910 (0.01)</td>
<td>-0.065</td>
<td>3.066 (0.01)</td>
</tr>
<tr>
<td></td>
<td>(-0.901)</td>
<td></td>
<td>(-1.024)</td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>-1.175</td>
<td>2.520 (0.02)</td>
<td>-1.230</td>
<td>2.654 (0.01)</td>
</tr>
<tr>
<td></td>
<td>(-0.232)</td>
<td></td>
<td>(-0.246)</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>-0.431</td>
<td>1.049 (0.30)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.103)</td>
<td></td>
<td>(-0.103)</td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>1.055</td>
<td>1.800 (0.10)</td>
<td>1.028</td>
<td>1.747 (0.10)</td>
</tr>
<tr>
<td></td>
<td>(0.0605)</td>
<td></td>
<td>(0.070)</td>
<td></td>
</tr>
<tr>
<td>MOV</td>
<td>-0.586</td>
<td>1.091 (0.30)</td>
<td>-0.535</td>
<td>0.988 (0.40)</td>
</tr>
<tr>
<td></td>
<td>(-0.058)</td>
<td></td>
<td>(-0.059)</td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>-0.524</td>
<td>1.043 (0.30)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.052)</td>
<td></td>
<td>(-0.052)</td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.836 (0.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.246)</td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>-</td>
<td>-</td>
<td>-0.036</td>
<td>0.223 (0.90)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-0.117)</td>
<td></td>
</tr>
</tbody>
</table>

\(-2\log \lambda \) 61.136 59.785
\(F^*\) 8.566 8.248
the largest (negative) elasticity followed by the change in travel time variable, TT. The positive sign of TT confirms the expectation that as the move pushes up travel time to work it raises the probability of job-quitting. Similarly increased housing costs RR reduce the probability of quits. The negative coefficient of the education dummy, ED, confirms our hypothesis that lower levels of education correspond to a less structured labour market and hence higher propensity to job leaving. The effect of changes in family earnings is shown by the variable, FE. As expected, this variable yields a negative coefficient on the job quit probability. As we have pointed out in the last section, this variable represents the changes in labour force participation within the family. Thus, if less members of the same household work after the move than before, then the head will be more obliged to stay in his job. The effect of non-labour income received since move, INC is positive and significant at 10% level. This is also a plausible result although we do not know the source, quality and quantity of such income. It is a plausible result as long as such income would alleviate the financial restraints of job leaving. The effect of special training yields the expected (negative) sign, although the null hypothesis would have a rather high significance level (30%). The effect of the MOV variable is negative, implying that the process of actual move involves concomitant circumstances that tend to disturb past job attachments. However, this negative effect can only be supported by a rather high significance level for the null hypothesis of zero effect (about 30% level). We similarly obtain a negative coefficient for the housing tenure dummy, HT, implying that owners are less likely to leave their past jobs. Again this implication is not supported with a sufficiently low significance level for the zero effect hypothesis.
Nevertheless, the model as a whole seems to be performing satisfactorily as judged by the log-likelihood ratio \((-2\log \lambda = 61.14\) ), which gives a highly significant value for the model when compared with the tabulated \(\chi^2(9) = 29.67\) at 0.0005 significance level. Our \(F^*\)-ratio = 8.57, which is significant at the level 0.001.

In the next trial we shall estimate a model where we keep the MOV variable but delete the special training variable SPT, and the housing tenure variable HT. We shall replace these variables by the amount of gross earnings per week before the move, GE and NCH the number of children, variables. The results of the calculations for this second model are presented in the right hand column of Table (6.4) above. Notice that, as before, the rent plus rates RR, the travel time TT, the family earnings FE, the AGE and the education ED variables are highly significant. They also seem quite stable although the AGE parameter has slightly decreased by .007, and the RR parameter by .004, while the ED parameter has decreased by .065 units. This slight instability of RR and AGE could be due to their relatively high correlation with the gross earnings variable, GE. Note that the partial correlation between AGE and GE is \(-0.36\) and between RR and GE is 0.22 - which are higher than the other inter-correlations of the variables. This is, however, only an incomplete explanation as collinearities cannot all be detectable by first order partial correlation coefficients. On the whole the model is reasonably stable but the newly introduced variables GE and NCH seem to be quite insignificant, although the number of children, NCH, parameter has the expected (negative) sign. Looking at the goodness-of-fit measures we see that this model is slightly inferior to the previous one. Specifically, the MOV variable has not shown any improvement.

Thus, we have estimated the model which we hope does incorporate
the significant explanatory variables: change in rent plus rates, RR; change in travel time, TT; change of family earnings, FE; AGE; education level, ED and non-labour income received, INC. This model is described by the first column in Table (6.5) below. Notice that the estimated coefficient parameters for these variables are quite stable, especially when compared with our first model of Table (6.4) above. The likelihood ratio test \(-2\log \lambda = 57.69\) is again implying a very good fit at 0.0005 level of significance when compared with \(\chi^2(6) = 27.87\). Our \(F^*\)-ratio = 12.11 gives a significant value at 0.005 level when compared with \(F(6, 174) = 8.88\).

Now, we may run this model on the actual movers alone in order to test the effect of the employment opportunities in new residential areas. The reduction of the sample size to 127 cases is not too serious, so that it may not affect the consistency of the MLE estimates. The results of the computations are shown in the second column of Table (6.5) above. Notice that there are slight variations in the values of the parameter estimates - the largest effect being in the education variable. The signs and significance levels of the variables RR, TT, AGE and ED are interesting as before but the family earnings variable is now quite poor and the non-labour income dummy, INC, is also less interesting. However, it is a rather interesting result to see that the number of employers in the new area, EMP, has not only got the expected sign (positive) but that it is also very highly significant. This implies that for this combined sample the larger the number of employment opportunities in the new area, the more likely that a re-housed employee may quit his previous job. (We see that EMP has a positive elasticity (= 0.22)). This result is interesting as it could support the hypothesis of random spatial distribution of employment opportunities; see section (6.2), IV. It is also consistent with the
## TABLE (6.5)

Results for the Combined Sample of Males

<table>
<thead>
<tr>
<th>Model</th>
<th>Combined Sample</th>
<th>Movers alone</th>
<th>Combined sample using ratio measure, TT</th>
<th>Combined sample using difference measure, TT × M = ( \hat{TT} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>( \hat{\beta} )</td>
<td>( t )</td>
<td>( \hat{\beta} )</td>
<td>( t )</td>
</tr>
<tr>
<td></td>
<td>(elast.)</td>
<td>(sig.)</td>
<td>(elast.)</td>
<td>(sig.)</td>
</tr>
<tr>
<td>CON</td>
<td>1.738</td>
<td>2.338</td>
<td>1.889</td>
<td>2.137</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>RR</td>
<td>-0.080</td>
<td>3.299</td>
<td>-0.087</td>
<td>2.900</td>
</tr>
<tr>
<td></td>
<td>(-0.377)</td>
<td>(0.001)</td>
<td>(-0.319)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>TT</td>
<td>0.036</td>
<td>2.894</td>
<td>0.034</td>
<td>2.503</td>
</tr>
<tr>
<td></td>
<td>(0.461)</td>
<td>(0.01)</td>
<td>(0.188)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>FE</td>
<td>-0.043</td>
<td>2.439</td>
<td>-0.011</td>
<td>0.484</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.02)</td>
<td>(0.020)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.058</td>
<td>2.901</td>
<td>-0.066</td>
<td>2.728</td>
</tr>
<tr>
<td></td>
<td>(-0.922)</td>
<td>(0.01)</td>
<td>(-0.892)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>ED</td>
<td>-1.210</td>
<td>2.66</td>
<td>-1.563</td>
<td>2.775</td>
</tr>
<tr>
<td></td>
<td>(-0.245)</td>
<td>(0.01)</td>
<td>(-0.275)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>INC</td>
<td>0.900</td>
<td>1.566</td>
<td>0.821</td>
<td>1.120</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.15)</td>
<td>(0.043)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>EMP</td>
<td>-</td>
<td>-</td>
<td>0.003</td>
<td>4.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.221)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>-2log( \lambda )</td>
<td>57.689</td>
<td>59.617</td>
<td>41.792</td>
<td>36.070</td>
</tr>
<tr>
<td>F*</td>
<td>12.108</td>
<td>11.874</td>
<td>8.675</td>
<td>6.733</td>
</tr>
</tbody>
</table>
finding of Mackay et al. (92) that the conurbations of Glasgow do not constitute a single labour market due to the costly process of residential mobility and long journey-to-work cost. Thus, relocation to a new area where there are a greater number of potential employers would have a significant positive effect on the job leaving propensities of the relocated workers.

Before moving to consider the other empirical models for male heads with weekly work hours greater than 30 (HRS = 1), it is appropriate to test other representations for the change in travel time variable, \( TT \); namely the ratio measure \( TT^* \) and the difference measure multiplied by the MOV variable, \( TT \). In the latter case we are ignoring the change in travel time for those males who have not yet moved into their already allocated council houses. Thus, we have computed two models including, respectively, \( TT \) and \( TT^* \) as they appear in Table (6.5) above. Notice that the performance of either of these two models is inferior relative to the original model with the simple difference measure, \( TT \) - as judged by \(-2\log\lambda\) and the \( F^*\)-ratio. However, it appears that the ratio model \( (TT^*) \) is the least satisfactory as the statistical significance of \( TT^* \) is relatively low and \(-2\log\lambda = 36.07\) (or \( F^* = 6.73\)) are much lower respectively than the corresponding values (57.69 and 12.11) of the simple difference model. Hence, within the range of variation of this data it seems that the simple difference measure, \( TT \) corresponds with the model with the best fit.

So far we have been indexing each MLE of the parameter coefficients with its associated point elasticity, which normally relates to very small changes in the mean of the variable in question. Now, to test the effect of specific finite changes we need to calculate arc elasticities corresponding to the different variables. The arc elasticity formula has been proposed by us in the last chapter; see
formula (5.44). The computation of this formula is very easy compared to the procedures of Richard and Ben-Akiva (121), and Westin (148). We have calculated these elasticities corresponding to specific percentage changes of the mean-variables in question, i.e. (5%, 10%, 20%, 30% and 50%). The results of the calculations are presented in Table (6.6) below using the change in rent plus rates, RR; change in travel time, TT; change in family earnings, FE; AGE; education level, ED; and non-labour income, INC. Note that there is no material change between the point elasticity values at the 5% arc elasticity values for all the variables. Even up to 20% change some arc elasticities do not differ considerably from point elasticities (e.g. INC, TT). It seems that the AGE variable's arc elasticity is the most rapidly increasing one (in absolute value). This is to be expected as AGE has got the largest point elasticity among the variables.

Let us now move to consider the empirical models for the cases with HRS = 1. Exactly the same tabular format of the combined cases' models will be adopted in the presentation of results here. We have a sample size of 157 cases with HRS = 1 (i.e. their labour supply hours per week is not below 30). However the proportion of quits in this case is only 0.254 compared to 0.334 of the combined model. Thus we expect some variability of results due to the elimination of a special quit-prone group, small though it was.

First consider Table (6.7) below, which contains the same variables as the first table (6.4) of the above discussed combined model, i.e. RR, TT, FE, AGE, ED, SPT, INC, MOV and HT. We find that, as before, the rent and rates variable, RR, the travel time variable, TT, the AGE and education, ED, variables stand out as highly significant with their a priori expected signs. However, the change in family earnings, FE is no longer significant. The housing tenure dummy, HT, and the MOV are
<table>
<thead>
<tr>
<th>Variable</th>
<th>Point Elasticity</th>
<th>Arc Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>RR</td>
<td>-0.377</td>
<td>-0.389</td>
</tr>
<tr>
<td>TT</td>
<td>0.460</td>
<td>0.467</td>
</tr>
<tr>
<td>FE</td>
<td>0.093</td>
<td>0.095</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.921</td>
<td>-0.967</td>
</tr>
<tr>
<td>ED</td>
<td>-0.245</td>
<td>-0.253</td>
</tr>
<tr>
<td>INC</td>
<td>0.053</td>
<td>0.054</td>
</tr>
</tbody>
</table>
again with negative coefficients but with high significance levels for the zero effects (null) hypotheses. The special training coefficient is neither significant nor interesting. As regards the performance of the model as a whole it has a very good fit as \(-2\log \lambda = 48.18 > x^2(9) = 29.67\) at .0005 level of significance. Similarly the \(F^*\)-ratio = 9.07 is bigger than \(F(9, 157) = 7.81\) at .001 significance level. However, we have estimated another model where we have removed the MOV and HT variables as shown in the adjacent column in Table (6.7) above. Notice that the results are quite stable, especially for the meaningful and significant variables, RR, TT, AGE and ED. However the non-labour income dummy, INC is less significant than before. The performance of the model is also very good despite the inclusion of the uninteresting variables FE and SPT. In the next model of Table (6.8) below we have re-inserted the MOV variable but removed the special training variable, SPT, and the family earnings variable, FE. In addition we have introduced two new variables: namely the number of children NCH, and gross earnings before the move, GE. The MLE of the coefficient parameters of RR, TT, AGE have slightly changed but the ED parameter estimate is quite stable. These changes could be attributed to the inclusion of gross earnings which has a positive correlation with AGE (0.36) and with RR (0.22), or due to some other hidden higher order correlations. These changes are none-the-less not too serious. The state of statistical significance for the important variables RR, TT, AGE and ED is still maintained and moreover the non-labour income variable, INC, is now more significant than before. The values of \(-2\log \lambda = 51.35\) and \(F^* = 8.60\) as before indicate very good fit the the model. Then we may remove the variables MOV (which have shown no improvement), GE (whose statistical significance is very low and its coefficient sign is perverse) and the number
### Table (6.7)

Results for Males who have worked more than 30 hours per week

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{}\beta$ (elast.)</th>
<th>t (sig.)</th>
<th>$\hat{}\beta$ (elast.)</th>
<th>t (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>1.562</td>
<td>1.740 (0.10)</td>
<td>1.408</td>
<td>1.584 (0.15)</td>
</tr>
<tr>
<td>RR</td>
<td>-0.072 (-0.382)</td>
<td>2.741 (0.01)</td>
<td>-0.075 (-0.406)</td>
<td>2.979 (0.01)</td>
</tr>
<tr>
<td>TT</td>
<td>0.038 (0.346)</td>
<td>2.749 (0.01)</td>
<td>0.041</td>
<td>2.970 (0.01)</td>
</tr>
<tr>
<td>FE</td>
<td>-0.005 (0.009)</td>
<td>0.236 (0.80)</td>
<td>-0.006</td>
<td>0.290 (0.80)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.056 (-1.008)</td>
<td>2.464 (0.02)</td>
<td>-0.056 (-1.008)</td>
<td>2.440 (0.02)</td>
</tr>
<tr>
<td>ED</td>
<td>-1.485 (-0.333)</td>
<td>2.706 (0.01)</td>
<td>-1.450 (-0.327)</td>
<td>2.684 (0.01)</td>
</tr>
<tr>
<td>SPT</td>
<td>0.122 (0.035)</td>
<td>0.257 (0.80)</td>
<td>0.068</td>
<td>0.145 (0.90)</td>
</tr>
<tr>
<td>INC</td>
<td>0.876 (0.052)</td>
<td>1.284 (0.20)</td>
<td>0.763</td>
<td>1.133 (0.30)</td>
</tr>
<tr>
<td>MOV</td>
<td>0.721 (-0.081)</td>
<td>1.115 (0.30)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HT</td>
<td>-0.373 (-0.024)</td>
<td>0.199 (0.90)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

-2logλ  | 48.184               | 46.685 |
F*       | 9.067                | 8.915 |
TABLE (6.8)

Results for Males who have worked more than 30 hours per week

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\beta}$ (elast.)</th>
<th>t (sig.)</th>
<th>$\hat{\delta}$ (elast.)</th>
<th>t (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>2.095</td>
<td>2.274</td>
<td>1.448</td>
<td>1.739</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.10)</td>
<td>(0.406)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>RR</td>
<td>-0.081</td>
<td>2.946</td>
<td>-0.075</td>
<td>2.993</td>
</tr>
<tr>
<td></td>
<td>(-0.424)</td>
<td>(0.01)</td>
<td>(-0.406)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>TT</td>
<td>0.047</td>
<td>3.150</td>
<td>0.041</td>
<td>2.994</td>
</tr>
<tr>
<td></td>
<td>(0.423)</td>
<td>(0.01)</td>
<td>(0.378)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.070</td>
<td>2.930</td>
<td>-0.056</td>
<td>2.462</td>
</tr>
<tr>
<td></td>
<td>(-1.232)</td>
<td>(0.01)</td>
<td>(-1.004)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>ED</td>
<td>-1.485</td>
<td>2.704</td>
<td>-1.448</td>
<td>2.721</td>
</tr>
<tr>
<td></td>
<td>(-0.327)</td>
<td>(0.01)</td>
<td>(-0.330)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>INC</td>
<td>1.114</td>
<td>1.613</td>
<td>0.788</td>
<td>1.788</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.15)</td>
<td>(0.048)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>MOV</td>
<td>-0.515</td>
<td>0.77</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.056)</td>
<td>(0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCH</td>
<td>-0.279</td>
<td>1.235</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.195)</td>
<td>(0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.002</td>
<td>1.196</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(0.30)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\log \lambda$ | 51.350 | 46.556

$F^*$ | 8.596 | 12.580
of children variable, NCH. The latter variable's coefficient has the expected (negative) sign though at rather high significance level of the null hypothesis. These removals lead us to the other model shown adjacently in Table (6.8) above. We may notice that the AGE, RR and TT coefficients have now become more comparable to their previous values before the inclusion of NCH and GE. The statistical significance of the INC coefficient parameter is now reasonably improved. The model as a whole is a very good fit as apparent from $-2\log \lambda = 46.56$ and $F^* = 8.60$. The likelihood ratio is again significant at 0.0005 significance level while the $F^*$ is significant at the 0.001 level. This model includes all of the interesting variables which appear to be consistent within the range of variation of our data.

It is also possible to repeat our trial of the travel time variable TT which is multiplied by the MOV variable. We have seen on the basis of the previous combined model that such a measure, $\tilde{TT}$ gives a relatively inferior model. Again we get the same impression about the relative superiority of the simple difference measure, TT, if we look at Table (6.9) below - the log-likelihood ration, $-2\log \lambda$ and the $F^*$-ratio are respectively 31.46 and 7.63 as compared to their corresponding values, 46.68 and 8.91 of the simple difference model.

We may now go to test our original five variables model (RR, TT, AGE, ED, INC) on actual movers alone. The results are also shown in Table (6.9) where we have introduced the employment opportunities variable, EMP. Note that the coefficient parameter estimates of the rent and rates variable, RR, the AGE variable and the education variable, ED, are quite stable when compared with those of the original five variables model for the joint model; even the change in the travel time, TT, coefficient is quite small. These variables maintain their high statistical
significance levels, but the INC variable is now rather poor. On the other hand, the coefficient parameter estimate for the employment opportunities variable, EMP, is positive as expected but with not more than 70% significance of the alternative nonzero effect hypothesis. It is also noticeable that the significance of the positive effect of the non-labour income, INC, is now quite low. In general the model provides a good fit as judged by $2 \log \lambda = 33.4$ and $F^* = 6.9$ when compared to their tabulated values at .005, and .01 levels of significance respectively.

As regards the point and arc-elasticities relating to the five-variables-model they are presented in Table (6.10) below. Note that the point elasticity for AGE is now even higher than that of the combined model. It is actually the largest point elasticity (= -1.004) followed by TT, the change in travel time's elasticity (= 0.378). As before we notice that for up to 10% increases in the means of these variables the arc elasticities do not change materially from the corresponding estimates of the point elasticities. The most rapidly changing arc-elasticity is that related to the AGE variable, i.e. there is absolute range of $(1.675 - 1.055) = 0.620$ units, compared to only 0.048 units for the TT mean-variable.

6.3.1 Recapitulation

To recapitulate, we have so far given separate treatment for two classes of data: combined cases (i.e. combining males who have worked more than 30 hours per week before move (HRS = 1) and those who worked less (HRS = 0) and the sub-sample of cases with HRS = 1. The main motivation has been our inability to get the MLE for the parameter coefficient of the HRS variable due to the non-convergence problem discussed on p. 247. However, although we have highly significant and
TABLE (6.9)

Males who have worked more than 30 hours/week

<table>
<thead>
<tr>
<th></th>
<th>Actual and intending movers using ratio measure $T^*$</th>
<th>Actual Movers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$ (elast.)</td>
<td>$t$ (sig.)</td>
</tr>
<tr>
<td>CON</td>
<td>1.154</td>
<td>1.517</td>
</tr>
<tr>
<td>RR</td>
<td>-0.058</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>(-0.353)</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>0.033</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-0.050</td>
<td>2.393</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>-1.095</td>
<td>2.308</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>0.794</td>
<td>1.293</td>
</tr>
<tr>
<td>EMP</td>
<td>0.001</td>
<td>1.001</td>
</tr>
</tbody>
</table>

- $2 \log \lambda$ | 31.459 | 33.383 |
- $F^*$ | 7.629 | 6.93 |
### Table 6.10

Arc Elasticities for Males who have worked more than 30 hours per week

<table>
<thead>
<tr>
<th>Variable</th>
<th>Point Elasticity (sig.)</th>
<th>Arc Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>RR</td>
<td>-0.406</td>
<td>-0.420</td>
</tr>
<tr>
<td>TT</td>
<td>0.378</td>
<td>0.384</td>
</tr>
<tr>
<td>AGE</td>
<td>-1.004</td>
<td>-1.055</td>
</tr>
<tr>
<td>ED</td>
<td>-0.330</td>
<td>-0.341</td>
</tr>
<tr>
<td>INC</td>
<td>0.047</td>
<td>0.049</td>
</tr>
</tbody>
</table>
meaningful MLE's for the effects of change in travel time, TT, change in rent and rates RR, AGE in years and the level of education, ED, yet when the models of the sub-sample cases (HRS = 1) have been considered we find that change in family earnings, FE is no longer interesting. In addition the employment opportunities variable, EMP is also less significant than with the combined cases model. Thus, it seems that the removal of cases with HRS = 0 has been felt at least rather meaningfully in the effect of the EMP variable. But the effect of the hours variable, HRS has, nevertheless, not been measured in a simple way. If we are not able to assess its effect by a single number we could do that by a simple geometrical description. This will be dealt with in the next section.

6.4 Geometrical Representation for the Effect of the Hours Dummy for Males

The concepts and techniques to be adopted here and in other relevant parts of this chapter have already been discussed in the last chapter and are not spelled out here. As the numerical non-convergence problem has deprived us of getting a single number of the effect of the hours dummy, HRS, for males we aim in this section to find a substitute. What we propose is to construct two frequency distributions for the estimated quit probabilities \( \hat{\theta}_i \) under the two classes of data (i.e. combined cases, and sub-sampled cases with HRS = 1). On the basis of these frequency histograms we shall construct smoothed \( S_B \) curves which are nicer and simpler. Then we shall plot these \( S_B \) curves up to scale in the same graph to compare their extent of skewness to the right - to what extent are the distributions of quit propensities concentrated near zero? The technical details of the curves' fitting are omitted
here, and could be seen in the fifth chapter (section (5.5)). Hence for each of the combined cases and sub-sample cases with HRS = 1, we shall construct an empirical histogram and a fitted $S_B$ curve. The empirical histograms merely show the relative frequency distribution of the estimated probabilities:

$$\hat{\theta}_i = \frac{\exp(\hat{\beta}'X_i)}{1 + \exp(\hat{\beta}'X_i)}$$

The $S_B$ curves are based only on the moments $\hat{\beta}'\hat{\mu}$ and $\hat{\sigma}^2 = \hat{\beta}'\hat{\Sigma}\hat{\beta}$; where $\hat{\mu}$ is the vector of sample means of the explanatory variables $X_i$, and $\hat{\Sigma}$ is estimated variance-covariance matrix.

Thus, we start with the combined cases (where HRS = 0, or 1) and utilize the six-variables model which only includes the significant variables; see Table (6.6). The outcome of our graphical experiment is produced in Figure (6.1) below. Similarly, we utilize the five-variables model of the sub-sample cases with HRS = 1 to arrive at Figure (6.2) below. Notice that the $S_B$ curves follow quite nicely the main trends of the empirical histograms. Next, by slightly modifying our computer routine we are able to plot these $S_B$ curves in a single graph, as shown in Figure (6.3) below. Note how simple and intuitive this figure looks. It shows that when the cases with HRS = 0 are excluded, this results in shifting the $S_B$ curve upwards and to the left - implying greater concentration of quit propensities near zero. Thus, the contrast between these two $S_B$ curves - which are plotted up to scale - gives a simple impression of the effect of the hours dummy, HRS for male heads.
Distribution of Individual Quit Probability for the combined sample of males.

Figure (6.1)
The distribution of estimated individual quit probabilities for the sub-sample of males with HRS = 1.

N = 157
$S_B$ curve for males with HRS = 1

$S_B$ curve for combined males.
6.5 Commentary on The Empirical Models for Females

Now, we shall turn to the discussion of female spouses' empirical models following the same tabular presentation method as that of male heads of section (6.3) above. The total sample size of female spouses who have been employed before the move is not very large (= 95). None of these responses has declared that the previous job was left due to employer's decision, and so they are all relevant for our job-quit analysis. The proportion of quits for this sample is 0.40 which is higher than that of 'combined' male heads (= 0.33). However, if we treat these estimates as binomial probabilities and apply a simple t-test for the difference between two binomial proportions \( P_1, P_2 \), we see that \( t = 1.15 \) which implies that the difference between \( P_1 = 0.40 \) and \( P_2 = 0.33 \), with sample sizes 95 and 175 respectively, is not statistically significant. At a later stage of this chapter we shall question the validity of the often repeated hypothesis that labour turnover is stronger among females than among males.

Turning to our empirical models we start the first model shown in Table (6.11) below. Apart from the constant term 'con' the variables included in this model are: change in rent plus rates, \( RR \); change in travel time, \( TT \); change in family earnings, \( EE \); Age in years, \( AGE \); level of education, \( ED \); receipt of non-labour income \( INC \); attainment of special training \( SPT \); the dummy for hours worked per week, \( HRS \); number of children, \( NCH \); and the mover descriptor \( MOV \). This time we are released from the previous problem of numerical non-convergence caused by the hours variable, \( HRS \), and we are able to get the MLE of its parameter coefficient. Note that the negative coefficient for this variable is strikingly significant as the zero effect hypothesis is attached with an extremely low significance level. This is, in fact, a central finding in this part of the study and we
### TABLE (6.11)

Results for Females

| Variable | TT used | | TT replaced by TT* |
|----------|---------|--------------------------------------------------|
|          | $\hat{\beta}$ | t (sig.) | $\hat{\beta}$ | t (sig.) |
| CON      | 4.140   | 2.855 (0.01) | 3.341 | 2.064 (0.05) |
| RR       | $-0.115$ | 2.543 (0.02) | $-0.118$ | 2.608 (0.01) |
| TT       | $-0.011$ | 0.414 (0.70) | $-0.020$ | 0.729 (0.50) |
| FE       | $-0.004$ | 0.169 (0.90) | $-0.004$ | 0.123 (0.90) |
| AGE      | $-0.020$ | 0.713 (0.50) | $-0.012$ | 0.473 (0.70) |
| ED       | 0.112   | 0.168 (0.90) | 0.011  | 0.016 (0.95) |
| INC      | $-0.526$ | 1.289 (0.20) | $-1.159$ | 1.870 (0.10) |
| SPT      | $-1.571$ | 1.866 (0.10) | $-1.579$ | 1.902 (0.10) |
| HRS      | $-3.630$ | 4.653 (0.001) | $-3.431$ | 4.321 (0.001) |
| NCH      | 0.058   | 0.149 (0.90) | - | - |
| CHD<5    | -       | - | 0.981 (0.084) | 1.395 (0.20) |
| MOV      | $-1.787$ | 1.695 (0.10) | $-1.781$ | 1.671 (0.10) |

$-2\log\lambda$ 59.480 61.003

$F^*$ 9.331 9.719
shall qualify on it later. As regards the other variables we see that
RR has got the expected (negative) sign and it is also highly signifi-
cant. The effect of SPT is negative and quite significant as expected
a priori but the education variable is now performing poorly. The MLE
of the age parameter coefficient has got the expected sign but it is
quite insignificant. It is also rather odd to have a negative coef-
ficient for the change in travel time variable, TT, which is after
all very insignificant. However, we find that the mover variable MOV is
now performing better than with males. Its parameter coefficient has
got a negative sign and it is reasonably significant. This implies that
the actual move into council housing has been associated with certain
effects which tended to disturb the past job attachments. On the other
hand, note that the non-labour income dummy, INC, has now got a negative
coefficient - as compared to a positive coefficients with males -
although in this particular model it is not very significant. The
effect of children number is positive as expected a priori but its statis-
tical significance is very low; similarly the effect of the family
earning variable, FE, is insignificant. These are the general remarks
which can be made on the basis of this model though the apparent anomalies
of the travel time variable, TT, and the non-labour income variable INC
need be considered as we shall attempt later. However, the model on the
whole gives a very good fit as judged by \(-2\log \lambda (= 59.59)\) and the \(F^*\)-ratio (= 9.33). The log-likelihood ratio is significant at the 0.0005
level and \(F^*\) is significant at the 0.001 level. The only difference
between this model and the second model in Table (6.10) above is the
replacement of the children number by the presence of children under
five years' dummy CH<5. This step has affected a slight improvement in
the general performance of the model - as judged by \(-2\log \lambda = 0.61.00,\nand \(F^* = 9.72. The significance of the special training variable SPT
has now improved and similarly for the non-labour income INC whose coefficient values have changed. The previously non-significant variables, TT, FE, AGE and ED remained still at poor states of significance. However, the important variables RR, HRS, SPT and MOV maintained quite stable coefficient parameter values. The newly introduced variable CH<5, for children below five years, has a better significance level than the children number NCH in the preceding model - but its significance is not very high.

To complete our cycle of variables testing, we run another model which introduces the housing tenure dummy HT, and the gross earnings before move, GE in place of CH<5 and ED. The results are shown in the first model of Table (6.12). This model only reassures us about the unimportance of the variables HT and GE. We still keep on getting reasonably stable parameter coefficients, specially for the significant variables: change in rent and rates, RR; non-labour income, INC; special training, SPT; the hours dummy HRS; and the mover descriptor MOV. We shall adopt the latter five variables as being the most important and delete all of the other variables. This leads us to our final model of Table (6.11) below. The MLE's of this model do not only support the relative stability of the model but they also produce marked reductions in the standard errors as apparent from the t-values. On the other hand, $-2\log \lambda = 58.68$ is highly significant at the 0.0005 significance level; and $F^* = 11.93$ is significant at 0.001. Thus the performance of the model is very good judging by these measures of fit.

Now we may look at the point-elasticities and the arc-elasticities as shown in Table (6.13). As expected we find the hours variable HRS having the largest elasticity ($= -0.565$) followed by the rent and rates variable RR ($= -0.343$). These (negative) elasticities measure the sensitivity of average quit probability to small changes in the means of
<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\beta}$ (elast.)</th>
<th>t (sig.)</th>
<th>$\hat{\beta}$ (elast.)</th>
<th>t (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>4.203</td>
<td>3.039</td>
<td>3.412</td>
<td>4.072</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>RR</td>
<td>-0.118</td>
<td>2.693</td>
<td>-0.123</td>
<td>3.073</td>
</tr>
<tr>
<td></td>
<td>(-0.322)</td>
<td></td>
<td>(0.343)</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>-0.006</td>
<td>0.237</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>-0.004</td>
<td>0.156</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-0.029</td>
<td>0.991</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.280)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.006</td>
<td>0.895</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>-1.329</td>
<td>1.594</td>
<td>-1.363</td>
<td>1.717</td>
</tr>
<tr>
<td></td>
<td>(-0.097)</td>
<td></td>
<td>(-0.102)</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>-1.331</td>
<td>1.663</td>
<td>-1.476</td>
<td>1.880</td>
</tr>
<tr>
<td></td>
<td>(-0.094)</td>
<td></td>
<td>(-0.106)</td>
<td></td>
</tr>
<tr>
<td>FT</td>
<td>-3.728</td>
<td>4.740</td>
<td>-3.543</td>
<td>4.902</td>
</tr>
<tr>
<td></td>
<td>(-0.579)</td>
<td></td>
<td>(-0.565)</td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>0.363</td>
<td>0.44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOV</td>
<td>1.608</td>
<td>1.492</td>
<td>1.813</td>
<td>1.790</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td></td>
<td>(-0.087)</td>
<td></td>
</tr>
<tr>
<td>-2logλ</td>
<td>60.465</td>
<td></td>
<td>58.676</td>
<td></td>
</tr>
<tr>
<td>$F^*$</td>
<td>9.687</td>
<td></td>
<td>11.935</td>
<td></td>
</tr>
</tbody>
</table>
the respective variables. The arc-elasticities on the other hand take account of finite reasonably large changes - here specified at 5%, 10%, 20%, 30% and 50%. As HRS has got the largest point elasticity its arc-elasticity is also the most markedly changing, followed by the arc-elasticity of RR. On the other hand, as the MOV variable has got the smallest point elasticity (= 0.087) it has also got the slowest changing arc-elasticity.

To sum up our findings for this section up to now, we have fitted several models of job-quit probability for the female spouses and ended up with a five variables model which seem to stand out as the most important. These are change in rent plus rates, RR; hours worked per week before move, HRS; special training for usual job, SPT; non-labour income received, INC; and the mover variable MOV. We had no problem with the inclusion of the hours variable HRS in the model as we had with the males' models. This variable HRS is rather outstanding and it deserves special treatment, as we shall attempt in the next section. However, the INC dummy has got a negative sign-counter to the a priori expectations. We have found no direct satisfactory way to explain this result on the basis of our data experimentation. However, we expect this non-labour income received to be of the Social Security type of payments which normally are paid to lower income families, e.g. child benefit, family income supplement, supplementary benefit, etc. Thus, in principle, those who receive such income are the more needy. In addition, if this income is not large enough to meet their financial obligations its mere receipt may only be indicative of need; and therefore non-tolerance of disrupted earnings due to quit.

We now move to re-consider the effect of the hours dummy variable
TABLE (6.13)

Arc Elasticities for the Female Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Point Elasticity</th>
<th>5%</th>
<th>Arc elasticities</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>-0.343</td>
<td>-0.343</td>
<td>-0.366</td>
<td>-0.391</td>
<td>-0.416</td>
<td>-0.469</td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>-0.102</td>
<td>-0.105</td>
<td>-0.108</td>
<td>-0.114</td>
<td>-0.120</td>
<td>-0.132</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>-0.106</td>
<td>-0.109</td>
<td>-0.112</td>
<td>-0.118</td>
<td>-0.124</td>
<td>-0.137</td>
<td></td>
</tr>
<tr>
<td>FT</td>
<td>-0.565</td>
<td>-0.587</td>
<td>-0.610</td>
<td>-0.659</td>
<td>-0.710</td>
<td>-0.823</td>
<td></td>
</tr>
<tr>
<td>MOV</td>
<td>0.087</td>
<td>0.090</td>
<td>0.092</td>
<td>0.097</td>
<td>0.102</td>
<td>0.111</td>
<td></td>
</tr>
</tbody>
</table>
This time we shall demonstrate its effect via a geometrical picture adopting the same principles as those of Figures (6.1), (6.2) and (6.3) below. In other words, we shall construct relative frequency histograms for the distribution of the estimated probabilities $\hat{\theta}_i$ and relate them to their corresponding $S_B$ curves. These considerations are dealt with in section (6.5) below.

6.6 Geometrical Representation of the Effect of the Hours Variable for the Female Spouses

This time the experiment is of a slightly different nature, but it is very simple. In the first case we shall drop the hours variable, HRS and in the second case we shall re-insert it. For each of the two cases we construct an empirical relative frequency histogram and an $S_B$ curve. Comparison of the $S_B$ curves under the two cases reveals diagrammatically the effect of the hours variable, HRS.

The outcome of the experiment is presented in Figures (6.4) and (6.5). The inclusion of the HRS is seen to make a marked structural change in the distribution of quit propensities as apparent in Figure (6.5). We can give yet a clearer picture if we plot the $S_B$ curves of Figures (6.4) and (6.5) up to scale in one graph, as it is shown in Figure (6.6). In this figure the effect of the HRS variable is quite sharp and decisive. Its related U-shaped curve simply implies that most female spouses have got quit propensities closer to zero or one. It is interesting to see the relevance of this finding with that of Heckman and Willis (69); see Chapter II, section (2.6). The latter authors have arrived at a similar U-shaped curve for the distribution of labour force participation probabilities for married females - which implied that most females have participation probabilities near zero or one.
Distribution of Quit Probabilities for female spouses when the hours dummy HRS is ignored.
Distribution of individual quit probabilities for females when the hours dummy is introduced.

**Figure (6.5)**
\[ S_B \text{ curve ignoring HRS variable.} \]

\[ S_B \text{ curve controlling for it.} \]

The contrast between these two curves signifies the effect of HRS.

Figure (6.6)
We have thought it appropriate to estimate two additional models for the cases with \( HRS = 1 \) and the cases with \( HRS = 0 \), respectively. The main aim of these two models is to construct empirical relative frequency histograms for the distribution of the estimated quit probabilities, \( \hat{\theta}_i \), and their associated \( S_B \) curves. As the sample sizes for these two groups are quite small (i.e. 53, and 42 for \( HRS = 1 \) and \( HRS = 0 \) respectively) we may not expect very interesting results from MLE's of the parameter coefficients; see the models in Table (6.14). Nevertheless the model of the \( HRS = 0 \) cases has surprisingly very good fit; specifically the income variable, INC, the AGE variable and the special training variables are quite significant. If we look at the other model with \( HRS = 1 \), we find a rather poor fit. However, the change in travel time variable, TT, now has got for the first time a positive coefficient as expected\(^\dagger\)—though its statistical significance is poor. The two models are used primarily to construct Figures (6.7) and (6.8) respectively. Each figure shows the relative frequency distribution of quit propensities for its corresponding sub-sample of cases, together with the \( S_B \) curves. The structures of the curves are easy to explain. In the case of the sub-sample \( HRS = 1 \) the sample proportion of quits is equal to 0.16 compared to 0.67 for the sub-sample \( HRS = 0 \). Thus, in the latter case, there is more concentration of quit propensities near one, (with a few exceptions near zero) while in the former case there is more concentration near zero.

Thus, if we plot the two \( S_B \) curves of Figures (6.7) and (6.8) in a single graph we should get a U-shaped curve as shown in Figure (6.9) below. The shape of this curve is highly comparable to that of Figure (6.6) above, and therefore the partitioning of the combined model into two models using \( HRS \) variable does explain the U-shaped phenomenon.

\(^\dagger\) Recall that we obtained a negative coefficient for TT before, which was, however, very insignificant. Now, the effect of TT is positive as expected but again insignificant. It seems that full-time females, if the sample is sufficiently large, could be found to respond in agreement with our a priori expectation /
Description of the distribution of individual quit probabilities for females with HRS = 1.

N = 53
Description of the distribution of individual quit probabilities for females with HRS = 0.

N = 42
The resultant $S_B$ curves for

(i) Females with HRS = 1

(ii) Females with HRS = 0
TABLE (6.14)

Two models for females; those who have worked more than 30 hours per week (HRS = 1), and those who have not.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Females with HRS = 0</th>
<th>Females with HRS = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$ (elast.)</td>
<td>$t$ (sig.)</td>
</tr>
<tr>
<td>CON</td>
<td>9.932 (0.05)</td>
<td>2.124 (0.05)</td>
</tr>
<tr>
<td>RR (-.112)</td>
<td>-0.084 (0.30)</td>
<td>1.070 (0.30)</td>
</tr>
<tr>
<td>TT (-.077)</td>
<td>-0.050 (0.30)</td>
<td>1.122 (0.30)</td>
</tr>
<tr>
<td>INC (-1.21)</td>
<td>-2.848 (0.10)</td>
<td>1.935 (0.10)</td>
</tr>
<tr>
<td>SPT (-.083)</td>
<td>-2.397 (0.15)</td>
<td>1.575 (0.15)</td>
</tr>
<tr>
<td>MOV (.040)</td>
<td>-2.109 (0.15)</td>
<td>1.587 (0.15)</td>
</tr>
<tr>
<td>CHD&lt;5 (-.045)</td>
<td>-0.581 (0.80)</td>
<td>0.325 (0.80)</td>
</tr>
</tbody>
</table>

$-2\log\Lambda$ 82.288 (N=42) 8.386 (N=53)

$F^*$ 41.275 1.620
More generally it can be shown that as the proportion of female spouses with HRS = 0 decreases this will result in depressing the right hand mode of the U-shaped curve and pushing up the left hand mode; similarly the converse can be established by symmetry. We shall demonstrate this point by increasing the proportion of HRS = 1 by 25% and then decreasing it by 25%. The geometrical description of this experiment is shown in Figures (6.10) and (6.11) below. The connected $S_B$ curve represents the original curve, while the disconnected curves represent the new curves after the change. The two pictures completely agree with our expectations.

Before concluding this section it is worthwhile exploring further the observation that female spouses with HRS = 1 has a much lower sample proportion of quit ($\approx 0.16$) as compared to male heads with HRS = 1 ($\approx 0.25$). This observation is contrary to the often repeated hypothesis that females have higher turnover rates than males. Nevertheless, Clowes$^{(30)}$ has observed in his modified organizational commitment model that:

"...when a comparison is made of labour turnover of males and females in the same factories it can be seen that there is no justification for the oft-repeated statement that labour turnover among women is higher than among men."

(Clowes$^{(30)}$, p. 252)

This observation is clearly compatible with our sample. Following our current procedure of geometrical representation we may illustrate this point by plotting the $S_B$ curve of male heads with HRS = 1 (Figure (6.2)) together with that of female spouses with HRS = 1 (Figure (6.7)). These two $S_B$ curves are also plotted up to scale as presented in Figure (6.12) below. See how flat the $S_B$ curve for male quits compared to that of

† (contd.)

expectation. The behaviour of part-time females needs more careful study on the basis of a larger sample, as we propose in our conclusions (Chapter VII below).
Original curve

the curve generated by a 25% fall in proportion of cases with HRS = 1.
The curve generated by a 25% increase in proportion of cases with $\text{HRS} = 1$.

Original curve.
Figure (6.12)
females - implying that for females with $HRS = 1$ quit probabilities are concentrated closer to zero than for males.

6.6.1 A Model for Female Actual Movers Alone

Before concluding this section it is desirable to test the last five-variables for the joint model on actual female movers alone. As the joint sample size is already not large enough ($= 95$), the exclusion of intending movers will reduce to 79. The smallness of the sample, as we know, will make the MLE's more open to the possibility of bias since the MLE's are only asymptotically efficient. Yet, a sample size of 79 is not really too small for the maximum likelihood procedure; (see Chapter III, p. 91). We may look at Table (6.15) below where we present the model for movers alone. This model incorporates the most important four variables $RR, INC, SPT, HRS$, together with the employment opportunities variable, $EMP$. Notice that the latter variable, $EMP$, is a reasonably significant variable and its coefficient parameter has got the expected (positive) sign. The rest of the variables have their expected signs but it appears that the standard errors of estimates are relatively high as judged by the low levels of the $t$-values. The rent and rates variable, $RR$, however, is the most significant variable. The model as a whole provides a good fit as judged by $-2\log = 67.90$ and $F^* = 13.30$.

6.7 Summary

We have reviewed in this chapter some empirical models of job-quit probability for male heads and female spouses - who have been employed prior to the move into council housing. We have defined and discussed
### TABLE (6.15)

**Female Actual Movers Model**

allowing for the employer variable (EMP).

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\beta}$ (elast.)</th>
<th>t (sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>1.030</td>
<td>0.982 (0.40)</td>
</tr>
<tr>
<td>RR</td>
<td>-0.149 (-0.240)</td>
<td>2.630 (0.01)</td>
</tr>
<tr>
<td>INC</td>
<td>-0.863 (-0.040)</td>
<td>1.636 (0.15)</td>
</tr>
<tr>
<td>SPT</td>
<td>-1.635 (-0.070)</td>
<td>1.386 (0.20)</td>
</tr>
<tr>
<td>FT</td>
<td>-1.129 (-0.106)</td>
<td>1.150 (0.30)</td>
</tr>
<tr>
<td>EMP</td>
<td>0.007 (0.405)</td>
<td>1.853 (0.10)</td>
</tr>
</tbody>
</table>

- $-2 \log \lambda = 67.896$ (N = 79)
- $F^* = 13.330$
the *a priori* expectations for the various explanatory variables which could be incorporated in the models. However, we faced a numerical problem which meant that we could not include the hours dummy, HRS, in a model of male quits. For this reason we have run our empirical models for the combined sample (i.e. where HRS =1, or 0) and then separately for the sub-sample with HRS = 1. In both cases we have obtained highly significant and meaningful results for the effects of: change in travel time, TT; change in rent and rates, RR; AGE and education level, ED. Moreover, in the combined model we have obtained a highly significant result for change in family earnings, FE. In the latter model we have an interesting result of the effect of employment opportunities in the new area, EMP. This variable has also had a highly significant (positive) coefficient parameter, implying that the larger the number of employment opportunities around the new locality, the more likely a worker will quit his previous job. As regards the effect of the HRS dummy, we have demonstrated its effect geometrically. This has been done by contrasting two fitted $S_B$ curves, one relating to the combined sample and the other relating to the sub-sample HRS = 1. This procedure revealed that in the latter case, quit probabilities are concentrated more closely near zero than in the former case.

The female spouses' models have shown that the most influential variables are: the rent and rates variable, RR; the hours variable, HRS; the special training variable, SPT; the mover variable MOV; and the non-labour income dummy, INC. The hours dummy HRS has been outstandingly significant. We have shown that the inclusion of this variable accounts for a U-shaped distribution of the distribution of job quit probabilities - implying that most female spouses have
quit propensities near zero or one. We have also drawn attention to the observation that female spouses with \( \text{HRS} = 1 \) have their quit propensities more closely concentrated near zero than male heads.
CHAPTER VII

A GENERAL REVIEW

(7.1) Introduction

The empirical analysis of job-quit probability as a function of several independent variables has been made in this study with special reference to the Glasgow labour market experience. The Glasgow Housing Corporation is currently engaged in the relocation of workers from what are regarded as condemned properties. However, as has been observed by Engelman (49) and others, this policy of relocation to council housing may adversely affect the labour market experience of working class families as long as explicit consideration is not made to the sensitivity of their job attachments to such residential shifting. The empirical findings of this study have originally arisen in an attempt to extend and improve on Engelman's (49) as regards the choice of a probability model, and choice of explanatory variables. At the same time we discuss the applicability of our disaggregate logistic model of quit behaviour to derive predictions related to changes in average quit probability (per unit time) of a given labour force, or turnover movements at the individual firm level. Thus, we attempt to improve on the current techniques of turnover analysis with special emphasis on voluntary terminations - although the elaborate extension to other forms of separations (e.g. layoffs and dismissals) should be straightforward. The data base of this study is a cluster random sample of households who have been assigned council houses in different parts of the city, some of whom have actually moved (actual movers) while another group - used as a control - have not moved at the time of the survey (intending movers).
The aim of this short chapter is to provide an overall view of our empirical findings, the associated policy implications or potentials of future research which can be drawn. Limitations in data, especially lack of data on job characteristics means that only some of the variables relevant for policy purposes can be analysed empirically. These points are outlined in the appropriate sections. We start with the basic problems of model choice and formulation, move to comment on the empirical findings and their implications, and finally outline the technical methodology which we have adopted and developed in the course of this study.

(7.2) Choice of Probability Model

The choice of the binary logistic model has been made essentially for its computational convenience as it automatically obeys the \(|0,1|\) restriction of the probability interval, thus redressing the limitations of the linear probability model of prediction bias and heteroscedasticity. Past experience has shown that the logistic model enjoys curvature properties which make it computationally more tractable than other non-linear probability models (e.g. Probit, Arctan, Gombit) although when compared to the logit, these models are virtually indistinguishable empirically.

At the same stroke we have considered the behavioural justification of the logistic model which is sometimes based on the random utility theory of choice. The latter theory is broadly outlined and the derivation of the logistic model mode on the basis of a binary choice situation. However, the applicability of the latter to quit probability requires certain restrictive assumptions which cannot be met in practice, e.g. that alternative offers be observed. On the other hand, depending on the distribution of ideocyncratic term - which is unknown - we get different mathematical probability models (e.g. logit, probit, etc.). Thus we have alternatively appealed to the statistical appropriateness and computational convenience of
the model. We have argued for the inclusion of non-pecuniary benefits, though it has not been possible to include them in the analysis.

(7.3) The Household Effect Problem

In the first place the clustered nature of the sample has drawn our attention to the problem of formulating the probability model in a form that allows for the conceptually interdependent responses of household members. It appears, from a selective review of the relevant literature, that the problem of the interdependent quantal responses of household members has not been adequately formulated. We have shown how the simultaneous logit model can be adapted to satisfy such problems, e.g. the similar problems of binary response, namely job-quits and labour force participation decisions. We have drawn some analogy between the simultaneous logit model and the simultaneous equations linear probability model which Ashenfelter and Heckman have adopted to represent labour force participation responses within a husband/wife household. The latter model is, however, potentially sensitive to the violation of the $[0,1]$ restriction of the probability interval, and to inefficient estimates due to the heteroscedasticity problem. It has been interesting to notice that both models satisfy a symmetry restriction of the interdependence measure in the husband/wife model of binary response - effect of husband on wife is equal to effect of wife on husband. The simultaneous logit satisfies this restriction automatically, while the Ashenfelter -Heckman (9) model imposes it by appealing to the Slutsky equation of income-compensated wage effects.

However, the effective utilization of the simultaneous logit model to formulate our quit behaviour problem on the basis of the household unit has been rendered impossible due to data shortage - as it becomes important to allow for changes in the labour force participation responses
of different household members. Our limited data has only allowed us to
test the sign of the interdependence parameter for the husband/wife labour
force participation model. In agreement with Ashenfelter and Heckman\(^9\)
Kniesner\(^82\) and Daniel\(^38\) our empirical test supports the complementarity
hypothesis of husband/wife participation decisions. The effect of the
actual move of families does not seem to disturb this trend.

(7.4) **Empirical Finding and Implications**

As a compromise for our inability to include all household members in
a single model due to the data limitation, and for lack of a priori evi-
dence of a zero "within family" dependence, we have chosen one male
member (i.e. head) and one female member (i.e. spouse) from each house-
hold. Then we have utilized the resultant two groups as two simple
random samples of males and females who have been employed prior to
the move. Thus, we specify the binary logistic model to estimate job-
quit probability as a function of several explanatory variables,
separately for males and for females. Parameter estimation and signi-
ficance testing are performed on the basis of maximum likelihood method,
while goodness-of-fit is measured through the likelihood ratio test together
with our \(F^*\)-ratio test which we have introduced by analogy with the con-
ventional Analysis of Variance set-up.

The results for both males and females indicate that the pre-move
weekly hours of work dummy (HRS) is a highly important determinant of
quit probability. As expected those who used to work less than 30
hours per week (i.e. part-time) are significantly more vulnerable to
quit than those who used to work more (full-time). It also seems that
the former class of workers are the less likely to be re-employed than
the latter class. Hence the possible adverse effects of relocation could
be more serious to the employment statuses of these relatively less
committed employees, especially females. Thus it should be desirable that the vulnerability of job attachments of this class of workers be explicitly allowed for. We particularly propose that the incidence of the move on this class of workers be examined more carefully in future research.

We have also established that the increase in travel-time due to the move has had a highly significant positive effect on the quit-probability of males, but the effect for females has not been significant. However, increases in rent and rates has had a significant negative effect on both male and female quits - implying that additional financial obligations in terms of housing costs tend to suppress the quit decisions of employees as the financial costs of a possible unemployment duration is less tolerable. This being the case, the state of job-satisfaction of these potentially suppressed quits due to the costly move, has to be examined in a future research.

Another interesting effect is provided by the number of employment opportunities within a given radius of the new location of actual movers. This variable, though crude, has provided a significant positive effect on the quit probability of the males' combined model and a reasonably positive effect for females as well. Thus the new locality or conurbation to which a worker is rehoused can allow the enforced residential mobility to release the job-mobility of those whose previous jobs become no longer tolerable - by providing new potential employment offers. This effect seems consistent with Mackay et al. (92) who established that the conurbations of the Glasgow local labour market constitute a series of loosely connected sub-markets due to the costly process of residential mobility and the costs of a long journey-to-work distance, so that the burden of unemployment may not be evenly distributed within the city. Thus we may derive the implication that those who have undergone considerable increase

† The model which combines males who worked less than 30 hours/week before move, and those who worked more.
in travel-to-work time due to the move, and moreover have been rehoused in an unemployment-prone conurbation should be doubly at a disadvantage. Given a considerable increase in travel-to-work time, these people tend to possess suppressed quit decisions when compared to those who have been compensated with a more favourable conurbation. Hence for policy considerations, if under special restrictive circumstances a worker has to undergo a substantial residential shift, then at least he should have the benefit of an employment favourable conurbation. However, it is also important to understand the resource constraints of council housing specially as regards their spatial scatter within the city in relation to the employment opportunities - which we propose for future research. As regards the effect of the actual move per se, it does not appear within the capacity of our data to have a strong positive effect on quit probability. To a reasonable significance level the female actual movers are more likely to quit than intending movers but the effect for males is less significant. It seems that the notification about the move and assignment of a new house takes effect in anticipation of the actual move, especially for males. Therefore, we propose that any potential help given to movers should be made well before the move actually takes place.

Moreover, there is a highly significant evidence that the quit propensity of some males tends to decrease with increase in the education level. A similar negative effect has been found for training (e.g. apprenticeship) on the females. It is worth recalling that the negative effect of increase in skill level on 'plant quarterly quit-rates' has been established by Mackay et al.\(^\text{(92)}\) for the 'tight' labour market of Birmingham, but not for the 'slack' labour market of Glasgow. However, they have also found that the journey-to-work distance is more restrictive to the geographical mobility of the less skilled workers than the more
skilled in the Glasgow labour market and Birmingham. Hence, our finding that the more qualified have a lower quit probability may partly reflect the differential impact of a given residential shifting which may be more important to the less qualified. On this basis it is recommended that special care be taken to the sensitivity of job-attachments of unskilled manual workers to substantial residential relocation.

The negative effect of age on job-quit probability appears to be highly significant for males but the effect for females has been rather poor. It has also been possible to test the effect of non-labour income received, which is described in our data by a crude dummy variable. This variable has produced a positive effect on male quits at a reasonable significance level, but a negative effect on females. The positive effect finding supports the intuitively appealing hypothesis that receipt of non-labour income makes out-of-the-job search more financially tolerable than the non-receipt. This hypothesis is usually adopted in job-search models, but the negative effect obtained for females seems odd and needs to be explained. One possible explanation is that, as this non-labour income is most likely received in a Social Security form, its receipt is indicative of financial stringency in lower income families. Hence, if such income is not received in sufficiently large amounts, its effect could be negative on the leaving propensity. However, while this is only a tentative explanation, we hope for a more rigorous empirical study to examine this effect in a future research. In particular it is interesting to examine and test the possibility of opposite response of male heads and married females - in the way we have got - to this variable. In addition to the above
variables we have tested the effects of children, gross earnings of the pre-move job, and housing tenure but our data have not captured them as significant.

(7.5) The Methodology of Turnover Analysis and its Policy-Orientation

In this study we have considered some interesting mathematical models of labour turnover which are based on the demographic structure of the life-table, like Silcock\(^{(128)}\) and Clowes\(^{(30)}\). Specifically the latter proposes a numerical policy-oriented model of turnover analysis which prescribes means for shifting the 'survival curve' of employees in a given firm upwards, thus reducing turnover. We have argued, however, that the curve to be shifted should be made responsive to the relevant employees' characteristics which are the object of the "screening/training policy" (using Clowes\(^{(30)}\) terminology). Thus we have asked the question: how does the expected number or proportion of employees leaving the organization respond to small (or finite) changes in the mean characteristics of the labour force? We have based our proposed answer on the probability distribution of employees' quit propensities at a given point of time. Hence, the appropriate curve which we have proposed as comparable to the survival or stability curves, is the one which describes the probability distribution of employees' quit propensities. An interesting analogy can be made with Mackay et al.\(^{(92)}\) where ordinary histograms for the frequency distribution of plant average quarterly quit-rates have been constructed and compared to reflect the effect of skill levels on the 'market quit-rate' of Glasgow and Birmingham. Similarly we construct histograms of quit-probabilities except that the latter are functions of several explanatory variables. Thus the shape of the histogram is determined by these explanatory variables. Then we fit a curve to this histogram which is sufficiently flexible and responsive
to changes in the underlying explanatory variables. We have found that the multivariate Normality assumption of the variables, and the logistic law for quit propensities are sufficient to derive the appropriate curve, which turns out to be Johnson's $S_B$ curve. This curve can provide two useful descriptive purposes: first it may describe the structure of quit propensities within a given labour force; secondly it can be used as a policy tool to help assess the potential sensitivity of this structure to possible changes in the relevant variables.

Thus, we have considered the measurement of elasticity of average quit probability to small or finite changes in the means of the explanatory variables, basing the analysis in terms of the $S_B$ curve. However as the exact point elasticity formula requires numerical integration, which is rather costly and complex, we have seen that a comparable formula directly based on the actual distribution of estimated probabilities provides a good approximation to the exact formula. In particular, the computational difficulties of arc elasticities has been substantially reduced in this study, as compared to Richard and Ben-Akiva (121) and Westin (148). We have exploited a certain mathematical property of the $S_B$ curve (RESULT (5.1); Chapter V) and derived an approximate computational formula to predict average binary response probability (i.e. quit) as a result of a given finite change in a variable's mean.

Interestingly this formula, which requires no numerical integration, has provided a good agreement with the exact, but rather complex, recipe which is based on the transformed $S_B$ curve.
The above outlined methodology has in fact been applied to our empirical models of males and females. The general impression which we have got is that the point elasticity estimates are generally low (less than unitary). Arc elasticities, which have been computed up to certain percentage changes of the explanatory variable, vary only slightly specially when point elasticities are very low. Moreover, we have constructed histograms of the individual quit probability estimates for the interesting models and fitted the smooth $S_B$ curves on these histograms to provide a simplified picture of them. In particular the effect of the pre-move weekly work-hours dummy (HRS) has been reflected very clearly in the shifting of the $S_B$ curve upwards and to the left for the males' models - implying that quit propensities would be concentrated closer to zero, when the work-rate increases.

The finding for females has been specially interesting. The allowance for the hours dummy (HRS) has resulted in a U-shaped distribution of their job-quit probabilities, which implies that most females have quit probabilities near zero or near one. Specifically this means that married females who work less than 30 hours a week are an average quit propensity closer to one, while those who work more per week have an average quit probability closer to zero. It is also interesting to recall that this U-shaped property has been established by Heckman and Willis\(^{(69)}\) for the distribution of labour force participation probabilities of married females in U.S.A., using the Beta curve.

Moreover, we have utilized the $S_B$ curves to compare the job-quit probabilities of full-time male workers with those of females. The females curve seem to be highly skewed to the right relative to that of the males - implying that the stability of the females job attachments
has been stronger for females than for males. This finding is similar to that of Clowes (30), and it should be examined more carefully in future research. It is also important to recognise that a full/part-time proxy cannot be simply interpreted as representing organizational commitment. Thus, it is another future research problem to examine the economic and sociological background of these two workers groups.

(7.7) Aggregation Bias and the Variance-Elasticity Formula of Binary Response Prediction

In this study we have examined the prediction bias which results from the adoption of the logistic 'naive formula'† of average binary response probability, as has been adopted by Viscusi (144) and Mefoff (100) to estimate average job-quit probability. We have established that if the linear combination of the explanatory variables' means is non-zero, then there is bound to be a prediction bias in the naive formula. Specifically, this bias is measurable directly by a simple function of the variables' covariance matrix, or the extent of individuals' heterogeneity.

This point has led us to consider the degree of responsiveness of average quit probability to small changes in the extent of employees' heterogeneity, when their characteristics means (or their linear combination) are fixed - i.e. the $\sigma$-elasticity of average quit-probability. This measure has originally been proposed by Westin (148) for aggregate binary prediction analysis on the basis of the $S_3$ curve. However, we have discovered that Westin's formula does not satisfy certain important criteria and therefore it should be incorrect. Alternatively, we have supplied what we believe is the correct formula and worked out its limiting properties.

† This formula is evaluated at the averages of the explanatory variables, while ignoring their variance-covariance structure.
(7.8) **AND FINALLY**

We hope that we have been able to shed some light on this problem despite the data shortage which has restricted the empirical analysis. In particular it could have been desirable to adopt a multinomial set-up for the responses of workers whereby an individual could be observed as working with the same employer, with a different employer, unemployed or not in the labour force. This may require a multinomial extension of the binary logit, and would certainly require a considerably large sample. It will also be interesting to test the household simultaneous choice model more fully and at a greater depth on the basis of richer data. We hope that we should be able to handle these problems in a future research and to introduce more extensions and refinements.
APPENDIX (A)

MATHEMATICAL NOTES
APPENDIX (A)

CHAPTER 4; SECTION (4.3)

Proof for the sign properties of $\psi$.

Given two dichotomous variables $Y, Z$ ($i = 1, ..., n$) and that $\theta_{ij} = \text{Prob}(Y=i, Z=j)$ $i,j = 0, 1$ whereas $\theta_{i.}$ and $\theta_{.j}$ are the marginal probabilities $\text{Prob}(Y=i)$ and $\text{Prob}(Z=j)$ respectively.

The independence (null) hypothesis can be expressed as

$$H_0 = \theta_{ij} = \theta_{i.} \theta_{.j}, \quad i,j = 0,1 \quad (A.1)$$

In fact although (A.1) implies that there are four restrictions, it is easy to check that they are mutually equivalent. For this reason the above hypothesis is tested against a chi-squared variable with a single degree of freedom; see Rao (118; p.403). Hence we can write (A.1) equivalently as a single restriction, as below:

$$H_0 = \theta_{11} = \theta_{1.} \theta_{.1} \quad (A.2)$$

where $\theta_{1.} = \theta_{10} + \theta_{11}$, and $\theta_{.1} = \theta_{01} + \theta_{11}$. Now, if we substitute for $\theta_{1.}$ and $\theta_{.1}$ in (A.2), multiply out and simplify, we find that

$$\theta_{11} = \theta_{1.} \theta_{.1} \iff \theta_{11} \theta_{00} = \theta_{10} \theta_{01}.$$

Hence, for independence we must have $\frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} = 1$ or equivalently

$$\psi = \log \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} = 0.$$

If $Y$ and $Z$ are mutually exclusive, or perfect substitutes, then we have $\theta_{11} = 0$. More generally, if $Y$ and $Z$ are substitutable, then the hypothesis is expressed as $\theta_{11} < \theta_{1.} \theta_{.1}$ which implies that
\[ \theta_{11} \theta_{00} < \theta_{10} \theta_{01} \]

OR, equivalently \[ \psi = \log \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} < 0 . \]

Similarly, if \( Y \) and \( Z \) are complementary then in the perfect case we have \( \theta_{11} = 1 \), and more generally \( \theta_{11} > \theta_{1*} \) implying that

\[ \theta_{11} \theta_{00} > \theta_{10} \theta_{01} \]

OR, equivalently \[ \psi = \log \frac{\theta_{11} \theta_{00}}{\theta_{10} \theta_{01}} > 0 . \]

Hence, relations (4.16) are verified.

CHAPTER 4, SECTION (4.4)

Considering the log-likelihood function:

\[ \chi = \log L = \sum_{i \in B} \beta x_i + \sum_{i \in C} \gamma s_i + n \psi - \sum_{i=1}^{n} \log \Delta_i \]  

(A.3)

where \( \Delta_i = \left(1 + \exp(\beta x_i) + \exp(\gamma s_i) + \exp(\beta' x_i + \gamma s_i + \psi)\right)^{-1} \),

\[ i = 1, \ldots, n. \]

see equation (4.28).

We may derive expressions for its first-order partial derivatives, and second-order partial derivatives which can be used to evaluate the maximum likelihood estimates numerically. The inverse of the second-order partial derivatives matrix (with the opposite sign) is an asymptotically efficient variance-covariance matrix for the MLE's.

We can re-write (A.3) as
\[ l = \sum_{s=1}^{p} t_s(1) \beta_s + \sum_{s=1}^{q} t_s(2) \gamma_s + n_1 \psi - \sum_{i=1}^{n} \log \Delta_i \]

where \( t_s(1) = \sum_{i \in B} x_{il} \), and \( t_s(2) = \sum_{i \in C} s_{il} \) \( (s = 0,1,\ldots, p) \).

Then the elements of the first order partial derivatives vector can be expressed as:

\[
\frac{\partial l}{\partial \beta_s} = t_s(1) - \sum_{i=1}^{n} \begin{pmatrix} \frac{\beta x_i}{e^{x_i}} + \frac{\beta x_{i+1}}{\Delta_i} \end{pmatrix} x_{is}
\]

\[
\frac{\partial l}{\partial \gamma_k} = t_s(2) - \sum_{i=1}^{n} \begin{pmatrix} \gamma s_i \frac{\beta x_i}{e^{x_i}} + \frac{\beta x_{i+1}}{\Delta_i} \end{pmatrix} s_{il}
\]

\[
\frac{\partial l}{\partial \psi} = n_1 - \sum_{i=1}^{n} \begin{pmatrix} \frac{\beta x_i + \gamma s_i}{\Delta_i} \end{pmatrix}
\]

\( (s, l = 0, \ldots, p) \).

The elements of the second-order derivatives matrix can be expressed as follows:

\[
\frac{\partial^2 l}{\partial \beta_k \partial \beta_k} = \sum_{i=1}^{n} \begin{pmatrix} \frac{\beta x_i + \gamma s_i}{\Delta_i^2} \end{pmatrix} x_{ik} x_{ik}
\]

\[
\frac{\partial^2 l}{\partial \gamma_k \partial \gamma_k} = \sum_{i=1}^{n} \begin{pmatrix} \frac{\gamma s_i + \beta x_i + \gamma s_i + \psi}{\Delta_i^2} \end{pmatrix} s_{ik} s_{ik}
\]
(i) The first method, the non-Bayesian, is based on the log-likelihood function:
\[ \log L = \sum_{j=0}^{1} \sum_{i=0}^{1} n_{ij} \log \theta_{ij} . \]

Then, McLaren's method is to use a quadratic approximation for the log-likelihood in the logistic transform \( h(\theta) = \log|\theta/(1-\theta)| \). Recall that it is possible to write
\[ \psi = \log \frac{\theta_{11}\theta_{00}}{\theta_{10}\theta_{01}} \]
\[ = \log(\theta_{00}/(1 - \theta_{0}(0))) - \log(\theta_{00}/(1 - \theta_{0}(1))) \]
\[ = h(\theta_{0}(0)) - h(\theta_{0}(1)) \equiv \]

\[ \frac{\partial^{2} \Delta_{i}}{\partial \psi^{2}} = -n \sum_{i=1}^{n} \left( \frac{\beta x_{i} + \gamma y_{i} + \psi}{(1 + e^{\beta x_{i} + \gamma y_{i}})} \right)^{2} \]
\[ \frac{\partial^{2} \Delta_{i}}{\partial \beta \partial \gamma_{k}} = -n \sum_{i=1}^{n} \left( \frac{(\beta x_{i} + \gamma y_{i} + \psi)(\gamma y_{i} - 1)}{(1 + e^{\beta x_{i} + \gamma y_{i}})} \right) \]
Then the 95% confidence interval, using the normality approximation can be written as

\[ h'h \pm 1.96 \sigma \psi \]

where \( b' = (1, -1) \) and \( h' = (h(\theta_0(0)), h(\theta_0(1))) \)

and \( \sigma \psi = \sqrt{b' V^{-1} b} \).

The (diagonal) information matrix, \( V \), is a 2 \times 2 matrix with elements

\[ \frac{\partial^2}{\partial h^2} \log L. \]

Thus:

\[ \frac{\partial^2}{\partial h^2} \log L = \left( \frac{\partial^2}{\partial \theta^2} \log L \right) \left( \frac{\partial \theta}{\partial h} \right)^2 + \left( \frac{\partial \theta}{\partial h} \right) \frac{\partial^2 \theta}{\partial h^2} \]

When evaluated at \( \theta = \hat{\theta} \) the second term vanishes since \( \frac{\partial}{\partial \theta} \log L = 0 \) at the MLE. So we get

\[ V_h = n \hat{\theta}^{-1} (1 - \hat{\theta})^{-1} \{\hat{\theta}(1 - \hat{\theta})\}^2 = n \hat{\theta}(1 - \hat{\theta}) \]

recalling that \( \partial h(\theta)/\partial \theta = \frac{1}{\theta(1 - \theta)} \). Hence \(^{\dagger}\)

\[ V^{-1} = \begin{pmatrix} n_0 \hat{\theta}_0(0)(1 - \hat{\theta}_0(0)) & 0 \\ 0 & n_0 \hat{\theta}_0(1)(1 - \hat{\theta}_0(1)) \end{pmatrix}^{-1} \]

\[ = \begin{pmatrix} \frac{1}{n_0 \hat{\theta}_0(0)(1 - \hat{\theta}_0(0))} & 0 \\ 0 & \frac{1}{n_0 \hat{\theta}_0(1)(1 - \hat{\theta}_0(1))} \end{pmatrix} \]

is the asymptotically efficient covariance matrix. Hence, multiplying

\(^{\dagger}\) Note that the off-diagonal terms of \( V \) are zeros since

\[ \frac{\partial^2}{\partial h_1 \partial h_2} \log L = 0, \]

where \( h_1 = h(\theta(0)0), \ h_2 = h(\theta(1)0). \)
out $b'v^{-1}b$, recalling that $b' = (1, -1)$, we get

$$
\sigma_\psi^2 = \frac{1}{\theta_0(0)^{-1}} + \frac{1}{\theta_0(0)^{n_0}} + \frac{1}{\theta_0(0)^{n_0 \cdot 1}} + \frac{1}{\theta_0(1)^{-1}} + \frac{1}{\theta_0(1)^{n_1}} + \frac{1}{\theta_0(1)^{n_1 \cdot 1}}
$$

as required.

By symmetry if we adopt the other equivalent definition

$$
\psi = \log(\theta_0(0)/(1 - \theta_0(0))) - \log(\theta_0(1)/(1 - \theta_0(1)))
$$

we get the corresponding formula:

$$
\sigma_\psi^2 = \frac{1}{\theta_0(0)^{-1}} + \frac{1}{\theta_0(0)^{n_0}} + \frac{1}{\theta_0(0)^{n_0 \cdot 1}} + \frac{1}{\theta_0(1)^{-1}} + \frac{1}{\theta_0(1)^{n_1}} + \frac{1}{\theta_0(1)^{n_1 \cdot 1}}
$$

(ii) The second is the Bayesian method due to Lindley. This method is based on a theorem of the Dirichlet distribution

$$
g(\theta) = \prod_{i,j} \theta_{ij}^{b_{ij}}
$$

that is, if $\sum_{i,j} c_{ij} = 0$, then $\sum_{i,j} c_{ij} \log \theta_{ij}$ has the distribution of

$$
\sum_{i,j} c_{ij} \log \chi^2(2 b_{ij})
$$

Then recalling that the posterior distribution is defined as

$$
\pi(\theta; n) = \prod_{i,j} \theta_{ij}^{a_{ij} + n_{ij} - 1}
$$

(A.4)

we utilize the above theorem to derive the marginal posterior distribution of

$$
\psi = \log \theta_{00} - \log \theta_{01} - \log \theta_{10} + \log \theta_{11}
$$
Now, Great Prior Uncertainty (GPU) is presented by taking $a_{ij}$ to be very small; in this case $a_{ij} = 0$ (all $i,j$).

The foregoing result (A.4) implies that $\psi$ has the marginal posterior distribution:

$$\log \chi^2(2n_{00}) - \log \chi^2(2n_{01}) - \log \chi^2(2n_{10}) + \log \chi^2(2n_{11}) \qquad (A.5)$$

Note that if $V$ is large enough then $\log \chi^2(V)$ can have the approximate normal distribution

$$\log \chi^2(V) \longrightarrow N(\log(V-1), \frac{2}{V}) .$$

Hence we can approximate the posterior mean of $\psi$ by

$$\log(2n_{00} - 1) - \log(2n_{01} - 1) - \log(2n_{10} - 1) + \log(2n_{11} - 1)$$

and similarly the posterior variance will be

$$\sigma^2_{\psi} = \sum \frac{1}{n_{ij}}$$

as adopted.

**CHAPTER (5): SECTION (5.5)**

**Expression for $E(\theta)$**

As we have remarked the first moment, $E(\theta)$, of the $S_B$ curve cannot be expressed in a simple closed form. The analytical expression given below is based on infinite series (Johnson, p.173).

Define the following terms:

$$a = \delta^{-1} \sum_{n=1}^{\infty} e^{-n^2/2\delta^2}$$

$$b = \cosh \left[ \frac{n(n-2 \gamma \delta)}{2\delta^2} \right]$$
\[ c = \text{sech}\left(\frac{n}{2\delta^2}\right) \]

\[ d = 2\pi\delta \sum_{n=1}^{\infty} e^{-\frac{1}{4}(2n-1)^2\pi^2\delta^2} \]

\[ e = \sin|(2n-1)\pi\gamma\delta| \]

\[ f = \text{cosech}|(2n-1)^2\pi^2\delta^2| \]

\[ g = 2 \sum_{n=1}^{\infty} e^{-n^2\pi^2\delta^2} \]

\[ h = \cos(2n\pi\gamma\delta). \]

Then,

\[ E(\theta) = \frac{\delta^{-1} + abc - def}{1 + gh} \]

(see Johnson (77: p. 173)).
APPENDIX (B) FORTRAN PROGRAMS I & II

***************
THE FIRST PROGRAM BELOW IS THE BASIC PROGRAM WHICH HAS BEEN
USED DURING THE STUDY
PROGRAM(I)
THIS PROGRAM CALCULATES THE M.L.E.'S OF A BINARY LOGISTIC PROB.

MODEL.
IT CALCULATES THE ASSOCIATED T-VALUES, ELASTICITIES AND MEASURES
OF GOODNESS OF FIT
IT CAN ALSO BE USED TO CONSTRUCT THE FREQUENCY DISTRIBUTION OF
ESTIMATED INDIVIDUAL PROBABILITIES AND THE CORRESPONDING JOHNSON
S.J FREQUENCY CURVE
IT WRITES ESTIMATED POINTS OF THE FREQUENCY CURVES IN ANOTHER FILE
WHICH IS LATER EXPLOITED BY PROGRAM.II. TO GET GRAPH PLOTS
THIS PROGRAM IS WRITTEN AS THAT IT CONTAINS 10 VARIABLES
BUT IT IS STRAIGHT FORWARD TO CHANGE THE NUMBER OF VARIABLES
PROVIDED THAT THEY ARE NOT TOO MANY (<15)

THE ARRAYS UTILIZED IN THE ANALYSIS ARE DECLARED BELOW

ALL REAL QUANTITIES ARE IN DOUBLE PRECISION
REAL*8 B(10),AT(10),FREQ(10),DERV1(10),DERV2(10,10),AVR(180,10)
*,AT(180),WPCE(25),SUMV1(10),VVE(10,10),STEP(10),FREQ(10),STUD
* (10),ARC(5),CHNG(10),AVX(10),CP(10,10),SERV1(10),COV1(10,10),DF
* (19),CORR(10,10),COV2(10,10),EPS(10),DETR(10,10),ELS(10),P(21),
*BY(41),BIT1(41),BIT2(41),ESP1(10),Q(41),WXP(10),BIT3(10),CY(41),
*BY(41),BY1(21),BY2(41),T(21),R(21),EY(21),EY(21),DY(157),RY(157)
*,VY(157),RATIO(10)
REAL*8 SUMD,WT,SSP,ESP,LIKE,SCALAR,CON1,CON2,TSSG,FIT,PROB,
*RSSG,ROD,PROP,AVP,SUP,TERM1,TERM2,AVVP,GP,EXPNT,AVG,STD,STD2,
*DEM,TVVP,TVVP,EXIT,LIKER,DOM,D,CON1,CON2,LOGR1,LOGR2,
*LOGR2,DP,REMEN,EVP,LIKER1,AVQ1,UX,TRZ,XPW,SM,CNS,5,CN4,
*SUN,TVP1,TVP2,AV,VP,TEST
INTEGER KVR(46),N1(21),M1(21)

FORMATS ARE USED TO WRITE DOWN INTERMEDIATE VALUES IN ADDITION
TO THE FINAL RESULTS
57 FORMAT('ITERATION NUMBER',I4)
01 FORMAT('EXPECTED(V); ACTUAL(V)=',1X,10F10.5)
02 FORMAT('DATA(V)=',1X,10F10.5)
03 FORMAT('TOTAL VARIATION; WITHIN CASE; BETWEEN CASES=',1X,10F10.5)
04 FORMAT('F-RATIO =',1X,4F10.5)
55 FORMAT('VARICANCE COVARIANCE MATRIX=',/(1X,10F9.3))
77 FORMAT(19(8F10.5),5F10.5)
66 FORMAT(' STATISTIC =',1X,10F10.5)
99 FORMAT('INVERSE OF MATRIX OF SEC. ORD. DERIV. /(2X,10F9.3))
99 FORMAT(5(8F10.5),F10.5)
98 FORMAT('VALUE OF DERIVED IEG.=; EEPI FOR JOINT EFFECTS=',1X,
*10F8.4)
98 FORMAT('EEPI & ERROR =',1X,10F10.5)
88 FORMAT('EC(1-P) & ERROR =',1X,10F10.5)
88 FORMAT('EC(P**2) & ERROR =',1X,10F10.5)
INITIAL VALUES ARE GIVEN AS ALL ZEROES FOR THE PARAMETERS (SEE
J. ANDERSON (1972, BIOMETRIKA))

DO 13 K=1,10
   B(K)=0.0
   AT(K)=0.0
   EXIT=0.0
   KOUNT=0
   Q(1)=0.0
   DO 235 L=1,40
   Q(L+1)=Q(1)+L*0.025
   P(1)=0.0
   DO 681 L=1,20
   P(L+1)=P(1)+L*0.05

9 FORMAT(I2,2I3,F10.5/2(8F10.5/),4F10.5)
9 FORMAT( [E(P(5%));E(P(10%));E(P(25%));E(P(50%)) =',1X,10F8.4)
2 FORMAT( EST(IE(P)) =',1X,10F10.5)
6 FORMAT( E(P) FOR DIFF. & CHANGE =',1X,4F10.5)
3 FORMAT( ARC ELAS. =',1X,10F10.5)
5 FORMAT( VALUE OF I(P) =',1X,10F10.5)
8 FORMAT( ANSWER; ERROR =',1X,10F8.4)
9 FORMAT( MAXIMUM LIKELIHOOD ESTIMATES =',1X,10F10.5)
9 FORMAT( NO. OF CHANGED VALUES OF 1ST. DERIVATIVES =',1X,16)
9 FORMAT( NO. OF ITERATIONS FOR THE MLE'S =',1X,16)

1 FORMAT( WEST-IN TOTAL VARIANCE =',1X,10F10.5)
0 FORMAT( WEST-F-TEST; WEST-CORR =',1X,10F10.5)

8 FORMAT( MATRIX OF SECOND ORDER DERIVATIVES =',2X,10F10.5)
8 FORMAT( VECTOR OF FIRST ORDER DERIVATIVES =',1X,F15.9)
7 FORMAT( CURRENT B VALUES =',5X,F15.5)
2 FORMAT( CURRENT VALUE OF THE LIKELIHOOD FUNCTION =',1X,F20.15)
2 FORMAT( R-SQUARE ; LIKE.IND; APP.F-RATIO =',1X,3F15.5)
3 FORMAT( MC-ELASTICITIES =',1X,10F10.5)
2 FORMAT( TOTAL S.S., RESID. MEAN. S. SQ. & REG. MEAN. SQ. =',1X,10F10.5)
1 FORMAT( NO. OF MOVERS AND RELEVANT CASES =',1X,3I6)

3 FORMAT( STANDARD ERROR [B] =',1X,F9.5)
3 FORMAT( EXPLAINED S.S. & RESIDUAL S.S. =',1X,3F10.5)
6 FORMAT( WEST-5%E(P);10%E(P);25%E(P);50%E(P);80%E(P),1X,3F10.5)
9 FORMAT( WEST-ELASTY. =',1X,10F10.5)
6 FORMAT( ElASTICITY =',1X,10F10.5)
3 FORMAT( P & AV(P) & AV(P(1-P)) =',1X,4F10.5)
6 FORMAT( EXPECTED(P)/SAMPLE(P) =',1X,F10.5)
7 FORMAT( MEAN(VARIABLE) =',1X,10F10.5)
0 FORMAT( LAMDA; PR(SUCCESSFUL PRED.) =',1X,3F8.4)
6 FORMAT( VARIANCE-ELASTICITY =',1X,F8.4)
0 FORMAT( AV(SSP); KINIEVE(P); TALVITIES =',1X,3F10.5)

5 FORMAT( EXPECTED<P)/SAMPLE(P) =',1X,F10.5)
5 FORMAT( -2LN(LAMDA) =',1X,F8.4)
1 FORMAT( COVARIANCE MATRIX FOR THE X'S =',1X,10F8.4)
7 FORMAT( CORRELATION MATRIX FOR X'S =',1X,10F8.4)
7 FORMAT( SQUARED CORRELATIONS FOR X'S =',1X,10F8.4)
2 FORMAT( COVARIANCE MATRIX FOR B*X =',1X,10F8.4)
3 FORMAT( CHARACTERISTICS VARIANCE =',1X,F8.4)

DO 13 K=1,10
   B(K)=0.0
   AT(K)=0.0
   EXIT=0.0
   KOUNT=0
   Q(1)=0.0
   DO 235 L=1,40
      Q(L+1)=Q(1)+L*0.025
   P(1)=0.0
   DO 681 L=1,20
      P(L+1)=P(1)+L*0.05

J. ANDERSON (1972, BIOMETRIKA)
DO 25 L=1,180
25 AY(L)=0.0
MOV=0
MOVQ=0
WT=1.0

THIS LOOP WILL DEFINE THE DEPENDANT VARIABLE AND THE SET OF INDEPENDENT VARIABLES

THOSE VARIABLES ARE RESPECTIVELY DEFINED INTO ARRAYS AY(180) AND AVR(180,10)

DO 3 I=1,497
,READ(4,101) KVR

THIS CONDITION (BELOW) DELETES THOSE WHO HAVE NOT BEEN EMPLOYED
IF(KVR(20).EQ.0) GOTO 3

THIS CONDITION DELETES FEMALES
IF(KVR(3).EQ.0) GOTO 3

THE NEXT ONE DELETES THE DISMISSED OR LAYED OFF
IF(KVR(28).EQ.1) GOTO 3

THE NEXT ONE DELETES THE PART-TIME WORKERS (IF REQUIRED)
IF(KVR(44).EQ.0) GOTO 3

THE NEXT DELETES THE 'INTENDING MOVERS' (IF REQUIRED)
IF(KVR(13).EQ.1) GOTO 3

THIS ONE DELETES THE MISSING VALUE CASES OF TRAVEL-TIME
IF(KVR(42).EQ.13000) GOTO 3
KOUNT=KOUNT+1

NOW WE CALL THE SUBROUTINE TO DEFINE THE VARIABLES
CALL VAR(KVR,AVR,KOUNT,180,AY,WT)

EXIT=EXIT+AY(KOUNT)
DO 17 K=1,10
17 AT(K)=AT(K)+AVR(KOUNT,K)*AY(KOUNT)

1 CONTINUE
BY NOW THE VARIABLES HAVE BEEN DEFINED
'KOUNT' IS NUMBER OF RELEVANT CASES
EXIT=EXIT/KOUNT
WRITE(6,1111) KOUNT
ITER=0
LABEL '14' IS RETURN POINT FOR THE ITERATIVE PROCESS
14 CONTINUE
SCALAR=0.0
SUMD=0.0
ITER=ITER+1
WE CHECK THAT IF ITERATIONS ARE MORE THAN 14 THEN WE STOP THE PROGRAM FROM RUNNING, WE THEN CHECK IT
IF(ITER.EQ.15) GOTO 88
IGRE=0
DO 1 K=1,10
SUMV1(K)=0
1 CONTINUE
IF A DAMPED ITERATIVE METHOD IS ADOPTED THEN IT MAY NOT BE
NECESSARY TO REPEAT INITIALIZING THE SECOND DERIVATIVES MATRIX
EACH TIME. THUS, THE EVALUATION PART CAN BE SUPPRESSED BELOW BY
PUTTING A 'C' IN THE FIRST COLUMN OF THE APPROPRIATE ASSIGNMENT
STATEMENTS.
DO 2 K1=1,10
DO 2 K2=1,10
2 DERV2(K1,K2)=0.0
THIS LOOP CALCULATES THE FIRST ORDER DERIVATIVES VECTOR AND THE
SECOND ORDER DERIVATIVES MATRIX FOR THE N.R. METHOD.
DO 22 I=1,KOUNT
   SSP=0.0
   DO 5 K=1,10
      SSP=SSP+AVR(I,K)*B(K)
   CONTINUE
   ESP=DEXP(SSP)
   SUMD=SUMD+DLOG(1.0+ESP)
   PROB=ESP/(1.+ESP)
RECALL THAT THE NEXT LOOP CAN BE SUPPRESSED IF A DAMPED METHOD
IS ADOPTED.
DO 66 K1=1,10
DO 66 K2=1,10
66 DERV2(K1,K2)=DERV2(K1,K2)-AVR(I,K1)*AVR(I,K2)*PROB*(1.-PROB)
   DO 7 K=1,10
      SUMV1(K)=SUMV1(K)+AVR(I,K)*PROB
   CONTINUE
DO 90 K=1,10
90 SCALAR=SCALAR+AT(K)*B(K)
'LIKE' IS THE CURRENT VALUE OF LIKELIHOOD FUNCTION, 'DERV1' IS THE
FIRST DERIVATIVES VECTOR, 'DERV2' IS SECOND DERIVATIVES MATRIX.
LIKE=SCALAR-SUMD
WRITE(6,557) ITER
WRITE(6,2222) LIKE
DO 8 K=1,10
8 DERV1(K)=AT(K)-SUMV1(K)
THIS WRITE STATEMENT WRITES THE CURRENT SECOND DERIVATIVES MATRIX.
WRITE(6,8888) DERV2
THE COMING STATEMENT WRITE THE FIRST DERIVATIVES VECTOR.
WRITE(6,888) DERV1
ID=10
N=10
IV=10
IFAIL=0
THE FOLLOWING ROUTINE INVERTS 'DERV2' DEFINED ABOVE.
CALL F01AAF(DERV2,ID,N,VEVE2,IV,WKPCE,IFAIL)
ROUTINE 'MULT' MULTIPLIES DERV2*DERV1=STEP.
CALL MULT(VEVE2,DERV1,STEP)
THE NEXT DO LOOP UPDATES THE PREVIOUS B VALUES: B(K)=B(K)-STEP(K)
AT THE SAME TIME CONVERGENCE IS TESTED BY AN 'IF' STATEMENT.
DO 39 K=1,10
   B(K)=B(K)-STEP(K)
   IF(BABS(DERV1(K)).GE.0.000000005) IGRE=IGRE+1
39 CONTINUE
THIS WRITES THE NUMBER OF PARTIAL FIRST ORDER DERIVATIVES WHICH DO NOT SATISFY THE CONVERGENCE CONDITION
WRITE(6,9) IGRE
IF(IGRE.EQ.0) GOTO 88
WE MAY ALSO WRITE THE CURRENT VALUES OF THE UPDATED ESTIMATES
WRITE(6,7777) B
THE ITERATION IS NOW REPEATED IF THE ABOVE CONVERGENCE TEST IS NOT SATISFIED
GOTO 14
 OTHERWISE THE M.L.E.'S ARE PRINTED
88 WRITE(6,99) B
DO 223 K=1,10
223 DERVI(K)=0.0
TSSQ=LIKE
LOGR1=LIKE
RSSQ=0.0
AVP=0.0
AVVP=0.0
RMQ=0.0
EVP=0.0
DO 295 K=1,10
FRED(K)=0.0
FRED2(K)=0.0
ELS(K)=0.0
295 CONTINUE
DO 202 K1=1,10
DO 202 K2=1,10
CP(K1,K2)=0.0
COV1(K1,K2)=0.0
COV2(K1,K2)=0.0
202 CONTINUE
DO 500 K=1,21
500 M1(K)=0
COM1=0.0
CON2=0.0
PROP=0.0
LIKER1=0
AVQ1=0
SUMD=0.0
THE COMING LOOP PERFORMS THE INITIAL STEPS OF COMPUTING THE VARIANCE/COVARIANCE MATRIX OF THE VARIABLES THEIR MEANS AND THE GOODNESS-OF-FIT MEASURES.
IT ALSO COMPUTES THE FREQUENCY DISTRIBUTION OF ESTIMATED PROBS.
DO 23 I=1,KOUNT
SSP=0.0
DO 231 K=1,10
231 SSP=SSP+AVR(I,K)*B(K)
ESP=DEXP(SSP)
PROB=ESP/(1.0+ESP)
DO 6 K1=1,10
DO 6 K2=1,10
6 CP(K1,K2)=CP(K1,K2)+AVR(I,K1)*AVR(I,K2)
RSSQ=RSSQ+(AY(I)-PROB)**2/PROB
RMQ=RMQ+(AY(I)-EXIT)**2/PROB
IF(PROB.LT.0.5) LIKER1=LIKER1+(1.0-PROB)
IF(PROB.GE.0.5) AVQ1=AVQ1+PROB
AVP = AVP + PROB
AVVP = AVVP + PROB*(1.0 - PROB)
DO 290 K = 1, 10
ELS(K) = ELS(K) + PROB*(1.0 - PROB)*AVR(I, K) + B(K)
FREQ(K) = FREQ(K) + AVR(I, K)
FREQ2(K) = FREQ2(K) + AVR(I, K)**2
290 CONTINUE
SUMD = SUMD + AY(I)
DELT = 0.0
IF (PROB.GE.0.5) DELT = 1.0
COM1 = COM1 + AY(I)*(1.0 - DELT) + (1.0 - AY(I)) * DELT
COM2 = COM2 + AY(I)*(1.0 - WT) + (1.0 - AY(I)) * WT
EVP = EVP + PROB**2

THIS LOOP FILLS IN THE FREQUENCY DISTRIBUTION OF THE ESTIMATED
PROBABILITIES WHICH ARE LATER PLOTTED IN A HISTOGRAM

DO 418 K = 1, 10
IF (PROB.GT.P(K).AND.PROB.LE.P(K+1)) M1(K) = M1(K) + 1
418 CONTINUE

EXPNT = 0.0
SSP = 0.0
AVP = AVP/KOUNT
LOGR2 = SUMD*DLOG(EXIT) + (KOUNT-SUMD)*DLOG(1.0-EXIT)
LIK1 = 1.0 - LOGR1/LOGR2
AVQ = (K/K-1)*(LIK1/(1.0-LIK1))
AVQ1 = AVQ1/KOUNT
LIK1 = LIKER1/KOUNT
AVQ1 = AVQ1 + LIKER1
SUMD = SUMD/KOUNT
LOGR = -2.0*(LOGR2-LOGR1)
AVP = AVVP/KOUNT
RSSQ = 1.0 - RSSQ/RMQ
PROP = EXIT/AVP
WT = AVP*KOUNT
DO 291 K = 1, 10
AVX(K) = FREQ(K)/KOUNT
SVR1(K) = AT(K)/FREQ(K)
ELS(K) = ELS(K)/WT
CHNG(K) = B(K)*AVX(K)

SSP = SSP + CHNG(K)

291 EPS(K) = CHNG(K)*AVVP/AVP
AGP = DEXP(SSP)/(1.0+DEXP(SSP))
THE COMING LOOP WRITE THE NEGATIVE OF THE INVERSE OF 'DERV2'
WHICH IS APPROX. = THE VAR/COVAR OF THE M.L.E.'S.
DO 19 K1 = 1, 10
DO 19 K2 = 1, 10
19 VEVE2(K1, K2) = -VEVE2(K1, K2)

NOW THE VAR/COVAR MATRIX IS WRITTEN DOWN:
THIS LOOP CALCULATES THE ASYMMPTOTIC STANDARD ERROR VECTOR FOR
THE M.L.E.'S.
IT THEN USES THEM TO COMPUTE 95% INTERVALS
DO 191 K = 1, 10
STEP(K) = DSORT(VEVE2(K, K))
STUD(K) = B(K)/STEP(K)
191 CONTINUE
WRITE(6,6666) STUD
WRITE(6,5555) VEVE2
WRITE(6,3333) STEP
CON3=1.0-CON1/KOUNT
CON1=1.0-CON1/CON2
WRITE(6,7999) IER
WRITE(6,1100) CON1,CON3
WRITE(6,293) SUMD,AVP,AVVF
WRITE(6,297) AVX
WRITE(6,296) EPS
WRITE(6,113) ELS
WRITE(6,294) PROP

THIS LOOP COMPUTES THE VAR/COVAR MATRIX FOR EXPLAN. VARIABLES
THE CORRELATION MATRIX AND COVAR. MATRIX FOR THE WEIGHTED
VARIABLES.

I.E THE 8X TERMS
DO 54 K1=1,10
DO 54 K2=1,10
COV1(K1,K2)=CP(K1,K2)/KOUNT-AVX(K1)*AVX(K2)
STD1=FREQ2(K1)/KOUNT-AVX(K1)**2
STD1=DSQRT(STD1)
STD2=FREQ2(K2)/KOUNT-AVX(K2)**2
STD2=DSQRT(STD2)
DEM=STD1*STD2
IF(STD1*STD2.EQ.0.0) DEM=1.0
CORR(K1,K2)=COV1(K1,K2)/DEM
COV2(K1,K2)=B(K1)*B(K2)*COV1(K1,K2)
DETR(K1,K2)=CORR(K1,K2)**2
EXPNT=EXPNT+COV2(K1,K2)
TERM=AGP*(AGP-1.0)+(AGP-0.5)
AGP1=AGP+TERH*EXPNT
WRITE(6,200) SSP,AGP,AGP1
WRITE(6,313) EXPNT
YXPNT=2.0*EXPNT
N=41
UX=DSQRT(3.1415926535+YXPNT)

THE NUMERICAL INTEGRATION IS DONE BY NAG COMPUTING QUADRATURE
ROUTINES.

THE FOLLOWING ROUTINE DEFINES 41 POINTS ON THE SB CURVE WHICH
ARE LATER USED TO CONSTRUCT A SMOOTH SB CURVE
CALL FUN(Q,SSP,YXPNT,N,BY,CY,BIT1,BIT2,UX)

WE INTEGRATE THE CURVE NUMERICALLY TO CHECK THAT IT IS A PROPER
PROB. DENSITY FUNCTION (OF COURSE NUMBERS WILL BE APPROXIMATE)
IFAIL=0
CALL D01GAF(Q,BY,41,SIM,TRZ,IFAIL)
WRITE(6,788) SIM,TRZ
IFAIL=0
CALL D01GAF(Q,CY,41,SIM,TRZ,IFAIL)
EV=SIM
WRITE(6,99998) SIM,TRZ
IFAIL=0
DO 415 I=2,40
DY(I)=CY(I)*(1.0-Q(I))
415 CONTINUE
DY(1) = 0.0
DY(41) = 0.0
CALL D01GAF(Q, DY, 41, SIM, TRZ, IFAIL)
WRITE(6, 99988) SIM, TRZ
IFAIL = 0
VP = SIM
SV = -SSP*(EV/VP)
WRITE(6, 666) SV
L = 0

THE NEXT LOOP IS USED TO CALCULATE THE ARC ELASTICITIES USING
OUR APPROXIMATE COMPUTATIONAL FORMULA
DO 502 K = 1, 10
L = K + 1
RATIO(1) = 0.05*B(L)*AVX(L)*AVVP + AVP
ARC(1) = B(L)*AVVP*(AVX(L) + 2.05/(RATIO(1) + AVP))
RATIO(2) = 0.10*B(L)*AVX(L)*AVVP + AVP
ARC(2) = B(L)*AVVP*(2.10*AVX(L)/(RATIO(2) + AVP))
RATIO(3) = AVP + 0.20*B(L)*AVX(L)*AVVP
ARC(3) = B(L)*AVVP*(2.20*AVX(L)/(RATIO(3) + AVP))
RATIO(4) = AVP + 0.30*B(L)*AVX(L)*AVVP
ARC(4) = B(L)*AVVP*(2.30*AVX(L)/(RATIO(4) + AVP))
RATIO(5) = AVP + 0.50*B(L)*AVX(L)*AVVP
ARC(5) = B(L)*AVVP*(2.50*AVX(L)/(RATIO(5) + AVP))
WRITE(6, 11133) ARC
WRITE(6, 11122) RATIO

502 CONTINUE

THIS PART RELATES TO THE EXPERIMENT WHICH WE DISCUSSED IN THE
5TH. CHAPTER ABOUT THE GEOMETRICAL DESCRIPTION OF THE POTENTIAL
EFFECT OF VARIOUS POLICY COMBINATIONS ON QUIT PROBABILITY
WE IMPOSE SPECIFIC % CHANGES ON THE MEANS OF THE VARIABLES
WXP(1) = SSP - 0.25*AVX(10)*B(10)
WXP(2) = WXP(1) + 0.25*B(10)*AVX(10)
WXP(3) = WXP(2) - 0.25*B(4)*AVX(4)
WXP(4) = WXP(3) + 0.25*B(5)*AVX(5)
WXP(5) = SSP + 0.25*AVX(10)*B(10)
WXP(6) = WXP(5) - 0.25*AVX(10)*B(10)
WXP(7) = WXP(6) + 0.25*AVX(4)*B(4)
WXP(8) = WXP(7) - 0.25*AVX(5)*B(5)
DO 602 K = 1, 10
CALL FUN(Q, WXP(K), YXPNT, N, BY, CY, BIT1, BIT2, WX)
IFAIL = 0
CALL D01GAF(Q, CY, N, SIM, TRZ, IFAIL)

602 CONTINUE

TVPI = EV*(1. - EV)
WRITE(6, 2121) TVPI
K = 10
TEST = ((TVPI - VP)/VP)*((KOUNT - K)/(K - 1))
WT = 1. - VP/TVPI
WRITE(6, 100) TEST, WT
DO 567 K = 1, 10
BIT3(K) = B(K)*AVX(K)*VP/EV
567 CONTINUE

WRITE(6, 789) BIT3
IFAIL = 0
SUM = 0.0
CALL FUN(P, SSP, YXPNT, 21, EY, BY1, BIT1, BIT2, WX)
DO 579 K = 1, 21
579 SUM = SUM + EY(K)
DO 680 K = 1, 21
EY(K) = EY(K) + KOUNT / SUM
M1(K) = M1(K) * SUM / KOUNT
680 CONTINUE
N1 = 1
WRITE(11, 899) NGR, M1, EY1
THIS PART RELATES TO THE SIMULATION EXPERIMENT OF THE
AGGREGATION BIAS PROBLEM (SEE CHAPTER 5 SEC(5, 3))
DO 72 I = 1, KOUNT
SSP = 0.
DO 722 K = 1, 10
722 SSP = SSP + AVR(I, K) * B(K)
PROB = DEXP(SSP) / (1. + DEXP(SSP))
QY(I) = PROB
RY(I) = SSP
CALL FUN(Q, SSP, YXPNT, N, CY, CY, BIT1, BIT2, WX)
CALL DO1GAF(Q, CY, N, SIM, TRZ, IFAIL)
VY(I) = SIM
72 CONTINUE
WRITE(10, 66677) RY
WRITE(10, 66677) VY
WRITE(10, 66677) QY
THIS PART DEALS WITH COMPUTATION OF MORRISON'S FORMULA OF THE
PROPORTION CORRECTLY CLASSIFIED (USING NUMERICAL INTEGRATION)
DO 800 K = 1, 20
P(K) = Q(K+20)
BIT1(K) = 1.0 / (1.0 - P(K))
BIT2(K) = DLOG(P(K)/(1.0 - P(K))) - SSP
BIT2(K) = -BIT2(K) ** 2 / YXPNT
BIT2(K) = DEXP(BIT2(K))
BY(K) = BIT1(K) * BIT2(K) / WX
800 CONTINUE
P(21) = 1.0
BY(21) = 0.0
DO 801 K = 2, 21
R(K) = Q(K)
BIT1(K) = 1 / R(K)
BIT2(K) = DLOG(R(K)/(1.0 - R(K))) - SSP
BIT2(K) = -BIT2(K) ** 2 / YXPNT
BIT2(K) = DEXP(BIT2(K))
CY(K) = BIT1(K) * BIT2(K) / WX
801 CONTINUE
CY(1) = 0.0
R(1) = 0.0
IFAIL = 0
CALL DO1GAF(P, BY, 21, SIM, TRZ, IFAIL)
VP = SIM
IFAIL = 0
CALL DO1GAF(R, CY, 21, SIM, TRZ, IFAIL)
EV = SIM + VP
WRITE(6, 802) EV, AVQ1
NEXT PART DEALS WITH COMPUTATION OF OUR ANALYSIS OF VARIANCE
SET-UP FOR THE COMPUTATION OF F*-RATIO

THE FOLLOWING QUANTITY IS 'THE EXPLAINED VARIATION'
\[ \text{REMN} = \text{EVP} - \text{KOUNT} \times \text{AVP}^2 \]

THE NEXT IS THE 'UNEXPLAINED VARIATION'
\[ \text{AVVP} = \text{KOUNT} \times \text{AVVP} \]

THE NEXT IS 'TOTAL VARIATION' USING AVERAGE PROBABILITIES
\[ \text{EVP} = \text{AVP} \times (1.0 - \text{AVP}) \]

THIS ONE IS THE SAME AS THE ABOVE, USING SAMPLE PROPORTION
\[ \text{TVVP} = \text{EXIT} \times (1.0 - \text{EXIT}) \]
\[ \text{DVP} = \text{TVVP} - \text{EVP} \]
\[ \text{PROP} = \text{KOUNT} \times \text{EVP} \]
\[ \text{WT} = \text{REMN} / \text{PROP} \]
\[ \text{NK} = 10 \]
\[ \text{SSP} = (\text{REMN} / (\text{NK} - 1)) / (\text{AVVP} / (\text{KOUNT} - \text{NK})) \]

NOW OUR F*-RATIO TEST WILL BE WRITTEN DOWN
\[ \text{WRITE}(6,104) \text{ SSP} \]

NEXT WE WRITE THE COVARIANCE AND CORRELATION MATRICES
\[ \text{WRITE}(6,311) \text{ COV1} \]
\[ \text{WRITE}(6,317) \text{ CORR} \]
\[ \text{WRITE}(6,318) \text{ DETR} \]
\[ \text{WRITE}(6,312) \text{ COV2} \]

NEXT THE LOG-LIKELIHOOD RATIO IS WRITTEN DOWN
\[ \text{WRITE}(6,310) \text{ LOGR} \]

THE FOLLOWING ARE THE OTHER MEASURES OF GOODNESS-OF-FIT WRITTEN
\[ \text{WRITE}(6,212) \text{ RSSQ, LIKER, AVQ} \]
STOP
END

BELOW ARE SOME SUBROUTINES WHICH ARE USEFULL
THE COMING ROUTINE DEFINES WHICH VARIABLES TO USE
SUBROUTINE VAR(IVR,X,KOUNT,KMAX,Y,WT)
DIMENSION IVR(46),X(KMAX,10),Y(KMAX)
DOUBLE PRECISION X,Y,WT,RR,RR2,RR1
VARIABLES ARE NOW TO BE DEFINES INTO ARRAY X(KMAX,10) FROM THE
ORIGINAL DATA
X(KOUNT,1)=1.0
RR=IVR(40)/WT
RR=RR/100
X(KOUNT,2)=RR
RR1=IVR(45)/WT
RR2=IVR(46)/WT
RR=RR2-RR1
X(KOUNT,3)=RR
RR2=IVR(42)/WT
RR2=IVR(43)/WT
RR=RR1-RR2
RR=RR/100
X(KOUNT, 4)=IRR(13)/WT
X(KOUNT, 5)=IVR(13)/WT
IL0W=1-(1-IVR(18))*(1-IVR(19))
X(KOUNT, 6)=IL0W/WT
X(KOUNT, 7)=IVR(27)/WT
X(KOUNT, 8)=IVR(16)/WT
X(KOUNT, 9)=IVR(6)/WT
X(KOUNT, 10)=IVR(30)/WT
THE NEXT CONDITION DEFINES THE 0,1 DICHOTOMOUS VARIABLE
IF(IVR(23).EQ.0) Y(KOUNT)=1.0
RETURN
END

THE COMING SUBROUTINE DEFINES THE ORDINATES OF THE N CHOSEN
POINTS ON [0,1] INTERVAL USING THE SB CURVE
SUBROUTINE FUN(Q, SSP, EXPP, N, Y1, Y2, BIT1, BIT2, WX)
DIMENSION Q(N), Y2(N), Y1(N), BIT1(N), BIT2(N)
REAL*8 Q, Y1, Y2, BIT1, BIT2, SSP, EXPP, WX
M=N-1
DO 22 I=2, M
BIT1(I)=1.0/(Q(I)*(1.0-Q(I)))
BIT2(I)=DLOG(Q(I)/(1.0-Q(I)))-SSP
BIT2(I)=-BIT2(I)**2/EXPP
BIT2(I)=DEXP(BIT2(I))
Y1(I)=BIT1(I)+BIT2(I)/WX
Y2(I)=Q(I)*Y1(I)
22 CONTINUE
Y1(1)=0.0
Y1(N)=0.0
Y2(1)=0.0
Y2(N)=0.0
RETURN
END

THIS ROUTINE CALCULATES THE PRODUCT OF A MATRIX TIMES A VECTOR
SUBROUTINE MULT(A, V1, V2)
DIMENSION A(10, 10), V1(10), V2(10)
DOUBLE PRECISION A, V1, V2
DO 16 K=1, 10
16 V2(K)=0.0
DO 66 J1=1, 10
DO 66 J2=1, 10
66 V2(J1)=V2(J1)+A(J1, J2)*V1(J2)
RETURN
END

THE NEXT IS A SUBROUTINE TO GET THE LARGEST VALUE OF N NUMBERS
FUNCTION SUP(DF, N)
REAL*8 DF(N), TOP, SUP
TOP=DF(1)
DO 11 K=1, N
IF(DF(K).GT.TOP) TOP=DF(K)
11 CONTINUE
SUP=TOP
RETURN
END
PROGRAM(II)

*************
THIS PROGRAM IS BASED ON THE EGNP55.0RMANUAL PACKAGE DUE TO
"KEN-PEAC" OF THE KING'S BUILDING, EDINBURGH UNIVERSITY
WE HAVE DEVELOPED THIS PROGRAM TO DO THE GRAPH PLOTTING OF OUR
SB CURVES AND THE FREQUENCY HISTOGRAMS OF THE ESTIMATED PROBS.
OF COURSE THE PROGRAM CAN BE ADJUSTED TO SUIT THE PROBLEM IN
QUESTION
THE STATEMENTS WHICH ARE NOT REQUIRED IN A GIVEN RUN COULD BE
SUPPRESSED BY PUTTING A 'C' AT THE BEGINNING OF THE LINE
REAL 4 EY(21),EF1(21),EY1(21),EY11(21),X(21),EF(21)
*,NM5(21),EY5(21),EF11(21)
INTEGER NM(21),NM11(21),NM1(21)
REAL 4 TX,TN,SIM,H,TY,TF,T3
INTEGER N,NT
NOTE THAT THE ABOVE FORMAT 101 IS USED TO READ THE DATA AND IT
IS EXACTLY THE ONE WHICH HAS BEEN USED TO WRITE THIS DATA BY
PRAGRAM(I) ABOVE

01 FORMAT(I2,21I3,F10.5/2(8F10.5/),4F10.5)
X(1)=0.
DO 33 J=1,20
L=J+1
33 X(L)=O+.05*J
THUS 21 POINTS HAVE BEEN DEFINED ON THE 0,1] INTERVAL
CALL STARTP(10,'TAG-EL-DIN U.R.B',16)
THIS PROG. PLOTS THREE GRAPHS IN THE SAME SCALE, N TIMES
DO 22 I=1,N
READ(4,101) N,NM,EY
READ(4,101) N,NM1,EY1
READ(4,101) N,NM11,EY11
NP=I
THIS ROUTINE ASSIGN A GRAPH TITLE
CALL GRTITL('OBS. VS. FITTED ',16)
DO 11 K=1,21
EF(K)=FLOAT(NM(K))
EF1(K)=FLOAT(NM1(K))
EF11(K)=FLOAT(NM11(K))
11 CONTINUE
NUMERICAL INTEGRATION IS REPEATED IN THIS PROGRAM USING THE
'SIMPSON' RULE SUBROUTINE WHICH WE DO IN THIS PROGRAM
CALL SIMP(EY,21,SIM,.05)
CALL SIMP(EY1,21,SIM1,.05)
CALL SIMP(EY11,21,SIM2,.05)
THE NEXT LOOP SCALES THE 'EMPIRICAL' FREQUENCIES TO THE PROB.
CURVE I.E. TH SB CURVE
DO 66 K=1,21
EF11(K)=EF11(K)/SIM2
EF1(K)=EF1(K)/SIM1
EY1(K)=EY1(K)/SIM1
EY(K)=EY(K)/SIM
EY11(K)=EY11(K)/SIM2
EF(K)=EF(K)/SIM
66 CONTINUE
N=21
THE DRAWING OF THE CURVES REQUIRES KNOWING THE LARGEST POINT ON
THE CURVES, AS DONE BELOW
TF=TOP(EY,N)
TY=TOP(EY1,N)
T3=TOP(EY1,N)
TX=TY
IF(TF.GT.TY) TX=TF
IF(T3.GT.TFX) TX=T3
IF(TY.GT.T3) TX=TY
CALL NEWPGE(NP,"A4",WIDE)
THE X-AXIS IS DONE BY THE NEXT ROUTINE
CALL XAXIS(0,.05,20,"QUIT PROB.",9,0.,0.)
NT=TX
NT=NT+1.
TN=FLOAT(NT)
THEN, THE Y-AXIS IS ALSO DONE
CALL YAXIS(0.,0.,TN,‘R(Q) & F(Q)’,11,0.,0.)
THESE SETPAR ROUTINES SPECIFY THE SIZES OF GAPS AND DASHES IN
THE CURVES TO BE MADE
CALL SETPAR(‘DASH’,1.)
CALL SETPAR(‘GAP’,2)
THE ROUTINES ‘CURVEK’ & ‘HIST’, DO THE CURVE AND THE HISTOGRAM
RESPECTIVELY
CALL CURVEK(X,EY,21)
CALL SETPAR(‘DASH’,1)
CALL SETPAR(‘GAP’,2)
CALL CURVEK(X,EY1,21)
CALL SETPAR(‘DASH’,0)
CALL SETPAR(‘GAP’,0)
CALL CURVEK(X,EY1,21)
CALL HISTGR(EF,21,0.,.05)
THE ERRX & ERRY ROUTINES DO THE LINES PARALLEL TO THE X AND
Y-AXIS
CALL ERRX(0.,TN,0.,1.)
CALL ERRY(1.,0.,0.,TN)
22 CONTINUE
CALL CLOSGR
END
THE FUNCTION BELOW SELECTS THE LARGEST NUMBER FROM N NUMBERS
FUNCTION TOP(A,N)
REAL=4 A(N)
X=A(1)
DO 33 K=1,N
33 IF(A(K).GT.X) X=A(K)
TOP=X
RETURN
END
THE SUBROUTINE BELOW APPLIES "SIMPSON RULE" TO INTEGRATE A CURVE

SUBROUTINE SIMP(EY,N,SI,H)

REAL*4 EY(N)
M=N-3
H=H/2

SUM=0.

DO 33 K=1,M

33 SUM=SUM+4*EY(2*K-1)+2*EY(2*K)

SUM=SUM+EY(1)+4*EY(N-1)+EY(N)
SI=SUM*H/3
RETURN
END
BIBLIOGRAPHY


5) Altham, P.M.E. (1976) "Discrete Variable Analysis for Individuals grouped into Families", Biometrika, 63, 2, 263.


20) Berkson, J. (1951) "Why do I prefer logits to probits?", Biometrics, 7, 327.


22) Berman, B. (1965); see Holt, C.C. (74), p. 103.


51) Fisk, P. (1977); A letter of private correspondence; Dept. of Statistics, J.C.M.B., Edinburgh University (16 August).


63) Hadley, G. "Linear Programming", Addison-Wesley Publishing Co.


72) Hey, G.D. and Mackenna, G.J. (1979) "To move or not to move?", Economica, 46, 175.


75) _______ (1970) "How can the Philips' Curve be moved to reduce both inflation and unemployment?", in Phelps et al. pp. 224.
81) Kerr, C. (1954) "The balkanization of the labour markets", Institute of Industrial Relations; University of California, Berkeley.


98) ________ and Reed, F. (1975) "Aggregate travel demand model from disaggregate behavioural models", Transportation Research Record, 534, 24.


110) _____ (1976) "Residential location and labour supply", J. of Political Economy, 84, No. 4, Suppl., 212.


