REGRESSION MODELS INVOLVING CATEGORICAL AND CONTINUOUS DEPENDENT VARIABLES

with

A STUDY ON LABOUR SUPPLY OF MARRIED WOMEN

by

YAT WING LAU

B.Sc., University Of British Columbia, 1973

A Thesis Submitted In Partial Fulfilment Of

The Requirements For The Degree Of

Master Of Science

in Commerce and Business Administration

in the Faculty

of

Commerce and Business Administration

We accept this thesis as conforming to the

required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 1975

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of COMMERCE AND BUSINESS ADMINISTRATION

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date JAN 9th 1976.

Abstract

This thesis is going to consider the inferences about the relationships that determine jointly a continuous variable and a categorical variable. These relationships can be considered separately into two models: a regression model and a probability model. The regression model can be estimated by ordinary least squares, or Zellner's two stage method. The probability model is estimated by the method of Nerlove and Press. Such relationships will be given more complex consideration.

This kind of model is applied in the analysis of an economic problem. It is to consider the labour supply of married women. Data are pooled from the Panel Study of Income Dynamics 1972. It is found that the age of the youngest child is the most significant factor to determine the number of hours worked by a married woman, and birth gap is the major effect in the probability of a wife having a child not older than six years of age.

Table of Contents

List of Tables

Acknowledgements

I would like to express my gratitude to my thesis committee members who offered me their advice, helpful criticisms, stimulating suggestions and assistance. Special thanks are due to Professor Press for the idea which made this thesis possible, his encouragement, and for the permission to use his logistic program to compute the data. Special thanks are also due to Professor Berndt for generously sharing his research work on the labour supply of married women, and providing access to information on the Panel Data. I also thank my mathematics teacher, Professor Nash, for his valuable comments.

I am indebted to Professor Doll for providing me a fellowship from the Centre for Transportation Studies during my first year study, and to Professor Press for offering me a research assistanship from his research fund of Chicago University in my second year. Without their financial support, the thesis in its present form could not have been achieved.

I would also like to express my gratitude to Miss Nancy Reid for sharing her early results which we did for the Economic and Statistic Workshop at the beginning of this year; to Miss Norine Smith, research assistant of Professor Berndt for her help to access data from the Panel Data file, and to Ms Valda Johnston for her beautiful editorial work.

Wing Lau

November of 1975.

Chapter I

Introduction

This thesis[1] is concerned with making inferences about relationships that determine jointly a continuous variable and a categorical variable. Given the discrete random variable, the continuous variable is related to a set of explanatory variables; also the discrete random variable is related to the same set of explanatory variables. For example, the timing of a married woman participating in labour force activity will likely depend upon the age of her youngest child. Let us assume that a housewife will work less hours when she has a child not older than 6 years. We will call z, a family constraint variable, if it is 1 when a family has a child of 6 years or younger, and 0 otherwise. The timing of her labour force activity in each case, that is for a given value of z, may relate to her hushand's income, her expected wage, her fecundity and so on. Also the probability of a family having a child not older than 6 years can be related to the same set of explanatory variables. We would like to know the joint probability of her timing in labour force activity as well as her youngest child not being older than 6 years. It is shown that we can estimate the parameters of conditional regression equations and the parameters of the discrete dependent variable regression individually. It is very easy to extend the problem to involve a polytomous dependent variable in regression rather than a dichotomous dependent variable.

To consider the above problem in terms of a regression model, the model can be extended into a set of regression equations including a set of conditional regressions and a set of discrete dependent variable regressions. For example, at year t, for a given value of z, the timing of a married woman in labour force activity can be expressed in terms of a set of explanatory variables, and the probability of a family having a child not older than 6 years is related to the same set of explanatory variables. Hence, we may consider such relations for certain time periods, such as from 1967 to 1971. It is found that to apply Zellner's method(1962) to estimate the parameters of conditional regression equations is more efficient than to estimate them equation-by-equation. Some interesting extensions of this model are mentioned.

The plan of this thesis is as follows. This thesis is divided into a theoretical part and an applicational part. The first part is a theoretical discussion composed of chapters 2, 3, and 4. In chapter 2 we describe the basic model and prove that the estimators of the parameters in conditional regression equations and the parameters in discrete dependent variable regression are independent. In estimating parameters of discrete dependent variable regression, we explain why we prefer logit rather than other functions. Chow's test(1960) is used for testing the equality of coefficients of two conditional regression equations. In chapter 3 we extend the basic model into a set of regression equations. The method of estimating equation-by-equation is mentioned. Zellner's seemingly unrelated regressions method is applied in order to get an efficient estimation. Following is the description of testing the aggregation bias. Chapter 4 is to propose some interesting extensions of this basic model.

The second part is an application concerning an economic problem of labour supply of married women. The data used are from the Panel Study of Income Dynamics 1972, which is collected by the Survey Research Centre of the University of Michigan. Data are focused primarily on change in family economic status. Data-collection technique is mainly on the household personal interview. The empirical studies are compounded by two models which are based on chapters 2 and 3. Chapter 5 is the description of our economic models, chapter 6 gives the results under the first model and chapter 7 gives the results under the second model. Chapter 8 will be the conclusion of the whole thesis.

Footnotes

[1] This is based on the preliminary work done by Nancy Reid and me in the Econometrics and Statistics Workshop this year at the University of British Columbia. Miss Reid is a master's student of the Institute of Applied Mathematics and Statistics of the University of British Columbia.

Chapter II

Basic Model

I The mathematical model

Let us begin with a very simple model which has two dependent variables, one of which is y, a continuous variable, and the other is z, a dichotomous variable having the value 0 or 1. For the given value of z, y is distributed normally, and is expressed as a function of a number of variables[1] $x^1, \ldots, x^k$. Since z is dichotomous, we denote the functional relations of y and the x's as the following:

$$E(y \mid z=1) = f(x^1, \ldots, x^k)$$

and

$$E(y \mid z=0) = g(x^1, \ldots, x^k)$$

There will be a variety of functions to satisfy the above relations. The simplest relationship between y and the x's is linear. So for n observations, we write each of them more formally under the linear hypothesis as:

$$(y(i) \mid z(i)=1) = a^1 x^1(i) + \ldots + a^k x^k(i) + u(i)$$

and

$$(y(i) \mid z(i)=0) = b^1 x^1(i) + \ldots + b^k x^k(i) + v(i)$$

$i=1, \ldots, n$ or in vector notation,

$$(y(i)\mid z(i)=1) = A'X(i) + u(i)$$

and

$$(y(i)\mid z(i)=0) = B'X(i) + v(i)$$

where $A=(a^1,\dots,a^k)'$, $B=(b^1,\dots,b^k)'$, and $X=(x^1(i),\dots,x^k(i))'$; u, and v denote variables which may take on positive or negative values. Usually u and v are called error terms or disturbance terms. In order to make the model simple, let us first assume u and v have the same distribution. We assume u and v are random and normally distributed with mean zero, variance var(u) and zero covariance, that is

$$
\begin{aligned}
E[u(i)] &= 0 & \\
E[u(i),u(j)] &= 0 & i \neq j \\
&= \mathrm{var}(u) & i = j
\end{aligned}
$$

Hence A, B and var(u) are unknown parameters. We may wish to estimate these parameters statistically on the basis of our sample observations, and to test hypotheses about them. Therefore, if we consider the conditional distributions of y, then when z=1, y is distributed normally with mean A'X and variance var(u), and when z=0, y is distributed normally with mean B'X and variance var(u).

For the dichotomous dependent variable, we may be interested in the probability that z will have the value 0 or 1. The probability of z being 1 can also be expressed as a function of $x^1, \dots, x^k$. So,

$$\text{Prob}(z=1) = h(x^1, \ldots, x^k)$$

Suppose that we want a relationship in which Prob(z=1) is a nondecreasing function of t with $F(-\infty)=0$ and $F(\infty)=1$; for the ith observation,

$$\text{Prob}(z(i)=1) = p(i) = F(t(i)) \qquad i=1,\ldots,n$$

where $t(i)=c^1x^1(i)+\ldots+c^kx^k(i)$, or in vector notation:

$$\text{Prob}(z(i)=1) = p(i) = F(C'X(i)) \qquad (1)$$

$F(C'X(i))$ is taken to be a cumulative distribution function[2]. Therefore, we know p(i) will lie between 0 and 1, and p(i) is a nondecreasing function of C'X, but may be decreasing in some variables, depending upon the signs of the components of C. Therefore, we will focus our interest on the estimation of C.

Before we step into the estimation of these parameters, we should discuss more about function F, because in the history of statistics, there was a long argument about F. In the following section, we will discuss several transformation functions and explain why the logistic function is chosen.

II Discrete dependent variable regression

Failure of linear approximation to the probability function:

If we use linear approximation to the probability function F(C'X), then we will observe that the function is well approximated in the centre, but poor for very large or small value of C'X. There are technical difficulties in using standard regression techniques on binary data[3]. First, for given observation x(i), z(i) is a Bernoulli random variable, so that the variance of the jth disturbance term depends upon j. Those disturbance terms are heteroscedastic, therefore, ordinary least-squares estimation will give inefficient estimators and imprecise predictions. Zellner and Lee(1965), Goldberger(1964) suggested the use of generalized least-squares to remove the heteroscedastic problem, but this failed, because it ignored the Bernoulli character of the errors, or did not guarantee that z(i) should lie between 0 and 1 for all i, and resulted in some negative variances. Furthermore, the transformation on generalized least-squares led into the numerical problem that if the independent variable is larger than 1, the transformation is undefined. Cox(1970) concludes that linear approximation to this function fails, since, "because the z(i)'s are normally distributed , no method of estimation that is linear in the z(i)'s will in general be fully efficient."

Probit analysis:

One reasonable approach is called probit analysis. Bliss(1934) was the first to use it. Finny(1947) applied this method in analyzing quantal (binary) responses in bioassay, Cornfield and Nathan Mantal(1950) applied it in calculating the dosage response curve, and Tobin(1955) applied it in economic surveys.

This method applies a grouping method for estimating the equation (1). $F(t)$ is considered as the cumulative distribution function of the standard normal distribution by using grouped data[4]. Hence $p(i)$ is estimated by sample proportion, i.e.

$$p^{+}(i) = r(i)/n(i) \qquad i=1,\ldots,n$$

where $r(i)$ is the number of elements in ith cell having value 1, and $n(i)$ is the total number of elements in the ith cell. We define probit as following,

$$\text{Probit}(p^{+}(i)) = t^{+}(i) + 5 \qquad (2)$$

where $t^{+}(i)$ is defined by $p^{+}(i) = F(t^{+}(i))$. F is the cumulative standardized normal distribution. One adds 5 in the transformation in order to get positive values for the transformed variable. Hence $\text{Probit}(p^{+}(i))$ is normally distributed with mean 5 and variance 1. So we can apply ordinary least-squares to the transformed data. Putting it into a regression equation, it will be

$$\text{Probit}(p^{+}(i)) = C'X(i) + e(i) \qquad i=1,\ldots,n \qquad (3)$$

where $e(i)$ with zero mean, zero covariance, and variance equal to $var(e(i))$.

Finally, we note with Press and Nerlove(1973) that: "For this probit analysis method to be useful, there should be several observations per cell($n(i)$ > for every i). Moreover, efficiency of estimation is lost in the ad hoc procedure associated with the added 5 in (2). Note also that there are computational difficulties associated with the use of the integrals in this procedure. Unequal numbers of observations per cell are inefficient, and cells with one or zero observations per cell are not useful."

Logit analysis:

Another method called logit analysis was introduced by Berkson(1944). Using cell frequency, F(t) is considered as the cumulative distribution function of the standard logistic distribution function; that is,

$$F(t) = 1/(1+\exp(-t))$$

where t is real; so

$$p(i) = 1/[1+\exp(-C'X(i))]$$

or,

$$\log(p(i)/(1-p(i))) = C'X(i)$$

Now we define $\mathrm{Logit}(p^+(i))$ as following,

$$\mathrm{Logit}(p^+(i)) = \log(p^+(i)/(1-p^+(i)))$$

where $p^+(i)$ is estimated by sample portion. C can be estimated from regression estimation. Bishop(1969), Goodman(1970), Press and Nerlove(1973) apply this method in dealing with contingency tables.

Other transformations:

Coleman(1964) has proposed an exponential model by choosing

$$p(i) = 1-\exp(-X'C)$$

The weakness of this function is that p(i) is not constrained to lie between zero and one unless all of the parameters are non-negative. Goodman(1972) made a comment about this transformation. He said, "Coleman's article did not show how to test whether his model fit the actual data, nor was he able to measure how well it fit. Furthermore, he did not show how to test the statistical significance of the contribution made by the various parameters in the model, nor could he measure their contribution's magnitude."

Angular transformations are very possible candidates, but those transformations are not as simple as the logistic transformation. So our selection is limited to probit and logit analysis.

The choice of transformation:

The above discussion shows, our choice will be either probit or logit analysis. Gunderson(1974), Buse(1972), Chambers and Cox(1967) have done some work about this problem. They found that the numerical difference between these two is very slight except at the two extremes. From the optimization point of view, it is sure that the maximum of a logistic function is the global maximum. If we consider the cost of computation, logit analysis is much better than probit analysis. On the other hand, there is the theoretical argument that the probit transformation is the appropriate one to use under the hypothesis of log-normally distributed tolerances. Berkson(1951) is very doubtful as to the validity of this hypothesis. He says that the practice of injecting an interpretation of "tolerance" into response data is objectionable; it can be misleading and harmful. He explains that if on the other hand the formulation is only that of a "mathematical model", to guide the method of calculation, then it would seem more objective and heuristically sounder not to create any hypothetical tolerances, but merely to postulate that the proportion of responses affected follows the integrated normal function. For these reasons, the logit analysis is preferred.

III Estimation

We recall the conditional distributions of y when z=1, is $N(A'X, var(u))$, and when z=0, y is $N(B'X, var(u))$ where $A = (a_1, \ldots, a_k)'$, $B = (b_1, \ldots, b_k)'$ and $X = (x_1, \ldots, x_k)'$; $prob(z=1) = F(X'C)$ where $C = (c_1, \ldots, c_k)'$. Thus the joint density of y and z will be f(y,z). If we express the joint likelihood function using matrix algebra, then

$$f(Y,Z|P,X) = h(Y|Z,P,X)g(Z|P,X) \qquad (4)$$

Where, $P=(A',B',C',var(u))'$, X is a nxk matrix, i.e. n observations and k dimensions. For each observation i,

$$h(y(i)|z(i),P,X(i))$$
$$=(2\pi var(u))^{-0.5}\exp\{[y(i)-(X(i)'Az(i)+X(i)'B(1-z(i))]^2/(2var(u))\}$$

and,

$$g(z(i)|P,X(i)) = F(X(i)'C)^{z(i)}[1-F(X(i)'C)]^{(1-z(i))}$$

Let L(P) be the joint likelihood function, L(P) is proportional to the products of h and g, that is

$$L(P) : \prod_{i=1}^{n} h(y(i)|z(i),P,X(i))g(z(i)|P,X(i))$$

$$\log(L) : \sum_{i=1}^{n}\{z(i)\log[F(X(i)'C]+(1-z(i))\log[1-F(X(i)'C)]\}$$
$$-(2var(u))^{-1}\sum_{i=1}^{n}\{y(i)-[A'X(i)z(i)+B'X(i)(1-z(i)]\}^2$$
$$-(n/2)\log(var(u))/2 - (n/2)\log(2\pi) \qquad (5)$$

Estimation of logit parameter:

This is just as Dempster(1972) points out that the joint density of Y and Z in (4) can be factorized into two functions, h and g, which depend on disjoint parameter sets. The maximum likelihood estimators of all the parameters can be found by maximizing these two functions separately. The function g is a log likelihood from a fixed logit model, and function h is just a multivariate general linear regression model.

There are several methods to estimate logit parameters.

Berkson(1955) introduced a method called "minimum Chi square" whose results are asympototically equivalent to the maximum likelihood estimation. Theil(1970) suggested the use of the generalized least-squares method. Both are applicable only in large samples or designed experiments, since in Theil's method one deletes those cells which contain only one or no observations, and in Berkson's method one requires more than one observation per cell. Goodman(1972) used the maximum likelihood estimation in logit analysis, and he found that he got a somewhat smaller variance from MLE than Theil's estimation from weighted least-squares. One disadvantage of MLE is that it takes more computational time.

This thesis adopts the method of maximum likelihood estimation and uses the computer program which is developed by Press and Nerlove(1973). The method can be summarized as follows:

$$L(g(Z|P,X)) = \prod_{i=1}^{n} g(z(i)|P,X(i))$$

$$= \prod_{i=1}^{n} [F(X(i)'C)]^{z(i)} [1-F(X(i)'C)]^{(1-z(i))}$$

Define T^+ as the sum of $X(i)z(i)$, where i runs from 1 to n. T^+ is a sufficient statistic for C, i.e. T^+ is the sum of those $X(i)$ for which $z(i)=1$. Hence C^+, the MLE of C must satisfy

$$[1+\exp(-X(i)'C^+)]^{-1}X(i) = T^+ = \sum_{i=1}^{n} X(i)z(i) \qquad (6)$$

Note log(L) is globally concave[5], so (5) provides an absolute maximum. Hence,

$$p^{+}(i) = [\,1+\exp(-X(i)'C^{+})\,]^{-1}$$

There are many numerical methods by which L can be maximized. The program developed by Press and Nerlove(1973) uses the Fletcher-Powell method of function minimization and the Davidon algorithm[6] for computing the inverse of the information matrix.

Estimation of conditional parameters:

In order to estimate A,B and var(u) by using MLE method, we set $\partial f/\partial A = 0$, $\partial f/\partial B = 0$, and $\partial f/\partial var(u) = 0$. For example, to get A^{+}, the estimator of A; since $(1-z(i))z(i) = 0$, we can solve A^{+} from $\partial f/\partial A = 0$, then we have,

$$A^{+} = [\,X^{1}{}'X^{1}\,]^{-1}X^{1}{}'Y^{1}$$

where X^{1} is nxk matrix, for each observation i, $X^{1}(i) = [\,x_{1}(i)z(i), \ldots, x_{k}(i)z(i)\,]'$, i=1,...,n and $Y^{1} = [\,y(1)z(1), \ldots, y(n)z(n)\,]'$. Similarly,

$$B^{+} = [\,(X^{2})'(X^{2})\,]^{-1}(X^{2})'Y^{2}$$

where X^{2} is nxk matrix, for each observation i, $X^{2}(i) = [\,x_{1}(i)(1-z(i)), \ldots, x_{k}(i)(1-z(i))\,]'$, i=1,...,n and $y^{2} =$

$[y(1)(1-z(1)), \ldots, y(n)(1-z(n))]'$. The estimated variance will be

$$Var^{+}(u) = (Y-Y^{+})'(Y-Y^{+})/n$$

where $Y^{+} = (X^{1})'A^{+}+(X^{2})'B^{+}$. If we prefer to use the unbiased estimator of $var^{+}(u)$ then

$$Var^{+}(u) = (Y-Y^{+})'(Y-Y^{+})/(n-k)$$

Those results are not strange to us. If we split our sample into two groups, one contains all z=1 and the other contains all z=0, and if we apply ordinary least-squares on each group, we will get the same results(Appendix A). We know A^{+} and B^{+} are unbiased. The covariance matrices of A^{+} and B^{+} are

$$cov(A^{+}) = Var^{+}(u)[(X^{1})'(X^{1})]^{-1}$$

$$cov(B^{+}) = Var^{+}(u)[(X^{2})'(X^{2})]^{-1}$$

Estimation of unequal variances of conditional model:

If we relax the condition that u and v have the same distribution, then as shown in Appendix A, we may observe that the estimations of A and B are same as before, but

$$Var^{+}(u) = (Y^{1}-(X^{1})A^{+})'(Y^{1}-(X^{1})A^{+})/(n^{1}-k)$$

$$var^{+}(v) = (Y^{2}-(X^{2})B^{+})'(Y^{2}-(X^{2})B^{+})/(n^{2}-k)$$

$$cov(B^{+}) = Var^{+}(v)[(X^{2})'(X^{2})]^{-1}$$

where n^1 is the total number of observations when $z=1$, and $n^2 = n - n^1$.

IV Hypothesis testing

Logistic model:

For large samples, we set hypotheses about C^+ by using the fact that C^+ is asymptotically normal. Its covariance matrix is obtained from the inverse of its information matrix $I(C^+)$, where

$$I(C^+) = [\partial^2 L(C^+)/\partial C^i \partial C^j] = [\partial^2 g(C^+)/\partial C^i \partial C^j] \qquad i,j=1,\ldots,k$$

Also, any hypothesis about C^+ can be tested by using a likelihood ratio test (Appendix B). The likelihood ratio, r is the ratio of the value of the likelihood function g maximized under the constraints of the hypothesis being tested to the value maximized without constraints. In large samples, the value of $-2\log(r)$ is distributed as Chi square with q degrees of freedom; q is the number of independent restrictions in the null hypothesis.

Conditional model:

Hypothesis about A is H: $A = A^0$, where A^0 is a given vector. For each component under H, we know

$$(a^{+i} - a^{0i})/(w^{ii}\ Var(u))^{0.5}$$

has asymptotic t-distribution with q^1 degrees of freedom, in which $q^1 > 0$; w^{ii} is the ith diagonal element of $((X^1)'X^1)^{-1}$, and q^1 is the difference between m, the number of observations when z=1, and k, the number of independent variables. Since t^2 is distributed as F distribution, we can test the hypothesis using F-test. The ratio is distributed with $F(1,q^1)$.

An alternative way to test the hypothesis is Hotelling's T^2 test. Since $(A^+-A^0)'((X^1)'X^1)^{-1}(A^+-A^0)/(var(u))$ is a Hotelling's T^2 where $T^2 = kq^1F(k,q^1-k+1)/(q^1-k+1)$ and $F'(k,q^1-k+1)$ is an upper tail of the probability function.

Similarly for B, H: $B^+=B^0$, where B^0 is given, for each component under H, we know

$$(b^{+i} - b^{0i})/(r^{ii}\ Var(v))^{0.5}$$

is distributed as $t(q^2)$ where $q^2=n-m-k > 0$, and r^{ii} is the ith diagonal element of $((X^2)'(X^2))^{-1}$. Similarly the hypothesis can be tested by using F-test, $F(1,q^2)$. Or, using Hotelling's T^2, Hotelling's T^2 is $(B^+-B^0)'((X^2)'X^2)^{-1}(B^+-B^0)/(var(u))$, where $T^2 = kq^2F(k,q^2-k+1)/(q^2-k+1)$.

Testing equality between two conditional distributions:

The conditional variable y given z may have the same distribution for different values of z, so we are going to test their equality. We apply Chow's test(1960) to test H: A=B=F, and rewrite the distributions into linear models.

$$Y^1(i) = (y(i)|z(i)=1) = X^1(i)'A + 0B + u(i)$$
$$y^2(i) = (y(i)|z(i)=0) = 0A + X^2(i)'B + v(i)$$

In there, we assume they have equal variance and zero covariance. Under the H then

$$Y^1 = X^{1\prime}F + U$$
$$Y^2 = X^{2\prime}F + V$$

so F is estimated as

$$F^+ = [(X^1,X^2)(X^1,X^2)']^{-1}(X^1,X^2)(Y^1,Y^2)'$$

Let E=(U,V)' then

$$E^{+\prime}E^+ = [(Y^1,Y^2)'-(X^1,X^2)'F^+]'[(Y^1,Y^2)'-(X^1,X^2)'F^+] \qquad (7)$$

E^+ is estimated from the entire sample, so $E^{+\prime}E^+$ has rank n-k. Under the alternative hypothesis A≠B, we have

$$U^{+\prime}U^++V^{+\prime}V^+ = (Y^1-X^1A^+)'(Y^1-X^1A^+) + (Y^2-X^2B^+)'(Y^2-X^2B^+) \qquad (8)$$

$U^{+\prime}U^+$ has rank q^1 and $V^{+\prime}V^+$ has rank q^2. U and V are independent, so the rank of $U^{+\prime}U^++V^{+\prime}V^+$ is $q^1+q^2=n-2k$.

$$\begin{bmatrix} Y^1-X^1F^+ \\ Y^2-X^2F^+ \end{bmatrix} = \begin{bmatrix} Y^1-X^1A^+ \\ Y^2-X^2B^+ \end{bmatrix} + \begin{bmatrix} X^1A^+-X^1F^+ \\ X^2B^+-X^2F^+ \end{bmatrix} \qquad (9)$$

$$||(Y^1-X^1F^+, Y^2-X^2F^+)||^2$$
$$= ||(Y^1-X^1A^+, Y^2-X^2B^+)||^2 + ||(X^1A^+-X^1F^+, X^2B^+-X^2F^+)||^2$$
$$+ \text{ cross product terms}$$

Since the cross product term is zero, so the square on the left of (9) is the sum of squares on the right, that is

$$||(Y^1-X^1F^+, Y^2-X^2F^+)||^2$$
$$= ||(Y^1-X^1A^+, Y^2-X^2B^+)||^2 + ||(X^1A^+-X^1F^+, X^2B^+-X^2F^+)||^2 \qquad (10)$$

or say,

$$Q^1 = Q^2 + Q^3$$

From the estimations of A, B and F, we get

$$(X^{1\prime}X^1+X^{2\prime}X^2)F^+ = X^{1\prime}Y^1 + X^{2\prime}Y^2 = X^{1\prime}X^1A^+ + X^{2\prime}X^2B^+$$

which implies

$$B^+-F^+ = -(X^{2\prime}X^2)^{-1}(X^{1\prime}X^1)(A^+-F^+) \qquad (11)$$

A^+-F^+ is a linear transformation of U and V, so we substitute the estimated functions of A^+ and F^+ in terms of U and V. Then under H we will have

$$A^+ - F^+ = -(X^{1\prime}X^1 + X^{2\prime}X^2)^{-1}(X^1, X^2)'(U,V)' \qquad (12)$$

Substituting (11) and (12) into equation (10), we claim Q^3 has rank k, since rank(Q^2) = n-2k and

$$\mathrm{rank}(Q^1) \le \mathrm{rank}(Q^2) + \mathrm{rank}(Q^3)$$

Therefore, the H can be tested by F ratio

$$F(k,n-2k) = \frac{(||X^1A^+-X^1F^+||^2+||X^2B^+-X^2F^+||^2)(n-2k)}{(||Y^1-X^1A^+||^2 + ||Y^2-X^2B^+||^2)k}$$

V Polytomous Variable

Generalizing the dichotomous variable to a polytomous variable in this model is very easy. The basic structure on estimation and hypotheses testing are mostly the same, therefore in this section, we just bring out the idea of this generalized model and its parameters estimations; hypotheses testing is omitted. Let us assume that the categorical variable z has more than two categories. The distribution of y(i) given $z(i)=a^j$ is normal, $N(X(i)'S^j, \mathrm{var}(u^j))$, where a^j is a scalar, S is a vector kx1, X(i) is a vector kx1; i=1,..., n, n observations, and j=1,...,q, q possible responses on z.

$$p^{ij} = \mathrm{Prob}(z(i)=a^j) = F(X(i)'R^j)$$

where R^j is a vector kx1. Define a transformation t(i,j) as

$$t(i,j) = \prod_{\substack{r\neq j\\ r=1}}^{n} (z(i)-a^{r})/(a^{j}-a^{r}) \qquad r,j=1,\ldots,q;\ i=1,\ldots,n$$

Hence $t(i,j)=1$ when and only when $z(i)=a^{j}$, otherwise $t(i,j)=0$.

Now we define,

i) Y to be a $nqx1$ vector, $Y = (Y^{1}, \ldots, Y^{q})$; where $Y^{j} = (t(i,j)y(1), \ldots, t(n,j)y(n))$ for $j=1,\ldots,q$;

ii) X^{*} to be a block diagonal matrix with dimension $nqxgk$, i.e.

$$X^{*} = \begin{bmatrix} X^{*1} & 0 & \ldots & 0 \\ 0 & X^{*2} & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & X^{*q} \end{bmatrix}$$

for each X^{*j},

$$X^{*j} = \begin{bmatrix} t(1,j)x^{1}(1) & \ldots & t(1,j)x^{k}(1) \\ & \ldots & \\ t(n,j)x^{1}(n) & \ldots & t(n,j)x^{k}(n) \end{bmatrix}$$

iii) S to be a $qkx1$ vector, $S=(S^{1}{}',\ldots,S^{q}{}')'$, and S^{j} is $kx1$ vector, $S^{j}=(s^{1j},\ldots,s^{kj})'$;

iv) U to be a $nqx1$ vector, $U=(U^{1}{}',\ldots,U^{j}{}')'$; $U^{j} = (u(i), \ldots, u^{j}(n))'$, $j=1,\ldots,q$. The disturbance variance-covariance is a direct product of a qxq diagonal matrix D and a nxn unit matrix

I, where $\mathrm{diag}(D) = [\mathrm{var}(u^1), \ldots, \mathrm{var}(u^q)]$.

v) $p^{ij} = \mathrm{Prob}(z(i)=a^j) = \exp(X(i)'R^j) / \sum_{s=1}^{q} \exp(X(i)'R^s)$, and $R = (R^1{}', \ldots, R^q{}')'$ for $R^j = (r^{j1}, \ldots, r^{jk})'$.

As before, we can obtain those parameter estimates by estimating two separated models. Hence, the conditional model is

$$Y = X^* S + U$$

Since $z(i)(1-z(i)) = 0$, this implies X^{*i} and X^{*j} are orthogonal when $i \neq j$ so

$$S^j = ((X^{*j}{}' X^{*j})^{-1} (X^{*j})' Y^j$$

$$\mathrm{Var}^+(u^j) = (Y^j - (X^{*j})' S^{+j})' (Y^j - (X^{*j})' S^{+j}) / (n-k)$$

$$\mathrm{Cov}(S^{+j}) = ((X^{*j})'(X^{*j}))^{-1} \mathrm{Var}^+(u^j)$$

In logistic part, the maximum likelihood function is

$$L = \prod_{i=1}^{n} \prod_{j=1}^{q} (p^{ij})^{t(i,j)}, \qquad \sum_{j=1}^{q} p^{ij} = 1, \qquad \sum_{j=1}^{q} t(i,j) = 1.$$

Also we know $T^j = \sum_{i=1}^{n} X^i t(i,j)$ is sufficient for R given X. So the MLE of R^j can be found by maximizing L subject to the sum of R^j for all j is 0, and R^{+j} must satisfy the equations

$$\sum_{i=1}^{n} [\exp(X^i{}' R^{+j}) / \sum_{s=1}^{q} \exp(X^i{}' R^{+s})] X^i = \sum_{i=1}^{n} X^i t(i,j)$$

We claim the solution to this problem yields a maximum. Since,

$$\log(L) = \sum_{j=1}^{q} T^{j}{}'r^{j} - \sum_{i=1}^{n} \log\Big(\sum_{j=1}^{q} \exp(X^{i}{}'R^{j})\Big)$$

$$\partial^2 \log(L)/\partial R^{m}\,\partial R^{m}{}'$$

$$= -\sum_{i=1}^{n} \frac{\{\sum_{j=1}^{q} \exp[X^{i}{}'(R^{j}-R^{m})]-1\}(X^{i}X^{i}{}')}{\{\sum_{j=1}^{q} \exp[X^{i}{}'(R^{j}-R^{m})]\}^2}$$

Hence the log(L) is concave because $(X^{i}X^{i}{}')$ is positive semi-definite for all i, and exponential funtion is positive.

Footnotes:

[1] Constant term a, is considered as a product of ax^1, where x^1 is a variable always having value 1.

[2] J. Press and M. Nerlove, Univariate and Multivariate Log-linear and Logistic Models, Dec. 1973, pp.10.

[3] J. Press and M. Nerlove, Univariate and Multivariate Log-linear and Logistic Models, Dec. 1973, pp.5.

[4] Group data means many observations per cell.

[5] $\log(L) = -\sum_{i=1}^{n} \{z(i)\log[F(X(i)'C)] + (1-z(i))\log[1-F(X(i)'C)]\}$, F is convex and log function is increasing so the composition function of log(F) is convex. The sum of convex functions is convex, but the negative convex function is concave. Therefore log(L) is concave.

[6] see Box, Davies and Swann(1969) Ch.4 pp. 38 - 39.

Chapter III

System of Equations Model

The basic model proposed in the previous chapter can be expanded in every dimension. This chapter and the following chapter will discuss several extensions of this basic model. Because of the limited scope of this research, the presentation is as follows: we will discuss a simple extension, called system of equations model a little bit more in this chapter, then in the following chapter we will just mention some interesting extension models and leave out all the details. System of equations model is defined as a set of regression or logistic equations. This set can be partitioned into certain number of disjoint groups and each group can form a basic model as we discussed in the previous chapter. In the basic model, our interest concentrates on the joint density of two dependent variables involving continuous variable y and categorical variable z. Here we are not only interested in the joint density of these two dependent variables, but we are also interested in the interaction effects between groups. For example, we may wish to analyze the joint density of the timing of a married woman in labour force and her child not older than 6 years of age within a period 1967-71, but it will be more interesting to consider this problem year by year and observe the interaction effects between years. Let us say, y for given z is distributed normally, so written in matrix notation they will be

$$(y^t(i) \mid z^t(i)=1) \text{ is } N(X^t(i)'A^t, \text{var}(u^t))$$

$$(y^t(i) \mid z^t(i)=0) \text{ is } N(X^t(i)'B^t, \text{var}(v^t))$$

and

$$\text{Prob}(z^t(i)=1) = F(X^t(i)'C^t)$$

for $t=1,\ldots,d$, d groups; $X^t(i)$ is kx1 vector $i=1,\ldots,n^t$. In order to estimate the joint densities of $y^t(i)$ and $z^t(i)$, we have to estimate A^t, B^t, C^t, $\text{var}(u^t)$, and $\text{var}(v^t)$. Let us start with an easy method.

I Estimation equation-by-equation

This method is very simple. The estimation is based on the assumption that the data are independent between groups. Hence, those parameters can be found by considering the whole problem as d separate basic models, and estimating those models one by one.

In most of cases, data across groups are correlated. Hence, this kind of estimation is not efficient. In the following section, we will discuss a method which handles the case when correlation across groups is taken into account.

II Dependence among groups

In the previous chapter we assert that the estimation can be separated into two parts, because the joint density function can be factorized into two functions which depend on disjoint parameter sets. So if we wish to consider the interaction effects between groups, we will observe those effects on the logistic part and the regression part individually.

A) Interaction effects between groups on conditional regressions:

Estimation:

There are several kinds of interaction effects between groups on conditional regressions. In this section we only consider a special one, and we will discuss more about it in chapter 4. Let us assume that the disturbance terms in different groups are highly correlated. Hence under this assumption, the estimators obtained by an equation-by-equation are not in general efficient. Zellner(1962) has proposed an efficient method called "Estimating Seemingly Unrelated Regressions". This method applies Aitken's generalized least-squares to the whole system of equations. For group t, we know:

$$(y^t(i) \mid z^t(i)=1) \text{ is } N(X^t(i)'A^t, \operatorname{var}(u^t))$$

$$(y^t(i) \mid z^t(i)=0) \text{ is } N(X^t(i)'B^t, \operatorname{var}(v^t))$$

then $(y^t(i) \mid z^t(i))$ is distributed normally with mean $X^t(i)'A^t z^t(i) + X^t(i)'B^t(1-z^t(i))$ and variance $\operatorname{var}[u^t(i)z^t(i)] + \operatorname{var}[v^t(i)(1-z^t(i))]$. So if we write in regression equation with matrix algebra then

$$Y^t = X^{*t}{}'S^t + U^{*t}$$

where X^t is a nx2k matrix, for each row observation i, $X^{*t}(i) = [X^t(i)'z^t(i), X^t(i)'(1-z^t(i))]$, Y^t is nx1 vector of observations on the t th group, U^{*t} is nx1 vector which $U^{*t} = [u^t(1)z^t(1) + v^t(1)(1-z^t(1)), \ldots, u^t(n)z^t(n) + v^t(n)(1-z^t(n))]'$, and t=1, ..., d, d groups. So put then together, it will be

$$\begin{bmatrix} Y^1 \\ Y^2 \\ \cdots \\ Y^d \end{bmatrix} = \begin{bmatrix} X^{*1} & 0 & \cdots & 0 \\ 0 & X^{*2} & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & X^{*d} \end{bmatrix} \begin{bmatrix} S^1 \\ S^2 \\ \cdots \\ S^d \end{bmatrix} + \begin{bmatrix} U^{*1} \\ U^{*2} \\ \cdots \\ U^{*d} \end{bmatrix}$$

or,
$$Y = XS + U \qquad (1)$$

where $Y = (Y^{1'}, \ldots, Y^{d'})'$ X is a block-diagonal matrix, in which diagonal is $(X^{*1}, \ldots, X^{*d})$, $S = (S^{1'}, \ldots, S^{d'})'$, $U = (U^{*1'}, \ldots, U^{*d'})'$ to apply Aitken's generalized least-square, we get

$$S^+ = (X'H'HX)^{-1}X'H'Hy = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

where H is an orthogonal matrix such that $E(Huu'H')=H\Sigma H'=I$, and $var(S^+) = (X'\Sigma^{-1}X)^{-1}$, where

$$\begin{aligned}\Sigma^{-1} &= Var^{-1}(U) \\ &= \begin{bmatrix} E^{11}I & \cdots & E^{1d}I \\ & \cdots & \\ E^{d1}I & \cdots & E^{dd}I \end{bmatrix} \\ &= (E^+)^{-1}I\end{aligned}$$

since $(n-2k)E^+ = (n-2k)\,var(U^{*ij}) = U^{*i}{}'U^{*j} = (Y^i - X^{*i}{}'S^{*i})'(Y^j - X^{*j}{}'S^{*j})$, i,j=,...,d, where S^{*i} is estimated from the basic models, $S^{*i} = (A^{*i'}, B^{*i'})$. Hence $(E^{+})^{-1}$ can be estimated. These estimators are more efficient because in estimating the coefficients of a single equation, the Aitken procedure takes account of zero restrictions on coefficients occurring in other equations. Zellner and Huang(1962) pointed out that these estimators have the optimal forecasting properties.

Hypotheses testing:

We may wish to test that the data in the groups are homogeneous in items of their regression coefficient vectors.

$$H: S^{1} = S^{2} = \dots = S^{d}$$

There are several ways to test this hypothesis, but only two are considered. One is as Zellner(1962) suggested that the test statistic can be employed by using a F-test as

$$F(2k(d-1), d(n-2k)) = \frac{d(n-2k)(S^{+})^{-1}D'[DVar(S^{+})D']^{-1}DS^{+}}{2k(d-1)[Y'(E^{+})^{-1}IY - Y'(E^{+})^{-1}IXS^{+}]}$$

where D, the matrix of the restrictions, with dimension (d-1)xd,

$$D = \begin{bmatrix} I & -I & 0 & \dots & 0 & 0 & 0 \\ 0 & I & -I & \dots & 0 & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & 0 & \dots & I & -I & 0 \\ 0 & 0 & 0 & \dots & 0 & I & -I \end{bmatrix}$$

such that DS=0.

Another method is to use the maximum likelihood ratio test which leads to the same result. The general idea of this test has been stated in previous chapter. A detail about applying this test in this hypothesis is shown in Appendix B.

If this hypothesis is true, then there is no aggregation bias in simple linear aggregation. Hence, the estimators taken from the entire sample will be statistically asymptotic equivalent to the parameters estimated from individual groups.

B) Interaction effects between groups on logit model:

As we said, each group can form a basic model as introduced in the previous chapter. Within each group, we have one conditional continuous variable and one qualitative variable, so in this special model we have d conditional continuous variables and d qualitative variables. Therefore, if the qualitative variables are unordered, then we may explore our interest into the more general case of any number of jointly varying dichotomous variables and the probability that a qualitative variable takes on a particular value. This will bring our attention into the relation between the log-linear model of contingency table analysis and the standard logistic model. In this section, we discuss neither this relation, nor the general model of several qualitative dichotomous variables, because they have been done by Nerlove and Press(1973). The model which is suggested by Nerlove and Press is assumed that:

1) all higher order interaction effects vanish,[1]

2) the second order interaction effects are constant and independent of the values of any of the exogenous variables.

3) the main effects are linear functions of the various exogenous explanatory variables.

Parameters are estimated by the maximum likelihood method. The computational algorithm[2] is based on the Fletcher-Powell method of function minimization, and the Davidon algorithm for estimating the inverse of the matrix of the second derivatives, the information matrix[3].

IV Further discussion

In the previous discussion, we only consider the disturbances in the regression equations are correlated. For each period of time, we can also consider the observations are serially correlated. This can be easily solved, since we know how to handle autocorrelation problem in regression. We may detect autocorrelation by conventional test and apply Theil BLUS procedure[4] to get better estimates in our first stage, then we apply Zellner's method to get our result.

Furthermore, it is no great trouble to generalize this model to involve polytomous variables instead of dichotomous variables. In chapter 2, we have discussed about how to generalize our basic model from dichotomous to polytomous. This model is just a bit beyond our basic model, so everything discussed in this chapter are still applicable to polytomous variables. Therefore, polytomous generalization is omitted.

Footnote:

[1] This assumption in the computer program of Nerlove and Press has now been eliminated and in an updated reversion of the program, higher order interaction effects are permitted.

[2] See Nerlove and Press(1973), Appendix A, esp. pp. 92-94.

[3] See Box, Davies and Swann(1969), ch. 4, esp. pp. 38-39, and the references cited therein.

[4] H. Theil, "The Analysis of Disturbances in Regression Analysis", J. Am. Statist. Assoc., vol. 60, pp. 1067 - 1079, 1965.

J. Koerts, "Some Further Notes on Disturbance Estimates in Regression Analysis", J. Am. Statist. Assoc., vol. 62, pp. 169 - 183, 1962.

H. Theil, "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances", J. Am. Statist. Assoc., vol. 63, pp. 242 - 251, 1968.

J. Koerts and A. P. J. Abrahamse, "On the Power of the BLUS Procedure", J. Am. Statist. Assoc., vol. 63, pp. 1227 - 1236, 1968.

Chapter IV

Model Extensions

I Lagged variables model

We suppose that the dependent variables are dependent, not only on the current value of X, but also on the previous value of X. For example, the number of hours worked by wife depends on the economic factors (head's unemployment, wife's wage, ...etc.) of this year as well as the economic factors of the last year. Similarly, the probability that a married woman will bear a baby is dependent upon the economic factors of this year as well as last year. Let us consider the very simple case of one variable, and assume there is multicollinearity in the problem. We may assume that all the coefficients exponentially decrease with respect to time. So let $w = xz + x(1-z)$ then

$$y^{t} = a + bw^{t} + b^{1} w^{(t-1)} + \dots + e^{t}$$

$$\log(p^{t} / (1-p^{t})) = r + sx^{t} + s^{1} x^{(t-1)} + \dots$$

where

$$b^{i} = b*(d)^{i}, \qquad s^{i} = s*(d)^{i}, \qquad i=1,2,\dots, \qquad 0<d<1$$

In regression, we have

$$y^{t} = a + bw^{t} + bdw^{(t-1)} + \dots + e^{t}$$

$$y^{(t-1)} = a + bw^{(t-1)} + bdw^{(t-2)} + \dots + e^{(t-1)}$$

$$y^{t} - dy^{(t-1)} = a(1-d) + b^{t}w + (e^{t} - de^{(t-1)}) \qquad (1)$$

Equation (1) can be estimated easily. Similarly in the logistic part, we have

$$\log[p^{t}/(1-p^{t})] - d*\log[p^{(t-1)}/(1-p^{(t-1)})] = r(1-d) + sx^{t}$$

Similarly, the logit function can be estimated.

II Model with constraints

In some cases we may know some parameters will have meaning only in a certain domain, or interrelationship of parameters may form a constraint. For example, we may consider such a case that the total working hours of the head and the wife must greater than a certain number. We know the maximum likelihood method is suitable to estimate such a model with constraints.

III Model with jointly dependent variables

We may consider inferences about relationships that determine jointly dependent discrete variables, which are both categorical and unordered. For example, the hushand of a married woman is employed or unemployed; she will bear a baby or will not. We may wish to relate the joint probability of these two events to a set of social factors. As discussed in chapter 3, in some cases logistic function can be considered as several qualitative polytomous variables. The solution of such function has been proposed by Press and Nerlove(1973). Therefore the model with jointly dependent variables can be considered as a special case of chapter 3.

IV Simultanous-equation model

We may consider our basic model is composed of two simultanous-equation systems. One system is formed by linear regression equations, and the other system is formed by logistic equations. In regression system, we can apply three-stage least squares which is proposed by Zellner and Theil(1962). This method is known to us. The logistic system has been solved by P. Schmidt and R. Strauss(1974). They are using the maximum likelihood approach and consider it is a special case of the model of Nerlove and Press(1973). Many social problems can be analyzed by such model. For example, the expenditure of a family will depend upon the hours worked by the wife, and other factors, and also the hours worked by wife will depend upon family expenditure and the other factors. The probability of a family going on vacation will depend on whether the wife bears a baby. So if we wish to know the joint probability of a family going on vacation and its annual expenditure, we may apply this model.

V Recursive model

This is the most interesting model proposed. This is a very new study area. There is no formal literature about this kind of model. In this model, the continuous variable and the discrete variable are inter-dependent. That is

$$E(y|X,z) = f(X,z) \qquad z=0,1 \qquad (2)$$

$$p = \text{Prob}(z=1|y,X) = [\,1+\exp(-C'X-dy)\,]^{-1} \qquad (3)$$

P. Schmidt and R. Strauss have discussed this problem, but they only considered it as a start of this topic. They consider how to maximize the likelihood function. Their suggestion does not seem to be novel and is expensive to compute. The following discussion can be considered as an initial step in attacking this model, and we hope it is a step in the right direction.

If we rewrite equation (3), then it becomes

$$w = \log[\,p/(1-p)\,] = C'X + dy \qquad (4)$$

Since z is a categorical variable, we will apply dummy variables in regression (2). Hence, equations (2) and (4) form a simultaneous regression equation system. This we can solve either by three-stage least-squares, or the full information maximum likelihood method. The full information maximum likelihood method is an expensive computational method, and it will be very difficult to obtain parameter estimates when the number of degrees of freedom is large. Three-stage least-squares is an extension of two-stage least-squares, which we mentioned in chapter 3. It is more efficient than two-stage[1] if the disturbances in various structural equations are correlated. Both methods are described in many econometric text books.

Footnote:

[1] A. Zellner and H. Theil, "Three-stage, Least-squares: Simultaneous Estimation of Simultaneous Equations", Econometrica vol. 30, pp. 54 - 78, 1962.

Chapter V

A Study on Labour Supply of Married Women

--- Model Description ---

I Introduction

The empirical literature on female labour supply, especially for married women, is not much. Related studies are J. Korbel(1962), J. Mincer(1963), G. Cain(1966), S. Hoffer(1973), R. Freeman(1973), and E. Berndt and T. Wales(1974).

In this chapter, our study is to observe the labour force participation of married women in different situations, and the determination of these situations in the United States over the five-year period 1967-71. This study will be divided into two parts: the first part is to study our economic problem using five years data, and the second part is to study the problem yearly. Our data is drawn from the University of Michigan Survey Research Centre Panel Study of Income Dynamics(1972). There were 2500 family units randomly chosen, and each family unit was re-interviewed annually over the 1967-71 time period. As E. Berndt and T. Wales(1974) pointed out this period, 1967-71 was of particular interest since the national unemployment rate for women aged 20 and over varied considerably from 3.8% and 3.7% in 1968 and 1969 to 5.7% in 1971; further, toward the end of this period an increasing emphasis was placed on eliminating discrimination against working women.

II Specification of models

Although this study will be divided into two parts, they still have the same economical structure. In the first part, the model is built upon chapter 2 and in the second part, the model is built upon chapter 3. 340 family units are selected. The analyses are based on these family units. Every year, these family units are partitioned into two groups. Group I contains all the family units which have a child of 6 years or younger, and group II contains the rest. It is very obvious that these two groups are disjointed, but it is not true that the number of elements in each group is fixed for every year, because the age of the youngest child is increasing. Suppose the youngest child of a family was 5 years old in 1967, and the family had no further new born child within the time period 1967-71. Then in 1967 and 1968, this family belongs to group I, but in 1969-71, this family belongs to group II. Therefore, group I of 1967 and group II of 1970 are not disjointed. In here we call a family constraint variable z to be 1 when a family has a child not older than 6 years, and 0 otherwise. Hence, group I contains all the family units in which z=1, and group II is when z=0. We assume that the labour force activity of the wife in group I will be less than those in group II. We wish to observe how the timing of these two groups of married women in labour force will be determined differently, due to the same economic factors(or independent variables). Also we wish to relate the probability that a family has a child not older than 6 years of age to the same explanatory variables. In the following discussion we will call the conditional regression equations, labour equations, and the logistic function, a probability function.

III Specification of variables

Dependent variables:

Categorical dependent variable:

In this study, our categorical variable is z, whose value is 1 when a family has a child not older than 6 years of age, otherwise the value of z is 0. At the beginning z was defined as $z=1$ when a family has a new born child otherwise $z=0$. This definition of z led into statistical insignificance in the models because of our lack of observation in 1970 and 1971, and most of the wives did nct work when $z=1$. Furthermore, a wife will work less not only because of having a baby, but also because of her commitment to her family work. Suppose she has a child who is not of school age, or even if her child is in grade I, she may like to lock after her child rather than to work outside.

Continuous dependent variable:

Our continuous variable y, is the timing in the labour force of a married woman: that is her annual worked hours.

Independent variables:

At the beginning, the explanatory variables are centered around the following economic factors: birth gap, predicted wage of wife, head's income including income from elsewhere, unemployment of head, fecundity, and the ratio of incomes over needs.

BirthGap -- Birth gap

Birth gap is defined as expected completed family size minus actual number of children in year t. Expected completed family size is the total number of children expected and decided by a couple. These figures can be found on the survey data so they are actual data. We will expect that the larger the birth gap, the greater the probability that z=1.

Wife wg -- Wife expected wage

The predicted wage of the wife is measured according to the result of E. Berndt and T. Wales(1974). A wife with a higher predicted wage would tend to keep on working more and would try to avoid having a baby, or she would like to go back to the labour force as soon as possible. Therefore, if our assumption is correct the predicted wage will be positively correlated with y, the number of hours worked by the wife. Following the same argument, we can also assume that it is positively correlated with the probabilty of z=1. Many married women are family oriented. They would like to work at home, or given more care to their children rather than to work outside, or to work a few hours for pleasure. Therefore, in those cases, the predicted wage of the wife will correlated with y, or Prob(z=1) negatively. Hence, the relationships between (Wife wg) and y; between (Wife wg) and Prob(z=1) are in quadratic shape. That introduces a new variable: the $(\text{Wife wg})^2$ in our models.

Head inc -- Head's wage plus transfer income

The variable, the head's income is very similar to the predicted wage of the wife. It varies in U-shape too. This variable includes head's wage, aid to dependent children, pensions, incomes from welfare, social security, unemployment or workmen's compensation and alimony or child support. All the sources except the head's wage are called transfer income. Usually a wife has to work more when her family income is low. She would try to re-enter labour force quickly. On the other hand, just because the family income is high, that does not necessarily mean that the wife will work less, or have a higher probability of having a baby, because a high income family may have a high transfer income and a low head's wage. In this case, the wife will work more, and put less time into her house work in order to make her family economy stable. Hence $(\text{Head inc})^2$ is included in the models.

Unemploy -- head's unemployment

The relationship between y and the unemployment of the head(given in days), or between Prob(z=1) and the head's unemployment(given in weeks), is unexpected. Normally we would think when a hushand is unemployed, the wife would have to work more and leave the family affair behind. Unexpectedly, we get a negative result in testing our model I on the pulled data (Appendix C). It was not because the entire economy was bad within the period 1967-71. This outcome may be explained in the following ways. Some of the heads may be seasonal workers with high wage such that it is not necessary for their wives to work more. Another possible reason is a family may move from one town to another town because the head can not find a job in his own town. This will cause the wife to lose her job. Sometime when a person changes his job, he would take this opportunity to have a longer vacation. This will also cause his wife to work less. When we plotted out the data, using the number of hours worked by wife against the unemployment of the head in days, we found that data are distributed more or less in a quadratic form. Therefore, we add $(\text{Unemploy})^2$ as another new variable.

Fecundit -- Wife's fecundity

Fecundity is an age variable defined as 45 minus the age of wife at time t. We consider the relationship between fecundity and y, or between fecundity and Prob(z=1) is quadratic. $(\text{Fecundit})^2$ is added because a younger woman has less family work and better physical ability to work more, but she does have a high probability to bear a baby, which forces her to work less.

Inc/need -- ratio of total incomes except wife's wage over needs

Incomes per needs is defined as the total family net real income[1] minus the wife's wage and divided by the family needs[2]. Even if this ratio is high it does not always mean that the family has a high income, since the number of dependents may be less. Similarly, even if the ratio is low that may be caused by large family size. Most likely, this ratio varies with y or Prob(z=1) in a non-linear pattern, so we also consider $(\text{inc/need})^2$ in our model.

Finally in our models, we have 11 explanatory variables. They are:

1.	BirthGap	---	birth gap
2.	Wife wg	---	predicted wage of wife
3.	$(\text{Wife wg})^2$	---	square of Wife wg
4.	Head inc	---	head's income plus transfer income
5.	$(\text{Head inc})^2$	---	square of Head inc
6.	Unemploy	---	unemployment of head
7.	$(\text{Unemploy})^2$	---	square of Unemploy
8.	Fecundit	---	fecundity
9.	$(\text{Fecundit})^2$	---	fecundity square
10.	inc/need	---	ratio of total incomes except wife's wage over needs
11.	$(\text{inc/need})^2$	---	square of inc/need

IV Data restriction

The data sample is obtained from the University of Michigan Survey Research Centre, Panel Study Of Income Dynamics (1972) which is based on 5 annual surveys. We restrict data in our analysis of the models by using the following constraints:

1. The husband and wife were present in the household in all 5 years, 1967-71.
2. The head was married for the first time.
3. The husband worked for at least 350 hours in each of the 5 years.
4. The wife was not older than 45 years old in 1971.
5. The head was less than 50 years of age in 1971.
6. The birth gap, the family expected size minus the actual family size, was positive in all 5 years.

These constraints are used to eliminate all the special cases so that our analysis will be based on a more reliable sample. Suppose the first two constraints are used to eliminate those abnormal households. In some cases the wife will have been married before, and have children from the earlier marriage. Hence these two constraints ensure that the children in the faimly belong to the couple. In most of the families, the head is responsible for the family economy. The family expenditure is mainly dependent upon his earning. Hence, constraint 3 is used to ensure the stability of family economy. An observation is valid or interesting only if the wife is of child bearing age, or the children are not old enough to be left alone while she works. This is the usage of constraint 4. Restriction 5 is used to prevent case such that a young girl is married to an old millionaire. Such observations are not interesting, because data are biased. Constraint 6 is to ensure that the birth gap has statistical meaning. This leaves us with

1700 observations for the 5 years period: that is 340 family units in total.

Footnotes:

[1] The total family net real income is defined as the total real income minus the cost of earning income, minus help from outside the family unit, if there are children under 18.(see definition in A Panel Study Of Income Dynamics)

[2] The Family needs is adjusted according to the US annual living need standard in year t.(see definition in A Panel Study Of Income Dynamics)

Chapter VI

Empirical Results of Model I

This model is estimated by using the model described in chapter 2. We have 1700 observations in total, that is 340 family units in all the 5 years. In considering the timing of the married women participating in the labour force, we split the sample into two categories according to the age of the youngest child in the family. If a family has a child not older than 6 years of age, which is z=1, we put it into group I; the others we put into group II. Hence, group I(under the condition that z=1) has 885 observations, and group II(under the condition that z=0) has 815 observations. From table II, we observe that on the average, the number of hours worked by a wife in group I is much less (about 47%) than those worked by a wife in group II. This supports our assumption that a wife with a child not older than 6 years of age will work less hours.

I. Results from the labour regression equations:

From Table I, we find that the labour equation of group I has a bigger constant term than the equation of group II. Suppose we keep all the explanatory variables fixed and let the wife's wage and its square vary with the hours worked by the wife; then we find in group I, the curve is concave upward but in group II, the curve is concave downward. From the figures shown on table I, we know that the hours worked by those housewives having children not older than 6 years of age, normally will not be affected by their wages. They will work more only when their wages are high, except for those who are ambitious to work. Those wives whose children are older than 6 years of age, will work only up to a certain number of hours. In the labour equation of group I, we find that 1) wife's wage, 2) the head's wage plus transfer income, 3) the ratio of incomes over needs and, and 4) the squares of 1), 2), and 3) are statistically significant in the regression of the number of hours worked by the wife on the explanatory variables. On the other hand in group II, the hours worked by the wife is significantly affected by 1) the head's wage plus transfer income, 2) the ratio of incomes over needs, 3) fecundity, 4) the squares of 1), 2) and 3), and 5) the square of the wife's wage.

Hypothesis testing -- the equality of two labour equations

In the above discussion, we observe that each labour equation has its own characteristic structure, but now we assume the equations are equal, in order to test the equality of these two equations. From Chow's test, we find $F(12,1676) = 11.7861$ that is far beyound 95% significance value ($F=2.30$), so we reject the hypothesis of equality.

II Results from the probability equation

From Table III, we find the birth gap, the head's income, fecundity and the square of fecundity have a strongly significant effect on the probability of z=1. This is a very good outcome because these major variables turn out as we expected in chapter 5. Also, the ratio of incomes over need, the wife's wage and the square of the wife's wage, do sightly affect the probability of z=1.

We scale the units of the head's income, the ratio of incomes over needs, the unemployment of the head, and the squares of these variables, to speed up the convergence rate in maximizing the maximum likelihood function; otherwise convergence is very difficult to obtain.

III Further estimation:

If we consider those statistically significant variables which we obtain from the results of the estimation of this model as reliable variables, we would like to re-estimate the model only with those variables. The following are the new estimates.

Regression model -- Labour equation

Group I, (y | z = 1)

HourWork

= 1193.0 - 633.7(Wife wg) + 182.0(Wife wg)^2 - 0.1772(Head inc) + 0.25x10^{-5}(Head inc)^2 + 5.789(inc/need) - 0.0034(inc/need)^2

Group II, (y | z = 0)

HourWork

$= 792.1 - 35.57(\text{Wife wg}) - 0.1676(\text{Head inc}) + 0.22\times10^{-5}(\text{Head inc})^2 + 6.197(\text{inc/need}) - 0.0032(\text{inc/need})^2 + 42.15(\text{Fecund})$

Logistic model -- Probability equation

$\text{Logit}(p^+(i))$

$= -1.883 - 0.3412(\text{BirthGap}) + 0.5426(\text{Wife wg}) - 0.1556(\text{Wife wg})^2 + 0.0494(\text{Head inc}) - 0.1616(\text{inc/need}) + 0.1895(\text{Fecundit}) - 0.0037(\text{Fecund})^2$

Table I

Parameter Estimates for Labour Equations

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y unconditional
constant	1699.6306* (73.384)	472.9470* (21.344)	1180.0698* (71.507)
BirthGap	-44.7253 (1.378)	-36.7626 (1.708)	6.6327 (0.392)
Wife wg	-704.2543* (3.274)	364.1801 (1.774)	-130.0699 (0.863)
(Wife wg)^2	190.1427* (3.896)	-114.0060* (2.582)	31.8822 (0.958)
Head inc	-0.2082* (14.829)	-0.1744* (13.536)	-0.1949* (21.289)
(Head inc)^2	0.2918×10^{-5}* (9.931)	0.2348×10^{-5}* (7.248)	0.2617×10^{-5}* (13.436)
Unemploy	3.0143 (1.176)	-0.7786 (0.275)	1.4487 (0.749)
(Unemploy)^2	-0.0220 (1.403)	-0.0123 (0.666)	-0.0188 (1.547)
inc/need	6.6705* (8.978)	6.1337* (14.408)	6.1123* (17.821)
(inc/need)^2	-0.4145×10^{-2}* (4.455)	-0.3154×10^{-2}* (8.795)	-0.3184×10^{-2}* (9.775)
Fecundit	-12.6039 (0.629)	43.2141* (3.130)	5.2290 (0.469)
(Fecundit)^2	-0.5377 (0.785)	-1.7835* (3.309)	-1.1625* (2.833)
Observations	885	815	1700
R^2	0.2538	0.3201	0.2789

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Table II

Mean and Standard Deviation of the Model I

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y unconditional
HourWork	771.519 (792.743)	1133.79 (761.958)	945.196 (798.696)
BirthGap	0.3785 (0.7627)	0.6025 (1.4130)	0.4859 (1.1277)
Wife wg	1.8497 (0.5728)	1.9598 (0.6316)	1.9025 (0.604)
(Wife wg)^2	3.7491 (2.4977)	4.2393 (2.8939)	3.9841 (2.705)
Head inc	8297.75 (4569.88)	8933.23 (5243.44)	8602.40 (4913.16)
(Head inc)^2	0.897×10^8 (0.209×10^9)	0.107×10^9 (0.175×10^9)	0.981×10^8 (0.194×10^9)
Unemploy	5.1388 (21.5523)	4.1931 (19.8563)	4.6854 (20.7558)
(Unemploy)^2	490.385 (3509.11)	411.371 (3028.43)	452.505 (3286.73)
inc/need	285.906 (132.113)	332.063 (173.459)	308.034 (155.012)
(inc/need)^2	0.992×10^5 (0.108×10^6)	0.140×10^6 (0.184×10^6)	0.119×10^6 (0.151×10^6)
Fecundit	15.9503 (5.8996)	11.1767 (6.9893)	13.6618 (6.871)
(Fecundit)^2	289.177 (175.181)	173.709 (189.114)	233.821 (190.871)
Observation	885	815	1700

Standard deviations are in parentheses.

Table III

Probability Function Estimates of the Model I

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-1.94222*	0.3079	6.308
BirthGap	-0.33526*	0.0412	8.134
Wife wg	0.50063*	0.2557	1.958
$(\text{Wife wg})^2$	-0.14714*	0.0564	2.609
Head inc	0.06981*	0.0158	4.429
$(\text{Head inc})^2$	-0.00050	0.0003	1.505
Unemploy	0.20558	0.3257	0.631
$(\text{Unemploy})^2$	-0.09577	0.1999	0.479
inc/need	-0.17239*	0.0747	2.307
$(\text{inc/need})^2$	0.00019	0.0084	0.022
Fecundit	0.18811*	0.0216	8.726
$(\text{Fecundit})^2$	-0.00357*	0.0008	4.627

Log of likelihood function = -947.522 after 11 iterations.

* significant level of 5% under H: parameters = 0.0

note:

1 Head inc = $1,000   1 $(\text{Head inc})^2$ = $1,000,000

1 inc/need = 100   1 $(\text{inc/need})^2$ = 10,000

1 Unemploy = 100 days of head's unemployment

1 $(\text{unemploy})^2$ = 10,000 days.

Chapter VII

Empirical Results of Model II

Before we begin to analyze the result, let us consider the annual hours worked by wives. From table V, we observe that in general, those wives whose children are older than 6 years of age will work more hours. The range is from 30.17% to 57.95%[1]. In comparing fecundities we know that on the average those wives in group I are younger and their wages are less than those in group II. These are consistent with the findings of Berndt and Wales(1974).

I Results of single equation estimation

From table IV, we find that the head's wage plus transfer income, and the ratio of incomes over needs are statistically significant in most of the years, but the birth gap is not significant in any year. The degrees of significance of the rest of the explanatory variables vary in different years. If we compare the functional structures year-by-year, we will observe that they have different shapes. Suppose we keep all the explanatory variables constant except the unemployment of the head and its square, and compare 1968 with 1969; then we will find in 1968, the curve is concave downward, but in 1969 it is upward. Downward or upward will give us different interpretations. An upward curve will point out that there are big increases in two extremes. A downward curve will indicate to us that the variable will be meaningful only in a certain domain, because we are not interested in negative values of the dependent variable.

Hypothesis testings -- the equality of the two labour equations in each year:

The hypothesis testings of the equality of the two labour equations in 1967 using a 5% critical point, and in 1970 using a 1% critical point are acceptable; in all other years the hypotheses are not acceptable. These tell us that for 1967 or 1970 we can aggregate all the data and express the labour relation for that year using a linear function, regardless of what the value of z is.

II Results of Zellner's seemingly least squares method

If heteroscedasticity is present, then the weighting in a single equation, is not appropriate. The reason is that in an equation-by-equation method, we use the least-squares procedure to obtain the parameter estimates, but all variables are given the same weight. This weight is unsatisfactory in our sample. Suppose we use the two stage Aitken method as introduced by Zellner(1962); then we find that the regression parameters so obtained are at least asymptotically more efficient than those obtained by an equation-by-equation method using ordinary least-squares.

The comparison of the results from this method with the results from single equation estimation

In table VI, the results show a significant reduction in the deviations of estimated parameters. The range for reduction is from 10% to 20%. We find that there are apperently different quadratic forms. Such differences are possible because we expect the two stage Aitken's estimators are more efficient than those from ordinary least-squares. In the two stage Aitken's method we assign different weights to the sample, so the result should be somewhat different than the result of the ordinary least-squares. Most of the changes do not affect the economical interpretation, but some do. Suppose that in the labour equations of group II in table VI E, there are different interpretation of the wife's wage and the head's unemployment. For example, if we keep all the independent variables fixed except the wife's wage and its square, we will find from the single equation method that the curve is concave downward, but from two stage method it is concave upward. Hence, if we adopt the result from the single equation, we will say that those wives whose children are older than 6 years of age will not be stimulated to work more by their expected wages. On the other hand, if we use the result from the two stage estimation, we will observe that those wives will work more in both extremes. Likewise, we find a similar difference in the head's unemployment from this table. Such examples can be easily found from the results of other years. Another significant difference is in two stage estimation, none of the hypotheses of equality tested is acceptable. It is not like the single equation estimation where we do have some statistical significance.

Testing for Aggregation Bias for Model II:

Before we go into the logistic model, let us consider the aggregation bias. Our testing hypothesis asserts that data in every year are homogeneous insofar as regression coefficient vectors are concerned. That is

H: coeff of 67 = coeff of 68 = ... = coeff of 71

There are two testing methods: the method suggested by Zellner using an F-test, and the likelihood ratio test. Here we use the F-test approach. From the test we find that for the labour equation of group I, $F(48,1640) = 3.4044$ and for the labour equation of group II, $F(48,1640) = 58.4047$. Both are rejected[2]. Therefore, we conclude that there is an aggregation bias involved in single linear aggregation.

The estimated labour equations:

The about discussion shows that there is a significant reduction in the deviations of estimated parameters by using Zellner's two stages estimation method. The estimated equations can be summarized as follows:

Group I, $(y \mid z = 1)$

1967:

$$\text{HourWork} = -24.88 + 206.3(\text{BirthGap}) + 360.0(\text{Wife wg}) - 50.99(\text{Wife wg})^2 - 0.2742(\text{Head inc}) \quad 0.621 \times 10^{-5}(\text{Head inc})^2 + 37.26(\text{Unemploy}) - 0.5963(\text{Unemploy})^2 + 10.54(\text{inc/need}) - 0.0095(\text{inc/need})^2 + 20.42(\text{Fecundit}) - 1.512(\text{Fecundit})^2$$

1968:

$$\text{HourWork} = 719.5 + 58.64(\text{BirthGap}) + 54.45(\text{Wife wg}) + 8.394(\text{Wife wg})^2 - 0.1846(\text{Head inc}) + 0.242 \times 10^{-5}(\text{Head inc})^2 + 6.746(\text{Unemploy}) - 0.0856(\text{Unemploy})^2 + 6.638(\text{inc/need}) - 0.0031(\text{inc/need})^2 + 4.671(\text{Fecundit}) - 1.278(\text{Fecundit})^2$$

1969:

$$\text{HourWork} = 1074.0 + 44.85(\text{BirthGap}) - 240.5(\text{Wife wg}) + 81.98(\text{Wife wg})^2 - 0.2054(\text{Head inc}) + 0.267 \times 10^{-5}(\text{Head inc})^2 - 5.1539(\text{Unemploy}) + 0.1846(\text{Unemploy})^2 + 9.697(\text{inc/need}) - 0.0073(\text{inc/need})^2 - 52.76(\text{Fecundit}) + 0.3216(\text{Fecundit})^2$$

1970:

$$\text{HourWork} = 1413.0 - 92.92(\text{BirthGap}) - 653.0(\text{Wife wg}) + 157.1(\text{Wife wg})^2 - 0.2830(\text{Head inc}) + 0.470 \times 10^{-5}(\text{Head inc})^2 + 4.514(\text{Unemploy}) - 0.0290(\text{Unemploy})^2 + 10.68(\text{inc/need}) - 0.0074(\text{inc/need})^2 -$$

$15.17(\text{Fecundit}) - 0.8335(\text{Fecundit})^2$

1971:

HourWork

$= 1330.0 - 58.39(\text{BirthGap}) - 473.8(\text{Wife wg}) + 130.3(\text{Wife wg})^2 - 0.2869(\text{Head inc}) + 0.648 \times 10^{-5}(\text{Head inc})^2 + 4.826(\text{Unemploy}) - 0.0239(\text{Unemploy})^2 + 8.738(\text{inc/need}) - 0.0062(\text{inc/need})^2 - 19.41(\text{Fecundit}) - 0.2104(\text{Fecundit})^2$

Group II, (y | z = 0)

1967:

HourWork

$= 1704.0 + 6.748(\text{BirthGap}) - 789.1(\text{Wife wg}) + 100.5(\text{Wife wg})^2 - 0.2643(\text{Head inc}) + 0.433 \times 10^{-5}(\text{Head inc})^2 - 1.174(\text{Unemploy}) - 0.0102(\text{Unemploy})^2 + 10.37(\text{inc/need}) - 0.0067(\text{inc/need})^2 + 29.11(\text{Fecundit}) - 1.692(\text{Fecundit})^2$

1968:

HourWork

$= 1207.0 - 11.95(\text{BirthGap}) - 524.5(\text{Wife wg}) + 73.18(\text{Wife wg})^2 - 0.2294(\text{Head inc}) + 0.363 \times 10^{-5}(\text{Head inc})^2 - 0.2577(\text{Unemploy}) + 0.0541(\text{Unemploy})^2 + 11.08(\text{inc/need}) - 0.0077(\text{inc/need})^2 - 23.43(\text{Fecundit}) + 0.2166(\text{Fecundit})^2$

1969:

HourWork

$$= -61.86 - 53.81(\text{BirthGap}) + 479.7(\text{Wife wg}) - 135.4(\text{Wife wg})^2 - 0.1750(\text{Head inc}) + 0.257\times10^{-5}(\text{Head inc})^2 + 6.129(\text{Unemploy}) - 0.0385(\text{Unemploy})^2 + 7.854(\text{inc/need}) - 0.0050(\text{inc/need})^2 + 75.90(\text{Fecundit}) - 3.208(\text{Fecundit})^2$$

1970:

HourWork

$$= 47.04 - 34.05(\text{BirthGap}) + 590.2(\text{Wife wg}) - 146.9(\text{Wife wg})^2 - 0.1548(\text{Head inc}) + 0.199\times10^{-5}(\text{Head inc})^2 - 3.178(\text{Unemploy}) - 0.0035(\text{Unemploy})^2 \quad 5.528(\text{inc/need}) - 0.0025(\text{inc/need})^2 + 74.95(\text{Fecundit}) - 3.702(\text{Fecundit})^2$$

1971:

HourWork

$$= 839.7 - 112.1(\text{BirthGap}) - 104.1(\text{Wife wg}) + 8.551(\text{Wife wg})^2 - 0.1368(\text{Head inc}) + 0.176\times10^{-5}(\text{Head inc})^2 + 3.280(\text{Unemploy}) - 0.0163(\text{Unemploy})^2 + 4.813(\text{inc/need}) - 0.0021(\text{inc/need})^2 + 29.40(\text{Fecundit}) - 1.030(\text{Fecundit})^2$$

III Results of probability functions

From table VII we find the birth gap is very significant in all the probability functions. The results in the years 1968 and 1969 are interesting. In 1968, the probability function is significantly affected by the birth gap, the head's wage plus transfer income, fecundity and the square of fecundity. In 1969, the function is affected by most of the variables, such as the birth gap, the head's unemployment, the ratio of incomes over needs, fecundity, the wife's wage and the squares of the head's unemployment and the wage's wage. One thing which has surprised us is that the head's wage plus transfer incomes does not significantly affect the probability function that happens in most of the years except 1968, and 1971. Moreover, we do not draw any fruitful conclusion from the results of the years 1970 and 1971.

Test for aggregation bias:

In considering the test of the aggregation bias for this part, we find the ratio of maximum likelihood is so big(Chi square(48) = 0.39×10^9) that we can not accept the hypothesis that there is no aggregation bias.

The estimated probability functions:

We conclude the estimation of probability functions for each year as follows.

1967:

$$\text{Logit}(p^+(i))$$
$$= -2.428 - 0.4167(\text{BirthGap}) + 0.3549(\text{Wife wg}) - 0.0962(\text{Wife wg})^2 + 0.0651(\text{Head inc}) + 0.488 \times 10^{-4}(\text{Head inc})^2 + 0.0283(\text{Unemploy}) - 0.0005(\text{Unemploy})^2 - 0.0221(\text{inc/need}) - 0.0281(\text{inc/need})^2 + 0.2835(\text{Fecundit}) - 0.0073(\text{Fecundit})^2$$

1968:

$$\text{Logit}(p^+(i))$$
$$= -3.055 - 0.4486(\text{BirthGap}) + 1.002(\text{Wife wg}) - 0.2201(\text{Wife wg})^2$$

$+ 0.1150(\text{Head inc}) - 0.0005(\text{Head inc})^2 - 0.0109(\text{Unemploy}) + 0.0001(\text{Unemploy})^2 - 0.3319(\text{inc/need}) - 0.0028(\text{inc/need})^2 + 0.2486(\text{Fecundit}) - 0.0049(\text{Fecundit})^2$

1969:

$$\text{Logit}(p^+(i))$$
$$= -2.390 - 0.4669(\text{BirthGap}) + 1.671(\text{Wife wg}) - 0.3913(\text{Wife wg})^2 + 0.0717(\text{Head inc}) - 0.0003(\text{Head inc})^2 + 0.0534(\text{Unemploy}) - 0.0011(\text{Unemploy})^2 - 0.4832(\text{inc/need}) + 0.0289(\text{inc/need})^2 + 0.1198(\text{Fecundit}) - 0.0002(\text{Fecundit})^2$$

1970:

$$\text{Logit}(p^+(i))$$
$$= -2.035 - 0.2968(\text{BirthGap}) + 1.258(\text{Wife wg}) - 0.3676(\text{Wife wg})^2 + 0.0887(\text{Head inc}) - 0.0017(\text{Head inc})^2 + 0.0016(\text{Unemploy}) + 0.411\times10^{-4}(\text{Unemploy})^2 - 0.2755(\text{inc/need}) + 0.0144(\text{inc/need})^2 + 0.0776(\text{Fecundit}) + 0.0016(\text{Fecundit})^2$$

1971:

$$\text{Logit}(p^+(i))$$
$$= -1.297 - 0.1922(\text{BirthGap}) + 0.2137(\text{Wife wg}) - 0.1016(\text{Wife wg})^2 + 0.1593(\text{Head inc}) - 0.0059(\text{Head inc})^2 - 0.0049(\text{Unemploy}) + 0.340\times10^{-4}(\text{Unemploy})^2 - 0.1830(\text{inc/need}) + 0.0075(\text{inc/need})^2 + 0.0500(\text{Fecundit}) + 0.0023(\text{Fecundit})^2$$

IV Further estimation

If we consider those statistically significant variables which we obtain from the results of the estimation of this model as reliable variables, we would like to re-estimate the model only with those variables. The new estimates are shown as follows:[3]

Regression model -- labour equation

Group I, $(y \mid z = 1)$

1967:

Hourwork

$= 592.6 - 0.2331(\text{Head inc}) + 0.51\times10^{-5}(\text{Head inc})^2 + 45.87(\text{Unemploy}) - 0.7770(\text{Unemploy})^2 + 8.262(\text{inc/need}) - 0.0064(\text{inc/need})^2$

1968:

HourWork

$= 1768.0 + 187.8(\text{Wife wg})^2 - 0.0763(\text{Head inc}) + 0.14\times10^{-5}(\text{Head inc})^2$

1969:

HourWork

$= 732.6 + 51.77(\text{Wife wg})^2 - 0.1448(\text{Head inc}) + 0.13\times10^{-5}(\text{Head inc})^2 + 3.327(\text{inc/need})$

1970:

HourWork

$= 640.8 - 0.2325(\text{Head inc}) + 0.37\times10^{-5}(\text{Head inc})^2 + 7.428(\text{inc/need}) - 0.0048(\text{inc/need})^2$

1971:

HourWork

$= 291.1 - 0.1459(\text{Head inc}) + 7.100(\text{inc/need}) - 0.0050(\text{inc/need})^2$

Group II, (y | z = 0)

1967:

HourWork

HourWork

$= 455.6 - 0.196(\text{Head inc}) + 0.26\text{x}10^{-5}(\text{Head inc})^2 + 8.959(\text{inc/need}) - 0.0061(\text{inc/need})^2$

1968:

HourWork

$= 332.5 - 0.1308(\text{Head inc}) + 8.561(\text{inc/need}) - 0.0060(\text{inc/need})^2$

1969

HourWork

$= 655.8 - 0.1046(\text{Head inc}) + 6.096(\text{inc/need}) - 0.004(\text{inc/need})^2$

1970:

HourWork

$= 649.7 - 0.1521(\text{Head inc}) + 0.20\text{x}10^{-5}(\text{Head inc})^2 + 5.469(\text{inc/need}) - 0.0028(\text{inc/need})^2 + 62.44(\text{Fecundit}) - 3.204(\text{Fecundit})^2$

1971:

HourWork

$$= 919.0 - 0.131(\text{Head inc}) + 0.16\times10^{-5}(\text{head inc})^2 + 4.425(\text{inc/need}) - 0.0021(\text{inc/need})^2$$

Logistic model -- Probability equation

1967:

$$\text{logit}(p^+(i)) = -1.475 - 0.4866(\text{BirthGap}) + 0.2291(\text{Fecundit}) - 0.0058(\text{Fecundit})^2$$

1968:

$$\text{Logit}(p^+(i)) = -1.004 - 0.6208(\text{BirthGap}) + 0.0972(\text{Fecundit})$$

1969:

$$\text{Logit}(p^+(i)) = -1.086 - 0.5104(\text{BirthGap}) + 0.0100(\text{Fecundit}) + 5.609(\text{Unemploy}) - 11.56(\text{Unemploy})^2$$

1970:

$$\text{Logit}(p^+(i)) = 0.3183 - 0.0849(\text{Wife wg})^2$$

1971:

$$\text{Logit}(p^+(i)) = 0.2879 - 0.0363(\text{Head inc})$$

footnote:

[1] 1967 - 30.17%, 1968 - 51.20%, 1969 - 57.95%, 1970 - 46.36%, 1971 - 47.41%

[2] Under 5% critial points the value of F(48,1640) is 1.49.

[3] Some variables, which were statistically significant when we estimated the model with all the variables, were not significant in re-estimation. We considered such variables as not statistically significant.

Table IV A

Parameter Estimates for Labour Equations 1967

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample unconditional
constant	1558.7693*	1267.7070*	1556.3296*
	(30.861)	(25.414)	(42.271)
BirthGap	111.2579	-3.7038	29.7965
	(1.169)	(0.088)	(0.806)
Wife wg	-704.1122	-184.1179	-471.3710
	(1.576)	(0.413)	(1.463)
(Wife wg)^2	176.6363	-18.0680	90.6479
	(1.916)	(0.199)	(1.369)
Head inc	-0.2652*	-0.2698*	-0.2926*
	(6.312)	(6.912)	(10.685)
Head inc ^{2}	0.5668×10^{-5}*	0.4198×10^{-5}*	0.5784×10^{-5}*
	(4.666)	(3.045)	(6.893)
Unemploy	42.6361*	-1.5681	5.1312
	(3.348)	(0.225)	(1.044)
(Unemploy)^2	-0.7200*	-0.0193	-0.0531*
	(3.366)	(0.547)	(1.967)
inc/need	8.9754*	10.1893*	10.1931*
	(4.273)	(5.604)	(7.500)
(inc/need)^2	-0.7210×10^{-2}*	-0.6520×10^{-2}*	-0.7719×10^{-2}*
	(2.384)	(2.683)	(4.114)
Fecundit	-4.2753	12.6560	-0.6339
	(0.086)	(0.291)	(0.020)
(Fecundit)^2	-0.7383	-1.3261	-1.0798
	(0.4665)	(0.945)	(1.032)
Observations	189	151	340
R^2	0.3444	0.4646	0.3493

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Chow test F(12,316) = 2.2980

Table IV B

Parameter Estimates for Labour Equations 1968

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample unconditional
constant	2402.0710*	1344.8320*	1577.3882*
	(49.929)	(28.548)	(45.707)
BirthGap	-77.0932	-30.6823	30.2229
	(0.850)	(0.711)	(0.864)
Wife wg	-1036.6126*	-542.6887	-537.8107
	(2.229)	(1.142)	(1.651)
(Wife wg)^2	251.1645*	66.3974	112.4804
	(2.505)	(0.671)	(1.620)
Head inc	-0.1902*	-0.2265*	-0.2213*
	(6.316)	(5.276)	(11.508)
(Head inc)^2	0.1797×10^{-5}*	0.3386×10^{-5}*	0.3169×10^{-5}*
	(2.249)	(2.096)	(7.597)
Unemploy	1.8356	21.6627	10.3385
	(0.147)	(1.270)	(1.175)
(Unemploy)^2	-0.0231	-0.2880	-0.1266
	(0.158)	(1.040)	(1.103)
inc/need	4.0753	11.2963*	8.6154*
	(1.904)	(7.345)	(7.627)
(inc/need)^2	0.9217×10^{-3}	-0.8178×10^{-2}*	-0.5580×10^{-2}*
	(0.291)	(4.919)	(4.094)
Fecundit	-3.7857	-41.8059	-19.3969
	(0.082)	(1.098)	(0.693)
(Fecundit)^2	-0.9524	1.1036	-0.4556
	(0.628)	(0.822)	(0.474)
Observations	181	159	340
R^2	0.3135	0.4371	0.3606

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Chow test F(12,316) = 7.3582

Table IV C

Parameter Estimates for Labour Equations 1969

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample unconditional
constant	1910.3313* (37.255)	24.0713 (0.4852)	949.3052* (25.880)
BirthGap	-51.3668 (0.618)	-67.4059 (1.357)	2.9322 (0.079)
Wife wg	-828.3737 (1.377)	635.7993 (1.183)	-58.3414 (0.154)
$(\text{Wife wg})^2$	224.4707 (1.575)	-182.5847 (1.594)	9.3886 (0.111)
Head inc	-0.1907* (6.379)	-0.1901* (4.018)	-0.1936* (9.857)
$(\text{Head inc})^2$	0.2277×10^{-5}* (3.511)	0.3083×10^{-5} (1.802)	0.2364×10^{-5}* (6.149)
Unemploy	-11.4448 (0.698)	2.2814 (0.257)	0.4696 (0.078)
$(\text{Unemploy})^2$	0.2350 (0.608)	-0.0215 (0.367)	0.5634×10^{-2} (0.126)
inc/need	7.3298* (3.270)	7.5424* (5.102)	7.7892* (7.103)
$(\text{inc/need})^2$	-0.4885×10^{-2} (1.582)	-0.5037×10^{-2}* (3.351)	-0.5187×10^{-2}* (4.165)
Fecundit	-40.0683 (0.873)	70.3725 (1.738)	10.5490 (0.380)
$(\text{Fecundit})^2$	0.0498 (0.003)	-2.7817 (1.667)	-1.6838 (1.648)
Observations	178	162	340
R^2	0.2817	0.3325	0.2991

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Chow test $F(12,316) = 3.6634$

Table IV D

Parameter Estimates for Labour Equations 1970

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample unconditional
constant	1680.9971* (32.395)	52.8269 (1.074)	1033.5461* (28.059)
BirthGap	-113.2762 (1.889)	-33.8650 (0.592)	-36.6936 (0.934)
Wife wg	-736.9992 (1.174)	704.6772 (1.504)	1.1532 (0.003)
(Wife wg)^2	167.3082 (1.055)	-178.8685 (1.760)	-15.3298 (0.187)
Head inc	-0.2874* (5.622)	-0.1588* (6.260)	-0.2036* (9.236)
(Head inc)^2	0.4527×10^{-5}* (2.514)	0.2108×10^{-5}* (3.881)	0.2823×10^{-5}* (5.415)
Unemploy	5.7146 (1.249)	0.1341 (0.015)	-0.8891 (0.247)
(Unemploy)^2	-0.0382 (1.486)	-0.0872 (0.953)	-0.6823×10^{-2} (0.317)
inc/need	9.3285* (5.651)	5.3371* (6.407)	6.2063* (8.751)
(inc/need)^2	-0.6144×10^{-2}* (3.468)	-0.2656×10^{-2}* (4.236)	-0.3129×10^{-2}* (5.277)
Fecundit	4.4495 (0.097)	75.9236* (2.161)	28.2306 (1.085)
(Fecundit)^2	-1.3858 (0.841)	-3.5510* (2.225)	-2.3265* (2.264)
Observations	173	167	340
R^2	0.3320	0.3444	0.3088

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Chow test F(12,316) = 2.7094

Table IV E

Parameter Estimates for Labour Equations 1971

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample unconditional
constant	1496.4514* (27.655)	587.5898* (11.748)	1151.3694* (30.555)
BirthGap	-62.9480 (1.022)	-89.8988 (1.520)	-53.3046 (1.311)
Wife wg	-896.2939 (1.471)	205.7050 (0.358)	-298.0087 (0.721)
$(\text{Wife wg})^2$	224.4905 (1.448)	-75.6943 (0.558)	66.7110 (0.662)
Head inc	-0.2370* (3.256)	-0.1402* (5.117)	-0.1772* (7.926)
$(\text{Head inc})^2$	0.3970×10^{-5} (1.050)	0.1815×10^{-5}* (3.034)	0.2457×10^{-5}* (4.481)
Unemploy	4.1735 (0.840)	-3.0460 (0.471)	3.3144 (0.882)
$(\text{Unemploy})^2$	-0.0218 (0.749)	0.0274 (0.569)	-0.0183 (0.761)
inc/need	8.4099* (4.689)	4.8945* (5.095)	5.4899* (7.327)
$(\text{inc/need})^2$	-0.6105×10^{-2}* (2.940)	-0.2270×10^{-2}* (3.329)	-0.2643×10^{-2}* (4.454)
Fecundit	-8.1950 (0.178)	36.3512 (1.086)	19.9307 (0.811)
$(\text{Fecundit})^2$	-0.4393 (0.251)	-1.1797 (0.697)	-1.6618 (1.580)
Observations	164	176	340
R^2	0.2811	0.2361	0.2394

* significant level of 5% under H: parameters = 0.0

Asymtotic t values are in parentheses.

Chow test $F(12,316) = 5.5301$

Table V A

Mean and Standard Deviation of the Model II, 1967

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y uncondition
HourWork	774.894 (832.148)	1008.69 (806.400)	878.726 (827.825)
BirthGap	0.4180 (0.6013)	1.0331 (1.6183)	0.6912 (1.2054)
Wife wg	2.0312 (0.6498)	2.08863 (0.6742)	2.0567 (0.6604)
(Wife wg)^2	4.5457 (3.0821)	4.8139 (3.2770)	4.6648 (3.1682)
Head inc	7812.65 (4311.02)	7522.16 (5073.70)	7683.64 (4660.26)
(Head inc)^2	$0.7952\text{x}10^8$ $(0.1608\text{x}10^9)$	$0.8215\text{x}10^8$ $(0.1333\text{x}10^9)$	$0.8069\text{x}10^8$ $(0.1490\text{x}10^9)$
Unemploy	3.2090 (10.6461)	4.9735 (25.3447)	3.9927 (18.6509)
(Unemploy)^2	123.038 (630.650)	662.838 (5019.12)	362.773 (3382.22)
inc/need	242.873 (116.078)	265.384 (146.320)	252.871 (130.657)
(inc/need)^2	$0.7239\text{x}10^5$ $(0.8825\text{x}10^5)$	$0.9170\text{x}10^5$ $(0.1078\text{x}10^6)$	$0.8096\text{x}10^5$ $(0.9776\text{x}10^5)$
Fecundit	16.1481 (6.0053)	15.0530 (7.5189)	15.6618 (6.7313)
(Fecundit)^2	296.635 (190.053)	282.748 (238.273)	290.468 (212.603)
Observations	189	151	340

Standard deviations are in parentheses.

Table V B

Mean and Standard Deviation of the Model II, 1968

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y uncondition
HourWork	749.293	1132.99	928.726
	(756.947)	(763.656)	(782.811)
BirthGap	0.3812	0.7862	0.5706
	(0.6178)	(1.5482)	(1.1665)
Wife wg	1.9666	2.0144	1.9889
	(0.5927)	(0.6335)	(0.6117)
$(\text{Wife wg})^2$	4.2169	4.4564	4.3289
	(2.6983)	(2.9939)	(2.8387)
Head inc	8552.08	8499.53	8527.51
	(5424.92)	(4702.69)	(5092.60)
$(\text{Head inc})^2$	$0.1024\text{x}10^9$	$0.9422\text{x}10^8$	$0.9858\text{x}10^8$
	$(0.2962\text{x}10^9)$	$(0.1099\text{x}10^9)$	$(0.2285\text{x}10^9)$
Unemploy	2.8488	2.6376	2.7500
	(12.6523)	(11.3457)	(12.0418)
$(\text{Unemploy})^2$	167.313	134.872	152.142
	(1072.79)	(697.642)	(915.526)
inc/need	264.033	303.673	282.571
	(121.435)	(149.635)	(136.594)
$(\text{inc/need})^2$	$0.8438\text{x}10^5$	$0.1145\text{x}10^6$	$0.9845\text{x}10^5$
	$(0.1080\text{x}10^6)$	$(0.1231\text{x}10^6)$	$(0.1161\text{x}10^6)$
Fecundit	16.5580	12.5031	14.6618
	(5.8900)	(6.9917)	(6.7313)
$(\text{Fecundit})^2$	308.669	204.906	260.144
	(183.222)	(203.140)	(199.365)
Observations	181	159	340

Standard deviations are in parentheses.

Table V C

Mean and Standard Deviation of the Model II, 1969

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y uncondition
HourWork	756.927 (781.734)	1195.58 (745.969)	965.932 (794.667)
BirthGap	0.3483 (0.6656)	0.5988 (1.4764)	0.4677 (1.1323)
Wife wg	1.8301 (0.5237)	1.9638 (0.6173)	1.8938 (0.5733)
(Wife wg)^2	3.6218 (2.2029)	4.2351 (2.8226)	3.9140 (2.5321)
Head inc	8609.85 (5525.41)	9097.13 (4544.49)	8842.02 (5080.19)
(Head inc)^2	0.1045×10^9 (0.3004×10^9)	0.1033×10^9 (0.1135×10^9)	0.1039×10^9 (0.2307×10^9)
Unemploy	2.6938 (8.6431)	2.7284 (15.9771)	2.7103 (12.6585)
(Unemploy)^2	81.5407 (377.597)	261.136 (2441.43)	167.112 (1706.86)
inc/need	284.146 (123.874)	328.210 (144.391)	305.141 (135.644)
(inc/need)^2	0.9600×10^5 (0.1017×10^6)	0.1284×10^6 (0.1273×10^6)	0.1115×10^6 (0.1156×10^6)
Fecundit	16.0955 (5.9296)	10.9877 (6.5579)	13.6618 (6.7313)
(Fecundit)^2	294.028 (176.428)	163.469 (172.537)	231.821 (186.158)
Observations	178	162	340

Standard deviations are in parentheses.

Table V D

Mean and Standard Deviation of the Model II, 1970

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y uncondition
HourWork	781.601 (807.945)	1143.93 (758.429)	959.571 (803.603)
BirthGap	0.3757 (0.9357)	0.3713 (1.1745)	0.3735 (1.0581)
Wife wg	1.7079 (0.46924)	1.9147 (0.6349)	1.8095 (0.56554)
$(\text{Wife wg})^2$	3.1358 (1.8479)	4.0669 (2.8769)	3.5931 (2.4500)
Head inc	8253.68 (3700.80)	9687.73 (5921.69)	8958.05 (4963.42)
$(\text{Head inc})^2$	0.8174×10^{8} (0.9086×10^{8})	0.1287×10^{9} (0.2492×10^{9})	0.1048×10^{9} (0.1875×10^{9})
Unemploy	9.2269 (32.2653)	3.3802 (15.3756)	6.3552 (25.5447)
$(\text{Unemploy})^2$	1120.17 (5777.52)	246.421 (1535.59)	691.002 (4275.75)
inc/need	312.040 (140.308)	364.401 (188.303)	337.759 (167.447)
$(\text{inc/need})^2$	0.1169×10^{6} (0.1181×10^{6})	0.1680×10^{6} (0.2356×10^{6})	0.1420×10^{6} (0.1868×10^{6})
Fecundit	15.8150 (5.8220)	9.3952 (6.0260)	12.6618 (6.7313)
$(\text{Fecundit})^2$	283.815 (164.463)	124.365 (141.679)	205.497 (172.990)
Observations	173	167	340

Standard deviations are in parentheses.

Table V E

Mean and Standard Deviation of the Model II, 1971

Variable	Group I (y | z = 1)	Group II (y | z = 0)	Whole sample y uncondition
HourWork	797.360 (789.206)	1175.35 (734.960)	993.026 (783.650)
BirthGap	0.3659 (0.9530)	0.2898 (1.1064)	0.3265 (1.0345)
Wife wg	1.6824 (0.5206)	1.8392 (0.5805)	1.7636 (0.5572)
(Wife wg)^2	3.09972 (2.0359)	3.7178 (2.4179)	3.4197 (2.2598)
Head inc	8283.83 (3356.62)	9668.88 (5530.66)	9000.80 (4657.06)
(Head inc)^2	0.7982×10^8 (0.6195×10^8)	0.1239×10^9 (0.2085×10^9)	0.1026×10^9 (0.1574×10^9)
Unemploy	8.2317 (31.4318)	7.0483 (26.3531)	7.6191 (28.8773)
(Unemploy)^2	1049.70 (5537.04)	740.217 (3553.56)	889.495 (4498.73)
inc/need	333.982 (140.231)	387.778 (199.672)	361.829 (175.389)
(inc/need)^2	0.1311×10^6 (0.1130×10^6)	0.1900×10^6 (0.2454×10^6)	0.1616×10^6 (0.1952×10^6)
Fecundit	15.0366 (5.7939)	8.5171 (5.9904)	11.6618 (6.7313)
(Fecundit)^2	259.463 (154.686)	108.222 (126.997)	181.174 (159.872)
Observations	164	176	340

Standard deviations are in parentheses.

Table VI A

The Comparison of the Two Stage Aitken and the Single Equation Least-squares Estimation of the Model II, 1967

Variable	(y | z = 1) Two Stage	Single Egn	(y | z = 0) Two Stage	Single Egn
constant	-24.8795 (536.694)	1558.7693 (602.6486)	1704.61 (555.908)	1267.707 (599.956)
BrithGap	206.277 (79.5456)	111.2579 (95.1705)	6.7483 (40.4629)	-3.7038 (42.2918)
Wife wg	360.029 (402.916)	-704.1122 (446.8012)	-789.138 (413.982)	-184.1179 (445.6689)
(Wife wg)^2	-50.9948 (83.6112)	176.6363 (92.1762)	100.529 (84.9228)	-18.0680 (90.5690)
Head inc	-0.2742 (0.03637)	-0.2652 (0.0420)	-0.2643 (0.0351)	-0.2698 (0.0390)
(Head inc)^2	0.621×10^{-5} (0.1×10^{-5})	0.567×10^{-5} (0.1×10^{-5})	0.433×10^{-5} (0.1×10^{-5})	0.420×10^{-5} (0.1×10^{-5})
Unemploy	37.2640 (9.8779)	42.6361 (12.7363)	-1.1740 (6.1164)	-1.5681 (6.9694)
(Unemploy)^2	-0.5963 (0.1677)	-0.7200 (0.2139)	-0.0102 (0.0309)	-0.0193 (0.0352)
inc/need	10.5357 (1.7873)	8.9754 (2.1005)	10.3676 (1.6288)	10.1893 (1.8183)
(inc/need)^2	-0.0095 (0.0025)	-0.0072 (0.0030)	-0.0067 (0.0022)	-0.0065 (0.0024)
Fecundit	20.4197 (44.1752)	-4.2753 (49.6485)	29.1138 (40.8465)	12.6560 (43.4798)
(Fecundit)^2	-1.5124 (1.4128)	-0.7383 (1.5827)	-1.6923 (1.3143)	-1.3261 (1.4035)

Standard deviations are in parentheses.

Chow Test:

Two Stage, $F(12,316) = 3.4416$

Single Egn, $F(12,316) = 2.2980$

Table VI B

The Comparison of the Two Stage Aitken and the Single Equation Least-squares Estimation of the Model II, 1968

Variable	(y | z = 1) Two Stage	Single Eqn	(y | z = 0) Two Stage	Single Eqn
constant	719.489 (497.420)	2402.071 (617.6222)	1207.37 (506.332)	1344.832 (580.5073)
BrithGap	58.6397 (67.7885)	-77.0932 (90.6618)	-11.9464 (39.5871)	-30.6823 (43.1462)
Wife wg	54.4460 (374.372)	-1036.6126 (465.0873)	-524.545 (413.052)	-542.6887 (475.2111)
(Wife wg)^2	8.3943 (81.0106)	251.1645 (100.2469)	73.1776 (86.7298)	66.3974 (98.8818)
Head inc	-0.1846 (0.0229)	-0.1902 (0.0301)	-0.2294 (0.0358)	-0.2265 (0.0429)
(Head inc)^2	0.242×10^{-5} (0.6×10^{-6})	0.180×10^{-5} (0.8×10^{-6})	0.363×10^{-5} (0.1×10^{-5})	0.339×10^{-5} (0.2×10^{-5})
Unemploy	6.7462 (8.6091)	1.8356 (12.4788)	-0.2577 (13.3125)	21.6627 (17.0604)
(Unemploy)^2	-0.0856 (0.0990)	-0.0231 (0.1458)	0.0541 (0.2178)	-0.2880 (0.2769)
inc/need	6.6379 (1.5732)	4.0753 (2.1409)	11.0835 (1.2870)	11.2963 (1.5380)
(inc/need)^2	-0.0031 (0.0023)	0.0009 (0.0032)	-0.0077 (0.0014)	-0.0082 (0.0017)
Fecundit	4.6712 (38.6399)	-3.7857 (46.3501)	-23.4310 (34.3664)	-41.8059 (38.0776)
(Fecundit)^2	-1.2779 (1.2737)	-0.9524 (1.5159)	0.2166 (1.2025)	1.1036 (1.3420)

Standard deviations are in parentheses.

Chow Test:

Two Stage, $F(12,316) = 9.2483$

Single Eqn, $F(12,316) = 7.3582$

Table VI C

The Comparison of the Two Stage Aitken and the Single Equation Least-squares Estimation of the Model II, 1969

Variable	(y | z = 1) Two Stage	Single Eqn	(y | z = 0) Two Stage	Single Eqn
constant	1074.15 (584.199)	1910.3313 (685.601)	-61.8565 (515.343)	24.0713 (606.944)
BrithGap	44.8480 (69.4932)	-51.3668 (83.1572)	-53.8075 (43.6123)	-67.4059 (49.6914)
Wife wg	-240.4760 (501.982)	-828.3737 (601.6902)	479.664 (446.460)	635.7993 (537.3614)
(Wife wg)^2	81.9762 (118.436)	224.4707 (142.5031)	-135.453 (95.4104)	-182.5847 (114.5202)
Head inc	-0.2054 (0.0246)	-0.1907 (0.0299)	-0.1750 (0.0387)	-0.1901 (0.0473)
(Head inc)^2	0.267×10^{-5} (0.5×10^{-6})	0.228×10^{-5} (0.6×10^{-6})	0.257×10^{-5} (0.1×10^{-5})	0.308×10^{-5} (0.2×10^{-5})
Unemploy	-5.1539 (12.7613)	-11.4448 (16.3863)	6.1288 (7.0696)	2.2814 (8.8619)
(Unemploy)^2	0.1846 (0.3000)	0.2350 (0.3862)	-0.0385 (0.0462)	-0.0215 (0.059)
inc/need	9.6972 (1.8008)	7.3298 (2.2416)	7.8545 (1.2175)	7.5424 (1.4784)
(inc/need)^2	-0.0073 (0.002)	-0.0049 (0.0031)	-0.0050 (0.0013)	-0.0050 (0.0015)
Fecundit	-52.7559 (40.7958)	-40.0683 (45.8751)	75.8987 (35.6987)	70.3725 (40.4924)
(Fecundit)^2	0.3216 (1.4042)	0.0498 (1.5686)	-3.2084 (1.4490)	-2.7817 (1.6690)

Standard deviations are in parentheses.

Chow Test:

Two Stage, $F(12,316) = 7.7696$

Single Eqn, $F(12,316) = 3.6634$

Table VI D

The Comparison of the Two Stage Aitken and the Single Equation Least-squares Estimation of the Model II, 1970

Variable	(y | z = 1) Two Stage	Single Eqn	(y | z = 0) Two Stage	Single Eqn
constant	1412.61 (559.445)	1680.997 (656.8402)	47.0384 (441.321)	52.8269 (517.0958)
BrithGap	-92.9189 (52.7555)	-113.2762 (59.9558)	-34.0479 (51.4447)	-33.8650 (57.1928)
Wife wg	-652.953 (523.368)	-736.9992 (627.373)	590.173 (387.140)	704.6772 (468.432)
$(\text{Wife wg})^2$	157.065 (132.528)	167.3082 (158.5716)	-146.911 (84.1822)	-178.8685 (101.6526)
Head inc	-0.2830 (0.0422)	-0.2874 (0.0511)	-0.1548 (0.0214)	-0.1588 (0.0254)
$(\text{Head inc})^2$	0.470×10^{-5} (0.2×10^{-5})	0.453×10^{-5} (0.2×10^{-5})	0.199×10^{-5} (0.5×10^{-6})	0.211×10^{-5} (0.5×10^{-6})
Unemploy	4.5144 (3.5196)	5.7146 (4.5743)	-3.1778 (6.8682)	0.1341 (9.1211)
$(\text{Unemploy})^2$	-0.0290 (0.0199)	-0.0382 (0.0257)	-0.0035 (0.0687)	-0.0872 (0.0915)
inc/need	10.6796 (1.3660)	9.3285 (1.6508)	5.5279 (0.6995)	5.3371 (0.8330)
$(\text{inc/need})^2$	-0.0074 (0.0015)	-0.0061 (0.0018)	-0.0025 (0.0005)	-0.0027 (0.0006)
Fecundit	-15.1719 (41.0756)	4.4495 (45.8121)	74.9506 (32.0058)	75.9236 (35.1332)
$(\text{Fecundit})^2$	-0.8335 (1.4878)	-1.3858 (1.6470)	-3.7018 (1.4362)	-3.5510 (1.5960)

Standard deviations are in parentheses.

Chow Test:

Two Stage, $F(12,316) = 26.339$

Single Eqn, $F(12,316) = 2.7094$

Table VI E

The Comparison of the Two Stage Aitken and the Single Equation Least-squares Estimation of the Model II, 1971

Variable	(y | z = 1) Two Stage	Single Eqn	(y | z = 0) Two Stage	Single Eqn
constant	1330.05 (586.100)	1496.4514 (664.1467)	839.732 (493.843)	587.5898 (566.7747)
BrithGap	-58.3862 (56.9849)	-62.948 (61.6013)	-112.064 (54.4209)	-89.8988 (59.1419)
Wife wg	-473.837 (531.652)	896.2939 (609.2837)	-104.060 (491.017)	205.7050 (575.294)
(Wife wg)^2	130.288 (135.215)	224.4905 (155.0258)	8.5510 (116.298)	-75.6943 (135.7505)
Head inc	-0.2869 (0.0619)	-0.2370 (0.0728)	-0.1368 (0.0241)	-0.1402 (0.0274)
(Head inc)^2	0.648×10^{-5} (0.3×10^{-5})	0.397×10^{-5} (0.4×10^{-5})	0.176×10^{-5} (0.5×10^{-6})	0.182×10^{-5} (0.6×10^{-6})
Unemploy	4.8257 (4.1312)	4.1735 (4.9665)	3.2804 (5.1358)	-3.0460 (6.4734)
(Unemploy)^2	-0.0239 (0.0240)	-0.0218 (0.0291)	-0.0163 (0.0380)	0.0274 (0.0481)
inc/need	8.7383 (1.5652)	8.4099 (1.7937)	4.8126 (0.8318)	4.8945 (0.9606)
(inc/need)^2	-0.0062 (0.0018)	-0.0061 (0.0021)	-0.0021 (0.0006)	-0.0023 (0.0007)
Fecundit	-19.4112 (42.7355)	-8.1950 (46.0473)	29.3953 (30.9769)	36.3512 (33.4791)
(Fecundit)^2	-0.2104 (1.6331)	-0.4393 (1.7509)	-1.0301 (1.5451)	-1.1797 (1.6928)

Standard deviations are in parentheses.

Chow Test:

Two Stage, $F(12,316) = 48.711$

Single Eqn, $F(12,316) = 5.5301$

Table VII A

Probability Function Estimates of the Model II, 1967

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-2.42797*	0.74520	3.25817
BirthGap	-0.41671*	0.09670	4.30938
Wife wg	0.35486	0.54061	0.65641
$(\text{Wife wg})^2$	-0.09620	0.11071	0.86890
Head inc	0.06513	0.04625	1.40828
$(\text{Head inc})^2$	0.488×10^{-4}	0.00135	0.03629
Unemploy	0.02830	0.01896	1.49258
$(\text{Unemploy})^2$	-0.00046	0.00031	1.48280
inc/need	-0.02214	0.22531	0.09826
$(\text{inc/need})^2$	-0.02807	0.03065	0.91577
Fecundit	0.28355*	0.05580	5.08141
$(\text{Fecundit})^2$	-0.00735*	0.00177	4.14357

Log of likelihood function = -191.969 after 10 iterations.

* significant in 95% under H: parameter estimates = 0.0

note:

1 Head inc = \$1,000   1 $(\text{Head inc})^2$ =\$1,000,000

1 inc/need = 100   1 $(\text{inc/need})^2$ = 10,000

Table VII B

Probability Function Estimates of the Model II, 1968

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-3.05536*	0.80555	3.79287
BirthGap	-0.44861*	0.10617	4.22535
Wife wg	1.00167	0.60746	1.64895
$(\text{Wife wg})^2$	-0.22005	0.12970	1.69655
Head inc	0.11504*	0.03902	2.94809
$(\text{Head inc})^2$	-0.00055	0.00090	0.61233
Unemploy	-0.01095	0.01661	0.65922
$(\text{Unemploy})^2$	0.00015	0.00022	0.66373
inc/need	-0.33192	0.24119	1.37618
$(\text{inc/need})^2$	-0.00278	0.03091	0.08988
Fecundit	0.24864*	0.05810	4.27937
$(\text{Fecundit})^2$	-0.00489*	0.00196	2.49738

Log of likelihood function = -176.713 after 13 iterations.

* significant in 95% under H: parameter estimates = 0.0

note:

1 Head inc = $1,000 1 $(\text{Head inc})^2$ =$1,000,000

1 inc/need = 100 1 $(\text{inc/need})^2$ = 10,000

Table VII C

Probability Function Estimates of the Model II, 1969

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-2.39040*	0.77684	3.07709
BirthGap	-0.46689*	0.10790	4.32698
Wife wg	1.67102*	0.69282	2.41191
$(\text{Wife wg})^2$	-0.39129*	0.15509	2.52292
Head inc	0.07168	0.03768	1.90244
$(\text{Head inc})^2$	-0.00032	0.00088	0.36407
Unemploy	0.05337*	0.02657	2.00867
$(\text{Unemploy})^2$	-0.00109*	0.00054	2.03169
inc/need	-0.48319*	0.20561	2.35006
$(\text{inc/need})^2$	0.02887	0.02326	1.24127
Fecundit	0.11981*	0.05352	2.23850
$(\text{Fecundit})^2$	-0.00025	0.00204	0.12416

Log of likelihood function = -174.969 after 20 iterations.

* significant in 95% under H: parameter estimates = 0.0

note:

1 Head inc = $1,000   1 $(\text{Head inc})^2$ =$1,000,000

1 inc/need = 100   1 $(\text{inc/need})^2$ = 10,000

Table VII D

Probability Function Estimates of the Model II, 1970

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-2.03529*	0.76525	2.65963
BirthGap	-0.29682*	0.09165	3.23849
Wife wg	1.25770	0.73180	1.71864
$(\text{Wife wg})^2$	-0.36765*	0.17337	2.12059
Head inc	0.08875	0.05830	1.52238
$(\text{Head inc})^2$	-0.00170	0.00210	0.81031
Unemploy	0.00164	0.00909	0.18042
$(\text{Unemploy})^2$	$0.411\text{x}10^{-4}$	$0.834\text{x}10^{-4}$	0.49300
inc/need	-0.27551	0.15846	1.73869
$(\text{inc/need})^2$	0.01443	0.01493	0.96648
Fecundit	0.07757	0.05195	1.49326
$(\text{Fecundit})^2$	0.00157	0.00212	0.73948

Log of likelihood function = -172.197 after 15 iterations.

* significant in 95% under H: parameter estimates = 0.0

note:

1 Head inc = $1,000   1 $(\text{Head inc})^2$ =$1,000,000

1 inc/need = 100   1 $(\text{inc/need})^2$ = 10,000

Table VII E

Probability Function Estimates of the Model II, 1971

Variable	Coefficient	Asym stdv	Asym t-ratio
constant	-1.29702	0.72149	1.79768
BirthGap	-0.19225*	0.07349	2.61615
Wife wg	0.21367	0.71663	0.29816
$(\text{Wife wg})^2$	-0.10158	0.17430	0.58281
Head inc	0.15931*	0.07094	2.24571
$(\text{Head inc})^2$	-0.00592	0.00315	1.87641
Unemploy	-0.00493	0.00689	0.71503
$(\text{Unemploy})^2$	$0.340\text{x}10^{-4}$	$0.459\text{x}10^{-4}$	0.74210
inc/need	-0.18304	0.15392	1.18917
$(\text{inc/need})^2$	0.00748	0.01437	0.52037
Fecundit	0.04998	0.04688	1.06614
$(\text{Fecundit})^2$	0.00228	0.00206	1.10551

Log of likelihood function = -180.441 after 11 iterations.

* significant in 95% under H: parameter estimates = 0.0

note:

1 Head inc = $1,000   1 $(\text{Head inc})^2$ =$1,000,000

1 inc/need = 100   1 $(\text{inc/need})^2$ = 10,000

Chapter VIII

Conclusion

The proposed basic model, which involves discrete and continuous dependent variables, is estimated by separating the model into a simple regression model and a probability equation model. The regression model can be estimated by ordinary least-squares. It is suggested that the probability equation model be formulated as a logistic function and estimated by using the maximum likelihood method. While the basic model is extended into a system of equations, it can be separated into a system of regression equations and a system of logistic equations. The system of regression equations can be estimated by Zellner's two stage method in order to gain efficiency. The probability model can be estimated by the method of Nerlove and Press. It is interesting that, this basic model can be extended by considering the correlation of disturbances in the regression equations, to be a model with constraints, to be a model with jointly dependent discrete variables, or to be a simultanous-equation model. Those extended models are solvable. One extension of this model, which has not been solved with verification, is the recursive model. The basic model is more common in social science, although there is not much literature.

The economic model in this thesis is a study on the labour supply of American married women from 1967 to 1971. We find that the number of hours worked by married women is affected by the age of their youngest child very much, and slightly by their head's income. There is some effect from other social factors, such as the head's unemployment, and the ratio of incomes over needs, yet the significance of these factors varies from year to year. The birth gap has a significant effect on the probability of the wife having a child not older than 6 years of age. Therefore, the results tell us that the married woman's role in the labour market is quite dependent upon her family planning.

Appendix A

Least-Squares Estimation

Here we follow all the notations defined in chapter 2. Therefore the conditional regression will be written as following:

$$[Y^1, Y^2] = [X^1, X^2] \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} + [u, v]$$

We will call $Y = [Y^1, Y^2]$, $X = [X^1, X^2]$, and $E = [u, v]$. From generalized multivariate regression[1], we know

$$A^+ = (X'X)^{-1}X'Y$$

$$(X'X) = \begin{bmatrix} (X^1)' \\ (X^2)' \end{bmatrix} [X^1 \quad X^2] = \begin{bmatrix} (X^1)'(X^1) & 0 \\ 0 & (X^2)'(X^2) \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} (X^{1\prime}X^1)^{-1} & 0 \\ 0 & (X^{2\prime}X^2)^{-1} \end{bmatrix}$$

so

$$\begin{bmatrix} A^+ & 0 \\ 0 & B^+ \end{bmatrix} = \begin{bmatrix} (X^{1\prime}X^1)^{-1} & 0 \\ 0 & (X^{2\prime}X^2)^{-1} \end{bmatrix} \begin{bmatrix} X^{1\prime}Y^1 & 0 \\ 0 & X^{2\prime}Y^2 \end{bmatrix}$$

Hence,

$$A^+ = [X^{1\prime}X^1]^{-1}(X^1)'Y^1$$

$$B^+ = [X^{2\prime}X^2]^{-1}(X^2)'Y^2$$

$$Cov^+(u,v) = \begin{bmatrix} Var^+(u) & 0 \\ 0 & Var^+(v) \end{bmatrix} = E^+E^{+\prime}/n$$

Since,

$$E^+ = [Y^1 - X^1A^+ \quad Y^2 - X^2B^+]$$

then,

$$Cov^+(U,V) = \frac{1}{n} \begin{bmatrix} (Y^1-X^1A^+)'(Y^1-X^1A^+) & (Y^1-X^1A^+)'(Y^2-X^2B^+) \\ (Y^2-X^2B^+)'(Y^1-X^1A^+) & (Y^2-X^2B^+)'(Y^2-X^2B^+) \end{bmatrix}$$

$$= \begin{bmatrix} (Y^1-X^1A^+)'(Y^1-X^1A^+)/n & 0 \\ 0 & (Y^2-X^2B^+)'(Y^2-X^2B^+)/n \end{bmatrix}$$

For unbiased estimators of $Var^+(U)$ and $Var^+(V)$, we have

$$Var^{+}(u) = (Y^{1}-X^{1}A^{+})'(Y^{1}-X^{1}A^{+})/(n^{1}-k)$$

$$var^{+}(v) = (Y^{2}-X^{2}B^{+})'(Y^{2}-X^{2}B^{+})/(n^{2}-k)$$

where n^{1} is the total number of observations when $z=1$; $n^{2} = n-n^{1}$.

Footnote

[1] Press, J., Applied Multivariate Analysis, 1972, pp.220.

Appendix B

Likelihood Ratio Test for Micro Regression Coefficient Vector Equality

Under the hypothesis of chapter 3, H: $S^1 = \ldots = S^d$, The system of equations can be written as

$$\begin{bmatrix} Y^1 \\ \cdot \\ \cdot \\ \cdot \\ Y^d \end{bmatrix} = \begin{bmatrix} X^1 \\ \cdot \\ \cdot \\ \cdot \\ X^d \end{bmatrix} W + \begin{bmatrix} U^1 \\ \cdot \\ \cdot \\ \cdot \\ U^d \end{bmatrix}$$

or,

$$Y^* = X^*W + U^* \qquad \text{(B1)}$$

We define a transformation T, such that $E(TU^*U^{*\prime}T') = \mathrm{var}(U^*)I$. Let $TY^* = Y^o$, $TX^* = X^o$, and $TU^* = U^o$. Then the likelihood function, $L(U^o)$, under the hypothesis is

$$L(U^*) = (2\ \mathrm{var}(U^o))^{-dk/2} \exp(-U^{o\prime}U^o/(2\mathrm{var}(u^o))) \qquad \text{(B2)}$$

The maximum likelihood estimators for equation (B1) are

$$\mathrm{Var}^+(U^*) = U^{o+\prime}U^{o+}/dk$$

$$= (Y^o - X^oW^+)'(Y^o - X^oW^+)/dk$$

and

$$W^+ = (X^{o\prime}X^o)^{-1}X^{o\prime}Y^o$$

Hence if we rewrite equation (B2) in terms of these estimators, then

$$L(U^{*+}) = (2\ \mathrm{var}^+(U^*))^{-dk/2} \exp(-dk/2)$$

Likewise, we transform the variables in equation (1) of chapter 3 by T and express the maximum likelihood function in terms of U^+ then

$$L(U^+) = (2\ Var(U^+))^{-dk/2} \exp(-dk/2)$$

So, the estimated likelihood ratio, r is

$$r = L(U^{*+})/L(U^+)$$
$$= [Var^+(u^*)/Var^+(u)]^{-dk/2}$$

or,

$$-2\log(r) = dk\log[Var^+(u^*)/Var^+(u)]$$

Which is asymptotically distributed as Chi squares with (d-1)n degree of freedom.

Appendix C

Parameter Estimates for Labour Equation 1967 - 71 without $(\text{Unemploy})^2$

When unemployment of the head was considered as a linear relationship with the annual hours worked by the wife in the regression model, we found that they are negatively correlated. We tabulated the results of model I as following:

Variable	Group I (y | z = 1)	Group II (y | z = 0)
constant	1728.3230 (4.0485)	478.4102 (1.2482)
BirthGap	-42.8763 (1.7460)	-36.5538 (2.8869)
Wife wg	-714.5180 (11.0354)	360.3229 (3.0861)
(Wife wg)^2	192.8221 (15.6143)	-113.3932 (6.6019)
Head inc	-0.2069 (217.8773)	-0.1740 (182.9149)
(Head inc)^2	0.2892×10^{-5} (97.1418)	0.2337×10^{-5} (52.2196)
Unemploy	-0.2405 (0.0486)	-2.5008 (4.7875)
inc/need	6.5660 (78.8038)	6.1298 (207.5199)
(inc/need)^2	-0.0040 (19.0210)	-0.0032 (77.3457)
Fecundit	-13.2484 (0.4368)	43.1432 (9.7703)
(Fecundit)^2	-0.5110 (0.5569)	-1.7794 (10.9088)
observation	885	815
R^2	0.2521	0.3197

Asymtotic t values are in parentheses.

Bibliography

Berkson, J., (1951), "Why I Prefer Logits to Probits", Biometrics, December 1951, pp. 327-339.

Berkson, J., (1953), "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function", American Statistical Association Journal, September 1953, pp. 565-599.

Berkson, J., (1955), "Maximum Likelihood and Minimum X^2 Estimates of the Logistic Function", Journal of American Statistical Association, Vol.50, 1955, pp. 130-161.

Berndt E. R. and T. J. Wales, (1974), "Determinants of Wage Rates for Married Women: Results from Panel Data", Discussion Paper No.74-05, Department of Economics, University of British Columbia, March 1974.

Berndt, E. R. and T. J. Wales, (1974), "Labour Supply and Fertility Behaviour of Married Women: An Empirical Analysis", Discussion Paper 74-27, University of British Columbia, December 1974.

Bishop, Y. M. M., (1969), "Full Contingency Tables, Logits, and Split Contingency Tables", Biometrics, Vol. 25, 1969, pp. 383-400.

Bliss, C. I., (1934), "The Method of Probits", Science, Vol.79, No.2037, January 1934, pp.38-39.

Bliss, C. I., (1934), "The Method of Probits -- A Correction", Science, Vol.79, No.2053, May 1934, pp. 409-410.

Box, M. J., D. Davies, and W. H. Swann, (1969), Non-linear Optimization Techniques, Oliver and Boyd Ltd., Edinburgh, 1969.

Buse, A. (1972), " A Technical Report on Binary Dependent Variables as Applied in the Social Sciences", A Commissoned Research Project of the Alberta Human Resources Research Council, Edmonton, Alberta, 1972.

Chambers, E. A., and D. R. Cox, (1967), "Discrimination between Alternative Binary Response Models", Biometrika, Vol. 54, 1967, pp. 573-578.

Chow, G. C., (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", Econometrica, Vol.28, No.3, July 1960, pp. 591-605.

Cornfield, J. and N. Mantel, (1950), "Some New Aspects of the Application of Maximum Likelihood to the Calculation of the Dosage Response Curve", Journal of the American Statistical

Association, Vol. 45, 1950, pp. 181-210.

Cox, D. R. (1970), The Analysis of Binary Data, Methuen, London, 1970.

Dempeter, A. P., (1971), "An Overview of Multivariate Data Analysis", Journal of Multivariate Analysis I, 1971, pp. 316-346.

Dempster, A. P., (1972), "Aspects of the Multinomial Logit Model", Multivariate Analysis III, Edited by P. R. Krishnaiah, Academic Press, 1972, pp. 129-142.

Finney, D. J., (1947), Probit Analysis, Cambridge University Press, Cambridge, England, 1947. (3rd edition, 1971).

Goldberger, A. S., (1964), Econometric Theory, Wiley, New York, 1964.

Goodman, L. A., (1970), "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications", Journal of the American Statistical Association, Vol.65, No.329, March 1970, pp. 226-256.

Goodman, L. A., (1972), "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables", American Sociological Review, Vol.37, 1972, pp. 28-46.

Gunderson, M., (1974), "Retention of Trainees -- A Study With Dichotomous Dependent Variables", Journal of Econometrics 2, 1974, pp. 79-93.

Johnson, J., (1972), Econometric Methods, 2nd Edit, Mcgraw Hill, 1972.

Nerlove, M. and S. J. Press, (1973), Univariate and Multivariate Log-Linear and Logistic Models, Santa Monica, Calif.: Rand Corporation Report R-1306, 1973.

Neter, J., and E. S. Maynes, (1970), "On the Appropriateness of the Correlation Coefficient with a 0, 1 Dependent Variable", Journal of the American Statistical Association, Vol.65, No.330, June 1970, pp. 501-509.

Press, S. J., (1972), Applied Multivariate Analysis, Holt, Rinehart and Winston, New York, 1972.

Schmidt, P. and P. Strauss, (1974), "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach", University of Caroline, 1974.

Schmidt, P. And P. Strauss, (1975), "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach", Econometrica, Vol.43, No.4, July 1975, pp. 745-755.

Survey Research Centre, University of Michigan (1972), A Panel

Study of Income Dynamics, Ann Arbor, 1972.

Theil, H. (1969), "A Multinomial Extension of the Linear Logit Model", International Economic Review, Vol.10, No.3, October 1969, pp. 251-259.

Theil, H., (1970), "On the Estimation of Relationships Involving Qualitative Variables", American Journal of Sociology, Vol.76, 1970, pp. 103-154.

Theil, H. (1970), Principles of Econometrics, New York: John Wiley and Sons, 1970.

Tobin, J., (1955), "The Application of Multivariate Probit Analysis to Economic Survey Data", Cowles Foundation Discussion Paper No. 1, December 1, 1955.

Wonnacott, R. J. And T. H. Wonnacott, (1970), Econometrics, John Wiley and Sons Inc., 1970.
Zellner, A. and H. Theil, (1962), "Three-Stage Squares: Simultaneous Estimation of Simultaneous Equations", Econometrica , Vol. 30, No. 1, January 1962, pp. 54-78.

Zellner, A., (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of American Statistical Association, Vol.57, 1962, pp. 348-368.

Zellner, A. and D. S. Huang, (1962), "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations", International Economic Review, Vol.3, No.3, September 1962, pp. 300-313.

Zellner, A., (1963), "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results", Journal of American Statistical Association, Vol.58, 1963, pp. 977-992.

Zellner, A. and T. H. Lee, (1965), "Joint Estimation of Relationships Involving Discrete Random Variables", Ecometrica, Vol.33, No.2, April 1965, pp. 382-394.