FAMILY LABOUR SUPPLY AND LABOUR FORCE PARTICIPATION DECISIONS

by

JULES J.M. THEEUWES

Licentiate in Commercial and Consular Sciences,
University Faculties Saint Ignatius, Antwerp, 1966

Licentiate in Economic Sciences
University of Louvain, 1970

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE

REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

ECONOMICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August, 1975

Department of ECONOMICS

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

Date SEPT 12th, 75

ABSTRACT

The main objective of this study is the empirical estimation of family labour force participation functions. The appropriate estimation procedure for a model involving choice among multiple discrete alternatives requires a statistical technique different from ordinary least squares. In this study I use the binomial and multinomial logit model to estimate parameters affecting the probabilities of choosing a particular labour force alternative.

A theoretical contribution of this thesis to the econometric literature is the development of a procedure which, in the context of the multinomial logit model, allows one to test whether decision making is sequential or simultaneous. This procedure is applied in testing whether the family chooses simultaneously among possible alternatives or whether one partner decides first about participation and the other partner decides conditional upon the first. Using a Bayesian dicrimination technique it is found that the simultaneous decision model is more probable *a posteriori* than the sequential model.

A substantial portion of the empirical research in this study involves the estimation and comparison of family labour force participation and labour supply decisions. I attempt to discriminate statistically between the hypothesis that the parameters of supply and participation are either the same or that they are different and conclude that the hypothesis of different parameters is more probable, *a posteriori*.

In addition, the comparison of the parameters of family labour supply and labour force participation leads to interesting results, e.g., the substitution effect on both participation and supply behaviour of husband and wife. Another use of the estimated labour supply and labour force participation functions involves combining them to form unconditional labour supply functions. It is indicated that unconditional labour supply functions could be useful to evaluate the combined effect on supply and participation of a labour market policy.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENT

This thesis would not have been possible without the support of a great number of people. For their advice and encouragement throughout the course of this study, I wish to thank the members of my dissertation committee : Ernst R. Berndt, John G. Cragg, Terence J. Wales. I am especially grateful to Ernst R. Berndt for limitless hours of discussion on all aspects of the dissertation. In addition the technical advice of Keith Wales on aspects of computer programming has been most helpful.

I gratefully acknowledge the financial support from Canada Council 1971 - 1975, and the computer time from the Department of Economics at the University of British Columbia. I also wish to thank Ernst R. Berndt and Terence J. Wales for employment as research assistant. Thanks are also due to Sharron King for typing the final manuscript.

Finally I wish to acknowledge with special thanks my wife Magda who has supported and encouraged me througout my graduate studies.

CHAPTER I

INTRODUCTION

The amount of hours supplied by an individual in the labour market can be viewed as the result of two sequential decisions. First there is the decision whether to participate in the labour force. Second, given that the individual decides to enter the labour force, he/she then decides the actual number of hours to work. The first decision is a decision at the extensive margin, the second at the intensive margin. Aspects of the first decision are usually investigated in "labour force participation studies". The determinants of the choice at the intensive margin are the subject of "labour supply studies". With a few exceptions both sets of studies have developed in an unrelated way.

Most labour force participation studies are aimed at either explaining determinants of the substantial increase in labour force participation by married women in the post war period (Cain [1966], Mahoney [1961], Mincer [1962], Sweet [1973]); or at investigating cyclical behaviour of labour force participation rates for various age and sex groups (Barth [1968], Mincer [1966], Officer and Anderson [1969], Wachter [1972, 1974], Fair [1971], Cragg [1973]). An exhaustive treatment of labour force participation can be found in the voluminous study by Bowen and Finegan [1969].

Labour supply studies, on the other hand, seem to have been motivated mainly by the need to predict the disincentive effects of various personal income tax schemes such as progressive income tax (Break [1957], Kosters [1963], Wales [1973]), or a negative income tax proposal (Boskin [1967], Cain and Watts [1973], Green and Tella [1969]). More recent supply studies treat labour supply or leisure demand as a part of a system of consumer demand functions and are mainly interested in estimating parameters of the underlying utility function (Gussman [1972], Wales and Woodland [1974a, 1974b], Ashenfelter and Heckman [1974]).

A first objective of this thesis is to estimate and compare parameters of both the labour force participation decisions and the labour supply decisions for the same cross-section sample of families. A common theoretical framework is developed for both kinds of decisions (in Chapter II) and then labour force participation and labour supply functions are empirically estimated in Chapters IV and V respectively. The decision unit studied in this thesis is the family. This contrasts with most of the previous studies which have concentrated on choice at the individual level.

The empirical investigation of labour force participation decisions attempts to explain the determinants of the choice of the individual or family <u>between a finite number of distinct</u>

alternatives. For an individual the alternatives could be the choice to be in or to be out of the labour force. For a two-person family one can see a choice between four alternatives: both husband and wife working, husband only working, wife only working, and none working. The special nature of the dependent variable in labour force participation decisions requires an appropriate empirical technique. In this thesis I use (the multinomial extension of) the logit model developed by Theil [1969], Cragg and Uhler [1970, 1971], Cragg and Baxter [1970], and McFadden [1974]. A logit model allows me to explain the probability that a particular alternative will be chosen by a family. This probability is defined as a function of a set of independent variables.

Such a special statistical technique is not required for the empirical study of labour supply decisions. In this case the dependent variable, say annual hours of work, varies continuously within a wide range. The usual regression analysis will presumably lead to satisfactory results for the study of labour supply.

The empirical study of labour force participation choices is furthermore complicated by the fact that one does not observe the (potential) market wage rate for non-labour force participants. Economic theory predicts the importance of the market wage rate as a determinant of labour force participation choice. Thus one should include a wage rate in specifying the labour force participation function. A wage rate is observed for individuals in the labour market but clearly not for those out of the labour force. In an attempt to circumvent this problem I try to predict the potential market wage for non-labour force participants using a wage equation. This wage equation defines the wage rate as a function of a set of observed socio-demographic variables and is estimated over the sample of labour force participants. This prediction procedure and the problems associated with it are discussed in Chapter III.

Both labour force participation decisions and labour supply decisions are functions of wage rates and income variables. From an aggregative viewpoint therefore a change in "the market wage rate" or "income" will have two kinds of effects: (i) a number of individuals will enter or leave the market, (ii) individuals already in the market will adjust their supply of hours. In order to evaluate the total effect of, say, an economic policy such as a negative income tax, which changes both "income" and "wage rate", it would seem desirable to measure the combined response in a given population at the internal and external margin. Douglas [1934], in one of the earliest empirical labour supply studies, combined the wage elasticities of hours supplied with the wage elasticity of labour force participation into an estimate of the "most probable elasticity of the short time supply of labour". More recent studies (Hall [1973], Boskin [1973], Kalachek and Raines [1970]) combine the response at the internal and external margin using an expected value formula, i.e., the product of the probability of choosing a particular alternative, given the wage rate or income, times the predicted number of hours worked if that particular alternative is chosen, again given the wage rate or income. This "expected value" will give a more accurate idea of the total aggregate effect on hours worked of an economic policy which changes the "wage rate" and/or "income", than as is traditional, looking only at labour supply functions. This is especially true if labour force participation decisions are sensitive to wage rate and/or income changes. These matters will be discussed in Chapter VI of this thesis.

CHAPTER II

ECONOMIC THEORY OF LABOUR SUPPLY AND LABOUR FORCE PARTICIPATION

1. Introduction

Ever since Lord Robbins' [1930] seminal contribution, economists have tended to treat the choice between leisure and work as an application of the utility maximization paradigm[1]. Within this framework a labour supply function is defined as dependent upon prices and income. At the same time certain restrictions, imposed by the utility maximization assumption on the parameters of the supply function, are derived. I will show below that labour force participation decisions can also be discussed in this framework.

Static utility maximization thus leads to important theoretical predictions that are very useful in guiding empirical research. This is the basic reason why I develop a theoretical model for the labour force participation and labour supply choices of a family, based on the assumption of static utility maximization. But one should be well aware of the shortcomings of this assumption particularly in the case of labour supply and labour force participation decisions. The utility maximization paradigm is only valid if the family is free to choose any labour force participation alternative or labour supply pattern it desires, within its time constraint. This is not necessarily true in reality. There exist important social and institutional constraints on the labour market which severely restrict a family's choice set. For instance total hours of work in a particular industry is frequently the result of a collective agreement defining standard work week and regulating overtime work. Such institutional arrangements could constrain the individual's choice of working hours (unless both choices happen to be in agreement).[2] Part time work and multiple job holding are possibilities which could sometimes offset this social constraint depending on how easily an individual can find them in his/her labour market. In any case, one might expect that, for at least part of the sample, the observed labour force participation or labour supply choice does not correspond to what a family would have chosen without the social constraints.

A second shortcoming of the static utility maximization assumption is its neglect of dynamic considerations in labour force participation and labour supply decisions of a family. In a static framework wages and (non-employment) income are treated as exogenously given. In a dynamic context both variables are seen as the result of investments in human and non-human capital and thus become endogenous variables. The basic assumption of dynamic utility theory is that a family plans its labour force participation, labour supply, and (human and non-human) capital accumulation paths over its lifetime, maximizing an intertemporal utility function subject to a lifetime wealth constraint (Hicks [1958], Tintner [1938a, 1938b, 1939], Lluch and Morishima [1973]). Consequently labour force participation, labour supply, wage rates and income are determined simultaneously. Intertemporal relationships studied in dynamic utility theory are certainly relevant for the study of family choice behaviour. However, the theoretical predictions that can be derived in this framework depend substantially on (sometimes restrictive) assumptions about the functional form of the intertemporal utility function, on the relationship between the market rate of interest and the subjective rate of time preference and on the existence of perfect capital markets (for both human and non-human capital).[3] Although not explicitly incorporated in this thesis, I will sometimes rely on dynamic utility theory in cases where its possible implications are helpful to explain certain empirical results.

Static utility maximization also neglects the influence of uncertainty and of search and information costs on labour force participation and labour supply behaviour. For instance, even if an individual desires to supply a positive amount of hours, search and information costs might offset the expected benefits of joining the labour force.[4] Again I will occasionally supplement the complications of static utility theory with explanations derived from other paradigms if this is helpful in understanding empirical findings.

The role of the theoretical model in this thesis is essentially to provide a structure to organize the empirical investigations and to use as a reference in explaining the results. This pragmatic approach avoids the necessity of specifying an all embracing theoretical model without, however, losing completely the benefits of some form of theoretical guidance.

The starting point for the theoretical model then is the assumption that the family maximizes a utility function defined over the husband's leisure, the wife's leisure and "all other consumption goods". It should be noted that the existence of a family utility function depends on some very restrictive conditions such as non-jointness in consumption, independence of preferences and an optimal rule for reallocation of income (Samuelson [1956]). However, external consumption effects are the essence of family life and so the conditions are presumably not fulfilled. Its existence is nevertheless usually accepted in the study of family labour supply (e.g. Ashenfelter and Heckman [1974], Diewert [1971], Wales and Woodland [1974a, 1974b]) and I will follow this procedure. However, in the case of family labour force participation decisions I will suggest an alternative model in contrast with the model derived from the family utility assumption (Section 3.4 of this Chapter).

To formalize the basic theoretical model, the family is assumed to maximize:

(1.1) $U(C, L_m, L_f)$ with respect to $C \geq 0$, $L_m \geq 0$, $L_f \geq 0$,

(1.2) subject to: $pC + w_m L_m + w_f L_f \leq (w_m + w_f)H + A'$ or (dividing both sides by p) ,

(1.2') subject to: $C + \upsilon_m L_m + \upsilon_f L_f \leq (\upsilon_m + \upsilon_f)H + A = J$ and

(1.3) subject to: $H - L_m \geq 0$,

(1.4) subject to: $H - L_f \geq 0$.

The subscripts "m" and "f" indicate respectively husband and wife and

C: consumption goods, a composite commodity,

p: a price index for the composite commodity C,

L_i: leisure time, i = m, f.

w_i: money wage, i = m, f.

υ_i: real wage (defined as w_i/p), i = m, f.

H: total amount of time available in the period under consideration,

A': non-employment income in money terms,

A: non-employment income in real terms,

J: "whole income" (Becker [1966])

I ignore savings (which can only be treated adequately in a dynamic framework) and assume non-satiation for commodity and leisure consumption. Consequently (1.2') will become an equality.

The time constraints (1.3) and (1.4) are crucial in considering the labour force participation problem. If R_i ($i = m, f$) denotes worktime then the Lagrangian can be written as:

$$(1.5) \qquad L(C, L_m, L_f, R_m, R_f) = U(C, L_m, L_f) + \lambda[J - C - \upsilon_m L_m - \upsilon_f L_f] + \mu_m[H - L_m - R_m] + \mu_f[H - L_f - R_f] \, .$$

Applying the Kuhn-Tucker conditions to (1.5) will lead to two distinct solutions,

(i) interior solution, i.e.,

$$(1.6) \qquad R_i > 0 \; , \quad \mu_i = 0 \; , \quad \frac{U_{L_i}}{U_C} = \upsilon_i \; , \qquad i = m, f \, .$$

(ii) corner solution, i.e.,

$$(1.7) \qquad R_i = 0 \; , \quad \mu_i > 0 \; , \quad \frac{U_{L_i}}{U_C} = \upsilon_i + \frac{\mu_i}{\lambda} > \upsilon_i \; , \qquad i = m, f.$$

The interior solution (1.6) is the usual point of departure for the labour supply studies. Hereafter I will call these studies "conditional" labour supply studies because the samples over which they are estimated are usually restricted to labour force participants (Section 2). The corner solution (1.7) leads to labour force participation studies estimating the probability of a positive supply of hours (Section 3). I will define the function that combines the choices at the internal and external margin as an unconditional labour supply function since it is estimated over the whole sample of both participants and non-participants. Some alternative methods of defining this function are discussed in Section 4.

2. Conditional Labour Supply Functions

2.1 Theoretical Restrictions

To highlight the theoretical developments in conditional labour supply studies I will discuss the case where both husband and wife are working. This simplifies, with a few alterations, to the case where only one of them is working.

If U is a well behaved utility function, the first order conditions of the Lagrangian (1.5) can be solved uniquely to obtain a set of demand functions[5]

(1.8.1) $C = C(\upsilon_m, \upsilon_f, J)$,

(1.8.2) $L_m = L_m(\upsilon_m, \upsilon_f, J)$,

(1.8.3) $L_f = L_f(\upsilon_m, \upsilon_f, J)$.

Furthermore, a set of restrictions on the income and price coefficients can be derived. Usually these restrictions are either imposed on a system such as (1.8) (e.g., Cournot and

Engel aggregation) or they are tested for as parametric restrictions (e.g., symmetry, homogeneity and the sign of the compensated own substitution effect).

More specifically in the case of demand for leisure one usually tests for at least one of the following restrictions at the sample points:

(i) negative sign of compensated own substitution effect

$$\left[\frac{\delta L_i}{\delta \upsilon_i}\right]_{\hat{c}} < 0 , \qquad i = m, f. \tag{1.9.1}$$

(ii) symmetry of compensated cross substituion effect

$$\left[\frac{\delta L_m}{\delta \upsilon_f}\right]_{\hat{c}} = \left[\frac{\delta L_f}{\delta \upsilon_m}\right]_{\hat{c}} \tag{1.9.2}$$

where the subscript $\hat{c}$ indicates the utility compensated term of the Slutsky equation.

Neither the slope of the demand curve for leisure, nor the gross or net substitutability or complementarity between the husband's and wife's leisure are predictable from pure demand theory.

2.2 Functional Form of Demand Equations

The functional form of the demand equations (1.8.1 to 1.8.3) is constrained only in a general way by pure theory: one should be able to integrate them "backwards" into a "sensible" utility function. This requirement usually excludes demand functions which are linear in the parameters of prices and income.[6]

Two basic methods have been utilized in empirical studies of conditional labour supply functions. One approach confines itself to functional forms for demand equations that are compatible with utility theory. This method may involve nonlinear estimation techniques for a system of equations.[7] Another approach attempts to approximate the parameters of the demand equations with functions that are linear in the parameters. Its relationship to utility theory is somewhat pragmatic. However, this method has been used extensively, primarily because of its econometric simplicity.[8] For the same reason I would adapt this procedure to estimate conditional labour supply functions (Chapter V).

Demand relations which are linear in the parameters can be obtained from the first order conditions of (1.5). Totally differentiating these first order conditions and solving for the demand vector will give

$$(1.10.1) \qquad dL_i = \left\{ \left[\frac{\delta L_i}{\delta \upsilon_i} \right]_C + (H - L_i) \left[\frac{\delta L_i}{\delta A} \right] \right\} d\upsilon_i + \left\{ \left[\frac{\delta L_i}{\delta \upsilon_j} \right]_C + (H - L_j) \left[\frac{\delta L_i}{\delta A} \right] \right\} d\upsilon_j + \left\{ \frac{\delta L_i}{\delta A} \right\} dA$$

or alternatively:

$$(1.10.2) \qquad dL_i = \{\left[\frac{\delta L_i}{\delta \upsilon_i}\right]_c - L_i \left[\frac{\delta L_i}{\delta A}\right]\} d\upsilon_i + \{\left[\frac{\delta L_i}{\delta \upsilon_j}\right]_c - L_j \left[\frac{\delta L_i}{\delta A}\right]\} d\upsilon_j + \{\frac{\delta L_i}{\delta A}\} dJ \; .$$

The procedure is the same for dL_j as left hand side variable. One can also solve for the labour supply vector:

$$(1.10.3) \qquad dR_i = \{\left[\frac{\delta R_i}{\delta \upsilon_i}\right]_c - R_i \left[\frac{\delta R_i}{\delta A}\right]\} d\upsilon_i + \{\left[\frac{\delta R_i}{\delta \upsilon_j}\right]_c - R_j \left[\frac{\delta R_i}{\delta A}\right]\} d\upsilon_j + \{\frac{\delta R_i}{\delta A}\} dA \; .$$

Assuming that the expressions between curly brackets are constant then one obtains upon integration, the following linear demand or supply functions:

$$(1.11.1) \qquad L_i = a_1 + a_2\upsilon_i + a_3\upsilon_j + a_4A \; ,$$

$$(1.11.2) \qquad L_i = a_1' + a_2'\upsilon_i + a_3'\upsilon_j + a_4'J \; ,$$

$$(1.11.3) \qquad R_i = b_1 + b_2\upsilon_i + b_2\upsilon_j + b_4A \; .$$

This functional form was used by Kosters [1963], Cohen, <u>et al</u>. [1970]. One can generalize this procedure assuming that the expressions between curly brackets (1.10.1) to (1.10.3) depend on the respective wage or income level. To do this I rewrite (1.10.3) as follows:

$$(1.12) \qquad dR_i = A \, d\upsilon_i + B \, d\upsilon_j + C \, dA \; .$$

In (1.12) A is the slope of the supply curve, the sign of B indicates gross complementarity (if positive) or gross substitutability (if negative) and C is the income effect. Furthermore, $A - CR_i$ is the compensated substitution effect while net complementarity or substitutability is identified as $B - CR_j$.

In order to account, in a simple way, for possible non-linearities in the wage and income terms I introduce the following assumptions:

$$A = a_1 + a_2 \upsilon_i + a_3 \upsilon_i^2 , \tag{1.13.1}$$

$$B = b_1 + b_2 \upsilon_j + b_3 \upsilon_j^2 , \tag{1.13.2}$$

$$C = c_1 + c_2 A + c_3 A^2 . \tag{1.13.3}$$

Substituting (1.13.1) to (1.13.3) into (1.12) and taking the integral will result in the "polynomial" supply curve:

$$R_i = \text{constant} + (a_1 \upsilon_i + a_2 \upsilon_i^2 + a_3 \upsilon_i^3) + (b_1 \upsilon_j + b_2 \upsilon_j^2 + b_3 \upsilon_j^3) + (c_1 A + c_2 A^2 + c_3 A^3) . \tag{1.14}$$

It is readily seen that A, B, and C as defined in (1.12) are equal to the partial derivatives of (1.14) with respect to respectively, υ_i, υ_j, A. This relationship can then be used to test for restrictions (1.9.1) and (1.9.2).

Clearly, still higher order polynomials could be used in defining A, B, and C. In practice however, quadratic expressions are the most commonly used forms for (1.14) (e.g., Rosen and Welch [1971], Berndt and Wales [1974b]). In estimating labour supply functions (Chapter IV) I have found that occasionally a third order polynomial term would be significant but not a higher order. For a reason to be explained below (Section 4 of this Chapter) I use the logarithm of hours supplied as a dependent variable. This form can be derived by assuming that the right hand side in equations (1.13.1) to (1.13.3) is multiplied by R_i.

In the actual estimation of labour supply functions I also include a number of socio-demographic variables (e.g., age, education, experience) as independent variables in order to control (partly) for taste variations with respect to labour supply in the cross-section of families (see Chapter V).

3. Labour Force Participation Functions

3.1 The Shadow Wage

If either the husband or the wife (or both) is not found participating in the labour force then a corner solution condition such as (1.7) must hold. This implies that at that level of family income the market wage υ_i is smaller than the "shadow wage" or "home wage" ($\frac{U_{\ell_i}}{U_c}$) .

This is as much information as one can hope to get out of non-linear programming theory. An attractive feature of the recent family production models[9] might be its usefulness to explain inequality (1.7) somewhat further. The household is assumed to consist of a consumption sector and a production sector. Utility maximization takes place in two stages. In the first stage the household is seen as minimizing the cost of producing household commodities given the household technology, factor prices and initial endowments of time. This results in a household cost function. In the second stage the household maximizes a utility function defined over the household commodities and subject to the cost function. The final solution of this two stage procedure yields equilibrium values for quantities consumed and for household shadow prices, e.g., the shadow price of time. To explain the latter, one must concentrate on the equilibrium conditions of the household production sector. In equilibrium a factor price must be the same in all sectors of production and must be equal to the value of the marginal product. The value of the marginal product is the product of the commodity price (an "internal" concept in this framework) and the marginal physical product (a given technological fact).

The inequality (1.7) is thus obtained, either because the family values the household commodity greatly (high implicit commodity price) or because it is very efficient in producing these commodities (high marginal physical productivity). In either case the shadow wage should empirically reveal itself as a function of the consumption and production of certain household commodities. For instance: the importance of young children for the labour force participation of their mother is a well investigated example that can be explained in these terms.

3.2 Restrictions on Labour Force Participation Functions

In the interior solution case an important restriction on the coefficients of the demand curves is the negativity of the compensated own substitution effect. For the labour force participation case a similar result can be shown, using the weak axiom of revealed preference which is implied by a utility maximization program.

Suppose a family is constrained to move along the same indifference curve. Suppose furthermore that in a first situation it faces the price vector $v^1 = [p, w_f, w_m]$ and in response to this, chooses consumption vector $x^1 = [C, H, L_m]$, where H is the total amount of time for the period. In a second situation it faces $v^2 = [p, w_f^*, w_m]$ and chooses $x^2 = [C^*, L_f^*, L_m^*]$, where $L_f^* \leq H$ (on the same indifference curve).

The family will minimize the expenditure for a given level of utility and so the following inequalities are implied by utility maximization theory:

$$(1.15) \qquad v^2x^1 \geq v^1x^1 \text{ , and } v^1x^2 \geq v^2x^2 \text{ ,} \qquad \text{or}$$

(1.16.1) $\quad (w_f^* - w_f)\, H \geq 0 ,$

(1.16.2) $\quad (w_f - w_f^*)\, L_f^* \geq 0 .$

Adding (1.16.1) and (1.16.2) together:

(1.16.3) $\quad (w_f^* - w_f)(H - L_f^*) \geq 0 ,$

or

(1.16.4) $\quad \Delta w_f \;\; \Delta R_f \geq 0$

where Δ is a difference indicator.

Everything else (especially other family income) remaining constant, a sufficient increase (decrease) in the own wage rate should induce the consumer to join (leave) the labour force. This relationship was also shown geometrically by Ben-Porath [1973].

3.3 Derivation of the Multinomial Logit Model

A. Idiosyncrasies and stochastic specification

The corner solution condition (1.7) implies that the individual is not participating because at the observed level of family income his shadow wage is greater than (or equal to) his market wage. There is however no way to determine how large this gap is. Furthermore, unless the individuals are completely homogeneous this gap will differ among families due to idiosyncrasies with respect to labour force participation patterns. Household production theory may suggest variables that seek to explain the systematic variation in the shadow wage (e.g., children). However, an unexplainable portion will remain, partly because some variables remain unmeasured and partly because the shadow wage is itself an unobservable (determined by utility considerations). In a cross-section sample one might observe two individuals with the same income, same market wage, same socio-demographic variables (identifying the systematic part of the shadow wage), but one choosing to be in, the other to be out of the labour market. Consequently if one would use a linear probability model (i.e., with 0-1 dependent variable) to explain labour force participation decisions then the error term might play an undesirable role in such an equation. The error term will "explain" a substantial proportion of the observed variation in choice if the idiosyncratic elements (i.e., the unobserved "taste" variations) are important.[10]

The above discussion explains intuitively the unsatisfactory error structure encountered in using the conventional regression model for discrete choices. This corresponds to the technical shortcomings of the error term in the linear probability model.[11] In spite of these shortcomings the linear probability model has been used extensively in the analysis of labour force participation decisions (Cain [1966], Mahoney [1961], Boskin [1973]). A more satisfactory model to explain an individual's choice among distinct alternatives must, however, make explicit the effects of an individual's idiosyncratic preferences.

B. The binomial logit model

I start with a simple case. Assume that the sample of families is split up so that those households where the husband is working are only chosen. I am then only interested in the labour force participation of the wife. Utility maximization suggests that her labour force participation will be a function of her market wage, her shadow wage and other family income. Socio-demographic variables could be added in order to capture systematic variations in the unobservable shadow wage or in "taste for labour force participation". Let g stand for the function representing the systematic part in the explanation of labour force participation choice. Because of the importance of the unobservable idiosyncrasies on discrete choices, a factor representing these idiosyncrasies has to be added in the explanation. To do this one could theorize that a random variable ε is drawn from an assumed distribution (frequently the normal or logistic). The outcome of the common part g and the idiosyncratic part ε will then indicate whether the individual will participate.

More formally one can assume that there exists a "latent" variable q, which is the sum of g and ε. In this simple case the variable q can best be understood as the "desired" amount of supply. If $q \leq 0$ the individual is not in the labour market and vice versa if $q > 0$.

Now we would like to predict if individual i facing g_i will be in or out of the labour force. Assume ε is distributed following a logistic function.[12] Then

$$(1.17.1) \qquad Pr(q_i \leq 0) = Pr(\varepsilon_i \leq -g_i) = \frac{1}{1 + e^{-g_i}} \quad ,$$

$$(1.17.2) \qquad Pr(q_i > 0) = 1 - Pr(q_i \leq 0) = \frac{1}{1 + e^{g_i}} \quad .$$

It is easily seen that

$$(1.18) \qquad \ln \frac{Pr(q_i \leq 0)}{Pr(q_i > 0)} = \ln \left(\frac{1 + e^{g_i}}{1 + e^{-g_i}} \right) = g_i \ .$$

So that we end up with a simple relation between the logarithm of the odds of non-participation over participation (or vice versa) and the function g.

Once g_i is estimated (see Chapter IV for a discussion of estimation problems) model (1.18) will show the (logarithm of the) odds that a family, facing the given market wage and income variables and with the observed socio-demographic characteristics, will choose a consumption vector $[C, L_m, L_f]$ instead of $[C^*, L_m^*, H]$, i.e., it determines a probabilistic rule that can be used to split up the sample in two regimes. Extension to more than two regimes will now be straightforward.

C. The multinomial logit model

Assume that a population of families are all maximizing the basic model (1.1) to (1.4). Because of family idiosyncrasies, they will have different opinions and tastes with respect to the labour force participation of the husband and/or the wife. Therefore this general utility maximization problem will specialize over the population into four "regimes":

$$(1.19.1) \qquad \text{Max } [U(C, L_m, L_f): \; C + L_m\upsilon_m + L_f\upsilon_f \leq H(\upsilon_m + \upsilon_f) + A \; ; \; C > 0, \; H - L_m > 0, \; H - L_f > 0] \; ,$$

$$(1.19.2) \qquad \text{Max } [U(C, L_m): \; C + L_m\upsilon_m \leq H\upsilon_m + A; \; C > 0, \; H - L_m > 0] \; ,$$

$$(1.19.3) \qquad \text{Max } [U(C, L_f): \; C + L_f\upsilon_f \leq H\upsilon_f + A; \; C > 0, \; H - L_f > 0] \; ,$$

$$(1.19.4) \qquad \text{Max } [U(C): \; C \leq A; \; C > 0] \; .$$

So again what is needed is a probabilistic rule that will split the sample into four regimes characterized by the following alternative vectors: (C, R_1, R_2), (C, R_1), (C, R_2), (C). Define these vectors as respectively a_1, a_2, a_3, a_4 and X as their collective set, i.e., $a_i \in X$.

As in the previous discussion I will theorize that the a_i chosen by a particular family is a systematic function of the market wage, income variables and of socio-demographic variables capturing systematic variations in shadow wage and "taste". It is also a function <u>in a random way</u> of a term capturing the familial idiosyncrasies with respect to labour force participation. More specifically:

$$\ell(a_i) = g_i(\upsilon_m, \upsilon_f, A, T) + \varepsilon \tag{1.20}$$

where T stands for the included socio-demographic variables. $\ell(a_i)$, $a_i \in X$ is then a probabilistic rule that will split the sample up into the four regimes, allocating each family in its most probable regime[13] i.e.,

$$p_X(a_i) = \Pr[\ell(a_i) \geq \ell(a_j);\ a_i,\ a_j \in X] \ . \tag{1.21}$$

Furthermore if

$$p_X(\bigcap_i a_i) = p_X(a_1) \ . \ . \ . \ p_X(a_4) \ . \tag{1.22}$$

Then

$$p_X(a_i) = \int_{-\infty}^{+\infty} \Pr[\ell(a_i) = t] \prod_{a_j \in X - a_i} \Pr[\ell(a_j) \leq t]\, dt \ . \tag{1.23}$$

I make the following distributional assumption (which amounts to assuming that ε in (1.20) is distributed with the Weibull distribution[14]):

$$\begin{aligned} \Pr[\ell(a_i) = t] &= g_i\ e^{g_i} && \text{if } t \leq 0 \\ &= 0 && \text{if } t > 0 \ . \end{aligned} \tag{1.24}$$

Then substituting (1.24) into (1.23) will yield:[15]

$$(1.25.1) \qquad p_X(a_i) = \int_{-\infty}^{0} g_i\ e^{g_i t}\ \Pi\ [\int_{-\infty}^{t} g_j\ e^{g_j \tau}\ d\tau]\ dt$$

$$= \int_{-\infty}^{0} g_i\ e^{g_i t}\ \Pi\ e^{g_j t}\ dt$$

$$= \int_{-\infty}^{0} g_i\ \exp\ \{ \sum_j g_j t \}\ dt \quad ,$$

$$(1.25.2) \qquad p_X(a_i) = g_i [\sum_j g_j]^{-1} \ .$$

If the functional form $\exp(Z'\beta)$ is used for g, with Z corresponding to $(\upsilon_m, \upsilon_f, A, T)$ and β a set of coefficients to be estimated for each alternative $a_j \in X$ then (1.25) will be identical with the multinomial logit model (Cragg and Uhler [1970, 1971], McFadden [1974], Theil [1969]) which is an estimable function:

$$(1.26) \qquad p_X(a_i) = \exp(Z'\beta_i)[\sum_{j=1}^{4} \exp(Z'\beta_j)]^{-1} \ .$$

Note that now (compare with (1.18)), (1.26) implies

$$(1.27) \qquad \ln \frac{p_X(a_i)}{p_X(a_j)} = Z'(\beta_i - \beta_j) \ .$$

Consequently,

$$(1.28) \qquad \frac{d \ln \frac{p_X(a_i)}{p_X(a_j)}}{d\ Z_k} = (\beta_{ik} - \beta_{jk}) \ .$$

It should be noted that identification of the parameters of model (1.26) requires a normalization rule, say

(1.29) $\beta_j = 0$.

Therefore (see (1.28)) β_{ik} can be interpreted as the marginal change in the logarithm of the odds of alternative i over the "normalized" alternative j. Using (1.27) we can derive[16]

(1.30) $$d \ln p_X(a_i) = (\beta_{ik} - \sum_{j=1}^{4} p_j \beta_{jk}) \, d Z_k \, ,$$

i.e., the change in the logarithm of the probability of an alternative due to a change in Z_k depends on the outcome of a comparison between the change in the odds of all the alternatives. So whereas the odds are a monotonic function of the independent variables as in equation (1.27), this is not necessarily true for the probability of an alternative (see equation (1.30)).

D. Restrictions on the selection probabilities

The model expressed in (1.26) is the one that I propose to estimate. A discussion of the econometrics of (1.26) is deferred to Chapter IV where the model is used empirically. Before leaving the subject, however, I would like to mention an important theoretical restriction on the selection probabilities of the logit model.

Define X as a set of more than two alternatives and Y as a subset of X consisting of a_1 and a_2 only. Then the axiom of independence of irrelevant alternatives assumes:

$$\frac{p_Y(a_1)}{p_Y(a_2)} = \frac{p_X(a_1|Y)}{p_X(a_2|Y)} \quad , \tag{1.31}$$

i.e., the odds of a_1 being chosen over a_2 in the multiple choice situation X, where both a_1 and a_2 are available, equals the odds of the binary choice of a_1 over a_2. If this axiom holds and if $p_Y(a_2) \neq 0$ or 1 then it can be proven[17] that $p_X(a_1)$ can be written as (1.26), i.e., as the multinomial logit model. This axiom thus underlies the multinomial logit model.

The axiom can easily be violated in reality. To take Debreu's [1960] example, let X consist of:

(a_1) a recording of the Debussy Quartet by the C quartet,

(a_2) a recording of the 8th Symphony by Beethoven by the B orchestra conducted by F,

(a_3) a recording of the 8th Symphony by Beethoven by the B orchestra conducted by K.

The following binary choice probabilities are observed:

$$p(a_1|a_1, a_2) = 3/5, \quad p(a_1|a_1, a_3) = 3/5, \quad p(a_2|a_2, a_3) = 1/2.$$

Then this axiom would predict for the multiple choice situation:

$$p(a_1|a_1, a_2, a_3) = 3/7 \; .$$

Thus, in the binary choice situation the individual would rather have Debussy, but in the multiple choice situation he would prefer

Beethoven, which is a counter-intuitive result. This example suggests therefore that application of the logit model should be limited to situations where the assumption that the alternatives are distinct and weighted independently is plausible (i.e., the alternatives cannot be aggregated). The proposed labour force participation model (see next section) presumably fulfils this requirement (for further discussion, see Section 4 in Chapter IV).

3.4 Two Alternative Family Labour Force Participation Models

A first model follows from the assumption of the existence of a family utility function such as the one defined in (1.1), i.e., $U(C, L_m, L_f)$. If the family maximizes this function subject to their budget and time constraints (equations (1.2) to (1.4)) then they will choose between the following four labour force participation alternatives (see above equation (1.19.1) to (1.19.4):

(1.32.1) $[C^1, L_m^1, L_f^1]$, i.e., both husband and wife working,

(1.32.2) $[C^2, L_m^2, H]$, i.e., husband only working,

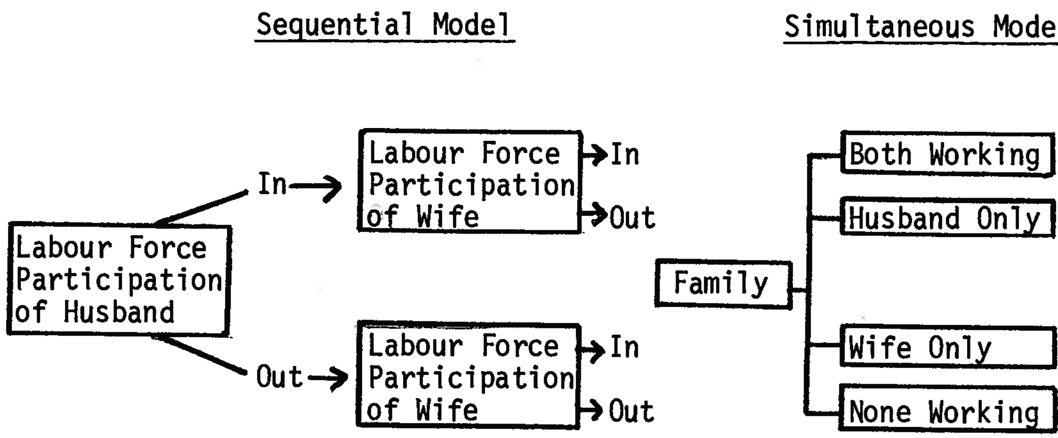
(1.32.3) $[C^3, H, L_f^3]$, i.e., wife only working,

(1.32.4) $[C^4, H, H]$, i.e., none working.

In this model the family, solves its utility maximizing program, considering all four alternatives simultaneously and chooses one. I call this the <u>simultaneous model</u>. In Chapter IV I use the multinomial logit model (1.26) to estimate the probability of a family choosing any of these four alternatives as a function of wages, income and socio-demographic variables. A second model follows from the following *a priori* considerations. For a variety of sociological reasons the husband appears always to be in the labour force unless he is either handicapped, retired, or in school. The husband does not seem to have much "choice" regarding labour force participation. On the other hand a social constraint in this respect does not appear to be as strong for the wife. Previous economic discussions on "primary" and "secondary" income earners in a family already hinted at this distinction caused by sociological constraints. (See Mincer [1966] for a summary of this literature.) Although they are a bit vague these considerations suggest a model in which the labour force participation "decision" of the husband is studied first. The wife's labour force participation choice is then estimated conditional upon her husband's "decision". I thus introduce a type of lexicographic ordering in the labour force participation decisions of the family. I call this the *sequential model*. In Figure 1 the simultaneous model and the sequential model are contrasted.

The sequential model is estimated using the binomial logit model (1.17.2). It is clearly possible to reverse the order of decision and to develop a sequential model in which the wife decides first and the husband's decision is conditional upon hers. I will also estimate this model (although it may seem to be somewhat unrealistic, *a priori*).

FIGURE 1



It is of considerable interest to contrast the estimates of the simultaneous and sequential models and to compare the values of their respective likelihood functions. In terms of the logit specification the sequential model is identical with the simultaneous model only under a very restrictive condition on the parameters of the labour force participation decision. (See Appendix at the end of this Chapter.) In Chapter IV I estimate and compare both the simultaneous and the sequential models.

4. Unconditional Labour Supply Functions

If changes in the market "wage rate" and/or in "income" have an effect on both the labour force participation and labour supply decisions of a sample of families then it may be of interest to obtain an idea of the combined effect of the changes at the internal and external margin. For instance the introduction of a negative income tax policy will change both the "wage rate" and "income" variables of certain families and therefore their labour force participation and labour supply behaviour. In what follows I will discuss two alternative ways of defining unconditional labour supply functions which can be used to summarize the total outcome of such a negative income tax policy.

The two alternative ways of defining an unconditional labour supply function follow from the hypotheses that the parameters of labour supply and labour force participation choice are either the same or different. In Chapter VI I will compare these two hypotheses. Furthermore I will combine the probabilities of the labour force participation choices of a family (estimated in Chapter IV) with the labour supply predictions derived from the conditional labour supply functions (estimated in Chapter V). This procedure (as will be seen below) assumes that the parameters of labour force participation and labour supply decisions are different.

4.1 Comparing the Parameters of Labour Supply and Labour Force Participation

In order to test the hypothesis that the parameters of labour force participation and labour supply are the same, I compare the estimation results using Tobin's [1958] limited dependent variable model with the results obtained with a variant of the limited dependent variable model developed by Cragg [1971][18]. Tobin's model restricts the parameters of labour force participation and labour supply to be the same whereas this is not the case for the Cragg model.

Tobin's model can be used if one assumes that labour force participation is simply truncated labour supply, i.e., one argues that the labour force participation problem only exists because one does not observe the "negative" supply of hours that people might desire to offer. To formalize this, assume that a desired labour supply variable q_t for individual t is defined by

(1.33) $$q_t = Z_t'\gamma + \varepsilon_t$$

whereby ε_t is independent and normally distributed with mean zero and variance σ^2. Z is a set of independent variables and γ a vector of coefficients. When the desired labour supply is negative, the variable that is actually observed, R_t, is zero. When q_t is positive then R_t is equal to q_t. Using the probit model the probability that R_t is zero is then defined as:

(1.34.1) $$\Pr(R_t = 0|Z_t) = C(-Z_t'\gamma|\sigma)$$

where C(.) designates the cumulative unit normal distribution.[19] The density for positive values of hours supplied is:

(1.34.2) $$\Pr(R_t|Z_t) = (2\pi)^{-1/2}\ \sigma^{-1}\ \exp\{-(R_t - Z_t'\gamma)^2/2\sigma^2\}\ .$$

The Tobin model thus only provides one set of parameters: γ/σ, restricted to be the same for both labour force participation and labour supply.

On the other hand it can be argued that for various reasons the restriction imposed by the Tobin model is not reaslistic. For instance, search and information costs might inhibit smooth transfer into the labour force even if a positive q_t is desired (Uhler and Kunin [1972]). Institutional constraints, e.g., standard work week, and limitations from the demand side, e.g., non-availability of part time jobs, might inhibit positive labour supply unless the individual desires to supply at least a certain amount. The existence of these "discontinuities" in going from zero to positive labour supply implies that the "continuous" model (1.33) is incorrect and that the parameters of labour force participation and labour supply choice are different. It would be more correct to separate the labour force participation and labour supply decision.

To do this I assume that a decision first has to be taken whether to participate. Suppose one decides to participate, then a decision is taken on how many hours to supply. The first decision might be represented by a probit model; the second by a standard regression model. However, the dependent variable in the labour supply case, i.e., R_t, can only take non-negative values. This non-negativity could be guaranteed by truncating the distribution of R_t at zero. The model then becomes:

$$\Pr(R_t = 0 | Z_{1t}, Z_{2t}) = C(-Z_t'\beta) , \tag{1.35.1}$$

$$\Pr(R_t | Z_{1t}, Z_{2t}) = (2\pi)^{-1/2} \sigma^{-1} \exp\{-(Y_t - Z_{2t}'\gamma)^2/2\sigma^2\} \frac{C(Z_{1t}'\beta)}{C(Z_{2t}'\gamma/\sigma)} , \text{ for } R_t > 0 \tag{1.35.2}$$

where Z_{1t} and Z_{2t} are vectors of independent variables for individual t and β and γ are vectors of coefficients. Tobin's model (1.34.1) and (1.34.2) is a particular form of (1.35.1) and (1.35.2) if $Z_{1t} = Z_{2t}$ and $\beta = \gamma/\sigma$.

In Chapter VI, I estimate and compare Tobin's and Cragg's model for the husband and for the wife, and test for differences in the parameters of labour supply and labour force participation.

4.2 Unconditional Labour Supply Functions: A Simulation

In Chapter IV, I estimate the probabilities of a family choosing among its labour force participation alternatives given a set of independent variables Z_1. This is done using either the simultaneous or the sequential model. In Chapter V, I estimate labour supply functions corresponding to the labour force participation alternatives. For the alternative "both working" I estimate a supply function separately for the husband and for the wife; for the alternatives "husband only working" and "wife only working", I estimate a supply function for respectively the husband and the wife.

Again, the dependent variable labour supply can only take non-negative values. To assure this restriction one can either use a truncated regression model (such as equation (1.35.2)) or one can assume (as I will do) that the logarithm of R_t is normally distributed. The general form for the labour supply functions is then:

$$\text{(1.36)} \qquad \ln R_{jt} = Z'_{2jt}\gamma_j + \varepsilon_{jt} \ , \quad t = 1, T \ , \quad j = 1, M$$

where ε_t is normally and independently distributed with mean zero and variance σ_j^2 and R_{jt} is the amount supplied (by either husband or wife) when alternative a_j is selected.

In order to derive unconditional supply functions define $P(a_{jt}|\upsilon_t)$ as the probability that family t will choose labour force participation alternative a_j given wage rate υ_t. Also define $E(R_{jt}|\upsilon_t, a_{jt})$ as the expected amount of hours supplied given υ_t and given that alternative a_j is chosen, i.e.,

$$\text{(1.37)} \qquad E(R_{jt}|\upsilon_t, a_{jt}) = \exp\{Z'_{2t}\gamma_j + \frac{1}{2}\sigma_j^2\} \ .$$[20]

Then an unconditional labour supply function for either husband or wife can be determined as an expected value, i.e.,

$$\text{(1.38)} \qquad E(R_t) = \sum_{j=1}^{4} P(a_{jt}|\upsilon_t) \, E\,(R_{jt}|\upsilon_t, a_{jt}) \ .$$

In the second part of Chapter VI I use definition (1.38) to simulate unconditional labour supply functions for the sample of families under investigation. I will also simulate the effects of changes in "income" by redefining (1.38) with income instead of wage rate as the independent variable.

5. The Problem of Unobserved Wage Rates

The demand and/or supply functions (equations (1.8.1) to (1.8.3)), derived in the case of an interior solution to the utility maximization problem, are defined as functions of the market wages and income.[21] These variables are observed for labour force participants so that they can be used in the supply equations. There exists, however, a serious problem for the estimation of labour force participation functions.

Utility maximization suggests that, given non-employment income, the labour force participation decision depends on the difference between the shadow wage and the market wage (equation (1.7)). The shadow wage is determined by household technology and by the preference structure for household commodities (see Section 3.1 above). The shadow wage is unobservable but I assume that it is functionally related to certain socio-demographic variables, e.g., age, education, age of children, etc. (see Section 3.3.A above).[22] The market wage is unobserved for non-participants. Since the market wage is, however, observed for labour force participants, it seems appropriate to use this additional information on part of the population to predict the missing information of the other part of the population. This procedure will be "correct" only if both participants and non-participants are structurally similar. I discuss the rather difficult issue of choosing a satisfactory predictor in Chapter II.

The labour force participation decision in Chapter III will then be defined as a function of the predicted market wage rate(s),[23] non-employment income, and a set of socio-demographic variables.

For reasons of comparability between labour force participation and labour supply decisions I also use predicted wage rates in the specification of the supply functions in Chapter V. In the same Chapter I also estimate supply functions using the observed wage rate and compare the supply and income elasticities for both specifications.

The data used in all empirical applications of this thesis are from the "Panel Study of Income Dynamics" [1972] from the Institute for Social Research, Survey Research Center at the University of Michigan. These data are described in Appendix A. The variables which are used in this thesis are defined in Appendix B.

FOOTNOTES TO CHAPTER II

1 To cite a few examples: Henderson and Quandt [1971, 29], Green [1971, 71-74], also Becker [1965].

2 In the literature on multiple jobholding (e.g., Bronfenbrenner and Mossin [1967], Moses [1962], Perlman [1966]) a distinction is made between individuals who become overemployed and those who become underemployed because of the standard work week constraint.

3 See Weiss [1972], Heckman [1972] for a discussion of these problems in the context of lifetime labour supply. Smith [1972] discusses possible predictions of dynamic utility theory for lifetime wage and labour supply profiles.

4 More specifically, see Uhler and Kunin [1972]. Also see Block and Heineke [1973] for a treatment of labour supply using a Von Neumann-Morgenstern approach.

5 Supply functions can easily be obtained using $H - R_i - L_i = 0$, $i = m, f$.

6 See Goldberger [1967, 101-104].

7 See Wales [1973], Wales and Woodland [1974a, 1974b], Ashenfelter and Heckman [1974].

8 This pragmatic approach was developed by Kosters [1963]. It is used by almost all the authors in a recent volume of labour supply studies edited by Cain and Watts [1973]. See also Cohen, Lerman, and Rea [1970].

9 See for instance, the models in the volume "New Approaches to Fertility", *Journal of Political Economy*, 81, 2, Part II, March/April, 1973. For a critical evaluation of this approach see Pollak and Wachter [1975].

10 See McFadden [1974, 307-312].

11 Goldberger [1964, 248-251]. One can, however, conveniently approach the linear probability model as a linear discriminant function. See Ladd [1966]. A technique to circumvent the problems of the linear probability model is to use "discretised" independent variables. See Bowen and Finegan [1969], Cohen, Lerman, and Rea [1970]. If the variables were initially continuous then this procedure entails a loss of information.

12 It is also frequently assumed that ε is unit normally distributed which then leads to the well known probit model. See Buse [1972] for a discussion of these and other models. The extension of the probit model for choices among more than two alternatives, although feasible, is less practical computationally than the extension of the logit model.

13 The way in which I derive the multinomial logit model is completely analogous to Luce and Suppes [1965, Chapter V] and McFadden [1974]. I avoid calling $\ell(a)$ a "random utility function" as they do because of the potential confusion in nomenclature.

14 Cfr. McFadden [1974, 111].

15 Cfr Theorem 32 in Luce and Suppes [1965, Chapter V].

16 Cfr Theil [1969, 254].

17 Cfr McFadden [1974, 109-110], also Luce [1959, Chapter I].

18 Both Tobin's and Cragg's model use the probit specification to express the probability of an alternative. Their models are developed for

two alternative choices only. It is possible to formulate an extension to the multinomial logit model that allows for a test on whether the labour supply and labour force participation are the same. See footnote on page 832 in Cragg's [1971] article. I choose however, to use the limited dependent variable models because computer programs were readily available for these tests.

19 E.g., $C(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp\{-t^2/2\}dt$.

20 Let R be the vector of observations on the dependent variable (hours supplied in this case) and let $y = \ln R$ be the corresponding vector in the log of this variable. If it is furthermore assumed that

$$(1) \qquad y_t = Z_t'\beta + \mu_t , \qquad t = 1, T$$

with μ_t normally and independently distributed with mean zero and variance σ^2, then it follows that y is normally distributed with mean $Z'\beta$ and variance-covariance $\sigma^2 I$ and R is lognormally distributed with the same parameters. The moments of the lognormal distribution are given by the following formula (e.g., see Press [1972, 139-140]):

$$(2) \qquad E(R_t)^k = \exp\{kZ_t'\beta + \frac{1}{2} k^2 \sigma^2\}$$

so that the mean of R_t would be

$$(3) \qquad E(R_t) = \exp\{Z_t'\beta + \frac{1}{2} \sigma^2\} .$$

In practice, however, β and σ^2 are unknown and in that case the distribution of y depends on the distribution of the estimators of β and σ^2 (e.g., see Raiffa and Schlaifer [1961, Chapter 13]). For convenience reasons I replace, throughout this thesis, the theoretical β vector in (3) with the b vector of least squares coefficient estimates and the theoretical σ^2 value with the least squares variance estimator s^2.

21 It would be more correct to say that labour supply is a function of income and of market wages corrected for the marginal tax rate. See, e.g., Wales [1973], Wales and Woodland [1974a]. I will introduce the effect of income tax considerations on labour supply in Chapter V.

22 Gronau [1973], Heckman [1974] have attempted to estimate the shadow wage as a function of similar socio-demographic variables.

23 Again it would be theoretically more appropriate to correct the wage rate for the marginal tax rate. However, since the predicted wage rate, which I have to use in this case, is only a crude indicator (see Chapter II, Section 3.3) of the potential wage offer that a non-labour force participant might receive, I do not pursue it further. In the case where I use the observed wage I introduce the marginal tax rate adjustment (see Chapter V).

APPENDIX TO CHAPTER II

The Relationship Between the Simultaneous and the Sequential Model

To illustrate the relationship between the sequential and the simultaneous labour force participation model for a family I first introduce some new notation. I use the indices A, B, C, D, for respectively the alternatives: both working, husband only working, wife only working and none working in the simultaneous model. For instance, the probability of "both working", using the logit specification (1.26) will be

$$\text{(A.1)} \qquad p(A) = \frac{e^{Z'\beta_A}}{e^{Z'\beta_A} + e^{Z'\beta_B} + e^{Z'\beta_C} + e^{Z'\beta_D}} \quad .$$

For the sequential model, p(1) and p(2) denote the probability of the husband being respectively in or out of the labour force. Further, p(11|1), p(12|1) denote respectively the probability of the wife being in or out of the labour force, given that her husband is in the labour force. By analogy I define p(21|2) and p(22|2) when the husband is out of the labour force. Therefore the probability of "both working" in the sequential model, again using the logit specification, will be

$$\text{(A.2)} \qquad p(11|1)\ p(1) = \frac{e^{Z'\beta_{11}}\ e^{Z'\beta_1}}{(e^{Z'\beta_{11}} + e^{Z'\beta_{12}})(e^{Z'\beta_1} + e^{Z'\beta_2})}$$

$$(A.2') \qquad p(11|1)\,p(1) = \frac{e^{Z(\beta_{11} + \beta_1)}}{e^{Z'(\beta_{11}+\beta_1)} + e^{Z'(\beta_{12}+\beta_1)} + e^{Z'(\beta_{11}+\beta_2)} + e^{(Z'(\beta_{12}+\beta_2))}} \, .$$

Now I investigate under what conditions the following equalities hold:

$$(A.3.1) \qquad p(A) = p(11|1)\,p(1)$$

$$(A.3.2) \qquad p(B) = p(12|1)\,p(1)$$

$$(A.3.3) \qquad p(C) = p(21|2)\,p(2)$$

$$(A.3.4) \qquad p(D) = p(22|2)\,p(2) \, .$$

To do this I rewrite the logit specification in (A.2') using the fact that in the binomial logit model the coefficients for the two alternatives are equal to each other except for the sign (see (1.17.1) and (1.17.2), i.e.,

$$(A.4.1) \qquad \beta_{12} = -\beta_{11}$$

$$(A.4.2) \qquad \beta_2 = -\beta_1 \, .$$

Thus (A.2') becomes

$$(A.5.1) \qquad p(11|1)\,p(1) = \frac{e^{Z'(\beta_{11} + \beta_1)}}{e^{Z'(\beta_{11}+\beta_1)} + e^{Z'(-\beta_{11}+\beta_1)} + e^{Z'(\beta_{11}-\beta_1)} + e^{Z'(-\beta_{11}-\beta_1)}} \, .$$

Similarly for the other three probabilities of the sequential model

(A.5.2):

$$p(12|1)\ p(1) = \frac{e^{Z'(-\beta_{11} + \beta_1)}}{e^{Z'(\beta_{11}+\beta_1)} + e^{Z'(-\beta_{11}+\beta_1)} + e^{Z'(\beta_{11}-\beta_1)} + e^{Z'(-\beta_{11}-\beta_1)}} ,$$

(A.5.3):

$$p(21|2)\ p(2) = \frac{e^{Z'(\beta_{21} - \beta_1)}}{e^{Z'(\beta_{21}+\beta_1)} + e^{Z'(-\beta_{21}+\beta_1)} + e^{Z'(\beta_{21}-\beta_1)} + e^{Z'(-\beta_{21}-\beta_1)}} ,$$

(A.5.4):

$$p(22|2)\ p(2) = \frac{e^{Z'(-\beta_{21} - \beta_1)}}{e^{Z'(\beta_{21}+\beta_1)} + e^{Z'(-\beta_{21}+\beta_1)} + e^{Z'(\beta_{21}-\beta_1)} + e^{Z'(-\beta_{21}-\beta_1)}} .$$

Note that I use the equality $\beta_{22} = -\beta_{21}$ in (A.5.3) and (A.5.4).
If one writes out the logit specifications for p(A), p(B), p(C), and p(D) similar to (A.1) it will easily be seen that a sufficient condition for (A.3.1) to (A.3.4) to hold is that

(A.6) $$\beta_{11} = \beta_{21} .$$

Note that (A.6) does not follow from the independence of irrelevant alternatives axiom (1.31). (A.6) is a much more restrictive assumption. Using (A.6) it follows that

(A.7.1) $$Z'\beta_A = Z'(\beta_{11} + \beta_1) ,$$

(A.7.2) $$Z'\beta_B = Z'(-\beta_{11} + \beta_1) ,$$

(A.7.3) $$Z'\beta_C = Z'(\beta_{11} - \beta_1) ,$$

(A.7.4) $$Z'\beta_D = Z'(-\beta_{11} - \beta_1) .$$

(A.6) implies that the parameters of the labour force participation choice of the wife are the same whether her husband works or does not work. It also implies that the other sequential model where the wife decides first about her labour force participation and then her husband is the same as the sequential model with the husband first and the wife second. This demonstrates analytically that the sequential model and the simultaneous model provide different hypotheses about family labour force participation behaviour.

CHAPTER III

PREDICTING POTENTIAL MARKET WAGE RATES

1. Introduction

In order to predict the potential market wage obtained by non-labour force participants if they were to enter the labour force, I use a least squares prediction technique.[1] This technique exploits the empirical relationship existing in the sample of labour force participants between the observed wage rate and certain observed socio-demographic characteristics. I assume that the same empirical relationship holds for the sample of non-participants in such a way that I can use the relevant socio-demographic characteristics observed for this population to predict the unobserved wage rate.

The assumption that the same systematic relationship holds for both the sample of participants and non-participants is crucial in order to justify this prediction technique and therefore requires some attention. Before discussing the issue of structural differences I elaborate first on the choice of "relevant" socio-demographic variables to be included in the wage equation.

1.1 Specification Problems

A central feature of some recent studies of wage and income differentials is the concept of rate of return on human capital investment.[2] In general terms human capital theory predicts that costs occurred because of investments (e.g., in institutional schooling or in on-the-job-training, etc.) should be compensated for by relatively higher wage rates after the investment. Therefore variables measuring "schooling" and "experience" become important variables in wage regressions. But very frequently other variables such as sex, race, union membership, occupation are also included (Hill [1959], Adams [1958], Hall [1973], Berndt and Wales [1974a]). In what follows I will do the same and will not restrict the specification of the wage equations to only those independent variables suggested by "human capital investment" theory but will also test for the significance of a set of other variables which I think to be relevant in explaining variations in the observed wage rate.

I estimate wage equations separately for husbands and wives. The specification, however, differs between these two equations because the amount of information available in the sample is much smaller for the wives than for the husbands (from whom the interview was taken). These informational gaps could cause estimation problems. If the variables missing in the specification of the wife's wage equations are significantly related to the dependent variable and are furthermore correlated with at least one of the included variables then the coefficients of the included variables will be biased.[3] Clearly, some of the coefficients of the husband's equations can also be biased as one is never sure of the exact specification. However, I presume that the danger of biased coefficients is greater for the wife's regression as there is less information available. Some indication of the bias in the estimated coefficients of the wife's wage equation, caused by the lack of information, can be obtained by comparison with the husband's coefficients. This comparison is clearly only reliable if it can be assumed that the husband's and wife's equations are structurally the same.

1.2 Prediction Problems

As mentioned above, the appropriateness of using a least squares predictor for the unobserved wage offer depends crucially on the assumption that the sample of non-participants is structurally the same as the sample of participants. Two potential situations could invalidate this assumption:

(A) The corner solution inequality (equation (1.7), Chapter I) indicates that for non-participants the shadow wage is greater than the market wage. The shadow wage was furthermore found to be equal to the value of the marginal product in home production which in turn is a function of household productivity and of tastes for household commodities (Chapter II, Section 3.1). If it could be assumed for instance, that the value of the marginal product does not differ between the sample of participants and non-participants then it would follow from the corner solution inequality that participants are on an average, higher market wage earners than non-participants. There is no firm theoretical expectation for this to hold; it is, however, empirically possible. If it were true the two samples could be structurally different and thus invalidate the proposed prediction technique.

(B) A second reason for structural difference can be deduced from the fact that the market wage is affected by experience. Therefore individuals who specialize in home production, for a substantial length of time (which seems to be true for a portion of the non-participants at any point in time) are presumably depreciating their market skills. If this is empirically true then the wage equation for participants would tend to overpredict the potential wage offer of the non-participants to this extent.

Because structural differences between the sample of participants are empirically possible the prediction technique might not be appropriate. This important matter will be discussed further in Section 3 of this Chapter. First, I will describe the process of obtaining predictors for the husband's wage (Section 2.2) and for the wife's wage (Section 2.3).

2. The Wage Equations

2.1 Description of the Sample

The sample used for the wage equation is a subset of the original sample as described in Appendix A. This subset is obtained by restricting the sample to (i) families with both husband and wife present. Either one must have worked at least once in the sample period (1967-71); (ii) families where husband and wife were the same all five years. This is introduced to make the education variable, observed only in 1967 and 1971, useful over the whole sample period; (iii) observations where none of the variables, to be used in the wage equations, is missing.[4]

The numerical importance of each restriction clearly depends on the order in which they are introduced. In the order mentioned above they cause, respectively 29, 45,[5] and 4 percent to be dropped. This leaves 22 percent or 1123 families observed for each of the five years between 1967 and 1971. If all the observations for which the husband is found working are pooled over all five years, the total sample size is 5076. The similar pooled sample size for the wives is 2880 observations.

Most empirical wage and income distribution studies[6] assert that income or wages tend to be lognormally distributed. In order to determine whether the observed wage distribution for husbands and wives fits the lognormal distribution better than the normal distribution, I calculate a χ^2 and a Kolmogorov-Smirnov goodness of fit test statistic[8] comparing the observed wage distribution with a theoretically expected lognormal or normal distribution.

Table I summarizes these tests for the husbands' and wives' wage distribution (using five intervals only for the wage distribution). Neither the χ^2 test nor the Kolmogorov-Smirnov test allows me to accept the null hypothesis that the husbands' and wives' wage distribution is either normally or lognormally distributed. However, comparing the values of the two test statistics for the normal and the lognormal, indicates that the observed wage distributions fit the lognormal better than the normal (certainly for the wives' wage distribution).

TABLE I - GOODNESS OF FIT TEST ON

WAGE DISTRIBUTIONS

	HUSBAND			WIFE		
		EXPECTED			EXPECTED	
INTERVAL[2]	OBSERVED	NORMAL $\mu=3.5$[3] $\sigma^2=2.4$	LOGNORMAL $\mu=2.5$ $\sigma^2=1.4$	OBSERVED	NORMAL $\mu=2.1$ $\sigma^2=1.8$	LOGNORMAL $\mu=1.2$ $\sigma^2=1.5$
(0.0-2.5)	1809	1733	2046	2144	1709	2089
[2.5-5.0)	2467	2024	2076	638	1031	645
[5.0-7.5)	589	1091	645	63	137	110
[7.5-10.0)	125	213	198	17	3	25
[10.0-∞)	86	15	111	18	0	11
χ^2		very large	163.5		very large	32.5
degrees of freedom			6			3
Kolmogorov-Smirnov		.10	.05		.15	.02[1]
Sample size		5076	5076		2880	2880

(1) accept Ho at .05 significance level.[9]

(2) an open interval is denoted "(", a closed "[".

(3) μ indicates the value of the mean and σ^2 the value of the variance used in calculating the expected frequency distribution. These values are derived from the observed distribution.

It should be emphasized that even though the observed wage rate tends to be lognormally rather than normally distributed, this is not sufficient information to conclude that the basic assumptions of the linear regression model[7] are more closely approximated with a semi-logarithmic model such as

$$(2.1.1) \qquad w_t = e^{X_t'\beta_1} e^{\mu_{1t}}$$

than with an arithmetic model such as

$$(2.1.2) \qquad w_t = X_t'\beta_2 + \mu_{2t}$$

Whether (2.1.1) is more appropriate than (2.1.2) depends on the conditional distribution of the dependent variable given the set of independent variables. The results on the distribution of the observed wage rate suggest that it would be of interest to estimate and compare the wage equations using both the semi-logarithmic and aritmetic specification.

2.2 The Husband's Wage Equations

The regression results on the husband's wage equation are presented in Table II for the arithmetic wage rate as dependent variable and in Table III for the semi-logarithmic wage rate. The definition of each variable is given in Appendix B.

In order to obtain a notion of the importance of certain groups of variables I use Theil's decomposition of the multiple correlation coefficient. The incremental contribution of each

TABLE II - HUSBAND WAGE RATE EQUATIONS

INDIVIDUAL YEARS AND POOLED (ARITHMETIC SPECIFICATION)

	1967	1968	1969	1970	1971	POOL
CONSTANT	- 2.75	- 2.40	- 2.84	- 2.49	- 3.91	- 2.99
	(.81)	(.82)	(.85)	(.95)	(1.20)	(.40)
GRADE 9/12	.51	.35	.32	.34	.24(n.s.)	.35
	(.16)	(.16)	(.16)	(.17)	(.21)	(.08)
TECH	.80	.38(n.s.)	.53	.51	.41(n.s.)	.54
	(.23)	(.23)	(.23)	(.25)	(.30)	(.11)
COLL	1.01	.86	.74	1.12	.86	.92
	(.24)	(.23)	(.23)	(.25)	(.30)	(.11)
BA	2.02	1.82	1.96	2.03	1.87	1.94
	(.32)	(.31)	(.30)	(.33)	(.38)	(.14)
PHD	2.78	2.68	2.95	3.30	3.31	3.00
	(.38)	(.37)	(.36)	(.38)	(.45)	(.17)
ACHIEVE	.01(n.s.)	.05	.03(n.s.)	.02(n.s.)	.04(n.s.)	.03
	(.02)	(.02)	(.02)	(.02)	(.03)	(.01)
RACE	- .28(*)	- .25(n.s.)	- .21(n.s.)	- .22(n.s.)	- .08(n.s.)	-.20
	(.17)	(.16)	(.16)	(.17)	(.21)	(.08)
IQ	.10	.09	.09	.08	.15	.10
	(.03)	(.03)	(.03)	(.03)	(.04)	(.01)
CATH	.14(n.s.)	.12(n.s.)	.26(*)	.07(n.s.)	.16(n.s.)	.14
	(.15)	(.81)	(.15)	(.16)	(.20)	(.07)
JEW	-.04(n.s.)	-.38(n.s.)	.86	.45	.07(n.s.)	.36(*)
	(.40)	(.39)	(.39)	(1.03)	(.51)	(.19)
AGE	.12	.10	.14	.14	.16	.14
	(.03)	(.03)	(.03)	(.04)	(.05)	(.02)
AGE SQ	-.0013	-.0010	-.0015	-.0015	-.0018	-.0015
	(.0004)	(.0004)	(.0004)	(.0004)	(.0006)	(.0002)
RISKAVOID	.10	.11	.11	.13	.17	.13
	(.04)	(.04)	(.04)	(.04)	(.05)	(.02)
URBAN	.30	.29	.16(n.s.)	.28	.45	.29
	(.12)	(.12)	(.12)	(.13)	(.16)	(.06)
SOUTH	-.24(*)	-.36	-.33	-.53	-.42	-.37
	(.14)	(.13)	(.13)	(.14)	(.17)	(.06)
4-9 YRS ON JOB	.05(n.s.)	.31	.23(n.s.)	.05(n.s.)	.33(*)	.21
	(.16)	(.14)	(.14)	(.16)	(.19)	(.07)
10-19 YRS ON JOB	.35	.53	.40	.39	.65	.45
	(.16)	(.16)	(.16)	(.18)	(.22)	(.08)
$\geq$ 20 YRS ON JOB	.41	.67	.45	.56	.73	.57
	(.21)	(.19)	(.19)	(.20)	(.24)	(.09)

TABLE II CONTINUED

	1967	1968	1969	1970	1971	POOL
MANAG	1.29 (.28)	1.84 (.26)	1.41 (.25)	1.77 (.26)	1.42 (.32)	1.55 (.12)
PROF	1.10 (.20)	1.20 (.30)	1.27 (.29)	1.27 (.30)	.71 (.35)	1.09 (.14)
SKILL	.95 (.22)	1.03 (.21)	.88 (.21)	.66 (.27)	.87 (.27)	.91 (.10)
CLERK	.52 (.26)	1.15 (.25)	.83 (.25)	.94 (.23)	.47 (.32)	.70 (.12)
SEMISKILL	.42(*) (.24)	.81 (.23)	.41(*) (.22)	.48 (.23)	.57 (.29)	.51 (.11)
UNSKILL	.48(*) (.26)	.58 (.25)	.28(n.s.) (.25)	.41(n.s.) (.27)	.14(n.s.) (.32)	.35 (.12)
SECOND JOB	-.20(n.s.) (.14)	-.46 (.15)	-.33 (.16)	-.52 (.18)	-.46 (.20)	-.38 (.07)
UNION	.63 (.15)	.57 (.14)	.66 (.14)	.61 (.15)	.51 (.12)	.60 (.06)
Number of observations(2) #OBS(2)	1039	1034	1023	992	988	5076
R^2(3)	.354	.379	.397	.409	.325	.366
$\bar{R}^2$(4)	.338	.363	.381	.393	.307	.363
SE(5)	1.8	1.8	1.8	1.9	2.3	1.9
μ(6)	3.17	3.36	3.49	3.62	3.74	3.47
σ(7)	2.22	2.21	2.24	2.43	2.72	2.37

(1) The numbers in brackets are the standard errors. The coefficients are all significant at the 5% level unless followed by (*) which indicates significance at the 10% level. If a coefficient is followed by (n.s.) this denotes "not significant". Variables are explained in Appendix B.

(2) #OBS: number of observations

(3) R^2: coefficient of multiple correlation

(4) $\bar{R}^2$: adjusted R^2

(5) SE: standard error of the regression

(6) Mean of dependent variable

(7) Standard deviation of dependent variable

TABLE III - HUSBAND WAGE RATE EQUATIONS

INDIVIDUAL YEARS AND POOLED (SEMI-LOGARITHMIC SPECIFICATION)[1]

	1967	1968	1969	1970	1971	POOL
CONSTANT	-1.30 (.20)	-.75 (.19)	-.93 (.22)	-.86 (.21)	-1.15 (.23)	-1.07 (.09)
GRADE 9/12	.19 (.04)	.16 (.04)	.16 (.04)	.13 (.04)	.11 (.04)	.15 (.02)
TECH	.25 (.06)	.15 (.05)	.19 (.06)	.18 (.06)	.14 (.06)	.18 (.03)
COLL	.28 (.06)	.24 (.05)	.24 (.06)	.26 (.06)	.22 (.06)	.25 (.03)
BA	.54 (.08)	.48 (.07)	.52 (.08)	.49 (.07)	.43 (.07)	.49 (.03)
PHD	.60 (.09)	.58 (.09)	.67 (.09)	.64 (.09)	.69 (.09)	.63 (.04)
ACHIEVE	.01 (.005)	.01 (.005)	.01 (.006)	.01 (.005)	.01 (.006)	.01 (.002)
RACE	-.17 (.04)	-.11 (.04)	-.14 (.04)	-.12 (.04)	-.07(*) (.04)	-.12 (.02)
IQ	.03 (.007)	.02 (.007)	.02 (.008)	.02 (.007)	.03 (.008)	.03 (.003)
CATH	.06(n.s.) (.04)	.03(n.s.) (.03)	.05(n.s.) (.04)	.01(n.s.) (.03)	.05(n.s.) (.04)	.04 (.02)
JEW	-.006(n.s.) (.10)	.10(n.s.) (.09)	.18(*) (.10)	.04(n.s.) (.10)	-.09(n.s.) (.10)	.05(n.s.) (.04)
AGE	.05 (.008)	.03 (.008)	.04 (.009)	.04 (.008)	.05 (.009)	.04 (.003)
AGE SQ	-.0006 (.0001)	-.0004 (.0001)	-.0005 (.0001)	-.0005 (.0001)	-.0006 (.0001)	-.0005 (.00004)
RISKAVOID	.05 (.01)	.05 (.01)	.05 (.01)	.05 (.01)	.05 (.01)	.05 (.004)
URBAN	.12 (.03)	.11 (.03)	.07 (.03)	.13 (.03)	.15 (.03)	.12 (.01)
SOUTH	-.09 (.03)	-.12 (.03)	-.13 (.03)	-.16 (.03)	-.10 (.03)	-.12 (.01)
4-9 YRS ON JOB	.09 (.04)	.10 (.03)	.09 (.04)	.03(n.s.) (.04)	.10 (.04)	.09 (.02)
10-19 YRS ON JOB	.15 (.04)	.15 (.04)	.09 (.04)	.10 (.04)	.17 (.04)	.13 (.02)

TABLE III CONTINUED

	1967	1968	1969	1970	1971	POOL
≥ 20 YRS ON JOB	.21	.43	.09(*)	.14	.17	.15
	(.05)	(.07)	(.05)	(.04)	(.05)	(.02)
MANAG	.57	.56	.57	.63	.63	.60
	(.07)	(.06)	(.06)	(.06)	(.06)	(.03)
PROF	.53	.43	.50	.49	.42	.48
	(.07)	(.07)	(.07)	(.07)	(.07)	(.03)
SKILL	.46	.38	.48	.46	.50	.46
	(.05)	(.05)	(.05)	(.05)	(.05)	(.02)
CLERK	.36	.35	.46	.33	.40	.39
	(.06)	(.06)	(.06)	(.06)	(.06)	(.03)
SEMISKILL	.29	.30	.33	.31	.40	.33
	(.06)	(.05)	(.06)	(.05)	(.06)	(.02)
UNSKILL	.32	.18	.27	.27	.21	.26
	(.06)	(.06)	(.06)	(.06)	(.06)	(.03)
SECOND JOB	-.06	-.12	-.08(*)	-.11	-.09	-.09
	(.04)	(.03)	(.04)	(.04)	(.04)	(.02)
UNION	.29	.26	.27	.24	.22	.26
	(.04)	(.03)	(.03)	(.03)	(.04)	(.02)
# OBS	1039	1034	1023	992	988	5076
R^2	.531	.523	.484	.528	.509	.510
$\bar{R}^2$	.518	.510	.470	.515	.496	.507
SE	.45	.40	.45	.42	.43	.43
μ	.967	1.05	1.08	1.11	1.14	1.07
σ	.64	.57	.61	.61	.61	.61

(1) Same comments as Table II

TABLE IV - THEIL'S DECOMPOSITION OF

$R^{2(*)}$ APPLIED TO THE HUSBAND'S POOLED WAGE EQUATIONS

	ARITHMETIC SPECIFICATION			LOGARITHMIC SPECIFICATION		
	ABSOLUTE	RELATIVE		ABSOLUTE	RELATIVE	
(1) CONSTANT	.0069	2.		.0133	3.	
(2) EDUCATION	.0746	20.		.0688	13.	
(3) AGE	.0174	5.	} 9	.0306	6.	} 9
(4) YRS ON JOB	.0104	4.		.0146	3.	
(5) OCCUPATION	.0469	13.		.1501	29.	
(6) UNION	.0101	3.		.0276	5.	
(7) SOCIO-ECON	.0260	7.		.0444	9.	
TOTAL	.1923	54.		.3494	68.	
MULTICOLLINEARITY EFFECT	.1735	46.		.1604	32.	
R^2	.3658	100.		.5098	100.	

(*) cfr. footnote (10) or Theil [1971, 181]. Each class is defined as the sum of the marginal contribution of the variables mentioned.

(2) education = grade 9/12 + tech + coll + ba + phd

(3) age = age + age sq

(4) yrs on job = 4-9 yrs on job + 10-19 yrs on job + $\geq$ 20 yrs on job

(5) occupation = manag + prof + skill + clerk + semiskill + unskill

(7) socio-econ = achieve + race + iq + cath + jew + riskavoid + urban + south + second + union

(1), (6) correspond to the same variable independent variable[10] is then consolidated into groups of selected independent variables. This is shown in Table IV. There it can be seen that "education" and "occupation" contribute most to the explanation of the wage rate. It can be seen that the higher R^2 in the semi-logarithmic wage rate regression is explained for an important part, by the more efficient (i.e., relatively smaller standard errors) estimation of the occupational variables.

I now discuss the independent variables in more detail. This discussion relates to the empirical wage regression literature. It is somewhat outside the main line of argument of this study. Therefore the rest of this section can be omitted without loss of continuity.

The coefficients of the education variables correspond to the predictions of human capital investment theory and can be interpreted as returns on investment in schooling. Such an interpretation is fully justified if education is specified continuously in terms of years of schooling. When the semi-logarithmic wage equation is used, one can interpret the coefficients of a YRS SCHOOL variable[11] as measuring the internal rates of return (Mincer [1970, 1974]). I use the same specification as in Table III, but substitute the YRS SCHOOL and YRS SCHOOL squared variables for the education dummies. This gives the following partial results

$$\ln w = -1.13 - \underset{(.006)}{.0119}\ \text{YRS SCHOOL} + \underset{(.0003)}{.0026}(\text{YRS SCHOOL})^2 \quad . \tag{2.2}$$

These parameter estimates imply internal rates of return going from .4 percent for individuals having finished only five grades to 8.2 percent for Ph.D.'s.[12]

Recent human capital studies (e.g., Taubman and Wales [1973], Griliches and Mason [1972]) have stressed the importance of controlling for ability on the estimation of educational coefficients. In these wage equations I try to control for ability using the IQ variable. When I leave out the IQ variable each education coefficient would in general, increase with a factor equal to at least one standard error. This implies an increasing bias for higher educational levels. This positive bias is caused by the positive effect of ability on both the wage rate and the educational level.[13] A similar bias is presumably also present for the wife's equation, because no measure of her ability is available (see Section 2.3).

Another human capital investment variable is post-schooling investment. This is usually measured by age. Theoretical human capital investment models predict concave earnings profiles (Ben Porath [1967, 1970], Rosen [1972]).[14] The significance of the squared age term supports this prediction. The age-wage profile reaches its maximum at about 45 years for both the pooled arithmetic and semi-logarithmic specification (respectively, 45.3 and 44.5). For the individual years, however, the arithmetic specification tends to predict a turning point at a later age than the logarithmic specific-ation. The human capital literature also predicts different age-wage profiles for different educational levels (Becker [1964]).

If one introduces age-education interaction terms into the regression one observes only a significant interaction for the BA and the Ph.D. level with AGE and AGESQ. Only for husbands with a Ph.D. is there evidence that the wage profile peaks at a much later age (around 50 years) than other husbands. The partial results of the regressions, specified as in Tables II and III but including the age-education interaction terms, are

$$(2.3.1)\quad \text{WAGE} = \underset{(.4)}{-1.9} + \underset{(.02)}{.093}\ \text{AGE} - \underset{(.0002)}{.001}\ \text{AGESQ} + \underset{(.07)}{.301}\ \text{AGE BA} - \underset{(.0007)}{.0032}\ \text{AGESQ BA}$$

$$+ \underset{(.06)}{.465}\ \text{AGE PHD} - \underset{(.0006\ )}{.004}\ \text{AGESQ PHD}$$

$$(2.3.2)\quad \text{ln of WAGE} = \underset{(.09)}{-.95} + \underset{(.003)}{.040}\ \text{AGE} - \underset{(.00005)}{.0005}\ \text{AGESQ} + \underset{(.01)}{.044}\ \text{AGE BA}$$

$$- \underset{(.0002)}{.0004}\ \text{AGESQ BA} + \underset{(.015)}{.0602}\ \text{AGE PHD} - \underset{(.0001)}{.0005}\ \text{AGESQ PHD}\ .$$

Other variables capturing returns on post-educational human investment are the experience variables measuring time on the present job. A significant increase in the wage rate is obtained for a husband having been on the same job more than five years and more than ten years. The increase for having more than twenty years experience at the same job is not significantly different from the increase already obtained after ten years. If I test the null hypothesis that these increments are the same, the calculated F-statistics equals 1.67 (arithmetic specification) or .22 (semi-logarithmic) whereas the critical value at the 5 percent level is 3.84.

A very important set of variables explaining the wage rate are the occupational dummies (especially in the semi-logarithmic specification). Although the occupational wage differences appear to correspond to a social status ordering (e.g., see Duncan et al. [1961]) it is interesting to test whether these differences are statistically significant. In Table V I bring together the results on the F-statistics for several restrictions imposed on the occupational dummy variables. The restrictions are introduced by re-estimating the pooled wage equation as specified in Tables II and III but with restricted occupational groups consolidated into one occupational level.

TABLE V

F-STATISTICS FOR RESTRICTIONS IMPOSED ON OCCUPATIONAL GROUPS - HUSBAND'S POOLED WAGE EQUATIONS

	Calculated F-Statistics	
Restriction	Arithmetic Specification	Semi-logarithmic Specification
PROF = SKILL	1.94	.38
CLERK = SKILL	4.17	10.25
PROF = CLERK = SKILL	4.31	6.58
CLERK = SEMISKILL	2.78	4.91
SEMISKILL = UNSKILL	2.50	11.44

Critical value for F (1, 5076) = 3.84 (5 percent) or 6.64 (1 percent), for F (2, 5076) = 2.99 (5 percent) or 4.60 (1 percent).

The strong confirmation of the SKILL = PROF restriction is surprising.

Finally, the SECOND JOB variable merits some further explanation. The dependent variable is defined as the logarithm of the ratio of labour income over hours worked. The marginal wage however, is not constant over the whole range of hours because it is an aggregate of income earned on standard time, overtime and on moonlighting jobs. Therefore the negative coefficient of the SECOND JOB variable indicates that moonlighters have lower average wages than non-moonlighters.

A test to determine whether the county unemployment rate (in the form of a set of dummies) had an influence on the male wage rate (in the years 1968 to 1971) does not lead to significant results. Another test for the years 1970-71 to study the effect of the industry (also in the form of dummies) on the wage rate gives a significant negative coefficient in the case of the agricultural industry only.

Guided by the goodness of fit tests discussed in the previous section one would expect that if the wage rate tends to be lognormally distributed then the arithmetic specification should be less adequate in explaining the right hand tail of the wage distribution. A supporting indication of this is found in comparing the fit (i.e., the R^2) of the arithmetic and semi-logarithmic specification using different truncation points of the right tail. Doing this the R^2 increases substantially in the arithmetic case but remains virtually constant in the semi-logarithmic case as shown in Table VI.

TABLE VI

R^2 FOR RESTRICTED SAMPLES - HUSBAND'S

POOLED WAGE EQUATIONS

Restrictions on Pooled Sample	Arithmetic Specification	Semi-logarithmic Specification	# OBS
All Observations	.37	.51	5076
Observations With Wage Rate < $25 Only	.45	.52	5069
Observations With Wage Rate < $20 Only	.48	.52	5063
Observations With Wage Rate < $15 Only	.50	.52	5052

2.3 The Wife's Wage Equations

The results on the wage equations for the wife are presented in Table VII for the arithmetic specification and in Table VIII for the semi-logarithmic equation. Table IX summarizes the results on the incremental contributions of the independent variables. As can be seen "education" and "occupation" (especially in the semi-logarithmic specification) contribute most to the fit.

An important difference between the husband's equation and the wife's equation should be stressed again. Because the interviews were conducted with the husband some variables are not observed for the wife. Of these missing variables especially IQ, YRS ON JOB, SECOND JOB, and UNION are important. ACHIEVE and RISKAVOID are also missing

TABLE VII - WIFE'S WAGE RATE

EQUATIONS INDIVIDUAL YEARS AND POOLED (ARITHMETIC SPECIFICATION)[1]

	1967	1968	1969	1970	1971	POOL
CONSTANT	-1.24(*) (.71)	-.22(n.s.) (.57)	-.23(n.s.) (.64)	.49(n.s.) (.75)	-1.82(n.s.) (1.35)	-.76 (.36)
GRADE 12	.07(n.s.) (.20)	.20(n.s.) (.15)	.38(*) (.16)	.21(n.s.) (.17)	.24(n.s.) (.28)	.21 (.09)
TECH	.27(n.s.) (.28)	.28(n.s.) (.20)	.19(n.s.) (.22)	.05(n.s.) (.23)	.70(*) (.39)	.31 (.12)
COLL	.35(n.s.) (.31)	.42(*) (.21)	.57 (.24)	.74 (.25)	.88 (.42)	.62 (.13)
BA	.74(*) (.39	1.04 (.27)	1.45 (.31)	1.65 (.32)	1.77 (.55)	1.29 (.16)
PHD	1.76 (.56)	.90 (.41)	1.39 (.47)	1.00 (.46)	3.29 (.78)	1.64 (.29)
ACHIEVEH	-.006(n.s.) (.03)	.03(n.s.) (.02)	.005(n.s.) (.02)	.01(n.s.) (.02)	.05(n.s.) (.04)	.02(n.s.) (.01)
CATH	.47 (	.37 (.14)	.18(n.s.) (.15)	.33 (.16)	.25(n.s.) (.26)	.30 (.08)
AGE	.10 (.03)	.03(n.s.) (.03)	.06 (.03)	.02(n.s.) (.03)	.12 (.06)	.07 (.02)
AGE SQ	-.001 (.0004)	-.0001(n.s.) (.0003)	-.0006(*) (.0003)	-.001(n.s.) (.0004)	-.0014 (.0007)	-.0007 (.0002)
RISKAVOIDH	.12 (.04)	.08 (.03)	.06(n.s.) (.04)	.02(n.s.) (.04)	-.02(n.s.) (.06)	.06 (.02)
URBAN	-.007(n.s.) (.14)	.07(n.s.) (.09)	.005(n.s.) (.11)	.22(*) (.11)	.41 (.19)	.15 (.06)
SOUTH	-.16(n.s.) (.14)	-.18(*) (.10)	-.29 (.11)	-.26 (.12)	-.22(n.s.) (.20)	-.22 (.06)
PROF	1.42 (.28)	1.37 (.20)	1.67 (.23)	1.55 (.22)	1.77 (.39)	1.59 (.12)
MANAG	2.80 (.53)	.21(n.s) (.48)	1.02(*) (.54)	.68(n.s.) (.46)	.32(n.s.) (.72)	1.01 (.25)
CLERK	.59 (.18)	.67 (.13)	.52 (.14)	.52 (.15)	.63 (.25)	.57 (.07)

TABLE VII CONTINUED

	1967	1968	1969	1970	1971	POOL
SKILL	.43(n.s.) (.57)	.72(n.s.) (.45)	.55(n.s.) (.44)	.52(n.s.) (.38)	.09(n.s.) (1.10)	.50 (.25)
SEMISKILL	.37(*) (.20)	.44 (.14)	.44 (.15)	.54 (.17)	.67 (.27)	.50 (.08)
# OBS	532	580	611	583	574	2880
R^2	.277	.338	.328	.328	.220	.248
$\bar{R}^2$	.253	.318	.309	.307	.196	.254
SE	1.47	1.12	1.28	1.30	2.17	1.52
μ	1.90	1.97	2.10	2.16	2.28	2.09
σ	1.70	1.36	1.53	1.56	2.42	1.76

(1) same comments as Table II. The variable names ending with the letter "H" refer to variables measured for the husband.

TABLE VIII - WIFE'S WAGE RATE EQUATIONS

INDIVIDUAL YEARS AND POOLED (SEMI-LOGARITHMIC SPECIFICATION)[1]

	1967	1968	1969	1970	1971	POOL
CONSTANT	-1.12	-.83	-.78	-.46(n.s.)	-1.24	-.96
	(.26)	(.25)	(.27)	(.31)	(.33)	(.12)
GRADE 12	.11(n.s.)	.14	.14	.16	.20	.15
	(.07)	(.06)	(.07)	(.07)	(.07)	(.03)
TECH	.15(n.s.)	.23	.11(n.s.)	.12(n.s.)	.34	.19
	(.11)	(.09)	(.09)	(.09)	(.10)	(.04)
COLL	.24	.27	.19(*)	.35	.37	.29
	(.11)	(.10)	(.10)	(.10)	(.10)	(.05)
BA	.44	.38	.48	.54	.36	.43
	(.14)	(.12)	(.14)	(.13)	(.14)	(.06)
PHD	.67	.42	.46	.45	.67	.52
	(.21)	(.18)	(.20)	(.20)	(.19)	(.09)
ACHIEVEH	-.003(n.s.)	.02	.005(n.s.)	.01(n.s.)	.02	.01
	(.009)	(.008)	(.008)	(.009)	(.009)	(.004)
CATH	.16	.15	.03(n.s.)	.13(*)	.10(n.s.)	.11
	(.07)	(.06)	(.06)	(.07)	(.06)	(.03)
AGE	.04	.02(n.s.)	.03	.008(n.s.)	.04	.03
	(.01)	(.01)	(.01)	(.01)	(.01)	(.006)
AGE SQ	-.0005	-.0002(n.s.)	-.0003	-.0001	-.0005	-.0003
	(.0002)	(.0001)	(.0001)	(.0002)	(.0002)	(.0001)
RISKAVOIDH	.05	.05	.04	.03(*)	.05	.04
	(.02)	(.01)	(.02)	(.02)	(.02)	(.007)
URBAN	.06(n.s.)	.09	.05(n.s.)	.14	.11	.10
	(.05)	(.04)	(.05)	(.04)	(.05)	(.02)
SOUTH	-.14	-.08(*)	-.11	-.11	-.13	-.11
	(.05)	(.05)	(.05)	(.05)	(.05)	(.02)
PROF	.73	.68	.79	.72	.71	.73
	(.11)	(.09)	(.10)	(.09)	(.10)	(.04)
MANAG	.77	.09(n.s.)	.58	.41	.15(n.s.)	.39
	(.20)	(.21)	(.23)	(.19)	(.18)	(.09)
CLERK	.48	.43	.39	.39	.38	.41
	(.07)	(.06)	(.06)	(.06)	(.06)	(.03)
SKILL	.49	.46	.38	.49	.08(n.s.)	.41
	(.21)	(.21)	(.19)	(.15)	(.27)	(.09)
SEMISKILL	.42	.39	.38	.44	.35	.40
	(.07)	(.06)	(.06)	(.07)	(.07)	(.03)

TABLE VIII CONTINUED

	1967	1968	1969	1970	1971	POOL
# OBS	532	580	611	583	574	2880
R^2	.391	.399	.344	.370	.382	.367
$\bar{R}^2$	.371	.381	.325	.352	.363	.3631
SE	.548	.495	.544	.535	.535	.532
μ	.403	.487	.534	.563	.590	.517
σ	.692	.628	.662	.663	.670	.666

(1) same comments as Table VII

TABLE IX - THEIL'S DECOMPOSITION OF

R^2 APPLIED TO THE WIFE'S POOLED WAGE EQUATIONS[1]

	ARITHMETIC		SEMI-LOGARITHMIC	
	ABSOLUTE	RELATIVE	ABSOLUTE	RELATIVE
(1) CONSTANT	.0012	1	.0131	4
(2) EDUCATION	.0356	14	.0384	10
(3) AGE	.0086	3	.0123	3
(4) OCCUPATION	.0723	28	.1642	45
(5) SOCIO-ECON	.0114	4	.0248	7
TOTAL	.1291	50	.2528	69
MULTICOLLINEARITY EFFECT	.1293	50	.1140	31
R^2	.2584	100	.3668	100

(1) comments are the same as Table IV, except that variables not defined in Tables VII and VIII are deleted from the definitions.

but the husband's variables could be used as proxies, assuming they reflect household, rather than individual, behaviour. RACE, IQ, and JEW, observed for the husband are not significant in the wife's equation. The effect of these missing variables is to bias the coefficients of the included variables, e.g., the bias in the coefficients of the education variable caused by a missing ability (IQ) variable is an obvious example.

With this possible bias in mind, I discuss now the estimation results in more detail. Again this discussion is somewhat outside the main line of argument of this study and the rest of this section can be omitted without loss of continuity.

The education coefficients yield results expected from human capital theory. Internal rates of return can be deduced from the following specification

$$\text{(2.4)} \qquad \ln \text{ of WAGE} = -.71 - \underset{(.02)}{.0544} \text{ YRS SCHOOL} + \underset{(.0008)}{.0042} (\text{YRS SCHOOL})^2 \, .$$

For the wife the rate of return ranges from -3 percent for individuals having completed only five grades to 9.7 percent for Ph.D.[15] It is somewhat hazardous to compare the wife's rates of return with the husband's, because of the missing IQ variable. Inclusion of the latter in the husband's equation reduces his rate of return (see Footnote 12).

The only other human capital element in the wife's wage equation is the set of AGE variables. The turning points predicted by the pooled equations are respectively 50.2 years (arithmetic) and 52.3 years (semi-logarithmic), which is later than for the husband. This can be expected if women enter the labour force later than men, e.g., because of childbearing and taking care of pre-school children. The turning points predicted in the individual year equations are, however, quite erratic and even insignificant in 1968 and 1970. That the age variable is less firmly established for the wife's equation should come as no surprise. The sample of married women is certainly heterogeneous with respect to post-educational human capital investment. As will be seen, the sample contains a much higher proportion of occasional labour force participants than the sample of husbands.

Evidence on differences in wage profiles for varying educational levels is also very dubious. In this sample the AGE PHD interaction terms are significant only in the arithmetic specification.[16] None of the other age-education interactions is significant in either specification.

The reversed order of magnitude of the coefficients of PROF and MANAG for the wife compared with the husband might cause some concern. This result could be explained by noting that the occupational classification is probably different for males than for females. Furthermore, a different set of restrictions, imposed on the occupational variables, holds for the wives (see Table X).

TABLE X

F-STATISTICS FOR RESTRICTIONS IMPOSED ON OCCUPATIONAL GROUPS - WIFE'S POOLED WAGE EQUATIONS

Restriction	Calculated F-Statistic Arithmetic Specification	Semi-Logarithmic Specification
SKILL = PROF	16.64	11.46
SKILL = CLERK	.09	0.0
SKILL = CLERK = SEMISKILL	.35	.07
MANAG = SKILL = CLERK = SEMISKILL	1.39	.07

The 5 percent critical value for F (1, 2880) = 3.84, for F (2, 2880) = 2.99, and for F (3, 2880) = 2.60. Therefore one can conclude that for the wives the occupational classification can, without loss of information, be compressed into only three occupational classes: professional workers, workers with some skills, and unskilled workers.

Of the remaining socio-demographic variables, attention should be drawn to the insignificance of the RACE dummy (significant in the husband's equation). The JEW dummy is already weak in the husband's equation, so the fact that it does not pass the critical level in the wife's equation is not surprising. Other variables that were insignificant were dummy variables indicating county's unemployment levels (tested for in a 1968-1971 pool) and dummy variables for the industry (tested for in a 1970-71 pool).

Comparisons were also made on the fit of the arithmetic and semi-logarithmic specification if the dependent wage variable is truncated at the right side. There are only four wives with a wage higher than \$15. If these four are deleted, the R^2 for the arithmetic specification increases from .26 to .32 while the R^2 of the semi-logarithmic specification remains unchanged at .37.

3. Prediction

3.1 Problems in Selecting a Predictor

The objective of this Chapter is to find a suitable predictor for the unobserved market wage of non-participants. This task presents several difficulties and potentially insoluble issues (i) the missing variable problem: it was indicated above that an important flaw in the wife's equation is the bias caused by ommitted variables. Yet another missing variables problem applies to both the husband and wife equations. Certain variables, included in the wage regressions discussed in the previous section are observed only if the individual is working. These variables are, in the case of the husband, the experience and the occupation dummies, the second job and the union dummy; in the case of the wife only the occupation dummies. The problem then becomes what to use as a predictor. Two options are readily apparent: (A) to use the equations discussed in the previous section but delete the unobserved variables; (B) to use a new equation obtained from regressing wage on only those variables that are observed for non-participants.

If the true model corresponds to the husband and wife equations presented in Tables II, III, VII, and VIII, then these least squares estimates can be used to obtain optimal (BLUE) predictors.[17] If, however, the least squares estimates corresponding to option (A) or (B) are used then our predictor will presumably be no longer unbiased. Biasedness is caused in option (A) by deleting relevant variables. The least squares estimates in option (B) could also be biased because of the missing variables. Correlation between the included and excluded variables could possibly reduce the bias in this case. In order to obtain some idea of the comparative predictive performance of either of these two biased predictors, I compare their respective predictions in a test in Section 3.3.

(ii) The sample problem: an even more severe problem in predicting unobserved market wages follows from an intuitively reasonable, though hardly provable consideration. Suppose that the sample used to estimate the wage regressions is different from the sample for which the potential wage rate must be predicted. If these two populations are indeed structurally different then it is not reliable to use the regression estimates derived in one population to predict the other population.

Arguments to explain the structural difference between the sample of working individuals (over which wage regressions are estimated) and the sample of non-labour force participants (for whom I need to predict the potential wage offer) were already mentioned above (see Section 1, this Chapter).

3.2 Test of Structural Difference

To obtain some notion of the existence and extent of these structural differences and of the resulting severity of the prediction bias I proceed by splitting all the observations on either the husbands or the wives up into three groups: (A) a group consisting of individuals who were in the labour force all five years; (B) a group consisting of individuals who were in the labour force at least one year but not all five years and finally (C) the ones that were never in the labour force.

It is reasonable to view these three groups as being "ordered" in terms of structural differences, i.e., A is "closer" to B than to C. If this is true then I can argue that if I find that A is structurally different from B then A is different from C and also A and B pooled are different from C. The ultimate objective of these tests is to assess the appropriateness of predicting potential market wage rates for non-participants using information on participants.

A first indication on structural differences between sample A and B can be found from a Chow test[18] on the null hypothesis that A and B belong to the same sample. The equations for husband and wife are specified as discussed in the previous section. The null hypothesis tested is that sample B is structurally equal to sample A. In the husband's case there are 504 individuals out of 5076 who did not work all five years but worked at least once. The calculated F-statistic for the arithmetic specification is 2.14, for the semi-logarithmic specification 5.14. The theoretical value of F (27, 5076) is smaller than 1.53 (5 percent significance level) or 1.81 (1 percent significance level). Consequently the null hypothesis is rejected.

Similarly for the wife I test that the sample of 1140 wives who did not work all five years but worked at least once are structurally the same as the sample of wives that worked all five years. The calculated F-statistics are 4.09 (arithmetic case) and 6.0 (semi-logarithmic case). The critical value for F (18, 2880) is smaller than 1.65 (5 percent level) or 2.01 (1 percent level). Again the null hypothesis that sample A and B are the same must be rejected.

It is not possible to test whether sample C is different from A and B as the wage rate is not observed for C. But in view of the definition of respectively A, B, and C and taking account of the fact that A and B are already structurally different, one certainly should be aware of the probability that C could be structurally different from A and B.

3.3 Prediction Test

To investigate the matter of prediction bias further, I present the results of using the least squares coefficients estimated over sample A to predict the observed wages of sample B. I am interested in the following issues:

(i) How good are the predictions from sample A onto sample B judged by some conventional criteria. This will possibly give an idea of the prediction bias when predicting potential wage offers for non-participants (as I will do below).

(ii) How does the predictive power compare for the semi-logarithmic versus the arithmetic specification. Also how do predictions compare using the least squares coefficients of a fully specified equation versus a partially specified one. With "full" I mean the specifications used in Section 2 of this Chapter; with "partial" I mean a regression including only those variables observed for non-participants. As was seen in Section 3.1 both methods usually lead to biased predictors.

The prediction results are summarized in Table XI. First of all, the predictions are less than fully satisfactory. Maybe they are still acceptable for the husband, but they are very poor for the wife. The possible implication of this result is compounded by the fact that I must predict a potential wage rate for wives in many more cases than for the husband. There is not much difference in terms of predictive ability between the arithmetic or semi-logarithmic specification except that the semi-logarithmic predictor does a bit better (smaller mean squared error and smaller inequality coefficient) in the husband's case. Neither is there much difference between the full or partial specification.

Together these tests do not provide great confidence on the reliability of the predictors. It does indicate, however, that any least squares predictor will probably be unsatisfactory, certainly

TABLE XI - GOODNESS OF FIT MEASUREMENTS COMPARING

PREDICTED WITH OBSERVED WAGE RATES FOR OCCATIONAL LABOUR FORCE PARTICIPANTS

	HUSBAND				WIFE			
	FULL		PARTIAL		FULL		PARTIAL	
	A.S.**	S.S.**	A.S.	S.S.	A.S.	S.S.	A.S.	S.S.
# OBS	504	504	504	504	1140	1140	1140	1140
# POSITIVE ERROR	344	300	171	169	572	556	348	321
% POSITIVE ERROR	68%	60%	33%	34%	50%	49%	31%	28%
CORRELATION COEFFICIENT	.48	.48	.49	.48	-.02	-.02	-.01	-.01
ROOT MEAN SQUARE ERROR	1.81	1.74	1.71	1.71	33.8	33.8	33.8	33.8
MEAN ERROR	.58	.44	-.35	-.31	1.25	1.25	.77	.71
REGRESSION COEFFICIENT	.9	1.3	.8	.8	-1.3	-1.7	-.6	-.4
THEIL's INEQUALITY COEFFICIENT(*)	.33	.31	.26	.27	.94	.95	.93	.93
BIAS(*)	.14	.06	.04	.03	.001	.001	.00	.00
VARIANCE(*)	.24	.48	.20	.21	.97	.97	.95	.95
COVARIANCE(*)	.62	.46	.76	.76	.029	.029	.05	.05
AVERAGE WAGE	2.7	.75	2.7	.75	1.9	.4	1.0	.4

(*) Cfr. Theil [1961] pages 31-42.

(**) A.S. = arithmetic specification, S.S. = semi-logarithmic specification for the wife's potential wage rate.

3.4 Choice of A Predictor

Given the results of the previous sections the final choice of which predictor to use must be somewhat subjective. I have chosen a predictor based on the coefficients of a semi-logarithmic regression estimated over the population of participants and specified to contain only those variables that are observed for non-participants, i.e., what was called above the "partial" specification. I choose the semi-logarithmic form because it relates better to the observed wage distribution than the arithmetic form (Section 2.1 above). I select the "partial" specification because if the included and excluded variables are correlated, this specification might capture more of the variation in the dependent variable than the "full" specification (which uses the results obtained in Section 2 above but deletes the unobserved variables).

The regressions used for prediction of the husband's and of the wife's wage rate are given in Table XIII. Note that some variables which were not found significant when the specification of Table IV or IX was used turn out to be significant in the "partial" specification of Table XII, e.g., KIDS, IQ H .

Using the regressions of Table XII I will "predict" wage rates both for non-participants and participants. This predicted wage variable is used in the labour force participation functions of Chapter IV and the labour supply functions of Chapter V. It can be interpreted

TABLE XII - HUSBAND'S AND WIFE'S POOLED
WAGE EQUATIONS (SEMI-LOGARITHMIC SPECIFICATION ESTIMATED FOR PREDICTION PURPOSES)(1)

	HUSBAND	WIFE
CONSTANT	- .91 (.1)	-.87 (.14)
GRADE 12	.18 (.02)	.19 (.03)
TECH	.19 (.03)	.23 (.05)
COLL	.24 (.03)	.42 (.05)
BA	.51 (.03)	.76 (.06)
PHD	.66 (.04)	.89 (.09)
ACHIEVEH	.007 (.003)	.02 (.004)
RACE	-.14 (.02)	-.1 (.03)
IQH	.02 (.003)	.02 (.006)
CATH	.07 (.02)	.16 (.03)
AGE	.06 (.004)	.03 (.006)
AGE SQ	-.0007 (.00004)	-.0003 (.0001)
RISK H	.06 (.004)	.06 (.007)
URBAN	.19 (.01)	.10 (.02)
SOUTH	-.18 (.02)	-.11 (.02)
KIDS	-.009 (.003)	-.01 (.005)
# OBS	5076	2880
R^2	.39	.28
SE	.48	.56
μ	1.07	.52

(1) same comments as Table II

as a indicator of the "permanent wage rate", i.e., as the long run level of the rental income from the individual's human capital.[19]

In order to simplify the prediction of the arithmetic wage rate using the semi-logarithmic regression I assume that the predicted arithmetic wage rate is lognormally distributed with mean $X'\hat{\beta}$ and variance covariance matrix $s^2 I$ where X is the matrix of observed variables, $\hat{\beta}$ is the vector of estimated coefficients in Table XII and s^2 is the squared standard error of the regressions in the same table. The prediction formula for any individual t is then[20]

(2.5) $$\text{predicted wage } (t) = \exp \{X_t'\hat{\beta} + \frac{1}{2} s^2\} .$$

FOOTNOTES TO CHAPTER III

1 In the preliminary investigations of labour force participation, I have used another prediction method. This method consisted of restricting the sample to those families where at least one wage rate is observed in the sample period (1967-71). Then I predicted a potential wage rate choosing the observed wage rate in a year of participation as close as possible to the year of non-participation (I also adjusted the observed wage rate for a 4 percent annual growth rate). This method, however, excludes from the sample an obviously interesting group in terms of labour force participation behaviour, viz, those husbands or wives that did not participate at all in the five year sample period. Because of this bias I have less confidence in the results of these models and therefore have dropped it after some experimentation.

2 See, e.g., Becker [1964], Mincer [1970], Ben-Porath [1967, 1970], Rosen [1972] for some of the theoretical foundations of this approach. See, e.g., Johnson and Stafford [1974], Mincer [1974] for some recent empirical applications.

3 See Theil [1971, 548-549].

4 The reason for most of the incomplete observations was that the county unemployment rate was missing. This information was to be obtained, not in the interview, but from state employment offices which frequently failed to answer.

5 Half of these are split-offs, i.e., new families formed during the interviewing period. Cfr. Appendix A.

6 Cfr. Mincer [1970] for a good summary in this respect.

7 Cfr. Theil [1971, 110-111], Assumption 3.3.

8 The χ^2 test statistic is based on the difference between observed and expected frequency in each interval. The Kolmogorov-Smirnov test utilizes the relative cumulative frequency of a theoretical and observed distribution and is based on the greatest divergence between these two. Cfr. Kendall and Stuart [1973, 436-482], Volume 2, Chapter 30.

9 This significance, however, disappears if finer class intervals are introduced.

10 Cfr. Theil, Chapter 4 especially page 181. Define R^2 as the multiple correlation coefficient of the regression containing all independent variables, and R_h^2 as the multiple correlation coefficient corresponding to the same equation specified without the h^{th} variable. Then the incremental contribution of the h^{th} variable is defined as $R^2 - R_h^2$ and the following equality holds

$$R^2 - R_h^2 = \frac{(1 - R^2)}{n - k} \; t_h^2 ,$$

t_h being the t-ratio of the h^{th} variable. The multicollinearity effect is defined as $R^2 - \sum_h (R^2 - R_h^2)$.

11 Cfr. Appendix B for coding of the YRS SCHOOL variable.

12 The formula for the internal rate of return derived from (2.2) is $r = - .0119 + .0052t$ so that for

0 - 5 grades	t = 3	r = .0037
6 - 8 grades	t = 7	r = .0245
9 -11 grades	t = 10	r = .0401
highschool	t = 12	r = .0505
technical training	t = 13	r = .0557
college dropouts	t = 14	r = .0609
BA	t = 16	r = .0713
PHD	t = 18	r = .0817

See Mincer [1974, 51-59]. His rates of return explaining the logarithm of earnings, however, are decreasing for increasing schoolings levels, even controlling for weeks worked. His regression, however, includes fewer variables. If ability is left out of the regression the internal rates of return on average, increase .4 percent to .5 percent for each educational level. The result that r is increasing with increasing education brings about the problem of when to stop education. Presumably an individual's finite lifespan makes him decide to stop investing and reap the benefits.

13 In its simplest form (assuming all other variables kept constant), Theil's [1971, 549] missing variable formula for this case is

$$\beta_{w,\,ed} = \beta_{w,\,ed.iq} + \beta_{w,\,iq.ed} \cdot b_{iq,\,ed} \;.$$

14 Note that these models divert from the classic paradigm of maximizing a (lifetime) utility criterion. Decision makers are assumed to maximize lifetime disposable income. This simplifies the optimization problem. Heckman [1972] has tried to carry out the classic utility

maximization procedure in this context but argues that this model is only tractable if one assumes specific functional forms for the utility and production functions.

15 Similar to footnote 12 the formula can be derived from (2.4) and is equal to

$$r = .0544 + .0084t$$

so that for

0 - 5 grades	$t = 3$	$r = -.0292$
6 - 8 grades	$t = 7$	$r = .0044$
9 - 11 grades	$t = 10$	$r = .0296$
highschool	$t = 12$	$r = .0464$
technical training	$t = 13$	$r = .0548$
college dropouts	$t = 14$	$r = .0632$
BA	$t = 16$	$r = .0800$
PHD	$t = 18$	$r = .0968$

16 $$\mathrm{WAGE} = \underset{(-1.8)}{-.64} + \underset{(.02)}{.0642}\ \mathrm{AGE} - \underset{(.0002)}{.0007}\ \mathrm{AGE}^2 + \underset{(.08)}{.2244}\ \mathrm{AGE\ PHD} - \underset{(.0008)}{.0022}\ \mathrm{AGE}^2\mathrm{PHD}$$

$$\ln \mathrm{WAGE} = \underset{(.13)}{-.94} + \underset{(.006)}{.0307}\ \mathrm{AGE} - \underset{(.0001)}{.0003}\ \mathrm{AGE}^2 + \underset{(.03)}{.0278}\ \mathrm{AGE\ PHD} - \underset{(.0003)}{.0003}\ \mathrm{AGE}^2\mathrm{PHD}$$

17 Cfr. Theil [1971, 125], Theorem 3.5.

18 See Chow [1960] or Fisher [1970]. A Chow test, however, requires homoskedastic disturbances in order to be valid. Johnston [1972] suggests that testing for homoskedasticity can be done by applying the standard

test for homogeneous variances to the dependent variable if one has plentiful cross-section data. (See Johnston [1972, 218].) This test statistic for homogeneous variance will be distributed approximately as $\chi^2(m - 1)$ under the hypothesis of homogeneous variances. (m is the number of intervals used to split up the dependent variable.) Applying this test statistic to the arithmetic wage rate gives a calculated χ^2 value of 46.2 for the husbands and 91.9 for the wives ($m = 16$). The theoretical χ^2 (15) value at the .5 percent level is 32.8 so that the null hypothesis of homoskedasticity cannot be accepted on the basis of this test.

19 Similar to Friedman's [1957] definition of permanent income. See especially page 21.

20 See Footnote 20, Chapter II.

CHAPTER IV

FAMILY LABOUR FORCE PARTICIPATION DECISIONS

1. Introduction

Two alternative models of family labour force participation decisions will be considered in this thesis. (See Chapter II, Section 3.4 above.) First the simultaneous decision model in which a family is described as choosing among four possible labour force participation alternatives: (i) both husband and wife working, (ii) husband only working, (iii) wife only working, and (iv) none working. This model is estimated using the multinomial logit technique (see Chapter II, Section 3.3.C above).

The second family labour force participation model is the sequential decision model. In this model the labour force participation decisions are taken sequentially: first the husband decides whether to join the labour force; then the wife decides conditional upon her husband's choice. I will also briefly mention the other possible sequential model where the wife decides first and her husband's choice is conditional upon hers. This second sequential model seems to be less realistic *a priori*. As will be seen the statistical results confirm this *a priori* expectation. I use the binomial logit technique (see Chapter II, Section 3.3.B above) to estimate the sequential model.

Logit models establish the probability of choosing a particular alternative as a function of a set of independent variables. As was found in Chapter II (Section 3.1) labour force participation choices depend on the difference between the shadow wage and the market wage, given the level of family income. To capture the unobservable shadow wage concept which is related to "tastes" for household commodities and to "productivity" in the household (see Chapter II, Section 3.1), I introduce a set of socio-demographic variables, e.g., education, race, children, etc. The own market wage is approximated with an estimate of the "permanent" wage rate (see Chapter III). Utility theory predicts a positive relationship between the change in the own wage rate and the decision to participate, keeping utility constant (see Chapter II, Section 3.2). The latter is achieved by controlling for the income effect. I measure the income effect with an asset income variable (see Appendix B) and then compare the coefficients on the own wage rate and the asset income variable. Another element of family income, however, is the labour income that the other partner can earn. This effect will be approximated with the "permanent" wage variable of that partner.

In what follows I discuss first some technical issues related to the estimation of logit models (Section 2). Then I estimate and discuss the results of the simultaneous labour force participation model in Section 3 and of the sequential model(s) in Section 4. Section 5 concludes this chapter with a comparison of both family labour force participation models.

2. Estimation Problems of the Logit Model

The binomial logit model originated in biometric studies but has already been used extensively in economics, e.g., studies of the acquisition of durable goods, the choice of transportation modes, etc.[1] The multinomial logit model was developed by Theil [1969] and its statistical properties were extensively discussed by McFadden [1974]. The multinomial logit model has been used in economics to study corporate choice among long term financing instruments (Baxter and Cragg [1970]), the demand for cars (Cragg and Uhler [1970]), the structure of asset portfolios of household (Cragg and Uhler [1971]), the choice of transportation modes (McFadden [1974]) and occupational choice (Hall and Kasten [1974]).

The logit models may be estimated by maximum likelihood. (See, e.g., Theil [1970] for another estimation method.) Assuming that the choice for each family is independent of the choice of other families, the likelihood for the sample becomes

$$(3.1) \qquad e^L = \prod_{t=1}^{f_1} p_X(a_{1t}) \quad \cdots\cdot \quad \prod_{t=1}^{f_M} p_X(a_{Mt}) ,$$

where X is a set of M possible alternatives a_i $(i = 1, M)$, $p_X(a_{it})$ is the probability of family t $(t = 1, T)$ choosing alternative a_i and f_i is the number of families in the sample choosing alternative i so that $\sum_{i=1}^{M} f_i = T$, the sample size.[2] If I substitute the logit specification discussed in Chapter II (equation (1.26)):

$$(3.2) \qquad p_X(a_{it}) = \frac{e^{Z_t'\beta_i}}{\sum_{j=1}^{M} e^{Z_t'\beta_j}}$$

into equation (3.1) and differentiate the logarithm of (3.1) with respect to the β's then I obtain nonlinear normal equations. Since the matrix of second partials of the log-likelihood function is the negative of a weighted moment matrix of the independent variables, it is negative semi-definite and thus L is concave in β.[3] It has been shown that the nonsingularity of the Hessian depends, analogous to the least squares model on a full rank condition for the matrix of independent variables.[4] This condition will be satisfied provided the independent variables are not collinear. If the Hessian is non-singular then a vector maximizing the logarithm of (3.1) will be unique, provided a maximum exists. It has been shown[5] that the probability that a maximum likelihood estimator exists and that it is consistent and asymptotically normally distributed approaches one (under some fairly weak conditions) as the sample size approaches infinity.[6]

If L in (3.1) is strictly concave in β, non-linear gradient methods will yield the maximum likelihood estimates if the model is well specified (i.e., full rank condition and not all observations concentrated in one alternative). To estimate the binomial logit models I use two computer programs: CSP and PROLO. For the multinomial model I use CSP and THEIL.[7] All programs are based on the standard Gauss-Newton iterative method.

In terms of use of computer time the more recent CSP program usually converged substantially faster than THEIL and only slightly faster than PROLO. PROLO and THEIL on the other hand provide the user with results not available in CSP, such as goodness of fit statistics (see below), and the asymptotic variance covariance matrix. The latter is particularly useful to test the significance of differences between the coefficients belonging to different alternatives in the multinomial logit model. THEIL also offers the possibility of restricting the coefficients in and across the alternatives.

THEIL, however, is very time-consuming and experimenting with various specifications using this program is prohibitive.[8] My research strategy therefore is to derive a satisfactory model by means of the CSP program and to re-estimate the "final" model using THEIL. The numerical estimates of the coefficients and asymptotic standard errors were sufficiently close (up to the third digit) in both programs. This provides some confidence that the estimates of both programs are numerically reliable.

The goodness of fit statistics that will be given for the logit models in the next two sections are all transformations of a likelihood ratio. The likelihood ratio under consideration, compares the logarithm of the likelihood at its maximum for the given logit model with the logarithm of the likelihood of the same model with all coefficients except the intercept constrained to be zero. The logarithm of the likelihood of this constrained model can simply be written as

(3.3) $$L_{constr} = \sum_{j=1}^{M} f_j \ln (f_j/T) ,$$

where M is the number of alternatives, f_j is the number of occurrences of alternative j and T is the sample size. If L is the maximum of the logarithm of the (unconstrained) model, then a likelihood ratio test statistic can be defined for the logit model as

(3.4) $$\lambda = -2(L_{constr} - L) ,$$

which is asymptotically distributed as χ^2 with degrees of freedom equal to the total number of parameters in the model minus the number of alternatives. This difference is equal to the number of restrictions imposed. A pseudo R^2 can be defined as

(3.5) $$\text{pseudo } R^2 = 1 - \exp(-\lambda/T) ,$$

and a proportional pseudo R^2 as

(3.6) $$\text{prop pseudo } R^2 = \frac{1 - \exp(-\lambda/T)}{1 - \exp\{2(L_{constr} - L_{max})/T\}} ,$$

whereby L_{max} is the logarithm of the maximum possible probability. The proportional pseudo R^2 is then the ratio of variation accounted for over the total explainable variation. It is identical to the conventional R^2 in multiple regression.[9]

3. Description of the Sample

The sample used for the logit models in this Chapter is again drawn from the original Survey Research Center sample described in Appendix A. Observations were obtained by restricting the sample to (i) households with both husband and wife present during each of the five sample years; (ii) households where husband and wife are the same individuals during the sample period. This restriction allows me to use variables observed in only one year for all five years, e.g., education is only observed in the first and last sample year (1967); (iii) households for which the husband is married only once. This restriction allows me to use the variable "age of husband at first marriage" to determine the length of the present marriage relationship; (iv) households consisting of immediate family members only, i.e., nuclear families with husband, wife and children. I avoid thereby possible influences on the labour force participation decisions of husband and wife created by the presence of other working adults in the family. The influence of working children on the labour force participation decisions of their parents will be discussed; (v) households for which all variables to be used in the empirical investigation are present.

If the restrictions are introduced in the above mentioned order they cause respectively 25, 31, 7, 7, and 9 percent to be discarded from the original sample of 5062 families leaving 21 percent or 1083 households. At an early stage of the investigation I decided to estimate labour force participation functions for the pooled sample only. As each of the 1083 families is observed five times the pooled sample size has 5415 sample points. The decision to use only the pooled sample is motivated primarily by the fact that the alternative "wife only" occurs only a few times in each individual year (Cfr. Table XIII). As a consequence, the estimated parameters for this alternative in the individual years might depend too much on the particular characteristics of the few families involved.

TABLE XIII

DISTRIBUTION OF LABOUR FORCE PARTICIPATION CHOICES IN THE SAMPLE FOR INDIVIDUAL YEARS (IN ABSOLUTE NUMBERS)

Year	Both Working	Husband Only Working	Wife Only Working	None
1967	453	564	11	55
1968	481	528	15	59
1969	514	492	15	62
1970	485	502	19	77
1971	492	487	21	83
POOL	2425	2573	81	336

As can be seen from Table XIII there are some substantial changes in the distribution of the families labour force participation choices from year to year. Most of the variation takes place between the alternatives "both working" and "husband only working", which is of course, the change in participation of the wife. The alternatives "wife working" and "none working" grow in importance over the sample period, which can be explained by the importance of age as a determining factor for their occurrence.

I have also counted how many families choose the same alternative all five years. The results show that respectively 284, 308, 7 and 43 families (out of 1083) remain in respectively the "both working", "husband only", "wife only" and "none working" alternative all five years. This amounts to respectively 58 percent, 59 percent, 43 percent, and 63 percent of the pooled sample points listed on the bottom line of Table XIV.

It should be noted that the "none working" alternative is not the same as retirement. Although the status of the husband in the families choosing the "none working" alternative was almost always "retired" (except for a few husbands being unemployed or in school), there were a substantial number of "retired" husbands in the labour force too.[10] Out of the 5415 observations there were 539 husbands whose status was "retired"; 139 of which were in the labour force. My impression is that although for most of the husbands retirement seems to be a permanent decision, it is not impossible for them to join the labour force again.

The scanty evidence presented in the previous paragraphs indicates that there is presumably sufficient variation in the labour force participation choices of the individual families over the sample period to give some confidence to the results of the empirical investigation

TABLE XIV

DESCRIPTION OF THE POOLED SAMPLE IN TERMS OF

LABOUR FORCE PARTICIPATION CHOICES

Variables[1]	Both Working	Husband Only	Wife Only	None	All
# OBS	2425	2573	81	336	5415
Mean Of					
YRS SCHOOL H	11.1	10.7	8.0	8.9	10.8
YRS SCHOOL W	11.4	10.9	9.3	8.8	10.9
MARRIAGE	17.3	18.8	32.1	41.9	19.7
WAGE H (PRED)[2]	3.95	3.97	2.75	2.63	3.86
WAGE W (PRED)[2]	2.26	2.16	1.92	1.74	2.18
ASSET Y	1093.	1401.	1075.	1949.	1292.
Number of Observations With Dummy Variable = 1					
LIMIT H	231 (9%)	268 (10%)	64 (79%)	170 (51%)	733 (14%)
LIMIT W	21 (.9%)	56 (2%)	3 (4%)	24 (7%)	104 (2%)
KID 6	921 (38%)	1256 (49%)	15 (19%)	18 (5%)	2210 (41%)
SOUTH	957 (39%)	883 (34%)	35 (43%)	103 (31%)	1978 (37%)
URBAN	1506 (62%)	1559 (61%)	30 (37%)	201 (60%)	3296 (61%)
RACE	513 (21%)	451 (18%)	16 (20%)	30 (9%)	1010 (19%)
MORTG	1392 (57%)	1382 (54%)	16 (20%)	50 (15%)	2840 (52%)
RESERVE	1286 (53%)	1366 (53%)	36 (44%)	240 (71%)	2928 (54%)

1 Variables are explained in Appendix B

2 Using the regression predictor

explaining these choices as functions of certain exogenous variables.

Table XIV describes the variables for each labour force participation alternative. A few patterns are apparent: families in the alternative "none working" are on an average older (in terms of the age of marriage variable) and possess more wealth (see ASSET Y, RESERVE). As was seen above, most of these families are in the retirement phase of their life cycle. The families in the alternative "wife only" are mostly found in this alternative because of a handicapped husband (in almost 80 percent) of the cases. The families in the alternatives "both working" and "husband only" are generally better educated, younger and potentially higher wage earners than the other cases. It appears that a handicap for the wife and the presence of pre-school aged children is the crucial distinction between these two alternatives.

4. The Simultaneous Family Labour Force Participation Model

The estimation results for the multinomial logit model for the family's labour force participation decisions are presented in Table XV. The estimates are obtained by normalizing on the alternative "none working" and are therefore interpreted as the first derivatives of the logarithms of the odds of that alternative over the alternative "none working" (see equation (1.28), Chapter II). Furthermore differences in the coefficients between the columns for each variable are the first derivative of the logarithm of the odds of one alternative over the other. These column differences are presented in Table XVI. Asymptotic "standard errors" for these differences are calculated based on the asymptotic variance covariance matrix V.

The existence of significant differences between the various alternatives would seem to be an indication that the alternatives should not be aggregated and thus that the axiom of irrelevant alternatives (see Chapter I, Section 3.3.D) is valid for this model.

Education for husband and wife has a different effect on the labour force participation choice. More education seems to increase the odds of retirement for husbands while the reverse is true for wives (Table XV). In choosing between the "both working" and "husband only" alternative, the former seems to be more preferred by higher educated couples (Table XVI).

If the husband is handicapped then the odds increase substantially in favour of the "none working" alternative (Table XV) or in favour of the "wife only" choice (Table XVI), with the latter choice being more probable than the "none working" choice (Table XV). The LIMIT H variable seems therefore to be crucial in determining the labour force participation of the husband. (This is confirmed in the sequential model below.) If the wife is handicapped then the probability of the "both working" choice becomes smaller compared with the probability of either "none working" (Table XV) or "husband only" (Table XVI). It does not have any significant influence on the odds of the "wife only" choice (both tables).

TABLE XV

SIMULTANEOUS FAMILY LABOUR FORCE PARTICIPATION MODEL.

RESULTS OF MULTINOMIAL LOGIT MODEL.[1]

Variable[2,3]	Both Working (β_A)	Husband Only (β_B)	Wife Only (β_C)
CONSTANT	3.40 (.64)	3.73 (.63)	-.52 (n.s.) (1.06)
YRS SCHOOL H	-.12 (.03)	-.15 (.03)	-.09 (.05)
YRS SCHOOL W	.15 (.03)	.08 (.03)	.03 (n.s.) (.05)
LIMIT H	-1.50 (.17)	-1.60 (.16)	1.38 (.30)
LIMIT W	-1.13 (.39)	-.26 (.34)	-.83 (n.s.) (.65)
MARRIAGE	-.13 (.01)	-.10 (.01)	-.06 (.02)
KID 6	-.42 (**) (.30)	.55 (*) (.30)	.23 (n.s.) (.46)
SOUTH	.78 (.18)	.61 (.18)	.16 (n.s.) (.30)
URBAN	-.78 (.18)	-.83 (.18)	-1.17 (.30)
RACE	-.02 (n.s.) (.27)	-.42 (**) (.28)	.02 (n.s.) (.43)
WAGE H (PRED)	.83 (.16)	1.10 (.16)	-.12 (.25)
WAGE W (PRED)	.001 (n.s.) (.22)	-.41 (*) (.22)	1.22 (.29)
ASSET Y	-.00009 (.00003)	-.00002 (.00002)	-.0001 (**) (.00008)
MORTG	.77 (.19)	.52 (.19)	-.25 (.33)
RESERVE	-.25 (n.s.) (.20)	-.31 (**) (.19)	-.79 (.32)

TABLE XV CONTINUED

Variable[2]	Both Working (β_A)	Husband Only (β_B)	Wife Only (β_C)
NUMBER PARTICIPATING	2425	2573	81
NUMBER OF OBSERVATIONS	5415		

Log likelihood at maximum: - 4179.0

Likelihood ratio test (χ^2 with 42 d.f.) 1916.13

Pseudo R^2 = .30

Proportional pseudo R^2 = .35

1 Each coefficient is significant at the 5% level (t-test) except when followed by

(*) significant at 10% level

(**) significant at 20% level

(n.s.) not significant

2 Variables are defined in Appendix B

3 Numbers in parenthesis are asymptotic standard errors

TABLE XVI[1]

DIFFERENCES BETWEEN COEFFICIENTS IN SIMULTANEOUS FAMILY LABOUR FORCE PARTICIPATION MODEL

Variable	Both Working vs Husband Only ($\beta_A - \beta_B$)	Both Working vs Wife Only ($\beta_A - \beta_C$)	Husband Only vs Wife Only ($\beta_B - \beta_C$)
CONSTANT	-.3385 (**) (.2093)	3.9148 (.9463)	4.2533 (.9456)
YRS SCHOOL H	.0314 (.0111)	-.0256 (n.s.) (.0418)	-.0496 (n.s.) (.0415)
YRS SCHOOL W	.0566 (.0141)	.1113 (.0487)	.0558 (n.s.) (.0485)
LIMIT H	.1040 (n.s.) (.1010)	-2.8829 (.2686)	-2.9869 (.2673)
LIMIT W	-.8758 (.2695)	-.3086 (n.s.) (.6622)	.5672 (n.s.) (.6376)
MARRIAGE	-.0270 (.0037)	-.0739 (.0146)	-.0469 (.0145)
KID 6	-.9692 (.0746)	-.6550 (*) (.3767)	.3143 (n.s.) (.3760)
SOUTH	.1685 (.0692)	.6143 (.289)	.4458 (**) (.2859)
URBAN	.0538 (n.s.) (.0689)	.3905 (**) (.2826)	.3367 (n.s.) (.2819)
RACE	.3972 (.0887)	-.0414 (n.s.) (.3815)	-.4386 (n.s.) (.3807)
WAGE H (PRED)	-.2612 (.0389)	.9524 (.2314)	1.2136 (.2316)
WAGE W (PRED)	.4113 (.0626)	-1.221 (.2427)	-1.6324 (.2447)
ASSET Y	-.000073 (.000016)	.000015 (n.s.) (.000075)	.000088 (n.s.) (.000075)
MORTG	.2446 (.063)	1.0211 (.2987)	.7766 (.2986)
RESERVE	.0631 (n.s.) (.0674)	.5414 (*) (.2971)	.4784 (**) (.2963)

1 Same comments as Table XV

As is to be expected the odds of the "none working" choice over all other choices increases as the couple grows older (Table XV). This is clearly the effect of retirement. Also the odds of the "wife only" alternative over the two alternatives where the husband works, increases with years married (Table XVI). Presumably the husband, being on an average older than his wife, retires before her. Another reason might be that the possible handicap for the husband is age-related. "Both working" becomes less probable than "husband only" in later stages of the life cycle. I suspect this is a mixture of two reinforcing effects: a "life cycle effect", i.e., "both working is more probable for younger couples in general; and an "age cohort effect", i.e., the younger generation tends to go out working together more than the older generation. Given the limited timespan of the sample it is impossible to disentangle both effects.

The fact that the couple has a pre-school aged child is an important factor especially in the choice between "both working" and "husband only" (Table XVI). It is thus a determinant in the labour force participation decision of the wife. (This will be confirmed in the sequential model below.)

The odds of "none working" versus the other alternatives seem to increase for an urban environment and to decrease for a family living in the south (Table XV). Living in the south also increases the choice "both working" as compared with the alternatives where either only the husband or the wife works (Table XVI).

If the couple is non-white then the probability of finding them both in the labour force instead of only the husband will be higher than if the couple were white (Table XVI). Again "RACE" is mostly a determinant of the wife's labour force participation decision (as will also be seen in the sequential model).

I now turn to the economic variables. The (permanent) wage rate of the husband increases the odds of those alternatives where he is found working, i.e., "both working" and "husband only", compared with the alternatives where he is not in the labour force, i.e., "wife only" and "none working" (see both Tables). Because the income effect is very small or non-significant, this seems to conform to the theoretical expectation of a positive relation between wage changes and labour force participation, keeping utility constant (see Chapter I, Section 3.2). The same positive relation between the (permanent) wage rate and labour force participation choice usually holds for the wife as well. The wife's (permanent)wage increases the odds of choosing "wife only" over "none working" (Table XV), of choosing "both working" over "husband only" and of choosing "wife only" over "husband only" (Table XVI).

It is also interesting to observe the effect of a change in one partner's wage rate on the other partner's labour force participation choice. The first column of Table XVI shows the effect of the husband's wage rate on his wife's labour force participation decision and the second column shows the reverse. In both cases there is a negative effect indicating some sort of substitution effect, i.e., the higher the wage of one partner the lower the probability of finding the other partner in the labour force.

The level of asset income increases the odds of having nobody working compared with having both working (Table XV). A possible explanation of this result could be that couples where both are working are usually young couples (see above for the effect of years married) who are just beginning in terms of non-human capital accumulation. The couples where nobody is working are generally older (again see above for the effect of years married), are usually retired and presumably possess a stock of non-human capital which they have accumulated over their lifetime.

The influence of asset income on the choice between "both working" and "husband only" (see Table XVI) can be explained in the same life cycle framework. As the couples where only the husband works are usually older than the "both working" couples (see year married effect again), these couples presumably have accumulated some assets which eventually "enable" the wife to leave the labour force.

This life cycle explanation relating the assumption of accumulation of non-human capital as the life cycle goes on with the shift from "both working" to "husband only" to "none working" as the couple gets older is also explored in Table XVII. In this table, I present the average asset income as well as the frequency of the four possible labour force choices for consecutive age of marriage intervals.

TABLE XVII

RELATIONSHIP BETWEEN MARITAL AGE, LABOUR FORCE PARTICIPATION CHOICE, AND ASSET INCOME IN THE SAMPLE

Age of Marriage Interval	Average ASSET Y $	Standard Deviation	Percentage Both Working	Husband Only	Wife Only	None	Total =100%
(0 - 10][1]	617	1232	51	48	1	0	1457
(10 - 20]	1123	2176	48	51	1/2	1/2	1665
(20 - 30]	1757	3519	46	48	1	5	1289
(30 - 40]	1926	2891	42	47	3	8	611
(40 - 50]	1901	2332	12	31	10	47	311
(50 - ∞]	2398	2768	0	17	0	83	82
Total			2425	2573	81	336	5415

1. "(" = open interval, "]" = closed interval.

Table XVII confirms that for these sample average asset income is higher for older than for younger couples (but note the large standard deviation relative to the mean in each age group) and that "both working" occurs most for couples married less than 10 years and "none working" for couples who were married more than 50 years. All other age groups choose "husband only" more frequently. Again it should be stressed that Table XVII reflects "life cycle" effects as well as "age cohort" effects.

If the family has a mortgage debt then the odds of the "both working" alternative increase compared with all other alternatives (Table XV and XVI). Furthermore, it also increases the odds of "husband only" as compared with "wife only" and "none working" (Tables XV and XVI). Thus mortgage debt encourages labour force participation for husband and wife. The influence on the labour force participation of the husband seems to be more pronounced (this will be confirmed in the sequential model). Note, however, that one should not exclude the possibility of a causal effect running in the other direction, i.e., families have a mortgage debt because they are both working.

Having some savings (RESERVE) seems to be most important in the choice between "wife only" versus "none working". If the family has savings the probability increases that the wife will not join the labour force in this case (Table XV).

I have also tested the influence of a dummy variable indicating whether the children in the household had any income (mostly labour income) and have found that it has no significant influence on the labour force participation choice of their parents. This independence between the labour force participation decisions of the parents and children was also established by Bowen and Finegan [1969]. Dummy variables indicating the unemployment level in the county where the family lives also were found to have no significant effect.[11] Dummy variables indicating religious preference (Jewish, Catholic) were not significant.

As can be seen from Table XV most coefficients are not significant (in terms of asymptotic t-ratios). This implies only that the odds of the particular alternative over the alternative "none working" remain unchanged when the variable in question changes. It does not imply that the odds of that particular alternative over another alternative will remain unchanged. For instance, YRS SCHOOL W has no influence on the odds of "wife only" over "none working" (Table XV) but has an effect on the odds of "both working" over "wife only" (Table XVI). If one is only interested in the odds of all the alternatives over "none working" then it is possible to find a more parsimonious specification for that particular multinomial logit model. To do this I re-estimate the logit model as in Table XV but restrict the coefficients of the non-significant variables to zero.[12] This leads to the result presented in Table XVIII.

The restricted model is very similar to the unrestricted version in terms of coefficient values. It is worthwhile testing the null hypothesis that the restrictions are valid. This can be done using the likelihood ratio test which is asymptotically χ^2 distributed with degrees of freedom equal to the number of restrictions. The calculated statistic is $\lambda = -2(-4183 + 4179) = 8$, with degrees of freedom equal to $45 - 31 = 14$. The critical value at 5 percent of χ^2 (14) is 23.7, suggesting that one cannot reject the null hypothesis that the restrictions are valid. Therefore the more parsimonious model of Table XVIII is equal in informational content to Table XV in terms of comparing all alternatives with the "none working" alternative.

TABLE XVIII

SIMULTANEOUS FAMILY LABOUR FORCE PARTICIPATION MODEL.

RESULTS OF RESTRICTED MULTINOMIAL LOGIT

MODEL[1]

Variable[2]	Both Working (β_A)	Husband Only (β_B)	Wife Only (β_C)
CONSTANT	3.54 (.53)	3.89 (.53)	-.10 (.61)
YRS SCHOOL H	-.12 (.03)	-.15 (.02)	-.11 (.04)
YRS SCHOOL W	.13 (.03)	.08 (.03)	
LIMIT H	-1.46 (.16)	-1.56 (.16)	1.40 (.31)
LIMIT W	-.88 (.26)		
MARRIAGE	-.14 (.009)	-.11 (.008)	-.06 (.01)
KID 6	-.50 (.25)	.48 (*) (.25)	
SOUTH	.73 (.16)	.56 (.16)	
URBAN	-.77 (.17)	-.82 (.17)	-1.19 (.28)
RACE		-.39 (.07)	
WAGE H (PRED)	.82 (.11)	1.08 (.11)	
WAGE W (PRED)		-.42 (.06)	1.08 (.17)
ASSET Y	-.00007 (.00002)		
MORTG	.87 (.17)	.63 (.16)	
RESERVE			-.64 (.28)

TABLE XVIII CONTINUED

Variable	Both Working (β_A)	Husband Only (β_B)	Wife Only (β_C)
NUMBER PARTICIPATING	2425	2573	81

NUMBER OF OBSERVATIONS 5415

Log likelihood at maximum -4183.0

Likelihood ratio test (χ^2 with 28 d.f.) 1908.2

Pseudo R^2 = .30

Proportional Pseudo R^2 = .35

1 Each coefficient is significant at the 5 percent level (t-test) except when followed by (*) which means significant at the 10 percent level only.

2 Variables are defined in Appendix B

5. The Sequential Family Labour Force Participation Model(s)

The basic assumption of the family labour force participation model is that one partner decides first about his labour force participation and the other partner chooses conditionally upon the labour force participation choice of the first partner. The sequential model in which the husband first decides his participation and his wife decides conditional upon his choice would seem to be, *a priori*, an acceptable description of family labour force participation behaviour. In this sequential model I first use a binomial logit model to estimate the coefficients of the labour force participation choice for all the husbands in the sample. In a second stage I estimate two binomial logit models for the labour force participation choice of the wives: one for the sample of wives where the husband is in the labour force and another for the sample of wives where the husband is not participating.

If one combines the probabilities of labour force participation choice of husband and wife one evidently arrives again at the four alternatives discussed in the simultaneous model. For instance the probability of the husband participating times the probability of the wife participating gives the probability of "both working" in the sequential model (see also Figure 1, Chapter II).

An important objective of this Chapter is to compare the simultaneous model and the sequential model as alternative models to explain family labour force participation behaviour. To be consistent I will therefore use the same independent variables in both models.

As was discussed in the Appendix to Chapter II the differences between the simultaneous model and the sequential model hinge on the difference between the coefficients of the two binomial logit models for the wives. (Equation (A.6) of the Appendix to Chapter II.) It will therefore be important to compare estimates of these coefficients.

Besides the sequential model in which the husband decides first it is also possible to assume a sequential model in which the wife decides first. Although the model would seem to be *a priori* a less realistic description of family choice behaviour, I have estimated this model in order to be complete.

The results on the sequential model in which the husband decides first are given in Table XIX and the results on the other sequential model are in Table XX.

I will first discuss Table XX: the coefficients of the logit model for the wife's labour force participation choice estimated over the whole sample (Table XX, Column 1) are almost identical to the coefficients for the logit model estimated over the sample of wives with working husbands only (Table XIX, Column 2). The only exception is the coefficient of YRS SCHOOL H which is not significant in Table XX but is significant in Table XIX. Because of the similarity between these two columns and to avoid repetition I will only discuss the second column of Table XIX.

TABLE XIX

SEQUENTIAL FAMILY LABOUR FORCE PARTICIPATION MODEL (HUSBAND DECIDING FIRST). RESULTS OF BINOMIAL LOGIT MODELS[1]

Variable	Husband's Labour Force Participation Choice	Wife's Labour Force Participation Choice. Husband Working	Wife's Labour Force Participation Choice. Husband Not Working
CONSTANT	3.65 (.56)	-.32 (**) (.21)	-1.73 (**) (1.10)
YRS SCHOOL H	-.11 (.02)	.03 (.01)	-.09 (.05)
YRS SCHOOL W	.10 (.03)	.06 (.01)	.11 (*) (.06)
LIMIT H	-1.86 (.14)	.13 (.10)	1.22 (.36)
LIMIT W	-.38 (n.s.) (.32)	-.83 (.26)	-1.95 (.74)
MARRIAGE	-.10 (.009)	-.03 (.004)	-.04 (.02)
KID 6	.0009 (n.s.) (.25)	-.97 (.07)	.72 (n.s.) (.59)
SOUTH	.72 (.17)	.17 (.07)	.55 (**) (.34)
URBAN	-.56 (.16)	.06 (n.s.) (.07)	-1.31 (.37)
RACE	-.20 (n.s.) (.24)	.38 (.08)	-.44 (n.s.) (.37)
WAGE H (PRED)	1.08 (.14)	-.26 (.04)	.02 (n.s.) (.31)
WAGE W (PRED)	-.59 (.18)	.40 (.06)	1.00 (.40)
ASSET Y	-.00003 (**) (.00002)	-.00007 (.00002)	-.00016 (**) (.00010)
MORTG	.71 (.17)	.24 (.06)	-.25 (n.s.) (.39)
RESERVE	-.07 (n.s.) (.17)	.05 (n.s.) (.06)	-.60 (**) (.39)

TABLE XIX CONTINUED

Variable	Husband's Labour Force Participation Choice	Wife's Labour Force Participation Choice. Husband Working	Wife's Labour Force Participation Choice. Husband Not Working
NUMBER PARTICIPATION	4998	2425	81
NUMBER OF OBSERVATIONS	5415	4998	417
Log likelihood at maximum	-746.55	-3284.2	-154.0
Likelihood ratio (χ^2 with 14 d.f.)	1446	356	103
Pseudo R^2	.23	.07	.22
Proportional Pseudo R^2	.56	.09	.35

1 Same comments as Table XV

TABLE XX

Variable	Wife's Labour Force Participation Choice	Husband's Labour Force Participation Choice, Wife Working	Husband's Labour Force Participation Choice. Wife Not Working
CONSTANT	-.21 (n.s.) (.20)	3.97 (1.21)	3.74 (.67)
YRS SCHOOL H	.01 (n.s.) (.01)	-.06 (n.s.) (.05)	-.14 (.03)
YRS SCHOOL W	.07 (.01)	.18 (.06)	.05 (n.s.) (.03)
LIMIT H	.10 (n.s.) (.09)	-2.97 (.33)	-1.61 (.17)
LIMIT W	-.89 (.24)	-1.24 (*) (.72)	-.12 (n.s.) (.35)
MARRIAGE	-.04 (.003)	-.09 (.02)	-.10 (.01)
KID 6	-1.00 (.07)	-.65 (**) (.45)	.60 (*) (.32)
SOUTH	.21 (.07)	.82 (.36)	.69 (.20)
URBAN	-.05 (n.s.) (.07)	.16 (n.s.) (.34)	-.74 (.19)
RACE	.43 (.09)	-.003 (n.s.) (.004)	-.54 (*) (.31)
WAGE H (PRED)	-.20 (.04)	1.23 (.29)	.91 (.18)
WAGE W (PRED)	.41 (.06)	-1.52 (.27)	.05 (n.s.) (.27)
ASSET Y	-.00007 (.00002)	.00005 (n.s.) (.0001)	-.00003 (n.s.) (.00002)
MORTG	.27 (.06)	1.12 (.34)	.42 (.21)
RESERVE	.01 (n.s.) (.07)	.32 (n.s.) (.35)	-.32 (**) (.22)

TABLE XX CONTINUED

Variable	Wife's Labour Force Participation Choice	Husband's Labour Force Participation Choice.Wife Working	Husband's Labour Force Participation Choice . Wife Not Working
NUMBER PARTICIPATION	2506	81	2573
NUMBER OF OBSERVATIONS	5415	2506	2909
Log Likelihood at maximum	-3495.6	-187.23	-517.87
Likelihood ratio test (χ^2 with 14 d.f.)	485.5	540.88	1046.34
Pseudo R^2	.09	.19	.30
PROportional Pseudo R^2	.11	.78	.59

1 Same comments as Table XV

The second and third columns of Table XX present the results on the two "conditional" binomial logit models for husbands with respectively working and non-working wives. The estimation results for husbands with working wives (Table XX, Column 2) are however, dubious. When I use the PROLO computer program to estimate this model the logarithm of the likelihood becomes smaller at each iteration and eventually becomes smaller than the smallest possible number in FORTRAN programs (after three iterations). This indicates that the likelihood would maybe approach $-\infty$ if further iteration would be possible. This would certainly happen in a sample where all decision units choose one particular alternative and the other alternative was not chosen at all. I would suggest that the fact that in the sample in question 2425 out of 2506 husbands (i.e., 97%) choose to participate comes dangerously close to the situation where all would participate. The CSP program, however, does converge and these results are given in the second column of Table XX. However, because of this crucial difference in numerical results[13] I do not have much confidence in the values of these coefficients. These estimation problems also cast some doubt on the appropriateness of the sequential model in which the wife decides first. Because of this I will concentrate my further discussion on the sequential model shown in Table XIX.

In discussing Table XIX I first compare the coefficients of the husband's labour force participation choice with those of the wife's labour force participation choice, i.e., compare the coefficients in Columns 1 and 2 in Table XIX (remembering that the latter are similar to the coefficients in the first column of Table XX). Next I will compare the coefficients of the logit model for the wives whose husbands are working with the coefficients for the wives whose husbands are not working, i.e., compare Columns 2 and 3 in Table XIX.

5.1 The Husband's Labour Force Participation Versus The Wife's Labour Force Participation Decision

The probability of participation decreases for more educated men but it increases for more educated women. Bowen and Finegan [1969] and Cohen et al. [1970] found the same result for women but they established a positive education effect for men. However, they restricted their sample to prime-age males whereas my samples have no such age constraint The different result is therefore presumably due to the dominating effect of earlier retirement for more educated men in my sample. The educational level of the husband has no effect on the wife's labour force participation (see Table XX, Column 1) whereas the wife's education influences the husband's labour force participation positively.

A handicap reduces the probability of labour force participation to a considerable extent for the husband. The same is true for the wife but to a much smaller degree. Surprisingly enough, there is no effect of a handicapped husband on the wife's labour force participation and vice versa.

The negative effect of years married is also more considerable for the husband than for the wife. The presence of pre-school aged children reduces the probability of participation for married women considerably. This result has been much documented in numerous other studies (Cain [1966], Bowen and Finegan [1969], Cohen, et al. [1970]). It is again confirmed here together with the finding that the presence of pre-school aged children does not affect the husband's labour force participation.

The positive effect of a non-white dummy variable on the probability of participation for the wives has also been frequently established (Bowen and Finegan [1969], Cohen, et al.[1970], and especially Cain [1966]). This is also confirmed here, together with the result that it is not a relevant variable for the husbands. This was also the case in Cohen, et al. [1970], however, Bowen and Finegan [1969] found a significant racial difference in the "prime age males" group.

Ever since Mincer's [1962] contribution the participation studies have been interested in wage and income effects. Income effects are quite well researched but most of these studies, however, do not include a wage variable (Mincer [1962], Bowen and Finegan [1969], Cohen, et al. [1970], Cain [1966]). Recently Boskin [1973] estimated linear probability models for the labour force participation of men and women including among the independent variables a predicted wage variable similar to the one used in this thesis. In his study these wage variables were not significant except in one case: the negative effect of the husband's predicted wage on his wife's labour force participation equation. Therefore, the significance of the wage variables for the labour force participation model for both husbands and wives appears to be a new result in the study of labour force participation behaviour.

The own wage is again positively related to the probability of participation as was found already in the simultaneous model. Given the very small income effect, this positive wage effect corroborates the theoretically expected effect (Chapter II, Section 3.2). As in the simultaneous model the effect of the partner's wage is negative for both husbands and wives.

The level of asset income decreases the probability of participation for the wives, as was already found in previous studies (Cain [1966], Mahoney [1961], Bowen and Finegan [1969], Cohen, et al. [1970]). For "prime age males" both Bowen and Finegan [1969] and Cohen, et al. [1970] found a significant negative income effect, which is not confirmed in this study.

A mortgage dummy variable increases the probability of participation of husband and wife, but more so for the husband. This result was already suggested in the simultaneous model.

Most previous labour force participation studies have concentrated on the participation decisions of married women. This is understandable considering the observed substantial growth in their post-war participation rates. Because of its relative importance in the field of labour force participation studies I have tried to investigate the wife's labour force participation somewhat further. I have done this mostly by adding or changing variables to the basic specification used in Table XIX, Column 2. These additional results are thus for married women with a working husband.

First of all a dummy variable for the occurrence of a birth in the sample year has, as expected, a negative effect on the probability of participation.[14] I have investigated the effect of children further by splitting the KID 6 variable into a set of three dummy variables indicating whether the family has a child of less than or equal to two years; between two and four years; between four and six years. I find that the probability of participating decreases the most for the two to four age group and that this negative influence is substantially smaller for the four to six age group.[15] In another specification I replace the KID 6 variable with a variable counting the number of children in the household. This variable also has a significant negative coefficient. In the same equation a variable counting the number of children in high-school entered positively but was only significant at the 20 percent level.[16]

The MARRIAGE variable indicates decreasing probabilities of participation for the wife as the couple gets older. To explore this age-labour force participation relation somewhat further, I replace MARRIAGE with a set of dummy variables for the (30 - 40], (40 - 50], and (50 - 100) age cohorts of women. The results indicate that the logarithm of the odds becomes increasingly smaller for older age groups. A similar result was found by Cohen, et al.[1970]. A quadratic specification of the wife's age variable exhibits a peak in the probability of labour force participation at age 24[17].

I also have found that a series of dummy variables indicating intervals of the unemployment rate in the county where the family is living are insignificant[18] for the participation decisions of the wives.

5.2 Labour Force Participation for Wives with Husbands Working versus Labour Force Participation for Wives with Husbands Not Working.

As was proved above (see Appendix to Chapter II) a sufficient condition for the probabilities predicted by the sequential model to be equal to the probabilities predicted by the simultaneous model is that the vector of coefficients of the logit model for wives with working husbands (Table XIX, Column 2) is equal to the coefficient vector for the wives with non-working husbands (Table XIX, Column 3). Comparing column 2 and 3 in Table XIX, it is quickly seen that the coefficient vectors differ in many ways. There is a noticeable difference in the value of almost all the coefficients.

The condition that the two coefficient vectors should be equal in order for the probabilities of the simultaneous and sequential model (husband deciding first) to be the same, can also be tested statistically. To do this I use the likelihood ratio procedure to test the null hypothesis that the coefficient vectors are the same against the alternative hypothesis that they are different. The logarithm of the maximum likelihood under the null hypothesis is given in Table XX (column 1). The logarithm of the maximum likelihood under the alternative hypothesis is the sum of the logarithm of the maximum likelihoods given in respectively columns 2 and 3 of Table XIX. The computed value of the likelihood ratio test is 114.8. The critical $\chi^2(15)$ value is 32.8 at the .5% critical level. The null hypothesis is therefore decisively rejected.

The same procedure can be used to test the equality of the probabilities of the simultaneous model and the sequential model where the wife decides first. In this case the calculated test statistic is 82.9. Again the null hypothesis of equality of the two models must be rejected.

6. Choosing Between the Simultaneous and Sequential Models

Since it is difficult to choose between the simultaneous and sequential model on a priori grounds, it may be of interest to discriminate between them a posteriori. A Bayesian procedure can be helpful in this respect. Let h_1 denote the simultaneous model and h_2 the sequential model. Let β_1 and β_2 be the parameter vectors corresponding to each model. Also let $p(h_1)$ and $p(h_2)$ be the prior probabilities that either the simultaneous or the sequential model holds. $p(\beta_1|h_1)$ and $p(\beta_2|h_2)$ are the prior distributions for the parameters of each model. The joint posterior distribution of the models and their parameters, given the data Z, is

(3.7) $$p(\beta_i,h_i) \propto L(Z|\beta_i,h_i)\; p(\beta_i|h_i)\; p(h_i) \;, \text{ for } i = 1,2$$

where L is the likelihood function.

If one uses prior distributions of the form

(3.8) $$p(\beta_i|h_i)\; p(h_i) \propto C_i \;, \text{ for } i = 1,2$$

where C_i is a constant, then the (approximate) posterior probability for either model can be derived as[20]

$$(3.9) \qquad p(h_i|Z) \doteq (2\pi)^{K_i/2} \, |V_i|^{-1/2} \, \exp \{g(\beta_i^{\circ}|h_i, Z)\} \, C_i \; , \quad i = 1, 2$$

where K_i is the number of parameters in β_i, V_i is the asymptotic variance covariance matrix and g is the logarithm of the likelihood function. Both V_i and g are evaluated at the maximum likelihood estimates of β_i, i.e., β_i°.

Note that if instead of (3.8) the following prior distribution would be used

$$(3.10) \qquad p(\beta_i|h_i) \, p(h_i) \propto (2\pi)^{-K_i/2} \, |V_i|^{1/2} \; , \quad i = 1, 2$$

then (3.9) would reduce to

$$(3.11) \qquad p(h_i|Z) \doteq \exp \{g(\beta_i^{\circ}|h_i, Z)\} \; .$$

Discrimination would then be done on basis of the maximum values of the logarithms of the likelihood function. However, (3.10) is unacceptable as a respresentation of prior knowledge since it embodies knowledge of the likelihood function at its maximum.

Before I apply equation (3.9) to discriminate between the simultaneous and sequential models, I must specify the form of the likelihood function in the sequential model. (The likelihood function for the simultaneous model is defined in equation (3.1) above.) Using the same notation as in the Appendix to Chapter II the probabilities of a family choosing respectively, "both working", "husband only", "wife only", and "none working" are written in the sequential model as $p(11|1)\ p(1)$, $p(12|1)\ p(1)$, $p(21|2)\ p(2)$, $p(22|2)\ p(2)$ (see A.3.1 to A.3.4 in the Appendix to Chapter II). The likelihood function for the sequential model using this notation is then

$$(3.12) \qquad e^L = \prod_{t=1}^{f_1} p_t(11|1)\ p_t(1) \prod_{t=1}^{f_2} p_t(12|1)\ p(1) \prod_{t=1}^{f_3} p_t(21|2)\ p_t(2)$$
$$\prod_{t=1}^{f_4} p_t(22|2)\ p_t(2)$$

where the subscript t indicates the t^{th} family and f_1, f_2, f_3, and f_4 are the number of families in the sample choosing respectively the "both working", "husband only", "wife only", and "none working" alternative ($\sum_{i=1}^{4} f_i = T$). Upon rearranging (3.12) becomes

$$(3.13) \qquad e^L = \prod_{t=1}^{f_1+f_2} p_t(1) \prod_{t=1}^{f_3+f_4} p_t(2) \prod_{t=1}^{f_1} p_t(11|1) \prod_{t=1}^{f_2} p_t(12|1)$$
$$\prod_{t=1}^{f_3} p_t(21|2) \prod_{t=1}^{f_4} p_t(22|2)$$

i.e., the product of the likelihood functions of the binomial models for the husband's labour force participation decision, and for the labour force participation choice of the wives whose husband is in or out the labour force. Therefore by extension the (approximate) posterior probability of the sequential model will be equal to the product of the (approximate) posterior probabilities of the same binomial logit models.

The logarithm of the (approximate) posterior probability for the simultaneous model is equal to a constant plus -4146.5. The log of the (approximate) posterior probability for the sequential model in which the husband decides first, is equal to a constant -4166.75. Therefore a posteriori the simultaneous model is more probable than the sequential model. As was mentioned above the estimation of the simultaneous model in which the wife decides first, did not converge when I use the PROLO program. This would indicate that the logarithm of the (approximate) posterior probability would maybe approach $-\infty$. In this case this model would be highly improbable. If I use the results obtained from estimating this model with CSP then I can develop a discrimination procedure using the maximum values of the logarithms of the likelihood function (equation (3.11)). This implies, however, that an unacceptable prior (3.10) is used. Unfortunately CSP does not provide information on V_i so that I cannot calculate (3.9). The logarithm of the likelihood at the maximum for the three models is respectively, simultaneous model: -4179; sequential model (husband decides first): -4184.75; sequential model (wife decides first): -4200.70. On the basis of this criterion the simultaneous model is most probable and the sequential model in which the wife decides first is least probable a posteriori.

FOOTNOTES TO CHAPTER IV

1 See Buse [1972] for a discussion of the origins of the logit model and for references to applications in economics.

2 Cfr. McFadden [1974, 115], Equation 17. Note, however, that I assume there are no repetitions.

3 Cfr. McFadden [1974, 115], Equation 20.

4 Cfr. McFadden [1974, 111], Axiom 5.

5 Cfr. McFadden [1974, 116-120].

6 Cragg and Uhler [1971] mention a non-existence problem that might arise if the same alternative were chosen in all observations. See their footnote on page 343.

7 CSP (Cross Section Processor) Version 3 of June 1972 is a package of computer programs designed to carry out regression analysis on large bodies of cross-section data. It was written by M.G. Kohn with contributions by N.A. Barr (London School of Economics), Z.I. Brody (Hebrew University and Massachusetts Institute of Technology), R.E. Hall (Massachusetts Institute of Technology), and M.D. Hurd (Stanford). One of the programs in this package allows for binomial and multinomial logit analysis. PROLO and THEIL are part of a set of programs designed to execute multiple probit and logit analyses and extensions to them. These programs were written by J. Cragg [1968].

8 It took 29'49" computer time on an IBM 370-168 computer to estimate the simultaneous model presented in Table XVI.

9 This summary statistic was developed by Cragg and used in Cragg and Uhler [1970, 1971] and Cragg and Baxter [1970].

10 The information on the status of the wife is not available.

11 To test for the effect of the county unemployment rate I had to restrict the sample to the years 1968 to 1971 as the unemployment variable was only available in those years.

12 The THEIL program offers the possibility of restrictions within and across alternatives.

13 This kind of difference between PROLO and CSP and for that matter between THEIL and CSP occurred only in this particular case. Usually numerical differences between the programs were negligible (results were equal up to the third digit).

14 The coefficient and asymptotic standard error for the BIRTH variable was respectively -.38 and .11.

15 The coefficients and asymptotic standard errors are respectively -.53 (.07), -.78 (.10); -.34 (.11) for the (0 - 2], (2 - 4], and (4 - 6] age groups.

16 The coefficient of the number of children variable was -.08 (asymptotic standard error: .02) and for the number of children in highschool variable it was .11 (asymptotic standard error: .07).

17 The coefficients and asymptotic standard errors for the age groups are respectively -.16 (.08); -.35 (.10); -.88 (.12): for the (30 - 40], (40 - 50], and (50 - 100] age brackets. The coefficients and asymptotic standard errors for the AGE and AGESQ variables are respectively .0459 (.021) and -.00095 (.00024). This quadratic form has a maximum at 24.2 years.

18 To test this the sample must be restricted to the years 1968 to 1971 and to those families for whom the county unemployment rate was in fact observed. Sample size is then 3768 observations.

19 Discussion of this Bayesian discrimination technique can be found in Cragg [1971].

20 See Cragg [1971, 834-835]. The (approximate) posterior probability for the logit models is calculated in the PROLO and THEIL programs.

CHAPTER V

FAMILY CONDITIONAL LABOUR SUPPLY FUNCTIONS

1. Introduction

Once a family has decided on its labour force alternative its next step is to decide conditionally on the number of hours. More specifically if the family chooses the "both working" alternative it will have to determine the labour supply of the husband and the wife. Similarly, if it chooses "husband only" or "wife only" it will have to determine the hours to be supplied by respectively the husband and the wife.

Probabilities for the participation decisions were established in the previous Chapter. In this Chapter I will investigate the labour supply functions corresponding to the labour force participation alternative chosen.

2. Issues in Labour Supply Estimation

As discussed in Cahpter II, Section 2.2, two approaches to the estimation of labour supply functions exist in the economic literature. One approach is based on a rigorous application of the results of utility maximization theory. The strength of this approach is most apparent for a system of demand and supply equation as it enables the researcher to either impose or test for restrictions derived from utility maximization theory. This kind of approach was used for labour supply estimation with micro-data by Wales [1973], Wales and Woodland [1974a, 1974b], Ashenfelter and Heckman [1974], and with macro-data by Gussman [1972]. This method, however, may involve non-linear estimation techniques for a system of equations. A weakness of this approach lies in its treatment of "taste differences" which are assumed to be constant over the population.

Another approach to the estimation of labour supply functions attempts to approximate the parameters of the supply decision using functions that are linear in the coefficients. Its relationship to utility maximization theory is somewhat pragmatic (see Chapter II, Section 2.2). However, this method has been used extensively in the estimation of labour supply functions, primarily because of its econometric simplicity.[1] This method controls for "taste variations" in the sample by inserting socio-demographic variables in the regression.

The specific functional form that I use (see Chapter II, Section 2.2) is linear in the coefficients. It contains polynomial expressions in the wage and income variables so that supply and income elasticities are not restricted to being constants.[2]

The dependent variable in the labour supply equation, i.e., annual hours of work is by definition non-negative. If the family in its labour force participation choice decides to supply a <u>non-negative</u> amount of hours on the labour market, then this restriction should be imposed on the family's conditional labour supply function. Cragg [1971] suggested two methods of imposing the non-negativity restriction: (i) truncating the distribution of the dependent variable at zero, (ii) using a semi-logarithmic specification for the supply function. The former involves a iterative estimation technique[3] whereas the latter can be estimated with the usual regression techniques. For convenience I therefore choose the semi-logarithmic specification.

The parameters of the labour supply decisions can be determined either from a supply function with "hours worked" as a dependent variable or from a demand for leisure equation with "total number of hours minus hours worked" as a dependent variable. I choose to estimate supply functions because their interpretation is more straightforward and it simplifies the calculation of the unconditional supply functions in the next Chapter.

A fundamental objection to estimating supply functions at all is that the dependent variable "hours worked" is not in an individual's decision set but is instead determined by institutional constraints (standard work week) and by demand conditions (layoffs, spells of unemployment). In empirical investigations these problems are sometimes avoided by choosing a sample for whom these constraints are presumably non-binding, e.g., self-employed (Break [1957], Wales [1973]), individuals stating that they felt no constraint on their choice of hours (Wales and Woodland [1974a]). Frequently, however, these objections are either neglected or it is argued (Friedman [1962]) that overtime, moonlighting and part time jobs are readily available and that there exists sufficient room for individual choice. Whether the latter is true is clearly an empirical question. But even if it is accepted a further problem may arise if the marginal wage rate changes between these different time allocations, e.g., between normal time and overtime, and if only an average wage rate (i.e., total labour income divided by total hours of work) is observed, as is the case in my sample.[4] In this study as in most other labour supply studies (e.g., see Cain and Watts [1973]) the existence of possible institutional constraints on the labour supply decisions of some families in the sample may therefore affect the parameters of the estimated equations.[5]

For labour force participants an (average) wage rate is observed. Thus, this observed rate instead of the predicted wage rate (Chapter III) could be used. Since I will combine the labour force participation decision and the labour supply decision into an unconditional labour supply function, which is defined as a function of the wage rate (Chapter VI), consistency is desirable. I will therefore estimate labour supply functions with the predicted wage rates among the independent variables (see Section 4 of this Chapter) and will use these estimates in calculating the unconditional supply functions.

Prediction tests (Chapter III, Section 3.3) indicate that the predicted wage rate is at its best only a rough indicator of the observed wage rate. It would seem desirable to compare the supply estimates using the predicted wage rates with estimates obtained from observed wage rates. I will therefore also estimate the supply functions using the observed wage rates (see Section 3 of this Chapter).

The observed wage rate, however, is an <u>average</u> wage rate (i.e., income divided by hours), whereas the theoretically relevant variable is the marginal wage rate. The marginal wage rate can be different from the average wage rate, e.g., because of overtime premiums, multiple job holding with different wage rates, non-proportional income taxes, etc. The sample does not usually provide enough information to define the marginal wage rate. Therefore, most labour supply studies (e.g., Kosters [1963], Rosen and Welch [1971]) use the average wage rate. Some recent studies try to approach the marginal wage rate, at least partially, by taking the effect of non-proportional income taxes into account (e.g., Diewert [1971], Wales [1973], Hall [1973]). In this thesis I have also applied this partial adjustment to the observed average wage rate using the same method as Wales and Woodland [1974a]. This method consists of correcting the average wage rate for the marginal tax rate. In order to determine the marginal tax rate, I assume joint filing and determine from the 1967 to 1971 United States federal income tax tables the tax bracket[6] and hence the marginal tax rate for a family with that level of federal income tax. It should be stressed again that this "corrected" average wage rate will still be different from the relevant marginal wage rate because the adjustment is only for federal taxes (neglecting state income taxes) and it does not take account of, e.g., overtime premiums, multiple wage rates, etc.[7]

There has been some discussion in the empirical labour supply literature concerning the income variable to be used to measure the income effect (e.g., see Kosters [1963]). Transfer income is sometimes related to supply in an unsatisfactory way, e.g., unemployment benefits, income from pension plans. This creates spurious correlation between income and supply. Non-labour income such as rents, interests, dividends, etc. is usually non-zero for only a small portion of the sample. Using a definition similar to the one used by Hall [1973] I define a full rental income variable as the sum of taxable asset income plus 12 percent of the value of the family car(s) plus 6 percent of the net (i.e., corrected for the outstanding mortgage debt) value of the family house(s). It is of interest to note that only 5 percent (295 out of 5415) of the sample had zero rental income defined in this way. (The same income variable was used in the labour force participation models of the previous Chapter.)

The sample over which I estimate the labour supply functions using the predicted wage rate is identical to the sample used for the labour force participation models and is described in Section 3 of Chapter IV. I also estimate the labour supply functions using the corrected observed wage rate over the same sample, except that I have excluded observations with a zero wage rate.[8] This reduced sample is summarized in Table XXI[9] in terms of the variables used in the labour supply regressions.

TABLE XXI

DESCRIPTION OF SAMPLE USED FOR LABOUR SUPPLY REGRESSIONS

	Both Working Husband	Both Working Wife	Husband Only
# OBS	2409	2409	2567
Average of:			
YRS SCHOOL (*)	11.13 (3.3)**	11.35 (2.4)	10.75 (3.7)
ACHIEVE	9.18 (2.5)		9.13 (2.6)
AGE (*)	40.13 (11.6)	37.58 (10.9)	42.05 (12.2)
KIDS	2.77 (2.2)	2.78 (2.18)	3.07 (2.2)
WAGE NET (*)	2.74 (1.4)	1.72 (1.2)	3.13 (1.7)
ASSET Y	1039 (1816)	1093 (1816)	1405 (2878(
# of Families With Dummy Variable = 1			
RACE	511 (21%)	511 (21%)	451 (18%)
JEW	77 (3%)	77 (3%)	84 (3%)
LIMIT (*)	225 (9%)	20 (.8%)	265 (10%)
KID Y	607 (25%)	607 (25%)	530 (21%)
URBAN	1503 (62%)	1503 (62%)	1556 (61%)
SOUTH	950 (39%)	950 (39%)	880 (34%)
EXP	1807 (75%)		1956 (76%)
SEH	304 (13%)		444 (17%)
CLERK W		913 (38%)	
UNSKILL W		553 (23%)	
UNION	717 (30%)		832 (32%)
MORTG	1386 (57%)	1386 (57%)	1378 (54%)
RESERVE	1278 (53%)	1278 (53%)	1363 (53%)
IRREC	117 (5%)	117 (5%)	141 (5%)
Hours Supplied: Average of:			
LOG HOURS (*)	7.65 (44)	6.75 (1.08)	7.65 (.50)
HOURS	2227.0 (647.7)	1227.18 (739.8)	2276.13 (687.5)

* These variables are either for the husbands or the wives depending on the column in which they appear

** Numbers between brackets are standard deviations. If these numbers are followed by % then they denote relative frequencies

3. Estimation Results: Labour Supply Functions With Observed Wage Rates

The results of the supply equations for the "both working" and "husband only" alternative are presented in Table XXII. The supply curve for the "wife only" alternative is as follows

$$\text{(4.1)} \quad \text{LOG HOURS} = \underset{(.31)}{6.31} - \underset{(.28)}{.62}\ \text{URBAN} + \underset{(.26)}{.69}\ \text{WAGE W} - \underset{(.03)}{.08}\ (\text{WAGE W})^2$$

$$- \underset{(.06)}{.13}\ \text{ASSET Y}$$

$$\#\ \text{OBS} = 80;\ R^2 = .13;\ \text{SE} = 1.09\ .$$

Whereas the number of years of schooling increases the probability that the husband will not be in the labour force (see Tables XV and XIX)[10] there is, however, a tendency for more educated husbands to work more hours than less educated husbands conditional on their being in the labour force (Table XXII). This would suggest the hypothesis that more educated husbands are in the labour force for a shorter period of their life cycle but work on the average longer when they are in the labour force.

The effect of education on the labour supply of the wives is also positive and seems to be more pronounced than for the husband (Table XXII). Furthermore, more educated women tend to have a higher probability of joining the labour force (Table XV and XIX). Therefore higher educated women spend probably more of their lifetime in the labour force compared with less educated women and work longer hours on average while they are in it. The positive effect of education on labour supply was also established by Cohen, et al. [1970] and Hill [1973].

TABLE XXII

CONDITIONAL LABOUR SUPPLY FUNCTIONS (USING OBSERVED WAGE RATES[1]

Variables	Both Working Husband	Both Working Wife	Husband Only
CONSTANT	6.58 (.11)	4.98 (.28)	6.29 (.10)
YRS SCHOOL[2]	.007 (.003)	.024 (.009)	.01 (.002)
ACHIEVE	.01 (.003)		
RACE	-.07 (.02)	.17 (.06)	-.08 (.02)
JEW	.13 (.05)	-.26 (.13)	
LIMIT[2]	-.18 (.03)	-.87 (.23)	-.19 (.03)
KIDS		-.06 (.01)	-.01 (.004)
KID Y	-.06 (.02)		
URBAN		.18 (.05)	.07 (.02)
SOUTH		.08 ** (.05)	
AGE[2]	.049 (.005)	.084 (.01)	.074 (.004)
AGESQ[2]	-.0006 (.0001)	-.0009 (.0002)	-.001 (.0001)
EXP	.24 (.02)		.22 (.02)
SEH	.17 (.03)		.16 (.03)
CLERK W		-.20 (.05)	
UNSKILL W		-.45 (.06)	

TABLE XXII CONTINUED

Variable	Both Working Husband	Wife	Husband Only
UNION			-.06 (.02)
MORTG	.04 (.02)	.14 (.04)	.05 (.02)
RESERVE	.05 (.02)	.18 (.05)	.08 (.02)
WINDFALL Y	-.08 (.04)		
WAGE H NET[7]	-.06 (.02)	-.17 (.02)	-.12 (.01)
WAGE H NET **2	-.003 (.001)		.004 (.001)
WAGE H NET **3		.0005 (.0002)	
WAGE W NET	-.05 (.01)	.32 (.06)	
WAGE W NET **2	.003 (.001)	-.09 (.01)	
WAGE W NET **3		.004 (.0006)	
ASSET Y[3]	.06 (.02)	-.025 (.01)	.01 (.003)
ASSET Y **2	-.008 (.003)		
ASSET Y **3	.0003 (.0001		
# OBS	2409	2409	2567
R^2	.24	.12	.36
$\bar{R}^2$	.23	.11	.35
S E[4]	.39	1.01	.40
μ[5]	7.65	6.75	7.65
σ[6]	.44	1.08	.50

Notes to Table XXII:

1 Each variable is significant at the 5 percent level (t-test).

2 These variables are either for the husbands or the wives depending on the column in which they appear.

3 The ASSET Y variable is divided by 1,000.

4 S E: standard error of the estimate.

5 μ: mean of dependent variable.

6 σ: standard deviation of dependent variable.

7 $**^2$ and $**^3$ denote exponential powers.

There is no significant effect of the racial dummy variable on the probability of labour force participation for the husband (Tables XV and XIX). The negative effect of the same variable on the labour supply of the husband is however, significant (Table XXIII). Non-white families are more probable to choose the "both working" alternative over the "husband only" alternative (Table XVI). Being non-white also increases the odds of participation for the wife (Table XIX) and the average number of hours supplied (Table XXII). Similar differences in the labour supply for white and non-white men and women were found by Ashenfelter and Heckman [1973] and Cain [1966]. Combining these results it would seem that non-white married men work less hours than white men (or are constrained to do so) but that non-white families will probably attempt to offset this negative effect as non-white wives join the labour force more frequently and work more hours than white wives.

A dummy variable indicating that the husband belongs to the Jewish religion has no effect on the labour force participation decisions within the family (Chapter IV, Section 4). The effect of this variable on the labour supply of the husband and the wife who are both working is significant. A Jewish husband will work relatively more and his wife relatively less than non-Jewish couples. This religious variable is not significant for the husband whose wife is not in the labour force. A dummy variable indicating that the husband was Catholic has an insignificant effect on both labour force participation and labour supply.

A handicap diminishes both the probability of participating for husband and wife (Table XIX) and also the average numbers of hours supplied if they are in the labour force (Table XXII).

The number of children has a negative influence on the supply of labour of the wives who are working together with their husbands and on the supply of the husbands working alone (Table XXII). A KID 6 dummy variable has similar effects on the labour supply. I choose the KIDS variable in the specification of TABLE XXII because this variable gives a slightly better fit than the KID 6 variable (in terms of R^2). The KID 6 variable also decreases to a substantial degree the odds of the wife participating (Table XIX).

A variable indicating whether some children in the family have any income (mostly labour income) has no significant effect on the labour force participation decision of the family (see Chapter IV, Section 4). The same variable, however, affects the labour supply of the husband, in families where both partners are working, negatively.

A significant quadratic form for the AGE variable was found in all three cases of Table XXII. Similar results were found by Cohen et al. [1970], Smith [1972]. The amount of hours supplied over the life time tends to peak around 41.2 years for the husband and at 46.6 years for the wife when they are both working. When the husband only is working his supply peaks around 36.9 years.[11] Measuring the life cycle hours profile with the MARRIAGE variable instead of the AGE would give similar results. The AGE variable, however, fits better in terms of R^2 and compares with previous established results in labour supply studies.

Table XXII shows that self-employed men and men on the same job more than 3 years tend to work more hours. Female clerks and unskilled female workers (who account together for 60% of all the working wives) tend to work fewer hours than the other occupational groups.

It was established earlier that the existence of a mortgage debt increases the odds of choosing the "both working" alternative and the "husband only" alternative as compared to the other alternatives. However, it increases especially the odds of the "both working" choice (see Tables XV and XVI). In terms of labour supply it is similarly the wife's hours which are mostly increased.

The RESERVE dummy variable was only relevant in the labour force participation choice of the family between the "wife only" and "none working" alternative (Table XV). It is hard to explain the positive influence of having reserves on the supply of labour. Reserves can either be caused by labour supply, i.e., hard-working couples can save more, or it can cause increased labour supply, i.e., couples are working longer hours in order to accumulate savings.

Before discussing the wage and income effects I will make a general comparison in terms of the socio-demographic variables between the three labour supply functions represented in Table XXIII. The differences among the coefficients of the husbands' labour supply functions and the wives' labour supply functions are substantial. On the other hand there is a remarkable similarity among the coefficients of the socio-demographic variables in the two supply curves for the husbands (Table XXII, Columns 1 and 3). An exception to this general tendency is the set of age variables and some variables which are significant in one but not in the other supply function, e.g., ACHIEVE, JEW, KIDS, KID Y, URBAN, UNION, and WINDFALL Y. There is a substantial difference in the wage and income coefficients; but this difference is not reflected in the supply and income elasticities for both samples of husbands (see below).

In discussing the effects of the WAGE and ASSET Y variables I am interested in identifying the income and the (compensated) substitution effects and in checking for the non-negativity of the latter. The sum of the income and (compensated) substitution effect is the slope of the supply curve. In the "both working" case I am furthermore interested in identifying whether the husband's and the wife's labour supply are net complements or net substitutes.[12]

As discussed above in Chapter II, Section 2.2, the functional form chosen for the supply function is of the following general form (see Chapter II, equation (1.14)):

$$\text{(4.2)} \qquad \ln R_i = \text{constant} + (a_1 v_i + a_2 v_i^2 + a_3 v_i^3) + (b_1 v_j + b_2 v_j^2 + b_3 v_j^3) + (c_1 A + c_2 A^2 + c_3 A^3) \text{ , } i, j = m, f, \quad i \neq j \text{ .}$$

Let A, B, C, be defined as the partial derivatives of (4.2) with respect to respectively v_i, v_j, and A. Then it is readily seen (see also Chapter II, Section 2.2) that A is the slope of the supply curve for i, that B indicates gross complementarity or substitutability between the supply of i and j and that C is the income effect for i's labour supply. Furthermore $(A - CR_i)$ is the compensated substitution effect (which should be positive), while net complementarity or substitutability is identified as $(B - CR_j)$. Because of the semi-logarithmic form of (4.2) the corresponding elasticities are easily calculated by multiplying A, B, C, $(A - CR_i)$, $(B - CR_j)$ with the appropriate wage or income variable.

These elasticities are tabulated in Tables XXIII and XXIV. They are calculated using the point estimates of the wage and income coefficients in, respectively, Table XXII and equation (4.1). The supply curve for men is negatively sloped in both the "both working" and "husband only" alternatives. The curve tends to be more elastic (less steep) in the latter alternative. In the "both working" case the supply curve for the wives is upward sloping up to the \$1.93 net wage rate, then slopes negatively above that wage until the \$14.29 net wage after which it has a postive slope again. When the wife only is working the supply curve is almost always positively sloped. For this case, the labour supply curve is also much more elastic than for the other wives.

Income effects are usually small except in the "wife only" case. Only for women is leisure usually a normal good. The compensated substitution effect is positive for most of the wives in the sample but for almost none of the husbands. In the "both working" case husband and wife's working hours are usually both gross and net "substitutes", i.e., if one partner's wage goes up the other partner will work less. This result follows from the husband's equation as well as from the wife's

TABLE XXIII

LABOUR SUPPLY ELASTICITIES (CORRESPONDING TO LABOUR SUPPLY FUNCTIONS USING OBSERVED WAGE RATES)[1]

	Elasticity Defined At The Means	Average of Elasticities for Sample Points	Number and % of Sample Points With Positive Elasticity
BOTH WORKING - HUSBAND			
supply elasticity	-.20	-.22	0
income elasticity	.05	.03	2343 (97%)
compensated substitution elasticity	-.49	-.50	0
gross cross-elasticity	-.06	-.05	14 (.6%)
net cross-elasticity	-.16	-.15	11 (.5%)
BOTH WORKING - WIFE			
supply elasticity	.05	-.03	1720 (71%)
income elasticity	-.03	-.03	0
compensated substitution elasticity	.00005	.03	1920 (80%)
gross cross elasticity	-.45	-.41	7 (.3%)
net cross elasticity	-.11	-.26	11 (.5%)
HUSBAND ONLY			
supply elasticity	-.26	-.27	1 (.03%)
income elasticity	.013	.013	2567 (100%)
compensated substitution elasticity	-.33	-.34	1 (.03%)
WIFE ONLY			
supply elasticity	.73	.48	78 (98%)
income elasticity	-.14	-.14	0
compensated substitution elasticity	1.02	.78	79 (99%)

1 the definitions of the various elasticies is given in the text

TABLE XXIV

DISTRIBUTION OF THE ELASTICITIES OF THE SUPPLY CURVE
(OBSERVED WAGE RATE)

	# of Sample Points				% of Sample Points			
	$\varepsilon \leq -1$	$-1<\varepsilon \leq 0$	$0<\varepsilon<1$	$1 \leq \varepsilon$	$\varepsilon \leq -1$	$-1<\varepsilon \leq 0$	$0<\varepsilon<1$	$1 \leq \varepsilon$
Both Working								
Husband	11	2398	0	0	.5	99.5	0	0
Wife	63	626	1717	3	3	26	70.9	.1
Husband Only	0	2566	1	0	0	100	0	0
Wife Only	1	1	78	0	1	1	98	0

equation. For 99% of the sample points both the coefficients of the husband's supply equation and of the wife's equation predict net substitutablity.

In general, the parameters of labour supply for the wives correspond much more to the "textbook" representation of a labour supply function, i.e., upward sloping curve, positive compensated substitution term and leisure being a normal good,than the supply parameters of the husband. The "weakness" of the supply curve for males in this respect is found quite frequently (Kosters [1963], Cain and Watts [1973], Cohen, <u>et al</u>. [1970]). The finding that the labour supply of husband and wife are (gross and net) "substitutes" in the sense that an increase in the wage of one partner induces the other partner to work less would seem to be an intuitively sensible result. This result, however, tends to be contradicted by Wales and Woodland [1974a] who found that an increase in one partner's wage usually increases the labour supply of the other partner.

I will postpone a comparison between the wage effects on labour force participation and labour supply decisions until the next section where I estimate labour supply functions using the same wage variable as in the logit models.

4. Estimation Results: Labour Supply Functions with the Predicted Wage Rate

The results of the labour supply equations using the predicted wage rate (see Chapter III) instead of the observed wage rate (as in Table XXII) are presented in Table XXV. The labour supply function for wives in the "wife only" alternative is as follows

$$\text{(4.3)} \qquad \text{LOG HOURS} = \underset{(.30)}{6.5} - \underset{(.28)}{.59}\ \text{URBAN} + \underset{(.16)}{.34}\ \text{WAGE W (PRED)} - \underset{(.07)}{.12}\ \text{ASSET Y}$$

$$\text{\# OBS} = 81, \quad R^2 = .09, \quad \text{S E} = 1.11\ .$$

If I compare the results in Tables XXII and XXV I can see some important differences. The YRS SCHOOL variable is no longer significant when the predicted wage rate is used. The RACE variable is not significant in the husbands' supply functions in Table XXV, whereas it is significant in Table XXII. Note however, that RACE also has no influence on the labour force participation decisions of the husband (Tables XV and XIX). The peaks in the quadratic lifetime supply profiles (Table XXV) are at 35.5, 46.8, and 39.2 years for respectively, husband and wife in the "both working" alternative and for the husband in the

TABLE XXV

CONDITIONAL LABOUR SUPPLY FUNCTIONS (USING PRE-
DICTED WAGE RATES)[1]

	Both Working		Husband
	Husband	Wife	Only
CONSTANT	6.09 (.20)	5.06 (.27)	6.04 (.15)
ACHIEVE	.01 (.003)		
RACE		.13 (.06)	
JEW		-.29 (.13)	
LIMIT[2]	-.17 (.03)	-1.10 (.23)	-.19 (.03)
KIDS		-.08 (.01)	-.009 (.004)
KID Y	-.05 (.02)		
URBAN		.26 (.05)	
SOUTH	.03(*) (.02)		
AGE[2]	.04 (.006)	.09 (.01)	.05 (.005)
AGESQ[2]	-.0006 (.0001)	-.001 (.0002)	-.0007 (.0001)
EXP	.22 (.02)		.23 (.02)
SEH	.21 (.03)		.23 (.02)
CLERK W		-.18 (.05)	
UNSKILL W		-.46 (.06)	
UNION			-.1 (.02)
MORTG	.03(*) (.02)	.12 (.05)	.03(*) (.02)

TABLE XXV CONTINUED

	Both Working Husband	Wife	Husband Only
RESERVE	.04 (.02)	.23 (.05)	.03(*) (.02)
WINDFALL Y	-.07 (*) (.04)		
WAGE H (PRED)			.37 (.09)
WAGE H (PRED) $**2$	-.008 (**) (.005)	-.02 (.002)	.07 (.02)
WAGE H (PRED) $**3$	.0008 (**) (.0005)		.0042 (.001)
WAGE W (PRED)	.62 (.18)	.06 (**) (.04)	
WAGE W (PRED) $**2$	-.20 (.06)		
WAGE W (PRED) $**3$	-.02 (.007)		
ASSET Y	.03 (.01)	-.02(*)	.0009 (n.s.) (.26)
ASSET Y $**2$	-.0013 (*) (.0007)		
# OBS	2425	2425	2573
R^2	.179	.096	.310
$\bar{R}^2$	.173	.091	.306
S E	.40	1.04	.42
μ	7.65	6.74	7.65
σ	.44	1.09	.50

1 Each variable is significant at the 5 percent level (t-test) except when followed by (*) significant at 10 percent level
(**) significant at 20 percent level
all the other comments are the same as for Table XXII.

"husband only" case, which is different from the age profiles found previously. There are furthermore some minor differences in the JEW, URBAN, and SOUTH variables between Tables XXII a XXV. Differences in the coefficients of the socio-economic variables are to be expected because the predicted wage rate is defined as (nonlinear) function of most of these variables (see Table XII).

In Tables XXVI and XXVII I have again calculated the supply elasticities similar to Tables XXIII and XXIV. The major difference between the parameters of the supply function derived from the equation using the predicted instead of the observed wage rate lies in the fact that for the predicted wage function many more sample points lie on a positively sloped supply curve, e.g., compare the supply elasticities in Tables XXIII and XXVI. The same tendency becomes clear from comparing Tables XXIV and XXVII. There is a general "shift" in the supply elasticities from negative elasticities (in Table XXIV) towards positive elasticities (in Table XXVIII). As the income elasticities remain roughly the same in both cases this tendency also implies that the theoretical restriction on positive compensated substitution elasticities is satisfied for many more sample points in the predicted wage case than the observed wage case.

If the predicted wage rate is seen as the "permanent" or "long run" (i.e., freed from transitory fluctuations) wage rate then the labour supply functions derived from this long run wage rate could also be interpreted as the long run trade-off schedule between leisure and income. The long run supply elasticity would then usually be between

TABLE XXVI

LABOUR SUPPLY ELASTICITIES (CORRESPONDING TO LABOUR SUPPLY FUNCTIONS USING PREDICTED WAGE RATES)

Elasticities	Elasticity Defined at the Means	Average of Elasticities for Sample Points	Number and % of Sample Points With Positive Elasticity
BOTH WORKING - HUSBAND			
supply elasticity	-.11	-.108	120 (5%)
income elasticity	.03	.02	2414 (99.5%)
compensated substitution elasticity	-.67	-.31	9 (.4%)
gross cross elasticity	.02	.07	1669 (.69%)
net cross elasticity	-.03	.004	1264 (.52%)
BOTH WORKING - WIFE			
supply elasticity	.13	.13	2425 (100%)
income elasticity	-.02	-.02	0
compensated substitution elasticity	.17	.19	2425 (100%)
gross cross elasticity	-.59	-.67	0
net cross elasticity	-.39	-.47	7 (.3%)
HUSBAND ONLY			
supply elasticity	-.003	.06	1571 (61%)
income elasticity	.001	.001	2573 (100%)
compensated substitution elasticity	-.01	.05	1540 (60%)
WIFE ONLY			
supply elasticity	.65	.65	81 (100%)
income elasticity	-.11	-.13	0
compensated substitution elasticity	.49	.98	81 (100%)

TABLE XXVII

DISTRIBUTION OF THE ELASTICITIES OF THE SUPPLY CURVE (PREDICTED WAGE RATES)

	# of Sample Points				% of Sample Points			
	$\varepsilon \leq -1$	$-1<\varepsilon \leq 0$	$0<\varepsilon<1$	$1 \leq \varepsilon$	$\varepsilon \leq -1$	$-1<\varepsilon \leq 0$	$0<\varepsilon<1$	$1 \leq \varepsilon$
Both Working								
Husband	0	2305	120	0	0	95	5	0
Wife	0	0	2425	0	0	0	100	0
Husband Only	0	1002	1564	7	0	39	60.7	.3
Wife Only	0	0	73	8	0	0	90	10

0 and 1 (Table XXVIII).

A positively sloped supply curve (and the positive compensated substitution effect) would imply that the predicted wage rate affects labour force participation and labour supply decisions much in the same direction. This similarity seems to be present for the wives in both the alternatives "both working" and "wife only" and for husbands in the "husband only" alternative. There seems to be a difference in the wage effect on the labour force participation and labour supply decisions of the husbands in the "both working" alternative, a positive wage effect on labour force participation (Table XIX) and a generally negatively sloped supply curve (Table XXVI).

Whereas in the observed wage case both the husband's and the wife's supply parameters indicated gross and net substitutability between their leisure times (Table XXIII) this tendency is much weaker for the predicted wage case. The wife's supply function still predicts gross and net substitutability, but the cross elasticities derived from the husband's equation are less clear in this matter (Table XXVI). Net substitutability is predicted from both the husband's and the wife's equation for 48 percent of the sample points in the predicted wage case, but for 99 percent of the total number of observations in the observed wage rate. For labour force participation decisions the cross wage effect generally indicated a "substitution" effect in participation decisions (Table XIX).

It was found in Table XVI that the wife's wage effect was especially important in the choice between "wife only" and "none working". It is therefore not so surprising that the supply curve is also more elastic for wives whose husbands are not working (Table XXVI).

In the second part of the next Chapter I will combine the results of the supply functions in Table XXV with the results of, respectively, the simultaneous and sequential labour force participation models (Tables XV and XIX) to calculate unconditional supply functions.

FOOTNOTES TO CHAPTER V

1 This "pragmatic" approach was first used by Kosters [1963] and since then by Rosen and Welch [1971], Cohen et al. [1970], Berndt and Wales [1974] and in a recent volume of labour supply studies edited by Cain and Watts [1973].

2 In the actual estimation I have attempted to establish the lowest possible order of the polynomial expression in the wage and income variables which still captured the non-linearities that were present in those variables. Box-Tidwell [1962] suggest a technique to estimate the functional form of independent variables. In the case of the supply functions their technique was helpful in indicating whether there was any non-linearity in the variable, but was less appropriate in actually determining the power of the exponent(s).

3 The truncated regression method for the labour supply equation will be used in a test comparing the parameters of labour supply and labour force participation. See the next Chapter.

4 Overtime wages and second job wages are observed for part of the sample only, viz, only in the 1969 to 1971 sample years and only for husbands who were employed at the time of the interview.

5 A possible way to approach this problem of institutional constraints on labour supply would be to treat the supply decision as a choice among the following alternatives: part time job, standard work week job, overtime, second job, etc. This model could then be estimated using the multinomial logit technique.

6 Taking the surcharge into account for the fiscal years, 1968, 1969, and 1970.

7 The correction of the observed wage rate for the marginal tax rate was introduced mainly to make the results comparable with some of the recent labour supply studies. I did not adjust the predicted wage rate in this way. I prefer to interpret the predicted wage as a "long run", "permanent" measure of the opportunity cost of time.

8 16 out of 2425 in the "both working" alternative, 6 out of 2573 in the "husband only" case.

9 Except for the alternative "wife only" which consisted of 80 observations (one dropped for non-positive wage). To summarize this sample in terms of the variables used in the labour supply regression: average of WAGE NET W: 1.73 (1.29); average of ASSET Y: 1076 (2059); number of families in URBAN environment: 29 (36%); average of LOG HOURS, 6.78 (1.13), HOURS, 1262 (770.1).

10 This is not due to the fact that they are full time students. There were 48 husbands in the sample (of 5415) who gave "student" as their status. 38 of them worked during the sample year (i.e., were in the labour force).

11 The peak in the hours profile seems to come before the peak in the wage profile (see Chapter III, Section 2). This evidence was also found in Smith [1972] and can theoretically be explained in a dynamic utility framework. In this kind of model it depends on the subjective rate of time preference being smaller than the market rate of interest (Weiss [1972]).

12 It would have been more appropriate to estimate the two supply equations in the "both working" case using a generalized least squares approach. Then I could have imposed the symmetry restriction. Because of the large sample size this approach was however computationally impossible.

CHAPTER VI

UNCONDITIONAL FAMILY LABOUR SUPPLY FUNCTIONS

1. Introduction

The objective of this Chapter is to combine changes at the internal and external margins caused by changes in the wage rate (and income). I propose to do this by calculating an unconditional labour supply function (see Chapter II, Section 4).

An unconditional labour supply function defined for a given population could aid a policy maker in evaluating numerically the combined outcome of changes in labour force participation choices and hours of work decisions that would result from changing a policy instrument. For instance, such a function might be useful to determine the combined effect caused by changes in the women's wage rate that would result from the Equal Rights Amendment.

Two methods of combining labour force participation and labour supply functions can be used. First one could argue that labour force participation is only "truncated labour supply" and that both functions really are the same. This common function, however, has a special characteristic in that its dependent variable (hours supplied) has a lower limit at zero, i.e., non-positive supply is never observed. If this viewpoint is accepted then Tobin's [1958] limited dependent variable model can be used to estimate unconditional labour supply functions. Tobin's model has previously been used by Rosett [1958] and Heckman [1974] to study labour force participation and labour supply behaviour of married women.

On the other hand it can be argued that participation and supply decisions are qualitatively different (see Chapter II, Section 4). In such a case one would expect parameters of the two decisions to differ. Cragg [1971] has extended Tobin's model into a more general model that allows for differences in the parameters. The labour force participation and labour supply functions are estimated separately, the former using a probit model, the latter using a truncated regression model. It can be shown that the Tobin model is a special case of Cragg's general model. (See Chapter II, Equation (1.352)). In a second part of this Chapter I compare unconditional labour supply functions for the husband and for the wife estimated using respectively Tobin's and Cragg's model. I then use the Bayesian method explained in Chapter IV (Section 6) to discriminate among these two models. As will be seen the hypothesis that the labour force participation and labour supply parameters are different will turn out to be more probable (<u>a posteriori</u>) than the hypothesis that they are the same. Therefore the separate estimation of labour force participation and labour supply functions, as was done in, respectively, Chapters IV and V, would seem to be preferable.

The third part of this Chapter will combine results of Chapters IV and V into unconditional labour supply functions using the expected value method proposed in Chapter II (Section 4.2).

2. Comparing the Parameters of Labour Supply and Labour Force Participation

The results for the Tobit, Probit, and Truncated Regression models for, respectively, the husbands and wives are presented in Table XXVIII. The sample over which these models are estimated is the same sample as the one used for the logit models in Chapter IV. It is noted that starting values for the coefficients of the Probit models (Columns 2 and 5 of Table XXVIII) are obtained from the converged estimates of the binomial logit models in Table XIX (Column 1) and Table XX (Column 1). It is also noted that the dependent variable for the Probit and Truncated Regression models is hours supplied and not the logarithm of hours supplied (as in the supply equations of the previous Chapter).

It was mentioned earlier (Chapter II, Equation (1.35.2)) that Cragg's limited dependent variable model reduces to Tobin's model if and only if

$$(5.1) \qquad \beta = \frac{\gamma}{\sigma}$$

i.e., if the coefficients of the Probit model are equal to the coefficients of the Truncated Regression model divided by the standard deviation of the regression ("SIGMA" in Table XXVIII). If (5.1) holds then clearly the coefficients of the Probit model should also be equal to the coefficients of the Tobit model divided by the standard deviation.

TABLE XXVIII

UNCONDITIONAL LABOUR SUPPLY FUNCTIONS, COMPARISON OF THE TOBIN AND CRAGG MODELS[1]

	HUSBAND			WIFE		
	Tobit	Probit	Truncated Regression	Tobit	Probit	Truncated Regression
CONSTANT	2032.8 (96.2)	1.9 (.28)	1917.6 (110.0)	1384.4 (22.0)	-.11 (.12)	635.4 (300.2)
YRS SCHOOL H	-13.7 (5.0)	-.06 (.01)	2.9 ns (5.4)	11.4** (8.0)	.009 (.007)	8.7 ns (13.9)
YRS SCHOOL W	15.1 (6.2)	..05 (.02)	2.2 ns (6.8)	48.5 (10.0)	.04 (.008)	8.1 (17.1)
LIMIT H	-598.1 (42.6)	-1.0 (.08)	-248.2 (51.2)	35.9 ns (67.5)	.06 ns (.05)	-94.1 ns (116.8)
LIMIT W	-202.0 (101.0)	-.20 (.17)	-158.3** (123.4)	-788.9 (184.6)	-.54 (.14)	-547.0** (420.0)
MARRIAGE	-28.0 (1.6)	-.05 (.004)	-9.2 (1.8)	-27.5 (2.5)	-.02 (.002)	-2.8 ns (4.1)
KID 6	-36.7 ns (33.5)	-.02 ns (.12)	51.5** (35.5)	810.5 (53.6)	-.62 (.05)	-495.0 (88.1)
SOUTH	155.1 (31.5)	.38 (.09)	89.5 (34.0)	166.7 (50.0)	.13 (.04)	54.3 ns (80.7)
URBAN	-208.0 (31.1)	-.32 (.09)	-136.9 (33.7)	52.3 ns (49.6)	-.03 ns (.04)	336.4 (83.0)
RACE	62.4** (40.4)	-.009 ns (.12)	-12.9 ns (43.3)	281.4 (63.5)	.26 (.05)	-65.6 ns (102.2)
WAGE H (PRED)	152.7 (17.3)	.59 (.07)	50.8 (18.4)	-169.8 (27.7)	-.12 (.02)	-202.7 (46.8)
WAGE W (PRED)	-82.1 (27.8)	-.35 (.09)	-23.4 ns (29.3)	280.7 (43.1)	.25 (.04)	121.1* (64.7)

TABLE XXVIII CONTINUED

	HUSBAND			WIFE		
	Tobit	Probit	Truncated Regression	Tobit	Probit	Truncated Regression
ASSET Y [3]	.033 (.006)	-.000013ns (.000013)	.033 (.006)	-.059 (.011)	-.00004 (.000009)	-.037* (.022)
MORTG	189.9 (28.9)	.37 (.09)	100.8 (31.0)	252.4 (45.8)	.17 (.04)	181.9 (74.3)
RESERVE	34.8 ns (30.7)	-.03 ns (.09)	77.3 (33.8)	116.8 (49.0)	.008 ns (.04)	374.9 (79.8)
SIGMA[2]	972.3 (10.4)		946.0 (15.0)	1384.4 (22.0)		1162.9 (65.1)
NUMBER PARTICIPATION	4998	4998		2506	2506	
NUMBER OF OBSERVATIONS	5415	5415	4998	5415	5415	2506
Log likelihood at maximum	-41320.	-739.5	-40260.	-23520.	-3494.6	-19930.
Pseudo R^2	.37	.24	.58	.09	.09	.08
Likelihood Ratio Test χ^2 with 14 d.f.	2538.6	1460.3	4329.6	516.7	487.6	197.2
Log of Posterior Distribution Constant +	-41312.	-737.6	-40248.0	-23510.	-3485.8	-19926.

1 Each coefficient is significant at the 5 percent level (t-test) except when followed by: (*) significant at 10 percent level; (**) significant at 20 percent level; (ns) not significant.

2 Estimate of standard deviation for the model.

3 Divided by 1000

A first way of establishing whether (5.1) will hold is by means of a rough comparison of the point estimates of the coefficients of the Probit and the Truncated Regression models (with the coefficients of the latter "standardized" by dividing them with the estimate of the standard deviation). In the husband's case this procedure reveals noticeable differences between the coefficients of YRS SCHOOL H, YRS SCHOOL W, LIMIT H, KID 6, SOUTH, MORTGAGE, and RESERVE. For the wife's case differences are found in the coefficients of YRS SCHOOL W, LIMIT H, MARRIAGE, URBAN, RACE, and RESERVE. For the wage and income variables I have tabulated (Table XXIX) the standardized coefficients of the Tobit and Truncated Regression models together with the coeffients of the Probit model (I included only the significant coefficients).

TABLE XXIX

COMPARISON AMONG THE WAGE AND INCOME COEFFICIENT IN THE TOBIN AND CRAGG MODEL

	Tobit	Probit	Truncated Regression
HUSBAND			
WAGE H (PRED)	.16	.59	.05
WAGE W (PRED)	-.08	-.35	
ASSET Y	-.00003		-.00003
WIFE			
WAGE H (PRED)	-.12	-.12	-.17
WAGE W (PRED)	.20	.25	.10
ASSET Y	-.00004	-.00004	-.00003

Table XXIX suggests that in general the positive effect of the own wage is more pronounced for participation decisions than for supply decisions (.59 versus .05 for husbands, .25 versus .10 for wives). The Tobit estimates are usually in between the probit and regression estimates.

A second method of comparing the hypothesis that labour supply and labour force participation parameters are the same and the alternative hypothesis that they are different consists of discriminating a posteriori between the two models under consideration. This Bayesian discrimination procedure was explained in Chapter IV (Section 6).

For the husband the logarithm of the posterior probability is equal to (a constant +) -41312. in the case of the Tobin model and equal to (a constant +) -40986. for Cragg's model (i.e., the sum of the posterior probability for the Probit and for the Truncated Regression model). For the wife the probabilities are respectively, (a constant +): -23510. (Tobin) and -23412. (Cragg). Therefore in both cases the hypothesis that the parameters of supply and participation are different is more probable a posteriori than the alternative hypothesis.

In conclusion, it would seem that the different methods of comparing the results of the Tobit model with the Probit and Truncated Regression models all tend to indicate that labour force participation is not simply "truncated supply" and that therefore separate estimations of the two models (as was done in Chapters IV and V) is a preferred procedure.

3. Unconditional Family Supply Functions

Suppose that a government policy changes the "wage rate" or "income" (e.g., through introduction of a negative income tax policy, through the Equal Rights Amendment). In response to this policy the probabilities for the four labour force participation alternatives will change as they are all functions of the "wage rate" and "income". Furthermore, the individuals who are in the labour market will alter their desired number of hours worked. A method of measuring the combined effect of the participation and supply changes caused by a change in the policy variables is to use the "expected value" formula. This method consists of multiplying the probability of participation (defined as a function of the policy variable) with the average number of hours supplied (also defined as a function of the policy variable) conditional on the participation alternative. (See Chapter II, Equation (1.38),

$$(5.2) \qquad E(R) = \sum_{j=1}^{4} P(a_j|q)\ E(R_j|q,\ a_j)\ .$$

$P(a_j|q)$ designates the probability of choosing alternative a_j (j = 1, 4) given policy parameter q. $E(R_j|a_j, q)$ stands for the expected number of hours supplied given q and given a_j. If q is the wage rate then (5.2) is the formula for the unconditional labour supply function (Hall [1973], Boskin [1973]).

To illustrate this method of combining the changes at the external and internal margin of the labour market, I apply equation (5.2) to the results obtained in Chapters IV and V. The probabilities of labour force participation are defined using either the results of the simultaneous model (Table XV) or the sequential model (Table XIX). The supply parameters are taken from Table XXV.

As policy parameters (i.e., the variable q in (5.2)) I use respectively, the husband's wage rate, the wife's wage rate and asset income. In order to define the participation functions and supply functions as a function of each of these policy parameters in turn, I must keep all the other variables "constant". To achieve this I assume that all the dummy variables in the equations take on zero value and that all the continuous variables remain at their average level.

Calculations for (5.2) as a function of respectively, the wage rates and asset income are performed in two steps. In a first step I compute the probabilities of the labour force participation choices as a function of the wage and income variables. Because the probabilities are a nonlinear function of the independent variables (see Equation (1.30) in Chapter II) it is of interest to simulate the distribution of the probabilities of the four alternatives for different wage and income levels. In a second step I combine these probabilities with the expected hours supplied derived from the conditional labour supply functions.

3.1 Distributions of the Probabilities of Labour Force Participation

The probabilities of the labour force participation alternatives defined as a function of the husband's wage rate, the wife's wage rate and asset income are shown in respectively, Tablex XXX, XXXI, and XXXII. The wage rates vary from zero to ten dollars. This range contains most of the sample points (see Table I, Chapter III). The asset income variable goes from zero to $10,000. The latter figure is approximately the mean plus four times the standard deviation. The three tables contrast results derived from the simultaneous and sequential models. As will be seen there are only minor differences for the results derived from these models.

Table XXX shows that increasing the husband's wage rate gradually from one to five dollars will cause a substantial shift in the relative probabilities of "both working" versus "husband only". For wage rates less than five dollars "both working" is the most probable alternative; above five dollars the "husband only" choice dominates. The results for the simultaneous and sequential models are very similar, except for the slightly higher probability of "none working" at low wages in the sequential model.

The results for the wife's wage rate (Table XXXI) show that for relatively high wage rates (seven dollars in the simultaneous model; ten dollars in the sequential model) the "wife only" alternative will be more probable than all other alternatives. For relatively low wages (up to one dollar in the simultaneous model; up to two dollars in the sequential model) the "husband only" choice is most probable. In between these two levels "both working" is the dominant alternative.

Table XXXII illustrates that asset income has only a minor influence on the probabilities of labour force participation. However, one can still observe a change in relative importance from the "both working" to the "husband only" choice (i.e., the wife leaving the labour force) as asset income rises to approximately $2,000. Note also that the probability of "none working" is increasing at a very slow pace as asset income increases.

TABLE XXX

PROBABILITIES OF LABOUR FORCE PARTICIPATION AS A FUNCTION OF THE HUSBAND'S WAGE RATE[1]

Husband's Hourly ($) Wage Rate	SIMULTANEOUS MODEL				SEQUENTIAL MODEL			
	B.W.*	H.O.	W.O.	N.	B.W.	H.O.	W.O.	N.
0	.60	.17	.09	.14	.39	.14	.09	.37
1	.65	.24	.04	.07	.53	.24	.05	.18
2	.64	.31	.01	.04	.57	.34	.02	.07
3	.60	.38		.01	.55	.42	.01	.02
4	.54	.45			.50	.49		.01
5	.48	.51			.44	.56		
6	.42	.58			.38	.62		
7	.36	.64			.32	.68		
8	.29	.70			.27	.73		
9	.24	.75			.22	.78		
10	.20	.80			.18	.82		

1 Probabilities are not indicated if very small (less than .005). Because of rounding errors the rows do not always sum to one.

* B.W. = "both working"; H.O. = "husband only"; W.O. = "wife only"; N = "none working".

TABLE XXXI

PROBABILITIES OF LABOUR FORCE PARTICIPATION AS A FUNCTION OF THE WIFE'S WAGE RATE[1]

	SIMULTANEOUS MODEL				SEQUENTIAL MODEL			
Wife's Hourly ($) Wage Rate	B.W.	H.O.	W.O.	N.	B.W.	H.O.	W.O.	N.
0	.33	.67			.30	.70		
1	.42	.57			.39	.60		.01
2	.52	.47		.01	.48	.51	0	.01
3	.62	.36	.01	.01	.57	.41	.01	.01
4	.69	.27	.03	.01	.65	.31	.03	.01
5	.70	.18	.12		.71	.23	.05	.01
6	.57	.10	.32		.73	.16	.11	
7	.33	.04	.63		.71	.10	.18	
8	.13	.01	.86		.64	.06	.29	
9	.04		.95		.53	.03	.43	
10	.01		.99		.40	.02	.58	

1 Same comments as Table XXX.

TABLE XXXII

PROBABILITIES OF LABOUR FORCE PARTICIPATION AS A FUNCTION OF THE LEVEL OF ASSET INCOME[1]

	SIMULTANEOUS MODEL			SEQUENTIAL MODEL		
Family Asset Income ($)	B.W.	H.O.	N.	B.W.	H.O.	N.
0	.55	.43	.01	.52	.47	.01
250	.55	.44	.01	.52	.47	.01
500	.54	.44	.01	.51	.47	.01
1000	.54	.45	.01	.50	.48	.01
1500	.53	.46	.01	.49	.49	.01
2000	.52	.47	.01	.49	.50	.01
5000	.47	.52	.01	.44	.55	.01
10000	.38	.61	.01	.35	.63	.02

1 The probabilities for the "wife only" case were always very small (less than .005). Because of rounding errors the rows do not always sum to one.

* B.W. = "both working"; H.O. = "husband only"; N. = "none".

3.2 Unconditional Supply Functions

I define first (2.6) as a function of the husband's wage rate. The husband is in the labour force if the family chooses either the "both working" or the "husband only" alternative. The probabilities for these two alternatives as a function of the wage rate are given in Table XXX. From the conditional labour supply functions in Table XXV (Columns 1 and 3) I calculate the expected amount supplied by the husband for each alternative given his wage rate. These expected amounts are given in Table XXXIII (Columns 1 and 2). It can be seen that at the same wage rate, husbands in the "both working" alternative supply on an average less hours to the labour market than the husbands in the "husband only" alternative.

TABLE XXXIII

CONDITIONAL AND UNCONDITIONAL LABOUR SUPPLY FUNCTIONS FOR THE HUSBAND[1]

Husband's Hourly ($) Wage Rate	CONDITIONAL SUPPLY (Yearly Hours)		UNCONDITIONAL SUPPLY (Yearly Hours)	
	B.W.	H.O.	SIM	SEQ
0	1863	1299	1334	926
1	1848	1758	1632	1411
2	1812	2114	1828	1751
3	1764	2316	1945	1938
4	1713	2370	1997	2016
5	1668	2324	2000	2027
6	1636	2239	1984	2008
7	1623	2173	1977	1997
8	1638	2180	2017	2035
9	1690	2317	2162	2180
10	1789	2677	2497	2520

1 B.W. = "both working"; H.O. = "husband only"; SIM = simultaneous model; SEQ = sequential model.

The unconditional labour supply functions for the husbands are calculated by multiplying the probabilities of "both working" and "husband only" (Table XXX) with the amount of hours in, respectively, Columns 1 and 2 of Table XXXIII. The unconditional function for the husband using the probabilities of the simultaneous model is given in Column 3 of this Table. Column 4 is derived using the probabilities of the sequential model. A remarkable feature of both unconditional functions for the husband is that they remain relatively constant (around 2,000 hours a year) as the wage rate goes from three to eight dollars. This constancy is caused by changes in the wage rate gradually shifting the probability weights from the "both working" supply parameters to the supply characteristics of the "husband only" case. This shift takes place because of the substitution effect in labour force participation (i.e., as one partner's wage goes up the other partner is less probable to be in the labour force).

Table XXXIV presents the unconditional labour supply function for wives. In Columns 1 and 2 I present the conditional supply functions of the wives in the "both working" and "wife only" alternative. The conditional supply function for the wives in the "wife only" case is already at 6,000 hours a year at the five dollar wage and increases even further thereafter. This result follows from the linear supply curve (Equation (4.3), Chapter V) that was estimated for the 81 wives with non-working husbands. Therefore I calculate the supply functions only up to five dollars. The last two columns show the unconditional supply function calculated from the probabilities of respectively, the simultaneous and the sequential model.

Up to five dollars the unconditional supply function is almost uniquely determined by the probability of choosing "both working" and by the conditional supply function given this coice. As the wife's

TABLE XXXIV

CONDITIONAL AND UNCONDITIONAL LABOUR SUPPLY FUNCTIONS FOR THE WIFE[1]

Wife's Hourly ($) Wage Rate	CONDITIONAL SUPPLY B.W.	W.O.	UNCONDITIONAL SUPPLY SIM	SEQ
0	1355	1090	443	403
1	1435	1528	607	554
2	1521	2142	801	736
3	1613	3003	1023	951
4	1710	4209	1320	1221
5	1813	5901	1946	1614

1 B.W. = "both working"; W.O. = "wife only"; SIM = simultaneous model; SEQ = sequential model.

labour force participation decision is very responsive to changes in her wage rate, the resulting unconditional labour supply function is very wage-elastic. This is in sharp contrast with the husband's unconditional supply function. The parameters of participation and supply relating to the "wife only" case will only have an impact on the unconditional supply function of the wife for wages above seven or eight dollars.

I also calculate unconditional "hours supplied - income" curves for the husbands and wives (Table XXV). The calculation procedure is the same as for unconditional supply function. Because neither the participation nor the supply functions are very sensible to changes in asset income I only contrast the value of the conditional

TABLE XXXV

CONDITIONAL AND UNCONDITIONAL "INCOME-HOURS WORKED FUNCTIONS[1]

Asset Income $	CONDITIONAL FUNCTIONS (Yearly Hours) B.W.	H.O.	W.O.	UNCONDITIONAL FUNCTIONS (Yearly Hours) SIM	SEQ
Husband:					
1,000	629	2393		1423	1472
10,000	709	2594		1856	1884
Wife:					
1,000	1548		1101	838	784
10,000	1268		361	481	449

1 Same comments as Tables XXXIII and XXXIV.

and unconditional "hours supplied-income" curves at $1,000 and $10,000. The combined outcome of changes at the internal and external margin due to changes in income shows that increasing income will encourage the families to withdraw the wife from the labour force and furthermore to reduce the hours supplied of those wives that remain in the labour force. Exactly the opposite is true for the husbands. In conclusion, it is perhaps appropriate to illustrate the use of unconditional supply functions in policy applications. I will do this by discussing possible changes in the families' labour market behaviour which could be caused by the Equal Rights Amendment. Suppose this amendment increases the wage rates that wives can obtain in the labour market. Using the results of this Chapter one would predict a shift in labour force participation choice from "husband only" to "both working" alternative (Table XXXII). This implies that a number of wives will join the labour force. This also implies that the husband might reduce his labour supply (compare Columns 1 and 2 in Table XXXIII). The combined result could be a substantial substitution of the wife's for the husband's labour supply within the family.

CHAPTER VII

CONCLUSIONS

The main objective of this study has been the empirical estimation of family labour force participation decision functions. The appropriate estimation procedure for a model involving choice among multiple discrete alternatives requires a statistical technique different from ordinary least squares. In this study I have used the binomial and multinomial logit model to estimate parameters affecting the probabilities of choosing a particular labour force participation alternative. While some previous studies have employed the binomial model, this study is the first to use the multinomial model in the context of family labour force participation decisions.

Although the main focus of this study has been empirical, I have also discussed and derived the binomial and multinomial logit models in the framework of static utility maximization theory. I have discussed only those theoretical aspects of discrete choice behaviour that were useful for interpretation in this essentially empirical study. However, further study of the theoretical properties of decisions involving choice among <u>discrete</u> alternatives is clearly desirable. Unfortunately, most of the theoretical literature focusses on <u>continuous</u> choices at the internal margin.

A theoretical contribution of this thesis to the econometric literature has been the development of a procedure which, in the context of the multinomial logit model, allows one to test whether decision taking is sequential or simultaneous. For example, this procedure was used to test whether the family chooses simultaneously among four possible participation alternatives or whether one partner decides first about participation and the other partner decides conditionally upon the first. Using a Bayesian discrimination technique it was found that the simultaneous decision model was more probable *a posteriori* than the sequential model. It would appear that this procedure for testing decision behaviour in discrete choice models could be further extended. It would then also be desirable to create computer software programs that would allow for direct testing of the theoretical restrictions in this context.

In the empirical portion of this thesis I devoted considerable attention to the problem of predicting the potential wage rate for men and women who are not observed in the labour force. Although this wage variable is extremely important for theoretical reasons, problems involving actual estimation of a wage predictor, have been largely neglected in previous studies. The parameter estimates used in this study to approximate the unobserved wage rate were suspected of being biased (because of missing variables and structural difference problems). Whether the predictors have any desirable statistical properties (e.g., minimum mean squared error) is unknown and is an appropriate subject for further research. It is somewhat surprising that the use of predicted wage rates based on biased estimates yielded such significant results.

A substantial portion of the empirical research in this study has been the estimation and comparison of parameters of family labour force participation and labour supply decisions. I have attempted to discriminate statistically between the hypothesis that the parameters of supply and participation are either the same or different and found that the hypothesis of different parameters is more probable, a posteriori. In addition, the comparison of the parameters of family labour supply and labour force participation decisions has lead to interesting results, e.g., the substitution effect on both participation and supply behaviour of husband and wife. Another use of the estimated labour supply and labour force participation functions involved combining them to form unconditional labour supply functions. It was indicated that unconditional labour supply functions could be useful to evaluate the combined effect on supply and participation of a labour market policy.

The use of the binomial and multinomial logit models in the empirical estimation of family participation functions has been relatively successful. The binomial logit model enabled me, while using a more appropriate statistical technique, to confirm and expand on results found previously in the area of labour force participation for married women. The satisfactory results obtained with the multinomial logit model in this study would seem to recommend its use in other labour supply problems.

One possible application concerns the study of a different set of labour supply alternatives, i.e., choice between a part-time job, a standard work week job, overtime, second job, etc. Because of institutional constraints in the labour market such a discrete choice model might be more appropriate than the traditional model of labour supply.

Another area of research would involve further exploration of the intertemporal aspect of the available panel data. Instead of assuming that the choice probabilities of all observations are independent, as was done in this study, one could assume that the five annual observations (1967-71) on the same family are temporarily related. McFadden [1974] has suggested such a variant of the logit model.

Although numerous empirical applications of the logit model could be suggested, the development of theoretical models of discrete choice behaviour to guide these empirical studies would seem to be a most urgent item for research.

BIBLIOGRAPHY

ADAMS, F.G. [1958], "The Size of Individual Incomes: Socio-Economic Variables and Chance Variation", *Review of Economics and Statistics*, 40, 390-398.

ASHENFELTER, O. and HECKMAN, J. [1973], "Estimating Labor Supply Functions", Chapter 7 of Cain and Watts [1973].

ASHENFELTER, O. and HECKMAN, J. [1974], "The Estimation of Income and Substitution Effects in a Model of Family Labour Supply", *Econometrica*, 42, 73-86.

BARTH, P.S. [1968], "Unemployment and Labour Force Participation", *Southern Economic Journal*, 34, 375-382.

BECKER, G.S. [1964], *Human Capital: A Theoretical and Empirical Analysis With Special Reference to Education*, New York: National Bureau of Economic Research.

BECKER, G.S. [1965], "A Theory of Allocation of Time", *Economic Journal*, 75, 493-517.

BEN-PORATH, Y. [1967], "The Production of Human Capital and The Life Cycle of Earnings", *Journal of Political Economy*, 75, 352-365.

BEN-PORATH, Y. [1970], "The Production of Human Capital over Time", *Education, Income and Human Capital*, edited by W.L. Hansen, 129-146. New York: National Bureau of Economic Research.

BEN PORATH, Y. [1973], "Labor Force Participation Rates and the Supply of Labor", *Journal of Political Economy*, 81, 697-704.

BERNDT, E. and WALES, T. [1974a], "Determinants of the Wage Rate for Married Women", Discussion Paper 74-05, Department of Economics, University of British Columbia.

BERNDT, E. and WALES, T. [1974b], "Determinants of Labour Force Participation and Timing of Births for Married Women, 1967-1971", Discussion Paper 74-27, Department of Economics, University of British Columbia.

BLOCK, M. and HEINEKE, J. [1973], "The Allocation of Effort Under Uncertainty: The Case of Risk-Averse Behaviour", *Journal of Political Economy*, 81, 376-385.

BOSKIN, M. [1967], "The Negative Income Tax and the Supply of Work Effort", *National Tax Journal*, 20, 353-367.

BOSKIN, M. [1973], "The Economics of Labor Supply", Chapter 4 in Cain and Watts [1973].

BOX, G. and TIDWELL, P. [1962], "Transformations of Independent Variables", *Technometrics*, 4, 531-550.

BOWEN, W. and FINEGAN, T.A. [1969], *The Economics of Labor Force Participation*, Princeton: Princeton University Press.

BREAK, G. [1957], "Income Taxes and Incentives to Work, An Empirical Study", *American Economic Review*, 47, 529-549.

BRONFENBRENNER, M. and MOSSIN, J. [1967], "The Shorter Work Week and the Labour Supply", *Southern Economic Journal*, 33, 322-331.

BUSE, A. [1972], *A Technical Report on Binary Dependent Variables as Applied in the Social Sciences*, Edmonton: Human Resource Council.

CAIN, G. [1966], *Married Women in the Labor Force: An Economic Analysis*, Chicago: University of Chicago Press.

CAIN, G. and WATTS, H. (Editors) [1973], *Income Maintenance and Labour Supply, Econometric Studies*, Institute for Research on Poverty Monograph Series. Chicago: Markham Publishing Company.

CHOW, G. [1960], "Test of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 591-605.

CRAGG, J. [1971], "Some Statistical Models for Limited Dependent Variables With Applications to the Demand for Durable Goods", *Econometrica*, 39, 829-844.

CRAGG, J. [1973], *Wage Changes and Labour Flows in Canada*, Ottawa: Prices and Incomes Commission.

CRAGG, J. and BAXTER, N. [1970], "Corporate Choice Among Long Term Financing Instruments", *The Review of Economics and Statistics*, 52, 225-235.

CRAGG, J. and UHLER, R. [1970], "The Demand for Automobiles", *Canadian Journal of Economics*, 3, 386-405.

CRAGG, J. and UHLER, R. [1971], "The Structure of the Asset Portfolios of Households", *The Review of Economic Studies*, 38, 341-357.

COHEN, M., REA, S. Jr., and LERMAN, R. [1970], "A Micro Model of Labor Supply", Bureau of Labor Statistics, Staff Paper 4, U.S. Department of Labor, Washington, D.C.: U.S. Government Printing Office.

DEBREU, G. [1960], "Review of R.D. Luce, Individual Choice Behavior: A Theoretical Analysis", *American Economic Review*, 50, 186-188.

DIEWERT, W.E. [1971], "Choice on Labour Markets and the Theory of Allocation of Time", Research Branch, Department of Manpower and Immigration, Ottawa.

DOUGLAS, P.H. [1934], *The Theory of Wages*, New York: Macmillan Publishing.

DUNCAN, O., HATT, P., NORTH, C. and REISS, A. [1961], *Occupation and Social Status*, New York: Free Press Publishing.

FAIR, R. [1971], "Labour Force Participation, Wage Rates and Money Illusion", *Review of Economics and Statistics*, 53, 164-168.

FISHER, F. [1970], "Test of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note", *Econometrica*, 38, 361-366.

FRIEDMAN, M. [1962], *Price Theory, A Provisional Text*, Chicago: Aldine Publishing.

GARFINKEL, I. [1973], "On the Estimation of Labour Supply Effects of A Negative Income Tax", Chapter 6 in Cain and Watts [1973].

GOLDBERGER, A. [1964], *Econometric Theory*, New York: Wiley and Sons, Inc.

GOLDBERGER, A. [1967], "Functional Form and Utility: A Review of Consumer Demand Theory", Workshop Paper 6703. Social Systems Research Institute, University of Wisconsin.

GREEN, C. and TELLA, A. [1969], "Effect of Nonemployment Income and Wage Rates on the Work Incentive of the Poor", *Review of Economics and Statistics*, 51, 399-408.

GREEN, J. [1971], *Consumer Theory*, Harmondsworth: Penguin Books Limited.

GREENBERG, D. and KOSTERS, M. [1973], "Income Guarantee and the Working Poor: The Effect of Income-Maintenance Programs on the Hours of Work of Male Family Heads", Chapter 2 in Cain and Watts [1973].

GRILICHES, Z. and MASON, W. [1972], "Education, Income and Ability", *Journal of Political Economy*, 80, S74-S103.

GRONAU, R. [1973], "The Effect of Children on the Housewife's Value of Time", *Journal of Political Economy*, 81, S168-S199.

GUSSMAN, T. [1972], "The Demand for Durables, Nondurables, Services, and Supply of Labour in Canada", Research Branch, Department of Manpower and Immigration. Ottawa.

HALL, R. [1973], "Wages, Income and Hours of Work in the U.S. Labor Force", Chapter 3 in Cain and Watts [1973].

HALL, R. and KASTEN, R. [1974], "Occupational Mobility and the Distribution of Income Among Young Men", mimeograph, Massachusetts Institute of Technology.

HECKMAN, J. [1972], "Estimates of A Human Capital Production Function Embedded in A Life Cycle Model of Labour Supply", mimeograph, New York: National Bureau of Economic Research.

HECKMAN, J. [1974], "Shadow Prices, Market Wages and Labour Supply", *Econometrica*, 42, 679-694.

HENDERSON, J. and QUANDT, R. [1971], *Micro-Economic Theory, A Mathematical Approach*, New York: McGraw-Hill Publishing.

HICKS, J. [1939], *Value and Capital*, Oxford: Clarendon Press.

HILL, E. [1973], "The Determinants of Labour Supply for the Working Urban Poor", Chapter 5 in Cain and Watts [1973].

HILL, T. [1959], "An Analysis of the Distribution of Wages and Salaries in Great Britain", *Econometrica*, 27, 355-381.

JOHNSON, G. and STAFFORD, F. [1974], "Lifetime Earnings in a Professional Labour Market: Academic Economists", *Journal of Political Economy*, 82, 549-570.

JOHNSTON, J. [1972], *Econometric Methods*, New York: McGraw-Hill Publishing

KALACHEK, E. and RAINES, F. [1970], "Labour Supply of Low Income Workers", Technical Studies, President's Commission on Income Maintenance Programs, Washington, D.C.: Government Printing Office.

KENDALL, M. and STUART, A. [1963], *The Advanced Theory of Statistics*, London: C. Griffin Publishing.

KOSTERS, M. [1963], "Effects of An Income Tax on Labour Supply", *The Taxation of Income From Capital*, edited by A.C. Harberger and M.J. Bailey, Washington, D.C., The Brookings Institution.

LADD, G.W. [1966], "Linear Probability Functions and Discriminant Functions", *Econometrica*, 34, 873-885.

LLUCH, C. and MORISHIMA, M. [1973], "Demand for Commodities Under Uncertain Expections", *Theory of Demand, Real and Monetary*, Oxford: Clarendon Press.

LUCE, D. [1959], *Individual Choice Behaviour, A Theoretical Analysis*, New York: Wiley and Sons, Inc.

LUCE, D. and SUPPES, P. [1965], "Preference, Utility and Subjective Probability" Chapter 19 in *Handbook of Mathematical Psychology*, edited by R.D. Luce, R.E. Busch, and E. Galanter. New York: Wiley and Sons, Inc.

MCFADDEN, D. [1974], "Conditional Logit Analysis of Qualitative Choice Behaviour", *Frontiers of Econometrics*, edited by P. Zarembka. New York: Academic Press.

MAHONEY, T.A. [1961], "Factors Determining Labor Force Participation of Married Women", *Industrial and Labor Relations Review*, 14, 563-577.

MINCER, J. [1962], "Labor Force Participation of Married Women", *Aspects of Labor Economics*, 63-106. Princeton: A Conference of the National Bureau of Economic Research.

MINCER, J. [1966], "Labor Force Participation and Unemployment: A Review of Recent Evidence", *Prosperity and Unemployment*, edited by R.A. Gordon and M.S. Gordon, 73-112. New York: Wiley and Sons, Inc.

MINCER, J. [1970], "The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach", *Journal of Economic Literature*, 8, 1-26.

MINCER, J. [1974], *Schooling, Experience, and Earnings*, New York: National Bureau of Economic Research.

MOSES, L.N. [1962], "Income, Leisure, and Wage Pressure", *Economic Journal*, 72, 320-334.

OFFICER, L. and ANDERSON, P. [1969], "Labour Force Participation in Canada", *Canadian Journal of Economics*, 2, 278-287.

A PANEL STUDY OF INCOME DYNAMICS [1972], Volumes I and II, Survey Research Center, Institute for Social Research, University of Michigan.

PERLMAN, R. [1966], "Observations on Overtime and Moonlighting", *Southern Economic Journal*, 33, 237-244.

POLLAK, R.A. and WACHTER, M.L. [1975], "The Relevance of the Household Production Function and its Implications for the Allocation of Time", *Journal of Political Economy*, 83, 255-278.

PRESS, J. [1972], *Applied Multivariate Analysis*, New York: Holt, Rinehart and Winston, Inc.

RAIFFA, H. and SCHLAIFER, R. [1961], *Applied Statistical Decision Theory*, Boston: Division of Research, Graduate School of Business Administration, Harvard University.

ROBBINS, L. [1930], "On The Elasticity of Demand for Income in Terms of Effort", *Economica*, 10, 123-129.

ROSEN, S. [1972], "Training and Experience in the Labor Markets", *Journal of Human Resources*, 7, 326-342.

ROSEN, S. and WELCH, F. [1971], "Labour Supply and Income Redistribution", *Review of Economics and Statistics*, 53, 278-282.

ROSETT, R. [1958], "Working Wives: An Econometric Study", *Studies in Household Behavior*, edited by T. Derberg, et al. New Haven: Yale University Press.

SAMUELSON, P.A. [1956], "Social Indifference Curves", *Quarterly Journal of Economics*, 70, 1-22.

SMITH, J. [1972], "The Life Cycle Allocation of Time in A Family Context", Ph.D. Dissertation, University of Chicago.

SWEET, J.A. [1973], *Women in the Labor Force*, New York: Seminar Press.

TAUBMAN, P.J. and WALES, T.J. [1973], "Higher Education, Mental Ability and Screening", *Journal of Political Economy*, 11, 28-55.

THEIL, H. [1961], *Economic Forecasts and Policy*, (second edition) Amsterdam: North-Holland Publishing.

THEIL, H. [1969], "A Multinomial Extension of the Linear Logit Model", *International Economic Review*, 10, 251-259.

THEIL, H. [1970], "On The Estimation of Relationships Involving Qualitative Variables", *American Journal of Sociology*, 76, 103-154.

THEIL, H. [1971], *Principles of Econometrics*, New York: Wiley and Sons, Inc.

TINTNER, G. [1938a], "The Maximization of Utility over Time", *Econometrica*, 6, 154-158.

TINTNER, G. [1938b], "Theoretical Derivation of Dynamic Demand Curves" *Econometrica*, 6, 375-380.

TINTNER, G. [1939], "Elasticities of Expenditure in the Dynamic Theory of Demand", *Econometrica*, 7, 266-270.

TOBIN, J. [1958], "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24-36.

UHLER, R.S. and KUNIN, R. [1972], "A Theory of Labour Force Participation", *Industrial Relations*, 11, 107-114.

WACHTER, M.L. [1972], "A Labour Supply Model for Secondary Workers", *Review of Economics and Statistics*, 54, 141-151.

WACHTER, M.L. [1974], "A New Approach to the Equilibrium Labor Force", *Economica*, 41, 35-51.

WALES, T.J. [1973], "Estimation of a Labour Supply Curve for Self-Employed Business Proprietors", *International Economic Review*, 14, 69-80.

WALES, T.J. and WOODLAND, A. [1974a], "Estimation of Household Utility Functions and Labour Supply Responses", Discussion Paper 74-07, Department of Economics, University of British Columbia.

WALES, T.J. and WOODLAND, A. [1974b], "Estimation of the Allocation of Time for Work, Leisure and Housework", Discussion Paper 74-19, Department of Economics, University of British Columbia.

WEISS, Y. [1972], "On the Optimal Lifetime Pattern of Labour Supply", *Economic Journal*, 82, 1293-1315.

APPENDIX A

DESCRIPTION OF THE ORIGINAL SAMPLE

The five year panel study, from which all the samples used in this thesis are drawn, was conducted by the University of Michigan Survey Research Center [1972]. This panel study employed personal interviews with heads of household as its major data-collecting technique. The sample also included single person household (roughly one third of the final sample). In the interview, demographic, employment, and income questions were asked usually for all family members; attitudinal, behavioural, and expectations questions were asked solely of the husband. These household data are then supplemented with area data gathered from the respondent's current county, state, or Standard Metropolitan Statistical Area.

The actual questionnaire begins by establishing the composition of the family. Then follows a large set of questions which can be grouped into nine sections: education, transportation, housing, employment of the head, housework and work for money by wife, income, sentence completion questions (intelligence test), feelings and leisure time use. The employment questions of the head were somewhat different depending on whether the head was employed, unemployed, or retired at the time of the interview.

The raw data obtained through the interviews were then edited. This was done not only to check on the numerical accuracy of certain questions or to obtain year to year consistency in the individual family data, but also to generate additional information using the raw interview data. This latter kind of editing was, for instance, frequently used to assign car, house, and mortgage values based on related questions such as age of the car or of the house, location of the house, etc. Editing was also important in the case of money income variables, e.g., assignment of the asset part and labour part of business income; calculation of the family taxes, etc. The coding of the attitudinal variables (e.g., risk avoidance, achievement motivation, intelligence quotient) was also part of the editing process.

The interviews took place for the first time in 1968 with the employment and income questions pertaining to 1967. The 1968 sample consisted of 4802 households. Some 1872 were chosen from a previous Survey of Economic Opportunity sample of households with income less or equal to the United States federal poverty line. Added to this were 2930 households from a cross-section sample of dwellings in the United States carefully chosen by the Survey Research Center so as to obtain an overall representative cross-section.

In the spring of the four following years the heads were re-interviewed, with 82.3 percent of the 1968 sample still present in the 1972 sample. If during the interviewing period 1968 to 1972, new families were formed containing a 1968 panel family member, then those new households (called "split-offs") were added to the sample insofar as possible. The number of families in the final sample after the 1972 interviewing wave is 5060, consisting of 3.972 original families interviewed all five years plus 1108 split-offs.[1]

In designing the interviews in each year a premium was put on year-to-year producibility of the important variables of the study.

This has been largely successful although some variables, or refinements thereof, exist only for certain years (usually sampled in 1969 to 1972 only).

Compared with the 1970 United States Census data, the sample is a representative subsample of United States families.[2] For a more detailed description of the sample and the sampling procedures one should consult "A Panel Study in Income Dynamics", Volume I and II [1972]. In Appendix B I define in more detail the variables which I have selected from this sample.

---

1 In reality there are 5062 units on the tape.

2 See "A Panel Study of Income Dynamics", Volume I [1972, 31-32].

APPENDIX B

DEFINITIONS OF VARIABLES

Comments

(i) A variable name ending with H refers to the husband. Similarly a W in last position refers to the wife. This convention will only be used if the variable exists for both husband and wife and if it is not clear from the context to whom it refers.

(ii) Wage rates and income variables are deflated using the Consumer Price Index (drawn from the United States Department of Labor, Handbook of Labor Statistics, Table 121 [1972]). These indexes are respectively: 1.0, 1.042, 1.098, 1.163, 1.213 for 1967 to 1971.

Variables

ACHIEVE	Score value (an integer between [0 - 16][1] on an index measuring achievement motivation (a higher score value indicates increasing motivation). The test consists of a set of 16 questions and is discussed in "A Panel Study of Income Dynamics" [1972]. It was administered to the husband for 1971 only.
AGE	Age of the individual.
AGESQ	Previous variable squared.
AGE BA	Equal to AGE*BA dummy.
AGESQ BA	Equal to AGESQ*BA dummy.
AGE PHD	Equal to AGE*PHD dummy
AGESQ PHD	Equal to AGESQ*PHD dummy.

ASSET Y The sum of taxable asset income plus 12 percent of the value of the family car(s) plus 6 percent of the value of the family house(s), all in real terms. Taxable asset income is defined as the sum of the asset part of farm, business, and other income (assigned amounts) plus the amount of rent, interest and dividends received. The value of the house is corrected for the outstanding mortgage debt.

BA Dummy = 1 if individual has a college degree.

BIRTH Dummy = 1 if the family had a child born in that year.

CATH Dummy = 1 if husband's religious preference is Catholic.

CLERK Dummy = 1 for clerical and sales workers.

COLL Dummy = 1 if individual attended college, but did not obtain a college degree.

EXP Dummy = 1 if the husband has been on his present job for four years or more.

GRADE 9/12 Dummy = 1 if individual completed [9 - 12] grades.

GRADE 12 Dummy = 1 if individual completed 12 grades.

IQ Score value (an integer between [0 - 13]) on a sentence completion test. As discussed in "A Panel Study of Income Dynamics" [1972], this purports to measure intelligence (a higher score value indicates a higher IQ). The test was administered to the husband in 1971 only.

JEW Dummy = 1 if husband's religious preference is Jewish.

KID 6 Dummy = 1 if the family has at least one child less than seven years old.

KIDS	Number of children of the husband.
KID Y	Dummy = 1 if children had any (labour, transfer, or asset) income.
LIMIT	Dummy = 1 if individual has a physical or nervous condition that limits the type or amount of work that he/she can do.
MANAG	Dummy = 1 for managers, officials, and proprietors.
MARRIAGE	Present age of husband minus age of husband at the time of his first marriage. Sample was restricted to contain only those husbands which married only once.
MORTG	Dummy = 1 if the family has a positive mortgage debt.
PHD	Dummy = 1 if individual has an advanced or professional college degree (i.e., beyond BA).
PROF	Dummy = 1 for professional, technical, and kindred workers.
RACE	Dummy = 1 if non-white.
RESERVE	Dummy = 1 if the family currently has savings equal or greater than two months' income.
RISKAVOID	Score value (an integer between [0 - 9]) of points given for having insurance, using seatbelts, having liquid savings, and being a non-smoker. This index supposedly indicates if the husband (or the family) is a riskavoider. (A higher score value indicates more riskavoiding behaviour).
SECOND JOB	Dummy = 1 if husband is a multiple job holder.
SEH	Dummy = 1 if the husband was self-employed.
SEMISKILL	Dummy = 1 for operatives and kindred workers.
SKILL	Dummy = 1 for craftsmen, foremen, and kindred workers.
SOUTH	Dummy = 1 if family lives in the southern region of the United States.

TECH Dummy = 1 if individual completed 12 grades plus some non-academic training.

UNION Dummy = 1 if husband belongs to a labour union.

UNSKILL Dummy = 1 for laborers and service workers, farm laborers.

URBAN Dummy = 1 if the family's address is inside the city limits of a city (population of 5,000 or more).

WAGE Annual wage income divided by annual hours of work. This average nominal wage rate is then deflated by the consumer price index.

WAGE NET The WAGE variable corrected for by the marginal federal income tax rate of the family (this procedure is described in Chapter V).

WAGE (PRED) This is a predicted real wage variable. The prediction method is described in Chapter III.

WINDFALL Y Dummy = 1 if the family has a windfall income of at least $500, e.g., a big settlement from an insurance company or an inheritance.

4-9 YRS ON JOB Dummy = 1 if husband is over 3.5 to 9.5 years on his present job.

10-19 YRS ON JOB Dummy = 1 if husband is over 9.5 to 19.5 years on his present job.

20 YRS ON JOB Dummy = 1 if the husband is over 19.5 years on his present job.

YRS SCHOOL Continuous education variable. The coding for this variable is as follows:

0 - 5 grades, assign 3 years of schooling

6 - 8 grades, assign 7 years of schooling

9 - 11 grades, assign 10 years of schooling

12 grades, assign 12 years of schooling

12 grades and technical training, assign 13 years of schooling

college, no degree, assign 14 years of schooling

college, no advanced degree, assign 16 years of schooling

college, advanced degree, assign 18 years of schooling

---

1 "[,]" denotes a closed interval.