Probabilistic Modelling of the Joint Labour Decisions of Husband and Wife in Farm Households

Hild-Marte Bjørnsen

Norwegian Institute for Urban and Regional Research

P.O.Box 44 Blindern, N-0313 Oslo, Norway

Tel: +47 22 95 89 65

Fax: +47 22 60 77 74

Email: marte.bjornsen@nibr.no

Web: www.nibr.no

Abstract

When working with micro data one sometimes encounters situations involving qualitative choices in addition to the continuous choices that are the traditional focus of empirical analysis. One such situation arises when we study the increase in multiple job-holdings among farm operators in western economies. In these analyses we are both interested in the qualitative choice of entering the off-farm labour market and the continuous choice of determining labour supply in all occupations. This paper presents a unified framework for formulating such a discrete/continuous choices that include randomness within the decision-making process. We consider the joint labour decisions of operator and spouse through an agricultural household model that combines the agricultural production, consumption, and labour supply decisions in a single framework. We develop a probabilistic decision rule for participating in off-farm work and derive the demand functions in farm production and labour supply functions in off-farm sector.

Introduction

Most articles concerning labour decisions of farm households are based on neoclassical theory. The central assumption of the theory is that of optimisation and the individuals/households are assumed to maximize their utility, which is a function of consumption goods and leisure time, subject to constraints on time, income and farm production. For an individual to be able to rank all his alternative allocations there are certain key assumptions that must be satisfied to give the desired properties to an individual's preference ordering. First of all, we must assume that the individual has unlimited information of all possible alternative allocations. Secondly, we must assume that he is able to rank all alternatives in a consistent and well-defined order, however alike or unlike they may be. This includes the assumption of *transitivity* which states that if *a* is preferred to *b* and *b* is preferred to *c*, then *a* must be preferred to *c*, and the assumption of *reflexivity* which states that all alternatives are preferred or indifferent to itself. These assumptions allow us to partition any given allocation into non-intersecting indifference curves that can be illustrated in a two-dimensional diagram where consumption of consumer goods is represented along the x-axis and consumption of leisure along the y-axis[1]. To give these indifference curves a particular structure we usually add assumptions of *non-satiation* (alternative *a* is preferred to alternative *b* if *a* contains more of at least one good and no less of any other), *continuity of the indifference curves* (changes in consumption of purchased goods can be fully compensated by changes in consumption of leisure), and *strict convexity* (the set of alternatives that are preferred to alternative *a* is strictly convex).

The neoclassical approach can be criticized for not giving a very adequate description of human behaviour. People don't usually have perfect information of the alternatives they can choose among and the choices they make are not necessarily consistent with their previous behaviour. It is not difficult to picture a situation where we get problems with ranking the alternatives we face. Even the simplest comparisons may cause uncertainties; given that I am feeling a bit peckish, do I prefer a sandwich or would I be better off by spending my last money on a pint of lager? Often we find that the choice alternatives are much more complex than this example. It is also highly probable that our preferences change over time. They may

[1] We treat the vector of all purchased consumption goods as one argument of the utility function and the household's leisure as a second argument. The household's preferences can then be presented as indifference curves in a two-dimensional diagram.

even change within a few minutes depending on what new information or shock that occurs. I may go from peckish to hungry within a short while and actually go for the sandwich! A household's preferences for leisure may be dependent on household characteristics like number and age of children, with personal characteristics like age and education, and with local labour market characteristics. In this case there exists exogenous variables that affect the choices the household makes.

Some argue that choice behaviour should be modelled as a *probabilistic process* to account for the individuals' uncertainty and for the observed inconsistency (Tversky 1972). This approach implies that we associate probabilities with either of the alternatives. If we are going to introduce probabilistic processes in our model, we need to consider what factors determine the probabilities. Shall we focus on the inconsistencies in the individuals' behaviour or are the measurement problems and inappropriate model specifications of greater concern? We must also distinguish between models with stochastic decision rules (and deterministic utility) and models with stochastic utility (and deterministic decision rule), respectively. In the first case we have a situation where an individual don't necessarily choose the alternative that brings him the higher utility, while in the second case the individual will choose the alternative that he *expects* will bring him the higher utility The second alternative is thus closer to the neoclassical models.

We can think of several reasons why we may not be able to model the labour decisions of the multiple-job holding households correctly. A common, and more or less unavoidable, problem is that we can't fully observe all the factors that enter into the individuals' utility functions. We may i.e. have insufficient knowledge of preferences for farming and attachment to the farm property (maintenance of family traditions ect). This kind of heterogeneity in preferences will lead to greater variance in our estimates. Another source of uncertainty may stem from insufficient recordings of observable variables like farm characteristics, household/individual characteristics, and regional characteristics. A seemingly important explanatory variable when estimating off-farm labour supply that more often than not seems to be absent from the reaserchers' data sets, is the individuals level of education. The lack of such variables contributes to increase the unexplained part of the model. According to Manski (1977) the different sources of uncertainty can be listed in four groups: *Non-observable characteristics* (when the vector of characteristics affecting the choice of the individual is only partially known by the modeler), *non-observable variations in individual utilities* (when the variance in preferences increase with increasing preference heterogeneity), *measurement errors* (when the amount of the observable characteristics is not perfectly known) and *functional misspecifications* (when the functional form of utility is not known with certainty).

Theoretical Model

We view the decision to participate in off-farm wage work by farm operators (O) and their spouses (S) through an agricultural household model that combines agricultural production, consumption and labour supply decisions into a single framework. The model bear resemblance to models used by Huffman and Lange (1989), Gould and Saupe (1989), Lass and Gempesaw (1992), and Weersink, Nicholson and Weerhewa (1997).

The population of farm household face the same choice set of alternative occupations. We assume that they are *statistically identical* and *independently distributed*. This assumption implies that their choices are governed by the same mulitnominal distribution and that the choices made by one household are independent of the choices made by others. The actual time allocations are allowed to vary between households.

I was planning to choose a random utility approach to model the off-farm labour decisions based on McFadden's conditional logit model. My aim was to let the households maximize utility by choosing among four alternative allocations:

D_1	both operator and spouse participate in wage work:	$M^O > 0,\ \ M^S > 0$
D_2	only operator participates in wage work:	$M^O > 0,\ \ M^S = 0$
D_3	only spouse participates in wage work:	$M^O = 0,\ \ M^S > 0$
D_4	neither of them participates in wage work:	$M^O = 0,\ \ M^S = 0$

where M^i , $i = O, S$ represents hours in off-farm wage work for operator and spouse. The optimization problem includes a discrete choice (whether to participate in off-farm work or not) and a continuous choice of finding the optimal allocation of time between the available options; farm work, off-farm work and leisure.

My intention was to find a way of expressing the indirect utility (U_n^* , $n = 1,2,3,4$) of each alternative as a function of the attributes of the *n*th alternative, $V_n(X_n)$, and a random term,

ε_n, which should capture the unobserved variations in tastes and characteristics of the households and unobserved attributes of the alternative

$$U_n^* = V_n(X_n) + \varepsilon_n$$

while the discrete choice could be expressed as

$$D_n = \begin{cases} 1 & \text{if } \; U_n^* = \text{Max}(U_1^*, U_2^*, U_3^*, U_4^*) \quad n = 1,2,3,4 \\ 0 & \text{otherwise} \end{cases}.$$

This caused me some trouble as I also wanted the decision rule to be stochastic. I have struggled with different kinds of specifications and have eventually ended up with what follows, a deterministic utility function and a stochastic decision rule. This does not alter the essence of the discrete and continuous choices the households face. *(I will consider this problem more closely during summer and may end up with reconsidering my choice of model.)*

We assume that a household maximize its utility subject to constraints on time, income, and farm production. Utility can be derived from purchased goods (G) and the household members' leisure time (L^i), and is affected by variables such as human capital characteristics (H^i) and other household and regional characteristics (Z_H) that are considered exogenous to current consumption decisions.

(1) $$U = U(G, L^O, L^S; H^O, H^S, Z_H), \qquad \partial U/\partial G > 0, \partial U/\partial L^i > 0, \;\; i = O, S$$

The utility function is assumed to be ordinal and strictly concave and is maximized subject to constraints on time, income and farm productivity. We let 24 hours be the total time endowment (T) in this model and assume that time for operator and spouse are heterogeneous. Time can be spent on leisure (home work, sleep, leisure time ect.) denoted L^i, farm work (F^i), and wage work (M^i), all measured in number of hours. By definition all farm households supply a positive amount of on-farm labour hours and we assume that both operator and spouse work on the farm. We also assume that all individuals enjoy a strictly positive number of non-working hours. Time spent in off-farm sector is a non-negative variable for both operator and spouse.

(2) $T = F^i + M^i + L^i$, $F^i, L^i > 0$, $M^i \geq 0$, $i = O, S$

The consumption of market goods at price P_G will be limited by available income earned from farm profits, net income from wage work, and other household income, V. The farm households are assumed to be competitive in input and output markets and farm profit is set equal to the price of farm output (P_Q) multiplied with output (Q) less the variable cost of production (RX), where R is the input price vector and X is the quantity of purchased farm inputs. Off-farm work is paid at the wage rate W^i. For simplicity, we don't include taxes in the model. The budget constraint of a household can then be represented by the following equation.

(3) $P_G G = P_Q Q - RX + W^O M^O + W^S M^S + V$

In the following we will treat consumer goods (G) as numeraire and set P_G equal to one.

The wage rates the operator and spouse face are assumed to depend on their respective human capital characteristics (H^i) and the local labour market conditions (Z_M).

(4) $W^i = W^i(H^i, Z_M)$, $i = O, S$

We assume flexible work schedules in off farm employment so the household members can maximize utility by offering an optimal number of off farm work hours and that the wage rates are independent of the hours worked.

The properties of the farm production function represent a third and final constraint on the household's consumption abilities. Farm output is a function of the labour hours put down in farm production (F^i) and a vector of purchased farm inputs (X), and is dependent on human capital characteristics (H^i) and farm specific characteristics (Z_F). While off farm wages were assumed to be independent of hours worked, marginal returns to farm labour are assumed to be diminishing. The production function is assumed to be strictly concave.

(5) $$Q = f(F^O, F^S, X; H^O, H^S, Z_F)$$

The discrete choice of the maximization problem is to choose the alternative ($i = 1,2,3,4$) that gives the household the highest utility while the continuous choice is to find optimal labour supply in farm and off farm sector.

The discrete choices of the utility maximizing problem can be expressed as four binary choice variables

(6)
$$D_1 = \begin{cases} 1 & \text{if } W^O \geq W^{OR} \text{ and } W^S \geq W^{SR} \\ 0 & \text{otherwise} \end{cases}$$
$$D_2 = \begin{cases} 1 & \text{if } W^O \geq W^{OR} \text{ and } W^S < W^{SR} \\ 0 & \text{otherwise} \end{cases}$$
$$D_3 = \begin{cases} 1 & \text{if } W^O < W^{OR} \text{ and } W^S \geq W^{SR} \\ 0 & \text{otherwise} \end{cases}$$
$$D_4 = \begin{cases} 1 & \text{if } W^O < W^{OR} \text{ and } W^S < W^{SR} \\ 0 & \text{otherwise} \end{cases}$$

where W^{iR} is individual i's reservation wage. $i=O,S$.

These four alternatives are mutually exclusive. The household can only choose one of the above alternatives at a time. We can add another constraint to the maximization problem.

(7) $$D_n D_m = 0 \quad \text{for all } n \neq m, \quad n, m = 1,2,3,4$$

Let us first assume that the utility function is deterministic, i.e. the household has full information on all arguments that's included in the maximization problem. We get the optimal quantities of off-farm labour supply (M^i), the variables in household consumption (G and L^i), and the variable inputs in production (F^i and X) by maximizing utility (1) subject to constraints on time (2), budget (3) and farm productivity (5). The Lagrange function can be expressed like

$$
\text{(8)} \quad
\begin{aligned}
£ = {} & U(G, L^O, L^S; H^O, H^S, Z_H) + \sum_{i=1}^{2} \lambda_i \left(T - F^i - M^i - L^i\right) \\
& + \gamma \left\{ P_Q f(F^O, F^S, X; H^O, H^S, Z_F) - RX + W^O(H^O, Z_M) M^O \right. \\
& \left. + W^S(H^S, Z_M) M^S + V - G \right\}
\end{aligned}
$$

Which gives the following first order conditions for an interior solution

(9) $$\frac{\partial £}{\partial G} = U_G - \gamma = 0$$

(10) $$\frac{\partial £}{\partial L^i} = U_{L^i} - \lambda_i = 0$$

(11) $$\frac{\partial £}{\partial F^i} = -\lambda_i + \gamma P_Q f_i' = 0$$

(12) $$\frac{\partial £}{\partial M^i} = -\lambda_i + \gamma W^i \leq 0$$

(13) $$\frac{\partial £}{\partial X} = \gamma (P_Q f_X' - R) = 0$$

We find the optimal solution by simultaneously solving the first-order conditions. We will have an interior solution for all choice variables with a possible exception of off-farm work hours that may be zero for either operator, spouse, or both. Neither operator nor spouse will work off the farm unless the wage rate they receive exceeds the marginal value of farm labour evaluated at the point of optimal allocation between farm work and leisure when no off-farm hours are supplied.

Equation (13) states that the marginal value of use of purchased input factors shall equal the marginal cost of these factors in optimum. From equations (9) to (12) we get that, assuming interior solution, in optimum the marginal value of farm work must equal the marginal value of off-farm work and equal the marginal rate of substitution between leisure and consumption, i.e. marginal value of time is equal in all employments. Consumption goods are set as numeraire ($P_G = 1$).

(14) $$P_Q f_i' = W^i = \frac{U_{L^i}}{U_G} \qquad i=O,S$$

If $W^i < P_Q f_i^{'}$ the individual will not supply off-farm hours and time is divided between farm work and leisure. If $W^i > P_Q f_i^{'}$ the individual will increase his/her off-farm hours, thereby increasing the marginal value of farm work until the marginal returns to both kinds of employment equal the marginal rate of substitution between the consumption goods.

Given the assumption that the farm household acts as price taker in input and output markets, the household model is recursive if an interior solution exists for all choice variables. Decisions on farm labour and purchased inputs are made first, and then consumption decisions on consumption goods and leisure.

Equations (11)-(13) are then the conditions for profit maximizing usage of farm inputs and can be solved independently of the other equations to obtain demand functions for farm inputs:

(15) $F^{i*} = D_F(W^O, W^S, R, P_Q, H^O, H^S Z_M, Z_F) \quad i = O, S$

(16) $X^* = D_X(W^O, W^S, R, P_Q, H^O, H^S Z_M, Z_F)$

To obtain the demand functions for leisure time we need equations (2), (3), (5), (9), and (10) conditional on (11)-(13):

(17) $L^{i*} = D_L(W^O, W^S, H^O, H^S, Z_M, Z_{H,}, V, \boldsymbol{p}) \quad i = O, S$

where $\boldsymbol{p} = P_Q Q^* - RX^* - W^M F^{M*} - W^S F^{S*}$

The off-farm labour supply functions are derived residually from the time constraint and will contain all exogenous variables in the constrained optimisation problem.

(18)
$$\begin{aligned} M^{i*} &= T - F^{i*} - L^{i*} \\ &= D_M\left(W^O, W^S, P_Q, R, V, H^O, H^S, Z_H, Z_F, Z_M\right) \; i = O, S \end{aligned}$$

If however optimal hours of off-farm work are zero for either operator or spouse, household decisions regarding production and consumption decisions must be made jointly, rather than recursively. Off-farm labour supply will still be made residually, but the unobservable wage rate for an individual not working off the farm will not be a determinant of the hours work by the other partner. Still, the supply function for this partner is conditional upon the participation decision by the partner without off-farm employment.

The reservation wage for participation in off-farm work for individual *i* is equal to the marginal value of his/her time when all time is allocated between farm work and leisure ($T = F^i + L^i$). This gives us the participation rule

$$(19) \qquad y^i = \begin{cases} 1 & \text{if } W^i \geq W^{iR} = P_Q f_i'|_{M^i=0} \\ 0 & \text{if } W^i < W^{iR} = P_Q f_i'|_{M^i=0} \end{cases} \qquad i = O, S$$

We assume that the off-farm labour demand equations for the operator and spouse can be expressed as a linear function of its arguments

$$(20) \qquad W^i = b_{iH} H^i + b_{iZ} Z_M + v_i \quad \text{if } W^i > W^{iR}|_{M^i=0} \quad i = O, S$$

where v_i is some random disturbance. The wage rate is observed only when the decision to work off the farm is made which occurs if the market wage exceeds the reservation rate. As the reservation wage equals the marginal value of farm work when zero off-farm work hours is supplied it will depend on non-wage and non-price variables exogenous to the household consumption (Z_H), production (Z_F) and labour supply (Z_M) decisions represented by the aggregate Z.

The supply of off-farm labour depends on the exogenous variables Z, the price variables P (where P is a vector $P = (P_Q, R, V)$), and the wage rates as shown in equation (18). As mentioned, the unobservable wage rate of an individual who is not participating in off-farm work will not be a determinant of the spouse's supply function. The off-farm labour supply functions of the four alternatives can be expressed as;

$$(21) \qquad M^O = \begin{cases} M_1^O = a_{OO}^1 W^O + a_{OS}^1 W^S + a_{ZO}^1 Z + a_{PO}^1 P & \text{if } W^O \geq W^{OR}, W^S \geq W^{SR} \\ M_2^O = a_{OO}^2 W^O + a_{ZO}^2 Z + a_{PO}^2 P & \text{if } W^O \geq W^{OR}, W^S < W^{SR} \\ M_4^O = 0 & \text{otherwise} \end{cases}$$

$$
(22)\qquad M^S = \begin{cases} M_1^S = a_{SO}^1 W^O + a_{SS}^1 W^S + a_{ZS}^1 Z + a_{PS}^1 P & \text{if } W^O \geq W^{OR}, W^S \geq W^{SR} \\ M_3^S = a_{SS}^3 W^S + a_{ZS}^3 Z + a_{PS}^3 P & \text{if } W^O < W^{OR}, W^S \geq W^{SR} \\ M_4^S = 0 \quad \text{otherwise} & \end{cases}
$$

By inserting for W^i from (20) we get the labour supply equations as functions of only the exogenous variables, H^i, P_Q, R, V, and Z_k, where $k = H, F, M$. We simplify the equations by letting all exogenous variables be represented by an aggregate *Y*, and get:

$$
(21')\qquad M^O = \begin{cases} M_1^O = a^1 Y^1 + \boldsymbol{u}_1 & \text{if } W^O \geq W^{OR}, W^S \geq W^{SR} \\ M_2^O = a^2 Y^2 + \boldsymbol{u}_2 & \text{if } W^O \geq W^{OR}, W^S < W^{SR} \\ M_4^O = 0 & \text{otherwise} \end{cases}
$$

$$
(22')\qquad M^S = \begin{cases} M_1^S = a^4 Y^4 + \boldsymbol{u}_4 & \text{if } W^O \geq W^{OR}, W^S \geq W^{SR} \\ M_3^S = a^3 Y^3 + \boldsymbol{u}_3 & \text{if } W^O < W^{OR}, W^S \geq W^{SR} \\ M_4^S = 0 & \text{otherwise} \end{cases}
$$

where $\boldsymbol{u}_n$, *n={1,2,3,4}*, is a random disturbance that stems from the off-farm labour demand equations. We assume that $\boldsymbol{u}$ is logistically distributed with distribution function $\frac{1}{1+\exp(-Y)}$, zero mean and variance equal to $\boldsymbol{p}^2/3$.

Each individual chooses his/her course of action on the principle that the other's behaviour is given. The probability of participating in off-farm work will then be defined by four conditional distributions.

$$
(23)\qquad \begin{aligned} \Pr\left(y^O = 1 \mid y^S = 1\right) &= \frac{1}{1+\exp\left(a^1 Y^1\right)} \\ \Pr\left(y^O = 1 \mid y^S = 0\right) &= \frac{1}{1+\exp\left(a^2 Y^2\right)} \\ \Pr\left(y^S = 1 \mid y^O = 0\right) &= \frac{1}{1+\exp\left(a^3 Y^3\right)} \\ \Pr\left(y^S = 1 \mid y^O = 1\right) &= \frac{1}{1+\exp\left(a^4 Y^4\right)} \end{aligned}
$$

This kind of conditional behaviour is not necessarily compatible. In order to make these conditional distributions unambiguously define a course of action, the condition $a^1Y^1 + a^3Y^3 = a^2Y^2 + a^4Y^4$ must be satisfied (Gourieroux 2000:57).

When each variable can take on only two values (0,1) the conditional probabilities can be expressed as functions of the joint probabilities as follows:

$$\Pr\{y^o = 1 | y^s = 0\} = \frac{\Pr\{y^o = 1, y^s = 0\}}{\Pr\{y^o = 0, y^s = 0\} + \Pr\{y^o = 1, y^s = 0\}} = \frac{P_{10}}{P_{00} + P_{10}} = 1 - \Pr\{y^o = 0 | y^s = 0\}$$

$$\Pr\{y^o = 1 | y^s = 1\} = \frac{\Pr\{y^o = 1, y^s = 1\}}{\Pr\{y^o = 0, y^s = 1\} + \Pr\{y^o = 1, y^s = 1\}} = \frac{P_{11}}{P_{01} + P_{11}} = 1 - \Pr\{y^o = 0 | y^s = 1\}$$

$$\Pr\{y^s = 1 | y^o = 0\} = \frac{\Pr\{y^s = 1, y^o = 0\}}{\Pr\{y^s = 0, y^o = 0\} + \Pr\{y^s = 1, y^o = 0\}} = \frac{P_{01}}{P_{00} + P_{01}} = 1 - \Pr\{y^s = 0 | y^o = 0\}$$

$$\Pr\{y^s = 1 | y^o = 1\} = \frac{\Pr\{y^s = 1, y^o = 1\}}{\Pr\{y^s = 0, y^o = 1\} + \Pr\{y^s = 1, y^o = 1\}} = \frac{P_{11}}{P_{10} + P_{11}} = 1 - \Pr\{y^s = 0 | y^o = 1\}$$

To prove that $a^1Y^1 + a^3Y^3 = a^2Y^2 + a^4Y^4$ is a necessary condition we start with the lemma:

(*) $$\frac{\Pr\{y^o = 0 | y^s = 0\}\Pr\{y^s = 0 | y^o = 1\}}{\Pr\{y^o = 1 | y^s = 0\}\Pr\{y^s = 1 | y^o = 1\}} = \frac{\Pr\{y^o = 0 | y^s = 1\}\Pr\{y^s = 0 | y^o = 0\}}{\Pr\{y^o = 1 | y^s = 1\}\Pr\{y^s = 1 | y^o = 0\}}$$

This equality can be proven by replacing the conditional probabilities with the joint probabilities:

$$\frac{\dfrac{P_{00}}{P_{00} + P_{10}}\dfrac{P_{10}}{P_{10} + P_{11}}}{\dfrac{P_{10}}{P_{00} + P_{10}}\dfrac{P_{11}}{P_{10} + P_{11}}} = \frac{\dfrac{P_{01}}{P_{01} + P_{11}}\dfrac{P_{00}}{P_{00} + P_{01}}}{\dfrac{P_{11}}{P_{01} + P_{11}}\dfrac{P_{01}}{P_{00} + P_{01}}}$$

We can now easily see that both the left and the right hand side of the equality equals $\frac{P_{00}}{P_{11}} = \frac{\Pr\{y^o = 0, y^s = 0\}}{\Pr\{y^o = 1, y^s = 1\}}$.

By replacing the conditional probabilities in (*) with the conditional logit distribution in (23) we obtain:

$$\frac{\{1-\Pr(y^o=1|y^s=0)\}\{1-\Pr(y^s=1|y^o=1)\}}{\Pr(y^o=1|y^s=0)\Pr(y^s=1|y^o=1)}=\frac{\{1-\Pr(y^o=1|y^s=1)\}\{1-\Pr(y^s=1|y^o=0)\}}{\Pr(y^o=1|y^s=1)\Pr(y^s=1|y^o=0)}$$

(**)
$$\exp(a^2Y^2)\exp(a^4Y^4)=\exp(a^1Y^1)\exp(a^3Y^3)$$
$$a^2Y^2+a^4Y^4=a^1Y^1+a^3Y^3$$

Which is the compatability condition.

We can now utilize this condition to express the joint probabilities. We know that the left hand side of (**) equals $\frac{P_{00}}{P_{11}}$ which implies

$$P_{00}=\exp(a^2Y^2+a^4Y^4)P_{11}$$

and from the relations between the conditional and joint probabilities we get

$$P_{10}=\frac{1}{\exp(a^2Y^2)}P_{00}=\frac{\exp(a^2Y^2+a^4Y^4)}{\exp(a^2Y^2)}P_{11}=\exp(a^4Y^4)P_{11}$$

$$P_{01}=\exp(a^1Y^1)P_{11}$$

while P_{11} can be expressed as

$$P_{11}=1-(P_{00}+P_{10}+P_{01})$$

This gives us the following probabilities for the four alternatives the household are faced with:

Alternative D_1: both operator and spouse work off-farm:

(24)
$$\Pr(y^o=1,y^s=1)=\frac{1}{1+\exp(a^1Y^1)+\exp(a^4Y^4)[1+\exp(a^2Y^2)]}$$

Alternative D_2: only operator works off-farm:

$$(25)\qquad \Pr\{y^O = 1, y^S = 0\} = \frac{\exp(a^4Y^4)}{1+\exp(a^1Y^1)+\exp(a^4Y^4)\left[1+\exp(a^2Y^2)\right]}$$

Alternative D_3: only the spouse works off-farm:

$$(26)\qquad \Pr\{y^O = 0, y^S = 1\} = \frac{\exp(a^1Y^1)}{1+\exp(a^1Y^1)+\exp(a^4Y^4)\left[1+\exp(a^2Y^2)\right]}$$

Alternative D_4: neither operator nor spouse works off-farm:

$$(27)\qquad \Pr\{y^O = 0, y^S = 0\} = \frac{\exp(a^2Y^2+a^4Y^4)}{1+\exp(a^1Y^1)+\exp(a^4Y^4)\left[1+\exp(a^2Y^2)\right]}$$

Estimation

Let Y^j , $j=1,2,\ldots,J$, represent the J possible values of the vector of explanatory variables in (22') and (23'), and call a "trial of type j" a run of the experiment performed under the conditions $Y=Y^{nj}$, $(n=1,2,3,4)$. Let there be m_j such trials (Gourieroux 2000:20).

In the bivariate case as outlined above we have observations on the pairs

$$y_{ij} = \left(y^O_{ij}, y^S_{ij}\right) \ , \ j=1,\ldots,J \ , \ \ i=1,\ldots m_j$$

which assumes the values $\{k_1,k_2\}$, $k_1=0,1$, $k_2=0,1$ with probabilities (as given in (24)-(27):

$$(24')\qquad P_{11j} = \frac{1}{1+\exp(a^1Y^{1j})+\exp(a^4Y^{4j})\left[1+\exp(a^2Y^{2j})\right]} = F^{11}\left(a^1Y^{1j}, a^2Y^{2j}, a^4Y^{4j}\right)$$

$$(25')\qquad P_{10j} = \frac{\exp(a^4Y^{4j})}{1+\exp(a^1Y^{1j})+\exp(a^4Y^{4j})\left[1+\exp(a^2Y^{2j})\right]} = F^{10}\left(a^1Y^{1j}, a^2Y^{2j}, a^4Y^{4j}\right)$$

(26’) $$P_{01j} = \frac{\exp(a^1Y^{1j})}{1+\exp(a^1Y^{1j})+\exp(a^4Y^{4j})\left[1+\exp(a^2Y^{2j})\right]} = F^{01}\left(a^1Y^{1j}, a^2Y^{2j}, a^4Y^{4j}\right)$$

(27’) $$P_{ooj} = \frac{\exp(a^2Y^{2j}+a^4Y^{4j})}{1+\exp(a^1Y^{1j})+\exp(a^4Y^{4j})\left[1+\exp(a^2Y^{2j})\right]} = F^{00}\left(a^1Y^{1j}, a^2Y^{2j}, a^4Y^{4j}\right)$$

The probability distribution functions are all strictly positive, $F^{k_1k_2} > 0, \quad \forall k_1, k_2$ and constrained by

$$\sum_{k_1=0}^{1}\sum_{k_2=0}^{1} F^{k_1k_2} = 1$$

and the compatability constraint $a^1Y^1 + a^3Y^3 = a^2Y^2 + a^4Y^4$.

The log of the likelihood function is then given by

(28) $$\log\left[L(y;a)\right] = \sum_{j=1}^{J}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1} m_{k_1k_2j}\log[P_{k_1k_2j}(a)]$$

The maximum likelihood estimator is obtained by writing the first-order conditions

$$0 = \frac{\partial\log(L)}{\partial a^n}$$
$$0 = \sum_{j=1}^{J}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1} m_{k_1k_2j}\frac{F_n^{k_1k_2}(j)}{F^{k_1k_2}(j)}Y_j^n$$

where $F_n^{k_1k_2}(j)$ indicates the partial derivative of $F^{k_1k_2}(j)$ with respect to the n-th coordinate.

The maximum likelihood estimator α_{ML} which solves the equation system converges asymptotically $(J$ fixed, $m_j \to \infty,\ \forall j)$ to $\mathbf{a}$,

$$\mathbf{a} = \begin{pmatrix} a^1 \\ a^2 \\ a^3 \\ a^4 \end{pmatrix}$$

and is asymptotically normally distributed $\alpha_{ML} \xrightarrow{asy} N\left(\mathbf{a}; E\left[-\frac{\partial^2 \log(L)}{\partial a^n \partial a^{\lambda\prime}}\right]^{-1}\right)$.

We find the asymptotic covariance matrix of α_{ML} by taking the second derivatives of the log likelihood function

$$\frac{\partial^2 \log(L)}{\partial a^n \partial (a^\lambda)'} = \sum_{j=1}^{J}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1} m_{k_1k_2j} \frac{\partial}{\partial (a^\lambda)'}\left[\frac{F_n^{k_1k_2}(j)}{F^{k_1k_2}(j)} Y_j^n\right]$$
$$= \sum_{j=1}^{J}\sum_{k_1=0}^{1}\sum_{k_2=0}^{1} m_{k_1k_2j}\left[-\frac{F_n^{k_1k_2}(j)F_\lambda^{k_1k_2}(j)}{(F^{k_1k_2})^2(j)} + \frac{F_{n\lambda}^{k_1k_2}(j)}{F^{k_1k_2}(j)}\right](Y_j^n)' Y_j^\lambda$$

According to Gourieroux (2000:93) a much simpler estimation method for the conditional logit model is given as follows:

If we let $p_{k_1k_2j} = \dfrac{m_{k_1k_2j}}{m_j}$ be the observed frequencies for each trial of type j, a^1 we can be

estimated simply by calculating the conditional frequency $\dfrac{p_{11j}}{p_{01j} + p_{11j}}$ and applying the

Berkson's method. Then the observed frequencies $p_{k_1k_2j}$ asymptotically converges to the true probabilities $P_{k_1k_2j}(a)$. I will not explain the Berkson method in any further detail.

Data

The data for my analysis can mainly be obtained from a yearly survey of Norwegian farm households (Account Results in Agriculture and Forestry) collected by the Norwegian Agricultural Economics Research Institute (NILF). The survey is one of the more comprehensive sources of farm statistics in Norway, but has so far had little widespread use in applied research. The survey dates back to the beginning of the 20th century and has since 1950 included approximately 1000 farm households representing different regions and principal productions (grain, dairy, livestock etc.). Participation in the survey is voluntary but restricted to farmers younger than the age of 67 (retirement age) and to farm households working at least 400 on-farm hours on a yearly basis. Farms that produce both grain and swine products, and dairy farms (pure dairy farms or dairy in combination with livestock production) have the highest representation both in absolute numbers and as share of the total population. Most farm households in the survey report between 1800 and 6000 on-farm work hours yearly, while a standard man-labour year in the agricultural sector is set like 1875 hours. There is no specific decision rule used for entering new households, but one aims to enter farm households that hold more or less the same characteristics (region/size/production) as the exiting farms. Somewhere between five and ten percent of the farm households are replaced each year, most commonly because the exiting households don't wish to continue being part of the survey.

The Account Results in Agriculture and Forestry is the most elaborate source of information on Norwegian farm households financial matters both in a regional and a production type of context. The survey includes data on daily or weekly labour hours for all household members, family members, and hired help and in all employments. On-farm labour compensation is calculated using the cost of hired help, added holiday allowances and social security payments, and off-farm income is divided between wage work and other income. The survey also includes data on total area of cultivated land and the division of land into different uses and the yield of and income from different agricultural crops, fruit, garden berries and vegetables. To allow for calculation of obtained prices from farm sales, the turnover from all farm products are registered. Also the household's consumption of own production is registered. All costs of production are reported in total figures for each and every production input. Finally, the survey includes detailed balance sheets information and profit and loss accounts for all households, including information on production grants, interest payments, tax payments, and investment grants. The data set does not contain information on personal and family characteristics such as education, martial status, and family size which have been found to be important explanatory variables in estimating farmers' off-farm work participation and off-farm labour supply (see e.g. Huffman 1980 and Weersink 1992).

Sample selection

I have not yet tried to estimate my model on the data described above. I will need to decide on what empirical model best suits my purpose and I plan to run some estimations during summer. I plan to extract a ten-year panel, 1991-2000, from the Account Results survey and to account for the lacking variables of this survey by matching the data with statistics from Statistics Norway to be able to include at least the household members' level of education and family size. I have not yet decided exactly what variables will be included in the estimations.

Comments and further challenges

As I mentioned at the beginning of this paper my intention was to specify a random utility model inspired by the work of Hanemann (1984) though probably following the outline of McFaddens conditional logit model. This proved to be a rather complex problem and the main obstacle is that my problem is to derive labour supply equations that does not directly enter the utility function.

My confusion was based on the selection of discrete choice variable, either:

<u>Choose the alternative that generates the higher utility:</u>

$$D_n = \begin{cases} 1 & \text{if} \quad U_n^* = \text{Max}(U_1^*, U_2^*, U_3^*, U_4^*) \quad n = 1,2,3,4 \\ 0 & \text{otherwise} \end{cases}$$

or

<u>choose the participation rule</u>:

$$y^i = \begin{cases} 1 & \text{if} \ W^i \geq W^{iR} = P_Q f_i^{'}|_{M^i=0} \\ 0 & \text{if} \ W^i < W^{iR} = P_Q f_i^{'}|_{M^i=0} \end{cases} \qquad i = O, S$$

I am uncertain whether these discrete choices can be integrated within a single framework for a two-stage estimation or not, where

Stage one: Calculate the probability of choosing alternative *n*

Stage two: Choose the alternative that yields the higher utility

A matter that needs further investigation is my choice of participation rule for off-farm labour supply. I doubt the definition of the reservation wage I have used is very adequate, particularly for the spouses. Another question is whether it is relevant for my problem to include the joint decisions of husband and wife if the off-farm labour decisions of the spouses are independent of the farm production decisions. I have been considering whether the tobit model might better serve my purpose, and I also have to extend the model framework to account for the panel structure of my data.

Literature

Anderson, S.P., A. De Palma, and J.-F. Thisse. (1992): *Discrete Chioce Theory of Product Differentiation.* Cambridge: MIT Press.

Gould, B.W, and W.E. Saupe. (1989): “Off farm labor market entry and exit.” *Amer. J. Agr. Econ.* 71, 960-967.

Gourieroux, C. (2000): *Econometrics of Qualitative Dependent Variables.* Cambridge University Press.

Greene, W.H. (1997): *Econometric Analysis.* Prentice-Hall, New Jersey.

Hanemann, W.M. (1984): “Discrete/Continuous Models of Consumer Demand.” *Econometrica* Vol. 52, Issue 3, 541-562.

Huffman, W.E., and H. El-Osta. (1998): “Off-Farm Work Participation, Off-Farm Labor Supply and On-Farm Labor Demand of U.S. Farm Operators.” *American Journal of Agricultural Economics*. 80, 1180-.

Huffman, W.E., and M.D. Lange. (1989): “Off-farm work decisions of husbands and wives: joint decision making.” *Review of Economics and Statistics*. 71, 471-480.

Lass, D.A., and C.M. Gempesaw II. (1992): “The Supply of Off-Farm Labor: A Random Coefficient Approach.” *Amer. J. Agr. Econ.* 74, 400-411.

Maddala, G.S. (1983): *Limited-dependent and qualitative variables in econometrics.* Cambridge University Press.

Manski, C. F. (1977): ”The Structure of Random Utility Models.” *Theory and Decision* 8: 229-254.

Tversky, A. (1972): “Elimination by Aspects: A Theory of Choice.” *Psychological Rewiev* 79: 281-299.

Weersink, A., Nicholson, C., and J. Weerhewa. (1998): “Multiple job holdings among dairy farm families in New York and Ontario.” *Agricultural Economics.* 18, 127-143.