DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

PROBABILITY FEEDBACK IN A RECURSIVE SYSTEM OF PROBABILITY MODELS

Quang H. Vuong

SOCIAL SCIENCE WORKING PAPER 443

September 1982

ABSTRACT

This paper presents a general model for qualitative endogenous variables that is defined by a recursive system of conditional probability models in which the probabilities of some outcomes may depend on the probabilities of posterior outcomes. The model is related to, but conceptually different from C. D. Mallar's (1977) simultaneous probability model. It has as special cases the multivariate logit model (M. Nerlove and S. J. Press (1973, 1976)) and the constrained nested logit model (D. McFadden (1981)). The model can also be used to analyze outcomes of some game situations. Two examples are in particular considered: a game against Nature and a Stackelberg game under uncertainty. Identification of the structural parameters in the first example is seen to be related to the classical problem of stochastic revealed preference as studied by M. K. Richter and L. Shapiro (1978).

PROBABILITY FEEDBACK IN A RECURSIVE
SYSTEM OF PROBABILITY MODELS

By Quang H. Vuong [1]
California Institute of Technlogy

In many economic situations the choice of an alternative depends, among other things, on the conditional probabilities of some outcomes given the alternatives of the choice set. Indeed, it is easily conceivable that any situation in which an individual faces a number of possible decisions that will produce outcomes with some probabilities specific to each decision, fits into this framework as long as these probabilities be known or correctly anticipated by the individual. Numerous examples of such a situation are readily found since they typically fall in the category of decision making under uncertainty.

For instance, consider the classical example of a firm that faces a random demand and that has to choose among a number of possible prices for its products. If the firm knows the various conditional probability distributions of its sales given the possible prices, then it is clear that the firm's decision about its price will depend on these conditional distributions. As another example inspired from the job-search literature, consider the case in which an unemployed individual has the choice among three job-search strategies which are (i) a random search, (ii) a self-directed search, and (iii) an indirect search (see e.g., J. M. Barron and O. W. Gilley (1981)). Since each job-search strategy is characterized by a specific probability of success, i.e., of finding a job, then the choice of the optimal strategy must obviously depend on these conditional probabilities.

This paper presents a general model for qualitative endogenous variables that is defined by a recursive system of (conditional) probability models in which the probabilities of some actions or decisions may depend on the conditional probabilities of posterior outcomes. Specifically, the explanatory variables in the conditional probability model of a variable depend, among other things, on the conditional probabilities of its posterior variables, and hence on the parameters of the conditional probability models associated with these posterior variables.

The main purpose of this paper is to illustrate with the help of various examples the generality of the recursive system of probability models with probability feedback. At the outset it is worth noting that this model is clearly a generalization of the standard recursive system of probability models as studied by G. S. Maddala and Lung-Fei Lee (1976), Lung-Fei Lee (1981), B. Ottenwaelter and Q. H. Vuong (1981), and Q. H. Vuong (1982a) among others, since the standard recursive system simply excludes any dependence of the conditional distribution of a variable on the conditional probabilities of its posterior variables. But we shall also show that, as special cases of the recursive system with probability feedback, one formally recovers some familiar probability models that have been studied in the econometric literature of qualitative dependent variables, such as the multivariate logit model (see, e.g.,

M. Nerlove and S. J. Press (1973, 1976), T. Domencich and D. McFadden (1975), T. Amemiya (1978)), and the constrained nested logit model (see, e.g., D. McFadden (1981)). Moreover, our general model and the simultaneous probability model proposed by C. D. Mallar (1977) are somewhat related. We shall, however, argue that they are conceptually quite different.

Finally, our model naturally arises in situations in which the structure underlying the individuals' responses is recursive and the individuals have some sort of probabilistic rational expectations about future outcomes. This is the case for the two examples that we shall study: a game against Nature and a Stackelberg game under uncertainty. For each example, random utility models à la D. McFadden (1974, 1981)) are used to derive the probabilistic choice models of the recursive system. Furthermore, since the problem of identifying the structural parameters of the game against Nature is related to the classical problem of stochastic revealed preference which was considered by e.g., M. K. Richter and L. Shapiro (1978), this problem receives particular attention. The paper therefore includes some general results on identification of the structural parameters of the recursive system with probability feedback. These results are then used to obtain some simple necessary conditions for identification in the two game examples.

The paper is organized as follows. In Section 1, the recursive system of probability models with probability feedback is formally introduced. For the sake of simplicity, we restrict our

attention to the case of only two endogenous qualitative variables. The model is then compared to C. D. Mallar's simultaneous probability model. In section 2, it is shown that the multivariate logit model and the constrained nested logit model are special cases of the present recursive model. In Section 3, two examples in a game setting (a game against Nature, and a Stackelberg game under uncertainty) are presented in order to motivate the general formulation. In Section 4, some general results on the identification of the structural parameters of the model are given and then applied to the game examples of Section 3. Finally, in Section 5, we briefly discuss the estimation of the model, and we indicate some topics for further research.

1. The Model

In this section, we formally introduce the model for the case with only two endogenous qualitative variables. Then we relate our model to C. D. Mallar's simultaneous probability model.

1.1 Definition and Notations

Let A and B be two qualitative variables with I and J categories respectively. The index sets of A and B are denoted by $\bar{I}$ and $\bar{J}$ where $\bar{I} = \{1,\ldots,I\}$, and $\bar{J} = \{1,\ldots,J\}$. The letters i and j will be used as indices of the categories of A and B, and the symbol "t" as indice of the individuals of a sample of size T. Let A_t and B_t be the two random variables that represent the "choices" of the t-th individual with respect to the variables A and B.[2] Hence A_t can have any value in $\bar{I}$ while B_t can have any value in $\bar{J}$.[3]

Let z_{ijt} and z_{jt} be respectively a $k_A \times 1$ vector and a $k_B \times 1$ vector. These vectors are interpreted as in the literature on probabilistic choice models (see, e.g., D. McFadden (1981)). Specifically, z_{ijt} combines measured attributes of the i-th alternative of A, of the j-th alternative of B, and measured characteristics of the t-th individual, while z_{jt} combines only measured attributes of the j-th alternative of B and measured characteristics of the t-th individual. It will be convenient to define Z_{ABt} as the $IJ \times k_a$ matrix of which the (i,j)-th row is z'_{ijt}, and Z_{Bt} as the $J \times k_b$ matrix of which the j-th row is z'_{jt}. Let α be an unknown parameter vector of size a, and β be an unknown parameter vector of size b.[4] These two parameter vectors are assumed to belong to the parameter spaces Γ_A and Γ_B , respectively.

The probability that the random variable A_t is equal to i given that B_t is equal to j for the t-th individual is denoted by $\Pr(A_t = i|B_t = j,t)$. Similarly, $\Pr(B_t = j|t)$ denotes the probability that the random variable B_t is equal to j for the t-th individual (see footnote 2). The recursive system of probability models with probability feedback is defined by:

For any $t = 1,\ldots,T$:

$$\Pr(A_t = i|B_t = j,t) = p_A(i|j,Z_{ABt},\alpha) \quad ; \quad \alpha \in \Gamma_A \tag{1}$$

$$\Pr(B_t = j|t) = p_B(j|\tilde{Z}_{Bt},\beta) \quad ; \quad \beta \in \Gamma_B \tag{2}$$

where $\tilde{Z}_{Bt}$ is a $J \times \tilde{k}_B$ matrix satisfying:

$$\tilde{Z}_{Bt} = \tilde{Z}_{Bt}(\Pr(A_1|B_1,1),\ldots,\Pr(A_T|B_T,T),Z_{Bt}) \tag{3}$$

and $\Pr(A_t|B_t,t)$ is the IJ-dimensional vector of which the (i,j)th component is $\Pr(A_t = i|B_t = j,t)$. The functions $p_A(.|.,.,.)$, $p_B(.|.,.)$, and $\tilde{Z}_{Bt}(.,\ldots,.)$ are assumed to be known.[5] Moreover, for any j, Z_{ABt}, and α, the function $p_A(.|j,Z_{ABt},\alpha)$ has, of course, to add up to one when summing over i, while for any $\tilde{Z}_{Bt}$, and β, the function $p_B(.|\tilde{Z}_{Bt},\beta)$ has to add up to one when summing over j.

Equation (3) is sufficiently novel to deserve an explanation. First, suppose that $\tilde{Z}_{Bt}$ does not depend on any of the conditional distributions $\Pr(A_t|B_t,t)$, $t = 1,\ldots,T$. It is clear that Model (1)-(3) reduces to a standard recursive system of probability models (see, e.g., the references previously given). Suppose now that $\tilde{Z}_{Bt}$ depends on Z_{Bt} and on the conditional probability distribution $\Pr(A_t|B_t,t)$ for the t-th individual. Then, from (2)-(3) it follows that for the t-th individual the probability distribution of B_t depends on the conditional probability distribution of A_t given B_t. Since A_t is posterior to B_t according to the recursive formulation (1)-(2), this feature captures the idea that a decision with respect to a variable actually depends on the (conditional) probabilities of posterior variables or later outcomes. Examples of such a case are given in Section 3. The formulation (3) is, however, more general since $\tilde{Z}_{Bt}$ is allowed to depend on all the conditional distributions $\Pr(A_t|B_t,t)$, $t = 1,\ldots,T$. It turns out that this is so for the multivariate logit model and the constrained nested logit model (Section 2).

In any case, since for any t, the distribution $\Pr(A_t|B_t,t)$ depends of α from (1), it follows from (3) that $\tilde{Z}_{Bt}$ depends on the parameter vector α of the conditional probability model (1), i.e.,:

$$\tilde{Z}_{Bt} = \tilde{Z}_{Bt}(\alpha) \quad (3')$$

Moreover, since α is unknown, then the "explanatory" variables $\tilde{Z}_{Bt}$ of the probability model (2) are unobserved. This raises some estimation issues which are briefly mentioned in Section 5, and more thoroughly studied for the logit case in Q. H. Vuong (1982b).

Before considering some examples, we now discuss the relationship between our model (1)-(3) and C. D. Mallar's simultaneous probability model.

1.2. A Comparison with C. D. Mallar's Simultaneous Probability Model

C. D. Mallar (1977) considers a simultaneous probability model for dichotomous variables in which the probability that an event occurs depends, among other things, on the probabilities that other events occur. To simplify the discussion, let us consider two events only. Let the variables A and B represent these two events. Then A is equal to 1 if the first event occurs, and equal to 0 (say) otherwise. The categories of B are similarly defined. Let p_{At} be the probability that A is equal to one, and p_{Bt} be the probability that B is equal to one for the t-th individual. C. D. Mallar's simultaneous probability model is defined by:

$$p_{At} = F(I_{At}) \quad (4)$$

$$p_{Bt} = F(I_{Bt}) \quad (5)$$

where the scalars or indices I_{At} and I_{Bt} satisfy the simultaneous system:

$$I_{At} = x'_{At}\,\beta_A + I_{Bt}\,\gamma_A \quad (6)$$

$$I_{Bt} = x'_{Bt}\,\beta_B + I_{At}\,\gamma_B \quad (7)$$

The function F(.) is any strictly increasing and continuous cumulative distribution function. The vectors x'_{At} and x'_{Bt} are observed explanatory variables. Since A and B are dichotomous, Equations (4)-(5) also define the probability that A is equal to 0, and that the probability that B is equal to 0 for the t-th individual.

For the sake of comparison with the recursive system (1)-(3), let us consider the special case where $\gamma_A = 0$. Then from (4) and (5) it follows that p_{At} only depends on $x'_{At}\,\beta_A$. Moreover, since $F^{-1}(.)$ exists, the index I_{At} is a function of p_{At}. Hence from (5) and (7) it follows that the probability distribution of B also depends on the probability p_{At}. In this sense, this special case of C. D. Mallar's simultaneous probability model resembles the recursive system with probability feedback (1)-(3). There is, however, an important difference between the two models. Indeed, while the recursive system (1)-(3) deals with the conditional probability distribution $\Pr(A|B,t)$ and the marginal probability distribution $\Pr(B|t)$, the simultaneous probability model (4)-(7) deals with the two marginal probability distributions $\Pr(A|t)$ and $\Pr(B|t)$.[6] As a consequence, while the recursive model (1)-(3) can uniquely define the joint probability distribution $\Pr(A,B|t)$, (i.e., the reduced form), the simultaneous model (4)-(7), by itself, cannot even in its general form $\gamma_A \neq 0$.[7]

2. Some Probability Models for Qualitative Dependent Variables

In this section, it is shown that the multivariate logit model and the constrained nested logit model are both special cases of the recursive system of probability models with probability feedback (1)-(3).

2.1. The Multivariate Logit Model

We shall consider only two endogenous qualitative variables A and B. Let $\Pr(A_t,B_t|t)$ be the joint probability distribution of the random variables A_t and B_t that represent the responses of the t-th individual with respect to A and B. The bivariate logit model directly specifies, for any t, the (parametric) family of joint distributions $\Pr(A_t,B_t|t)$ of interest (see, e.g., M. Nerlove and S. J. Press (1973, 1976)). Specifically, we have:

for any $t = 1,\ldots,T$, $i = 1,\ldots,I$, and $j = 1,\ldots,J$:

$$\log \Pr(A_t = i, B_t = j|t) = \mu_t^* + v_{ijt}'\,\theta \tag{8}$$

where

$$\mu_t^* = -\log\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\exp(v_{ijt}'\theta)\right] \tag{9}$$

and θ is an unknown parameter vector belonging to $\mathbb{R}^k$. The parameter vector θ may contain specific alternative parameters and/or interaction parameters between A_t and B_t or between these variables and individual characteristics. The k dimensional vector v_{ijt} combines as usual measured attributes of the i-th alternative of A, of the j-th alternative of B, and measured characteristics of the t-th individual. Moreover, it is assumed that the matrix V_{ABT} of which the (i,j,t)-th row is v_{ijt} is such that the parameter vector θ is identified.[8]

The bivariate logit model (8)-(9) is consistent with the Random Utility Maximization (RUM) hypothesis discussed by D. McFadden (1974, 1976, 1981). This hypothesis requires that the probability that the pair (i,j) be chosen by the t-th individual satisfy:

$$\Pr(A_t = i,\ B_t = j|t) = \Pr\left[U_{ijt} \geq U_{i'j't}\ ;\ \forall (i',j') \in \bar{I} \times \bar{J})|t\right] \tag{10}$$

where $\Pr(.|t)$ is the probability measure defined on the space $\mathbb{R}^{IJ}$ of utility functions $\{ U_{ijt};\ (i,j) \in \bar{I} \times \bar{J} \}$ of the individuals who have the measured characteristics of the t-th individual (for a formal exposition of the RUM hypothesis, see D. McFadden (1981)). Indeed from McFadden's (1974) argument for the univariate logit model, it follows that the choice probabilities associated with the bivariate logit model (4)-(5) can be derived from the random utility model:

$$U_{ijt} = v'_{ijt}\,\theta + \varepsilon_{ijt} \tag{11}$$

where for any t, the random error ε_{ijt}'s are independently and identically extreme-value distributed, i.e., with joint cumulative distribution function:

$$F(\varepsilon_{ijt}\ ;\ i \in \bar{I},\ j \in \bar{J}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \exp(-\exp(-\varepsilon_{ijt})) \tag{12}$$

To see that the bivariate logit model (4)-(5) is formally equivalent to a recursive system of logit models with probability feedback, it is convenient to partition the explanatory variables v_{ijt} into explanatory variables x_{ijt} that vary across i, and explanatory variables that do not.[9] Let the parameter vector θ be accordingly partitioned into γ and δ so that (11) becomes

$$U_{ijt} = x'_{ijt}\gamma + x'_{jt}\delta + \varepsilon_{ijt} \tag{11'}$$

The choice probabilities (8)-(9) can then be written as:

$$\log \Pr(A_t=i,\ B_t=j\,|\,t) = \mu^*_t + x'_{ijt}\gamma + x'_{jt}\delta \tag{8'}$$

where

$$\mu^*_t = -\log\left[\sum_{i=1}^{I} \sum_{j=1}^{J} \exp(x'_{ijt}\gamma + x'_{jt}\delta)\right] \tag{9'}$$

For any t, let us derive the conditional probability model for A_t given B_t, and the marginal probability model for B_t.[10] From (8')-(9'), we obtain:

$$\log \Pr(A_t=i\,|\,B_t=j,t) = \mu_{jt} + x'_{ijt}\gamma \tag{13}$$

$$\log \Pr(B_t=j\,|\,t) = \mu_t - \mu_{jt} + x'_{jt}\delta \tag{14}$$

where

$$\mu_{jt} = -\log\left[\sum_{i=1}^{I} \exp(x'_{ijt}\gamma)\right] \tag{15}$$

$$\mu_t = -\log\left[\sum_{j=1}^{J} \exp(-\mu_{jt} + x'_{jt}\delta)\right] \tag{16}$$

It is clear that the bivariate logit model (8)-(9) is equivalent to the recursive pair of univariate logit models (13)-(16) in the sense that any joint probability distribution $\Pr(A_t, B_t|t)$ satisfying (8)-(9) for some θ also satisfied (13)-(16) for some (γ,δ), and conversely, that any pair of probability distributions $\{\Pr(A_t|B_t,t),\ \Pr(B_t|t)\}$ satisfying (13)-(16) for some (γ,δ) defines a bivariate probability distribution $\Pr(A_t,B_t|t)$ that satisfies (8)-(9)

for some θ.

Let

$$z'_{ijt} = x'_{ijt} \quad ; \quad \alpha = \gamma$$
$$\tilde{z}'_{jt} = (-\mu_{jt}, x'_{jt}) \quad ; \quad \beta = (\rho, \delta')' \tag{17}$$

for some scalar ρ. From (15), it follows that μ_{jt} is a function of γ. Hence $\tilde{z}_{jt}$ is a (known) function of the parameters of the conditional probabiltiy model for A_t given B_t (see Equation (3')). Moreover, since the parameter vector $\theta = (\gamma', \delta')'$ of the bivariate logit model is identified (see footnote 7), it follows that the parameter vector γ of the conditional logit model (10) is identified. Thus, there is a one-to-one correspondence between γ and the T conditional probability distributions $\{Pr(A_t | B_t, t);\ t = 1,\ldots,T\}$. Hence μ_{jt} and therefore $\tilde{z}'_{jt}$ are functions of these T conditional distributions as stated by Equation (3).

Equation (13)-(17) therefore shows that the bivariate logit model is equivalent to a recursive pair of univariate logit models with probability feedback in which one coefficient (the ρ coefficient) is constrained to be equal to one.

2.2. The Constrained Nested Logit Model

Let us now consider the nested logit model which is one of the simplest qualitative choice models that avoids the Independence of

Irrelevant Alternatives (IIA) property implied by the logit model (see, e.g., D. McFadden (1981)). In order to make the comparison with the recursive pair of probability models (1)-(3), we shall actually consider the constrained two-level nested logit model. According to this model, an individual who faces a finite number of alternatives is assumed to choose first among sets of alternatives that are close to each other in the measured attribute space (level 2), and then to choose among the alternatives of the chosen set (level 1). Let the variables A_t and B_t represent the choices at level 1 and level 2 for the t-th individual or t-th population (see footnote 2). Given these definitions, each alternative of the initial choice set is associated with a pair of indices (i, j).[11]

The constrained two-level nested logit model is then defined by:

$$\log Pr(A_t = i | B_t = j, t) = \mu_{jt} + x'_{ijt}(\gamma/\rho) \tag{18}$$

$$\log Pr(B_t = j | t) = \mu_t - \rho\mu_{jt} + x'_{jt}\delta \tag{19}$$

where

$$\mu_{jt} = -\log\left[\sum_{i=1}^{I} \exp(x'_{ijt}\gamma/\rho)\right] \tag{20}$$

$$\mu_t = -\log\left[\sum_{j=1}^{J} \exp(-\rho\mu_{jt} + x'_{jt}\delta)\right] \tag{21}$$

The parameter $-\mu_{jt}$ is called the inclusive value of node j for the t-th individual, and the coefficient ρ is a measure of dissimilarity between alternatives of the same node. (The nested logit model (18)-(21) is constrained because the coefficient ρ is the same across nodes.)

The nested logit model is consistent with the RUM hypothesis (see, e.g., D. McFadden (1981)). In particular, the constrained nested logit model (18)-(21) can be derived from the random utility model (11') where the errors ε_{ijt}'s have the Generalized Extreme Value distribution defined by the cumulative distribution function:

$$F(\varepsilon_{ijt};\ i\in\bar{I},\ j\in\bar{J}) = \prod_{j\in\bar{J}} \exp\{ - \sum_{i\in\bar{I}} \exp(- \varepsilon_{ijt}/\rho)\}^{\rho} \tag{22}$$

with $0 < \rho \leq 1$.

From the comparison between Equations (13)-(14) and Equations (18)-(19), it follows that the main difference between the bivariate logit model and the constrained two-level nested logit model is that in the former model the parameter ρ is restricted to be equal to one. This property is sometimes used in empirical studies to test the IIA hypothesis implied by the logit model.[12]

Let

$$z'_{ijt} = x'_{ijt} \quad ; \quad \alpha = \gamma/\rho$$

$$z'_{jt} = (- \mu_{jt},\ x'_{jt}) \quad ; \quad \beta = (- \rho,\ \delta')' \tag{23}$$

From (20) and (23), it follows that the explanatory variables $\tilde{z}_{jt}$ of the probability model for B_t depend on the parameters α of the conditional probability model for A_t given B_t. Hence, as in the bivariate logit model, the vector $\tilde{z}_{jt}$ depends on the T conditional probability distributions $\Pr(A_1|B_1,1),\ldots,\Pr(A_T|B_T,T)$.[13] Therefore the constrained nested logit model is formally equivalent to a recursive system of logit models with probability feedback.

3. Two Game Theoretic Examples

In order to illustrate the applicability of the general recursive system with probability feedback (1)-(3), we now consider two examples in a game setting: a game against Nature, and a Stackelberg game under uncertainty. In the first example we shall use a random utility model to derive one of the two probability models of the resursive system, while in the second example, each of the two probability models will be motivated by an underlying random utility model.

3.1. A Game Against Nature

Let us consider the following simple game: an individual has to choose among a number of lotteries that will produce some outcomes with probabilities that are specific to each lottery. The individual derives some utility from each possible outcome, and knows the various probabilities associated with each lottery. It is thus natural to assert that the choice of the individual depends, among other things, on the conditional probabilities of the outcomes given the available lotteries.

As mentioned at the outset, numerous economic situations fit into that simple framework. Indeed the two examples given in the introduction clearly are illustrations of this game.[14] For instance, consider the job-search example in which an unemployed individual has the choice among the three job-search strategies that are (i) a random search, (ii) a self-directed search, and (iii) an indirect search. With the previous notations, let A be the dichotomous variable "labor force status" with the categories "employed" and "still unemployed", and let B be the trichotomous variable representing the individual's strategy choice. Then, the choice of the optimal job-search strategy obviously depends on the conditional probabilities of finding a job given the available strategies.

In this section, we shall be interested in the derivation of the probability model for B from a random utility model. It is, however, important to note that the conditional probability model for A given B will be considered as given. In other words, we shall not try to explain how this conditional probability model is obtained, and we shall, therefore, think of it as representing the probabilistic responses of Nature (say) to the individual's choices.

Let A_t be the random variable that indicates in which of the possible "states" of the world the t-th individual will be. (In the previous job-search example, A_t simply indicates whether the t-th individual will be employed or still unemployed.) Let B_t be the random variable that indicates which of the available strategies is chosen by the t-th individual. It is assumed that each individual t perfectly knows the conditional probability distribution $\Pr(A_t|B_t = j,t)$ associated with each strategy j. Hence the t-th individual knows the whole $IJ \times 1$ vector $\Pr(A_t|B_t,t)$.

On the other hand, the observer or econometrician knows all the T vectors $\Pr(A_t|B_t,t)$, but only up to some unknown parameter vector α. More precisely, the outsider knows the function $p_A(.|.,.,.)$ in Equation (1), and observes the $IJ \times k_A$ matrix Z_{ABt} of which the (i,j)-th row combines as before measured attributes of the i-th and j-th alternatives of A and B, and measured characteristics of the t-th individual.

Each individual is assumed to derive some utility from each state of the world. Specifically, the utility derived from state i by the t-th individual is:

$$U_{it} = z'_{it}\delta + \varepsilon_{it} \tag{24}$$

where z_{it} is a $d \times 1$ vector that combines measured attributes of the i-th state and measured characteristics of the t-th individual, δ is a commensurate vector of parameters unknown to the observer, and ε_{it} is a random component that represents the effects of unobserved variables. Let us note that in the present formulation the utility U_{it} derived from state i by the t-th individual does not depend on the strategy chosen by the individual.

We now assume that each individual maximizes his expected utility. Thus, individual t chooses strategy j if, and only if

$$E(U_{it}|j,t) \geq E(U_{it}|j',t) \qquad \text{for all } j' \text{ in } \bar{J} \tag{25}$$

where $E(.|j,t)$ indicates that the expectation is taken with respect to the conditional probability distribution $Pr(A_t|B_t = j,t)$. Or, equivalently, using (1) and (25), individual t chooses strategy j if, and only if,

$$\tilde{z}'_{jt}\delta + \tilde{\varepsilon}_{jt} \geq \tilde{z}'_{j't}\delta + \tilde{\varepsilon}_{j't} \qquad \text{for all } j'\text{s in } \bar{J} \tag{26}$$

where

$$\tilde{z}'_{jt} = \sum_{i=1}^{I} p_A(i|j,Z_{ABt},\alpha) z_{it} \tag{27}$$

$$\tilde{\varepsilon}_{jt} = \sum_{i=1}^{J} p_A(i|j,Z_{ABt},\alpha) \varepsilon_{it} \tag{28}$$

To complete the probability choice model for B, we assume that, for any t, the cumulative distribution function $F_t(.;\omega)$ of the random vector $\varepsilon_t = (\varepsilon_{1t},\dots,\varepsilon_{It})$ belongs to some family of distribution functions parameterized by the unknown parameters ω in some parameter space Ω. Hence, from (26), this distribution function

specified how the utilities of the individuals who have the same observed characteristics as those of the t-th individual are distributed. For instance, one may assume, as in Probit analysis, that the random vector ε_t has a multivariate normal distribution with mean zero and a covariance matrix Σ_t parameterized by some unknown parameters ω (see e.g., J. Hausman and D. Wise (1978)).[15]

Hence, from (27) it follows that the probability model for the choice of strategies is:

For any $t = 1,\dots,T$, and any $j = 1,\dots,J$

$$Pr(B_t = j|t) = Pr(\tilde{\varepsilon}_{j't} - \tilde{\varepsilon}_{jt} \leq (\tilde{z}'_{jt} - \tilde{z}'_{j't})\delta, \forall j' \in \bar{J}|t,\omega) \tag{29}$$

where $Pr(.|t,\omega)$ indicates that the choice probabilities are evaluated with respect to the distribution $F_t(.;\omega)$ of the random vector ε_t. It is noteworthy that, because of Equation (29), the random errors $\tilde{\varepsilon}_{jt}, j=1,\dots,J$ cannot be mutually independent. Hence the choice probabilities (29) cannot be obtained from a logit specification (see D. McFadden (1974)).

Let us now consider the system that consists of the two probability models (1) and (29). Let β be the vector $(\delta',\omega')'$. Equation (27) clearly shows that the explanatory variables $\mathcal{Z}_{Bt}$ depends on the J conditional probability distributions $Pr(A_t|B_t=1,t),\dots,Pr(A_t|B_t=J,t)$, i.e., on $Pr(A_t|B_t,t)$ (see Equation (3)). Hence the recursive system (1) and (29) associated with the present game is a special case of the general model with probability feedback (1)-(3).

Given the above probability model for A and B, an important issue is whether one can uniquely recover the structural parameters (α, γ, ω) by simply observing the final states and choices of T individuals. This problem, which is known in econometrics as the identification problem, will be considered in Section 4.

3.2 A Stackelberg Game under Uncertainty

In the previous example, each individual is playing against Nature. Since the probability model for Nature's responses is taken as given only one of the two probability models of the recursive system can be derived from a random utility model. Hence in order to derive both probability models from some random utility models, we now explicitly introduce two types of players. This leads to a Stackelberg game under uncertainty.

To fix ideas, it is useful to consider two examples of such a game. The first example is based on the following simple two-person game associated with the matrix:

Person B's moves

		1		2	
Person A's moves	1	U^A_{11}	U^B_{11}	U^A_{12}	U^B_{12}
	2	U^A_{21}	U^B_{21}	U^A_{22}	U^B_{21}

where U^A_{ij} and U^B_{ij} are the utilities derived from move i of A and move j of B. Suppose that, instead of the two players moving simultaneously, player B moves first, and that player B's move is known to player A when player A makes his move. Suppose also that none of the two players exactly knows the utility derived by his opponent although each player may observe some characteristics of his opponents as well as some attributes of the pair (i,j), such as, the monetary payoffs for both players that result from the combination (i,j). Hence, when player B chooses his optimal move, player B does not know for sure what will be player A's response. In other words, the reaction function of player A to the player B's move is probabilistic, and one has a Stackelberg game under uncertainty. Moreover, if one assumes that player B somewhat knows the conditional distribution of player A's moves given his moves, then it is intuitively clear that, for the observer who does not know exactly player B's utility, and hence who has to treat player B's move as random, the appropriate econometric model that explains which pair of strategies is chosen is a recursive system of probability models with probability feedback.

As a more concrete example for which the recursiveness of the system naturally arises, consider the election by a district (say) of an individual among a number of candidates each of whom has to favor and eventually pass one of a list of proposals or issues. Suppose that the district derives some (social) utility from each of the available issues and possibly also from which candidate is elected.

On the other hand, each candidate derives some utility, possibly from the issue that is finally passed, and certainly from which candidate is elected. However, a candidate once elected need not pass the issue that he favored before the election, and, instead, may choose to pass another issue. Hence for each candidate and for each issue there is only a probability that the issue will be passed by the candidate. It is clear that this is an illustration of a Stackelberg game under uncertainty, and that the appropriate econometric model for explaining which candidate is elected and which issue is passed will be a recursive system with probability feedback.

Let us now formally derive each of the probability models of the recursive system from some random utility model. As in a standard Stackelberg game, we distinguish two types of players: these are the "followers" and the "leaders". Let $\bar{I}$ and $\bar{J}$ be the set of actions that are respectively available to the followers and the leaders (see also footnote 3). Since in this game a state or final outcome results from both chosen actions, a state is naturally determined by a pair (i,j) in $\bar{I} \times \bar{J}$.

Each individual, follower or leader, derives some utility from each state (i,j). Specifically, we assume that the utility derived from state (i,j) by the r-th follower and s-th leader are respectively:

$$U^F_{ijr} = z^{F'}_{ijr}\gamma + \varepsilon^F_{ijr} \qquad (30)$$

$$U^L_{ijs} = z^{L'}_{ijs}\delta + \varepsilon^L_{ijs} \qquad (31)$$

where the c-dimensional vector γ and the d-dimensional vector δ are unknown to the observer. The vectors z^F_{ijr}, ε^F_{ijr}, z^L_{ijs}, and ε^L_{ijs} are interpreted as before. For instance, z^F_{ijr} combines measured attributes of the (i,j)-th state and observed characteristics of the r-th follower, while ε^F_{ijr} is a random component that represents the effects of unobserved variables. However, unlike the previous game against Nature, both indices i and j appear in the equations (30)-(31) since the utilities are now defined on the set of states $\bar{I} \times \bar{J}$.

In this example, an observation corresponds to one game played by a follower and a leader. Thus, an observation t is characterized by a pair of indices (r,s) where r and s respectively indicate the type of the follower and the type of the leader. Then, for any game t = (r,s), we let A_t and B_t be the variable that indicate which of the actions in $\bar{I}$ is chosen by the r-th follower, and which of the actions in $\bar{J}$ is chosen by the s-th leader.

We shall assume that for any t = (r,s), the random vectors ε^F_r and ε^L_s are mutually independent, where ε^F_r and ε^L_s are the two IJ-dimensional vectors of which the (i,j)-th component are respectively ε^F_{ijr} and ε^L_{ijs}. This assumption means that for any game (r,s), the two players of the game are drawn independently, one from the population of followers of type "r", and one from the population of leaders of type "s". Hence from (30)-(31), it follows that the random deviations ε^F_{ijr} and ε^L_{ijs} from the mean utilities $z^{F'}_{ijr}\gamma$ and $z^{L'}_{ijs}\delta$ are uncorrelated.[16]

Since for any game (r,s), the follower "r" takes the action j chosen by the leader "s" as given, then for any j, the follower "r" chooses action i if, and only if, for all i' in $\overline{I}$:

$$z^{F'}_{ijr}\gamma + \varepsilon^{I}_{ijr} \geq z^{F'}_{ijr}\gamma + \varepsilon^{F}_{ijr} \tag{32}$$

Let $F_r(.;\omega^F)$ be the cumulative distribution function of the random vector ε^F_r, where ω^F is some parameter vector in Ω^F unknown to the observer. Then from (32) it follows that the conditional probability model for the follower's responses to the leader's possible moves is:

$$Pr(A_t=i|B_t=j,t) = Pr(\varepsilon^F_{i'jr} - \varepsilon^F_{ijr} \leq (z^{F'}_{ijr} - z^{F'}_{i'jr})\gamma, \forall i' \in \overline{I}|r,\omega^F) \tag{33}$$

where $Pr(.|r,\omega^r)$ indicates that the choice probabilities (33) are evaluated with respect to the cumulative distribution function $F_r(.;\omega^F)$.

Let us now derive the probability model for the leaders' choices. Since each leader also observes the individual characteristics of his opponent, each leader knows the type of his opponent. We shall assume that leader "s" knows the choice probabilities (33) which characterize the probabilistic responses of follower "r" to each of leader "s" actions, i.e., that leader "s" knows the whole vector $Pr(A_t|B_t,t)$ where t = (r,s).[17] Thus if leader "s" maximizes his expected utility, leader "s" chooses action j if, and only if, for all j' in $\overline{J}$:

$$\tilde{z}^{L'}_{js}\delta + \tilde{\varepsilon}_{js} \geq \tilde{z}'_{j's}\delta + \tilde{\varepsilon}_{j's} \tag{34}$$

where

$$\tilde{z}^{L}_{js} = \sum_{i=1}^{I} Pr(A_t=i|B_t=j,t)z^{L}_{ijs} \tag{35}$$

$$\tilde{\varepsilon}^{L}_{js} = \sum_{i=1}^{I} Pr(A_t=i|B_t=j,t)\varepsilon^{L}_{ijs} \tag{36}$$

Since the outsider only observes some individual characteristics of the leader "s", and thus does not know U^L_{ijs}, the choice of leader "s" has to be considered as random. Let $F_s(.,\omega^L)$ be the cumulative distribution function of the random vector ε^L_s, where ω^L is a parameter vector in Ω^L unknown to the observer. Then the probability model explaining the leaders' choices is:

For all t=(r,s) and all j in $\overline{J}$.

$$Pr(B_t=j|t) = Pr(\tilde{\varepsilon}_{j's} - \tilde{\varepsilon}_{js} \leq (\tilde{z}^{L'}_{js} - \tilde{z}^{L'}_{j's})\delta, \forall j' \in \overline{J}|s,\omega^L) \tag{37}$$

where $Pr(.|s,\omega^L)$ indicates that the choice probabilities (37) are evaluated with respect to the cumulative distribution function $F_s(.;\omega^L)$. As in Section 3.1, it is worth noting that, because of (36), the random errors $\tilde{\varepsilon}^L_{js}$, j=1,...,J, cannot be mutually independent. Hence the choice probabilities (37) cannot be obtained from a logit specification.

It is now straightforward to see that the system of probability models (33) and (37) is a special case of the general recursive system with probability feedback (1)-(3). Indeed, let $\alpha = (\gamma', \omega^{F'})'$ and $\beta = (\delta', \omega^{L'})'$. From Equation (35) it follows that the explanatory variables $\tilde{Z}^{L}_{Bs}$ of the probability model (37) for B_t depend on the conditional probability distribution $\Pr(A_t | B_t, t)$ as asserted by Equation (3). Hence from Equation (33), the explanatory variable $\tilde{Z}^{L}_{Bs}$ depends on the unknown parameters α of the conditional probability model for A_t given B_t (see Equation (3')).

As in Section 3.1, an important issue is whether one can uniquely recover the structural parameters $(\gamma, \omega^F, \delta, \omega^L)$ from the outcomes of T games, i.e., by simply observing the actions chosen by T followers and T leaders. This problem will be considered in the next section.

4. Identification

We now derive some general results on the identification of the structural parameters in the general recursive system with probability feedback (1)-(3). As a special case, we shall recover the standard conditions for identification in an ordinary recursive system. Then we shall apply the general results to the identification of the structural parameters in the two game examples introduced in the preceding section.

From now on, it is assumed that the T pairs of random variables (A_t, B_t), $t = 1,\ldots,T$ are mutually independent. This

assumption, which is standard in the econometric literature on qualitative dependent variables, is, however, inessential for the results obtained in the following subsection. Its principal purpose is to complete the specification of the statistical model, i.e., of the joint distribution of the T pairs of random variables (A_t, B_t), $t = 1,\ldots,T$. Indeed, under this assumption, one has:

$$\Pr(\underline{A} = \underline{i} | \underline{B} = \underline{j}) = \prod_{t=1}^{T} \Pr(A_t = i_t | B_t = j_t) \tag{38}$$

and

$$\Pr(\underline{B} = \underline{j}) = \prod_{t=1}^{T} \Pr(B_t = j_t) \tag{39}$$

where $\underline{A} = (A_1,\ldots,A_T)$, $\underline{B} = (B_1,\ldots,B_T)$, $\underline{i} = (i_1,\ldots,i_T)$, and $\underline{j} = (j_1,\ldots,j_T)$. Since

$$\Pr(\underline{A} = \underline{i},\ \underline{B} = \underline{j}) = \Pr(\underline{A} = \underline{i} |\ \underline{B} = \underline{j}).\Pr(\underline{B} = \underline{j}) \tag{40}$$

it follows from (1), (2), (38), (39) that (α, β) uniquely defines the joint probability distribution $\Pr(\underline{A}, \underline{B})$.

4.1. Some General Results

The concept of identification that we shall use is that of T. J. Rothenberg (1971) and R. Bowden (1973). Specifically, let (α^o, β^o) be an admissible structure, i.e., an element of the parameter space

$\Gamma_A \times \Gamma_B$. Then, the structure (α^o,β^o) of the joint probability model for A and B is said to be identified given a sample of size T if there does not exist another admissible structure (α,β) for which the joint probability distribution of the T pairs $(A_t,B_t), t = 1,\ldots,T$ is identical to that of (α^o,β^o). Formally, (α^o,β^o) is identified if and only if

$$\nexists (\alpha,\beta) \in \Gamma_A \times \Gamma_B - \{(\alpha^o,\beta^o)\} \text{ such that}$$

$$\Pr(\underline{A},\underline{B};\alpha,\beta) = \Pr(\underline{A},\underline{B};\alpha^o,\beta^o) \tag{41}$$

where the equality (41) means that for any $(\underline{i},\underline{j})$, $\Pr(\underline{A} = \underline{i},\underline{B} = \underline{j};\alpha,\beta)$ is equal to $\Pr(\underline{A} = \underline{i},\underline{B} = \underline{j};\alpha^o,\beta^o)$.[18] Moreover, if condition (41) holds for any admissible structure (α^o,β^o), then the (joint) probability model for A and B is said to be identified.

It may be worth emphasizing two features of the preceding definition. First, this definition is that of global identification. Hence we shall not be interested in whether or not condition (41) only holds in a neighborhood of (α^o,β^o), i.e., in whether or not (α^o,β^o) is locally identified. Second, the preceding definition makes clear that the problem is that of identifying the structural parameters α and β given a sample of finite size. In other words, we shall not deal with the question of whether the structural parameters are identified when the sample size goes to infinity.

Given the recursive structure of the general model (1)-(3), it may be more useful to characterize the identification condition (41) by a condition on α^o, and a condition on β^o. This is the purpose of the next lemma of which the proof is straightforward and therefore omitted.

<u>Lemma</u>

Let (α^o,β^o) be an admissible structure of the general model (1)-(3). Then (α^o,β^o) is identified if and only if the following two conditions hold:

(i) $\nexists \alpha \in \Gamma_A - \{\alpha^o\}$ such that for some β^1 and β^2 in Γ_B

$$\Pr(\underline{A}|\underline{B};\alpha) = \Pr(\underline{A}|\underline{B};\alpha^o) \text{ and } \Pr(\underline{B};\alpha,\beta^1) = \Pr(\underline{B};\alpha^o,\beta^2) \tag{42}$$

(ii) $\nexists \beta \in \Gamma_B - \{\beta^o\}$ such that

$$\Pr(\underline{B};\alpha^o,\beta) = \Pr(\underline{B};\alpha^o,\beta^o) \tag{43}$$

Condition (43) is easily interpreted. Indeed it simply means that, given α^o, the structure β^o of the univariate probability model for B is identified. On the other hand, condition (42) is more complex. First, let us note that when there is no probability feedback, i.e., when we have an ordinary recursive system, condition (42) becomes equivalent to:

$$\nexists \alpha \in \Gamma_A - \{\alpha^o\} \text{ such that } \Pr(\underline{A}|\underline{B};\alpha) = \Pr(\underline{A}|\underline{B};\alpha^o) \tag{44}$$

In general, however, condition (44) implies, but is not implied by condition (42).

Moreover, it is important to note that Equation (44) actually means that for any $(\underline{i},\underline{j})$ in $\bar{I} \times \bar{J}$, $\Pr(\underline{A} = \underline{i} | \underline{B} = \underline{j}; \alpha)$ is equal to $\Pr(\underline{A} = \underline{i} | \underline{B} = \underline{j}; \alpha^o)$. Hence, condition (44) is not equivalent, but simply implied by the condition that the structure α^o of the (conditional) probability model for A be identified given the observed values $j_1^o,\ldots,j_T^o$ of $B_1,\ldots,B_T$. This latter condition is:

$$\nexists \alpha \in \Gamma_A - \{\alpha^o\} \text{ such that}$$

$$\Pr(\underline{A}|\underline{B} = \underline{j}^o,\alpha) = \Pr(\underline{A}|\underline{B} = \underline{j}^o,\alpha^o) \tag{45}$$

where $\underline{j}^o = (j_1^o,\ldots,j_T^o)$.

The characterization of identification given by the basic lemma is difficult to apply in practice since condition (42) involves two equalities. One may then rather use the following necessary or sufficient conditions which are easier to verify. These conditions directly follow from the lemma and the preceding discussion.

Corollary

Let (α^o,β^o) be an admissible structure of the general model (1)-(3).

(i) If (α^o,β^o) is identified, then condition (43) must hold.

(ii) If conditions (43) and (45) (or conditions (43) and (44)) are satisfied, then (α^o,β^o) is identified.

Thus a necessary condition for the identification of the general model (1)-(3) is that the probability model for B be identified for any α^o, while a sufficient condition is that the probability model for A be identified and that the necessary condition holds.

Finally, let us note that the preceding results are quite general since the probability models (1) and (2) need not satisfy any particular properties. An important special case, however, is that in which the functions $L_A(\alpha;\underline{i},\underline{j}^o)$ and $L_B(\alpha^o,\beta;\underline{j})$ are respectively globally concave in α and β for any $\underline{i},\underline{j}$, and α^o, where

$$L_A(\alpha;\underline{i},\underline{j}^o) = \log \Pr(\underline{A} = \underline{i}|\underline{B} = \underline{j}^o,\alpha) \tag{46}$$

$$L_B(\alpha^o,\beta;\underline{j}) = \log \Pr(\underline{B} = \underline{j};\alpha^o,\beta) \tag{47}$$

and $\underline{j}^o = (j_1^o,\ldots,j_T^o)$ is the observed value of $\underline{B}$. Indeed, many of the usual qualitative response models fall in this case as shown by J. W. Pratt (1981) who also gives general conditions for the global concavity of the log-likelihood function. If global concavity holds, then it directly follows from R. Bowden (1973)'s characterization of identification that establishing conditions (43) and (45) is equivalent to showing the strict concavity of $L_A(.;\underline{i},\underline{j}^o)$ and

$L_B(\alpha,.;\underline{j})$.[19] On the other hand, strict concavity will in general follow from global concavity if there are sufficient variations in the observations on A and B and in the values of the explanatory variables.

4.2. Examples

The previous general results can be applied to the problem of identification of the structural parameters in the two game examples of Section 3. Our principal purpose will be to derive a necessary condition for identification that is simpler, though weaker, than the condition that the probability model for B be identified (see previous corollary). The argument will be identical for both examples since in each example the probability model for B is derived from a random utility model, i.e., is a probabilistic choice model. Thus, to avoid repetition, we shall only discuss the first game example. In addition, the identification problem for this example is related to the classical problem of revealed preference.

In the game against Nature, an important question is whether one can uniquely recover the structural parameters α,δ, and ω from a finite number of observations on individuals' choices of strategy and individuals' final states. Since from Equation (24), the structural parameters δ completely characterize the mean utilities of the individuals, our discussion will be centered on the identification of these parameters.

It is then worth noting that this problem is the dual of the

one considered by M. K. Richter and L. Shapiro (1978). There, the problem was to determine what can be inferred about the (subjective) individual probabilities of some outcomes by observing the choices of an individual who maximizes expected utility with respect to his subjective probabilities and some unknown utility. While M. K. Richter and L. Shapiro showed that the subjective probabilities must, at most, satisfy a certain class of inequalities, the previous question about identification asks for a complete determination of the mean utilities. The fact that a positive answer can be given here, whenever the parameters δ are identified, actually results from our distinction between the individual's mean utility and the individual's utility.[20]

Recall that $\tilde{Z}_{Bt}$ is the $J \times d$ matrix of which the j-th row is the row vector $\tilde{z}'_{jt}$. Let $\tilde{Z}_B$ be the $JT \times d$ stacked matrix of which the t-th blocks is $\tilde{Z}_{Bt}$. Then, from Equation (29) which gives the probabilities of the individual's choices of strategy, it can readily be shown that a <u>necessary</u> condition for the probability model for B to be identified given α, is that the $JT \times (T + d)$ matrix M_B defined by

$$M_B = \left[I_T \otimes U_J \; ; \; \tilde{Z}_B \right] \tag{48}$$

be full column rank, i.e., of rank $T + d$, where I_T is the $T \times T$ identity matrix, and U_J is the J-dimensional vector of ones.

It is important to note that the previous condition on the rank of M_B is only necessary to the identification of the probability model for B. Whether this condition is also sufficient clearly depends on the functional form of the probabilities (29), i.e., on the choice of the cumulative distribution function $F_t(.;.)$ of the random utility components ε_t.[21] On the other hand, it is noteworthy that the above necessary condition holds irrespective of which cumulative distribution function $F_t(.;.)$ is chosen since the vectors $\tilde{z}_{jt}$ will always appear in (29) as differences. Also, the classical condition that the matrix of explanatory variables, i.e., Z_B, be of full column rank differs from the above necessary condition by the fact that this latter condition essentially requires T additional normalization constraints.

Thus, from the corollary of Section 4.1, it follows that a necessary condition for identification of the structural parameters α, γ, and δ is that rank $M_B = T + d$. A weaker necessary condition is that d be less or equal to T(J − 1), i.e., that the number of explanatory variables in the probability model for B be not greater than the number of observations times the number of available strategies minus one.

Let us now consider the problem of identifying the mean utilities $\bar{U}_{it}$. Since the mean utilities satisfy the parametric form $\bar{U}_{it} = z'_{it}\delta$, it follows from the previous discussion that a necessary condition for the identification of the mean utilities is that $d \leq T(J - 1)$. The general problem of revealed preference does not, however, restrict the functional form of the individual's utility. The previous necessary conditions nevertheless apply to the case in

which the mean utilities are unconstrained provided the parameter vector γ be suitably defined. Indeed, for any i and any t, we have:

$$\bar{U}_{it} = \bar{U}_{It} + (\bar{U}_{it} - \bar{U}_{It}). \tag{49}$$

It is clear that $\bar{U}_{It}$ cannot be identified by observing individual choices since $\bar{U}_{It}$ drops out of the decision rule (25). In other words, for any individual, we cannot identify, as might be expected, the levels of utility but only differences between utilities derived from various states or outcomes. Hence we can introduce the following T arbitrary normalization constraints:

$$\bar{U}_{It} = 0 \quad \text{for any } t = 1,\ldots,T \tag{50}$$

Then, let δ be the (I − 1)T-dimensional vector of which the (i,t)-th component is $\bar{U}_{it}$. If the parameter vector δ is identified, we shall therefore say that all the individuals' mean utilities are identified up to some additive (individual specific) constants. From the previous results, we obtain the following interesting result which, as mentioned earlier, does not depend on the choice of the distribution of the random utility components: a necessary condition that all the individuals' mean utilities be identified up to some additive constants is that $I \leq J$, i.e., that there are at least as many available strategies as possible outcomes.

A similar distribution-free result can also be obtained for the Stackelberg game under uncertainty of Section 3.2 when the leaders' mean utilities are left unconstrained. Specifically, a necessary condition that the leaders' mean utilities be identified up to some additive constants is that there are at least as many available moves for the leaders as available moves for the followers.

There does not seem to be, however, a simple distribution-free, necessary condition for the identification of the followers' mean utilities even though the probability model for the followers' responses is also a probabilistic choice model. This is so because the identification of the structural parameters γ must necessarily involve the probability model for the leaders' responses as stated by condition (42) of the basic lemma.

5. Concluding Remarks

In this paper we have introduced a general model for qualitative endogenous variables defined by a recursive system of probability models in which the probabilities of some outcomes may depend on the probabilities of posterior outcomes. The model was related to C. D. Mallar's simultaneous probability model, and various examples were given to illustrate the generality of the model. In particular, it was shown that the multivariate logit model and the constrained nested logit model are special cases of the model. The model is also readily applicable to the analysis of the outcomes of some game situations in which participants have some sort of rational expectations. Identification of the structural parameters of the model was then studied. General results on structural identification

were derived and used to obtain some simple necessary conditions for the identification of the utilities of the players in the two game examples that were considered.

An issue not addressed so far is the estimation of the structural parameters. An obvious way to proceed is to apply the maximum-likelihood technique, i.e., to maximize the joint log-likelihood function with respect to the unknown structural parameters. Under the usual assumption of (exogenous) random sampling, the joint log-likelihood function is given by:

$$L(\alpha,\beta) = \sum_{t=1}^{T} \log \Pr(A_t = i_t^o, B_t = j_t^o;\alpha,\beta) \tag{51}$$

where i_t^o is the observed A response for the t-th individual, j_t^o is the observed B-response for the t-th individual,

$$\log \Pr(A_t = i_t^o, B_t = j_t^o;\alpha,\beta) = \log \Pr(A_t = i_t^o | B_t = j_t^o,\alpha) + \log \Pr(B_t = j_t^o;\alpha,\beta), \tag{52}$$

and the terms on the right-hand side of (52) satisfy (1) and (2).

A simpler estimation method is to apply the two-step or sequential procedure that was considered by T. Amemiya (1978) in the context of a multivariate logit model. For the general model (1)-(3), this two-step procedure consists first in estimating by the maximum likelihood method the conditional probability model (1) treating as given the observed values of the endogenous variable B, i.e., in maximizing with respect to α the conditional log-likelihood function for A which is given by:

$$L_A(\alpha) = \sum_{t=1}^{T} p_A(i_t^o | j_t^o, Z_{ABt}, \alpha) \qquad (53)$$

This gives an estimate $\hat{\alpha}$ of α. In the second step, an estimate of β is obtained by maximizing the log-likelihood function for the observations on B assuming that $\alpha = \hat{\alpha}$, i.e., by maximizing:

$$L_B(\hat{\alpha}, \beta) = \sum_{t=1}^{T} p_B(j_t^o | \hat{Z}_{Bt}, \beta) \qquad (54)$$

where

$$\hat{Z}_{Bt} = \tilde{Z}_{Bt}(\hat{\alpha}) \qquad (55)$$

(see Equation (3')).

Asymptotic properties of this sequential estimator were studied in Q. H. Vuong (1982b) for the case in which the conditional probability model for A given B is a logit model, and the probability model for B is a logit model given α. In that paper, we also proposed an iterative sequential procedure, i.e., a procedure in which the parameters α and β are sequentially estimated at each iteration, that produces upon convergence an asymptotically efficient estimator of α and β. Although these results were derived for the case in which the probability models of the recursive system belong to the class of logit models, it is however likely that the results hold when the

component models belong to another class of probability models.

Finally, the present paper raises some interesting questions for further research. First, in the Stackelberg game under uncertainty considered earlier, the random utility components of the leader and the follower were assumed to be statistically independent. Thus, an important question is how one would relax such an assumption. Moreover, what would then be the interpretation of the introduced correlation between the players' random utility components within a rational expectation framework? Second, the recursiveness of the general model (1)-(3) implies that the model can be used in situations in which the structure underlying the data has the recursive property. This was certainly the case for the Stackelberg game. Therefore, another question not answered so far is how one would formulate a statistical model appropriate for the analysis of data obtained from situations in which players move simultaneously, such as in a Nash game. A related issue is: what would be the relationship between such a statistical model and the standard simultaneous model.

FOOTNOTES

1. I wish to thank David Grether; our conversations prompted much of the structure of the present model. I am also deeply indebted to Kim Border, Ed Green, John Link, and Marc Nerlove for helpful comments and suggestions.

2. The ramdomness of the variables A_t and B_t can be justified by postulating that the "choices" or responses of the t-th individual are random. A more classical interpretation is that A_t and B_t represent the choices of an individual randomly drawn from a population where observed characteristics are identical to those of the t-th individual.

3. It is important to note that this does not exclude neither the possibility that the sets $\bar{I}_t$ and $\bar{J}_t$ of alternatives in $\bar{I}$ and $\bar{J}$ that are available to the t-th individual depend on the t-th individual (more exactly on the characteristics of the t-th individual) nor the possibility that the set of alternatives in $\bar{I}$ that are available to the t-th individual depends on the t-th individual characteristics <u>and</u> the value of B chosen by the t-th individual. Indeed, this only depends on the specification of the probability functions in (1)-(2) and more precisely on which indices i and j correspond to probabilities a priori equal to zero.

4. Note that we do not necessarily have $a = k_A$. The present

formulation therefore allows for more unknown parameters in α than explanatory variables in the probability model (1). The additional unknown parameters might be due to some unknown variances as in Probit models (see e.g., J. Hausman and D. Wise (1978)) or to some unknown thresholds (see e.g., R. D. McKelvey and W. Zavoina (1975)). For the same reason, we do not necessarily have $b = \tilde{k}_B$, where $\tilde{k}_B$ is the number of explanatory variables in the probability model (2).

5. One may want to consider the more general case where the functions $\tilde{Z}_{Bt}$ are known up to some unknown parameters θ. However, as long as one does not specify how the unknown parameters β interact with Z_{Bt} and the conditional probabilities $Pr(A_t|B_t,t), t = 1,\dots,T$, in order to determine the probabilities $Pr(B_t|t)$, the formulation (2)-(3) is as flexible. In other words, there is no loss of generality in assuming the matrix functions $\tilde{Z}_{Bt}$ to be known.

6. One may argue that, if $\gamma_A = 0$ and if the first (say) explanatory variable in x'_{At} is the dummy variable B_t, then the model (4)-(7) becomes a recursive system with probability feedback. In this case, the probability p_{At} defined by (4) is interpreted as being the <u>conditional</u> probability that A is equal to one. given B_t. This is not. however, what C. D. Mallar's had in mind given his use of the adjective "marginal" throughout his paper. See also footnote 7.

7. To complete the simultaneous model (4)-(7), one may further assume that the variables A and B are independent, as C. D. Mallar does in the second part of his theorem. Let us mention, however, that this second part as well as the formula giving the asymptotic covariance matrix of his two-step estimator (1977, p. 1720, Eq. 8) are both incorrect.

8. In the logit case, global identifiability is equivalent to local identifiability. This simply follows from R. Bowden (1973) identifiability criterion and from the global concavity of the log-likelihood function (see e.g., S. J. Haberman (1974), D. McFadden (1974)). Hence a necessary and sufficient condition for the (global) identification of the parameter vector θ is that the log-likelihood function be strictly concave in θ. One can show that this holds if and only if the matrix $[I_T \otimes U_{IJ}\ ;\ V_{ABT}]$ is full-column rank, where I_T is the $T \times T$ identity matrix. For a direct proof, see Q. H. Vuong (1982a). See also Section 4.

9. The following discussion is similar to that of T. Amemiya (1978).

10 It is worth noting that the following argument applies as well to the recursive system defined by the conditional probability model for B_t given A_t, and the marginal probability model for A_t.

11. To simplify the notations, it is assumed that each of the J subsets of alternatives that are considered at level 2 have the same number I of alternatives. In general, the index set $\bar{I}$ at

the first level depend on the chosen subset j at the second level and on the t-th individual, i.e., $\bar{I} = \bar{I}_{jt}$, while the index set of the subsets at the second level depend on the t-th individual., i.e., $\bar{J} = \bar{J}_t$. See also footnote 3.

12. For another approach to testing the IIA property implied by the logit specification, see e.g., J. Hausman and D. McFadden (1980).

13. The argument used here is similar to, but not quite identical to the argument given in Section 2.1. Here, it is assumed that the parameters (γ,δ,ρ) of the constrained nested logit model are identified. Hence there is a one-to-one correspondence between (γ,δ,ρ) and the admissible probability distributions $\{Pr(A_t|B_t,t)\ ;\ \ t = 1,\dots,T\}$. It thus follows that the parameters γ/ρ of the conditional logit model for A given B are identified. This implies that μ_{jt} is a function of $(Pr(A_1|B_1,1),\dots,Pr(A_T|B_T,T))$.

14. Strictly speaking, in the random demand example, we need to assume that there is only a finite number of possible prices, and that for each price there is only a finite number of sales that are associated with strictly positive conditional probabilities.

15. Note that in the Hausman-Wise formulation, the covariance matrix $\sum_t$ is parameterized by some unknown parameters ω, and also by the unknown parameters δ appearing in (25). In such a case, one has to suitably define the parameter vector β of the model for B. (See also footnote 4.)

(See also footnote 4.)

16. This assumption essentially avoids problems raised by a sample selection bias (see e.g., J. Heckman (1979)). Formulation and estimation of a recursive system with probability feedback and sample selection bias will be considered in future work.

17. Such an assumption can be justified by assuming that leader "s" can (i) observe the individual characteristics of his opponent "r", (ii) knows the monetary attributes (say) of the state (i,j) for all (i,j), and (iii) knows the functional form $F_r(.;.)$ as well as the parameter vectors γ and ω^F. On the other hand, note that the outsider knows neither γ nor ω^F.

18. Actually, Equation (41) holds with probability one, where the probability measure is the one associated with (α^o,β^o). This qualification is, however, automatically satisfied by most qualitative response models, since in these models, the support of any admissible probability distribution is invariant with respect to (α,β).

19. Indeed, it can easily be shown that if log $f(y,\theta)$ is globally concave (resp. strictly concave) in θ, where $f(y,\theta)$ is the joint density of the observations, then the information integral function $H(\theta,\theta^o)$ (see R. Bowden (1973, p. 1070)) is globally concave (resp. strictly concave) in θ. The assertion then follows from R. Bowden's corollary (1973, p. 1071).

20. Such a distinction exists whenever, for some state i, the random component ε_{it} in Equation (24) has a non-zero variance. Then note that the Richter-Shapiro result on the (non) identification of the probabilities implies by duality that a necessary condition for the identification of the mean utilities is that. for some i, the variance of ε_{it} is not null. Indeed, if this were not the case, it would follow from Equation (28) that the variance of $\tilde{\varepsilon}_{jt}$ is null for any j so that Equation (26) would be identical to the one obtained with non-random utilities.

21. For instance, if the random errors $\tilde{\varepsilon}_{jt}$ appearing in Equation (29) can be assumed to be independently and identically extreme-value distributed so that the probability model for B is a logit model, then the necessary rank condition becomes also sufficient for global identification (see footnote 8). See, however, our remarks following Equation (29).

REFERENCES

Amemiya, T.: "On a Two-Step Estimation of a Multivariate Logit Model," Journal of Econometrics,8 (1978), 13-22.

Barron, J. M., and Gilley, O. W.; "Job Search and Vacancy Contact: Note," The American Economic Review, 71 (1981), 747-752.

Bowden, R.: "The Theory of Parametric Identification," Econometrica, 41 (1973), 1069-1074

Domencich, T. A., and McFadden, D.: Urban Travel Demand. Amsterdam: North Holland, 1975.

Haberman, S. J.: The Analysis of Frequency Data. Chicago: University of Chicago Press, 1974.

Hausman, J. A., and McFadden, D.: "A Specification Test for the Multinomial Logit Model," M.I.T. Working Paper, (1980).

Hausman, J. A., and Wise, D. A.: "A Conditional Probit Model for Qualitative choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, 46 (1978), 403-426.

Heckman, J. J.: "Sample Bias as a Specification Error," Econometrica, 47 (1979), 153-161.

Lee, L. F.: "Fully Recursive Probability Models in Multivariate Log-Linear Probability Models for the Analysis of Qualitative

Data," Journal of Econometrics, 16 (1981), 51-69.

Maddala, G. S., and Lee, L. F.: "Recursive Models with Qualitative Endogenous Variables," Annals of Economic and Social Measurement, 5 (1976), 525-545.

Mallar, C. D.: "The Estimation of Simultaneous Probability Models," Econometrica, 45 (1977), 1717-1722.

McFadden, D.: "Conditional Logit Analysis of Qualitative Choice Behavior," Frontiers in Econometrics, Ed. by P. Zarembka. New York: Academic Press, 1974.

____________: "Quantal Choice Analysis: A Survey," Annals of Economic and Social Measurement, 5 (1976), 363-390.

____________: "Econometric Models of Probabilistic Choice," in Structural Analysis of Discrete Date with Econometric Applications. Ed. by C. Manski and D. McFadden. Cambridge, Mass: MIT Press, 1981.

McKelvey, R., and W. Zavoina: "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," Journal of Mathematical Sociology, 4 (1975), 103-120.

Nerlove, M., and Press, S. J.: Univariate and Multivariate Log-Linear and Logistic Models, Santa Monica: Rand Corporation, 1973.

____________: "Multivariate Log-Linear Probability Models for the

Analysis of Qualitative Data," Discussion Paper Number 1, Center for Statistics and Probability, Northwestern University, 1976.

Ottenwaelter, B., and Vuong, Q. H.: "Modèles Conditionnels Log-Linéaires de Probabilité et Systèmes Récursifs," <u>Annales de l'INSEE</u>, 44 (1981), 81-120.

Pratt, J. W.: "Concavity of the Log-Likelihood," <u>Journal of the American Statistical Association</u>, 76 (1981), 103-106.

Richter, M. K., and Shapiro, L.: "Revelations of a Gambler," <u>Journal of Mathematical Economics</u>, 5 (1978), 229-244.

Rothenberg, T. J.: "Identification in Parametric Models," <u>Econometrica</u>, 39 (1971), 577-592.

Vuong, Q. H.: "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application," unpublished PhD Thesis, Northwestern University, 1982.

____________: "Probability Feedback in a Recursive System of Logit Models: Estimation," mimeo, California Institute of Technology, 1982.