GAMES IN ECONOMETRICS

WITH APPLICATIONS TO LABOR ECONOMICS

Thesis by

Paul Anders Bjorn

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1986

(Submitted September 25, 1985)

ABSTRACT

The essential starting point of this dissertation presents an alternative approach for formulating simultaneous equation models for qualitative endogenous variables. To be explicit, the endogenous variables will be generated as Nash equilibria of a game between two players, and the statistical model will be generated by invoking the random utility framework introduced by McFadden (1974, 1981). Contrary to the earlier simultaneous equations models (Heckman (1978)), the approach presented in Chapter II will not impose logical consistency constraints on the parameters. A distinctive feature of the model is that it extends the usual simultaneous model with structural shift to cases where the parameters need not satisfy the logical consistency conditions.

Following the game theoretic formulation set out in Chapter II, Chapter III proposes an alternative model where the equilibrium concept is that of Stackelberg. As in Chapter II, we will still assume that each player maximizes his own utility, with the statistical model again being derived using McFadden's random utility approach. A distinctive feature of this model is that it contains as a special case the usual recursive model for discrete endogenous variables.

With Chapters II and III as a theoretical background, the purpose of Chapter IV is to present an empirical study of the Nash and Stackelberg equilibrium models. The problem we examine concerns a married couple's joint decision whether or not to participate in the

labor market. We examine three competing specifications. Chapter V concludes this dissertation with a discussion of which of the three empirical models most adequately describes the joint labor force participation decision of a random sample of married couples. Since none of the three models are completely nested in each other, we are not able to employ any of the classical tests. As such, we use an alternative method developed by Vuong (1985) for choosing the most adequate model.

ACKNOWLEDGEMENTS

It is with pleasure that I thank those individuals who are largely responsible for my writing of this dissertation. First, and foremost, I thank my chairman, Quang Vuong, for providing the initial ideas for this dissertation. In addition, his patience, careful guidance and help over the last couple of years will always be appreciated. I also wish to thank David Grether and Richard McKelvey for valuable comments on various drafts of Chapters II and III. A special debt of gratitude goes to two of my classmates, Donald Lien and Ken McCue. Don was always more than willing to patiently answer what has seemed to be an endless number of questions. His ability to think precisely and intuitively always continues to amaze me. Ken's ablity with computers is unparalleled; the number of times he has freely provided computer assistance is too many to mention. Most of all, though, I would like to thank these two individuals, along with Jack Williamson and Tatsuyoshi Saijo, for providing levity and friendship when I needed it the most.

A special note of thanks goes to my typists over the last couple of years, Madeline Haddad, Linda Donnelly and especially Kathryn Kurzweil who was responsible for typing the last part of this dissertation. Her willingness to work late to meet my deadlines is appreciated. I would also like to express my thanks to our graduate secretary, Chris Smith, for her friendship and administrative support over the last four years and the National Science Foundation for financial support through grant SES-8410593 to Quang Vuong.

In addition, I would like to acknowledge my parents. Although they were never quite sure what I was doing in graduate school, my ability to endure is fundamentally due to them. Finally, I owe a debt of gratitude to Barbara Alexander. While perhaps not offering academic advice, her encouragement and willingness to put up with my schedule is greatly appreciated.

CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER I: INTRODUCTION

Over the last few decades, economists have become increasingly interested in studying economic decisions involving choice over a finite number of alternatives. For example, frequently analyzed behavioral phenomena are decisions over brands of consumer durables (see, e.g., Dubin and McFadden (1984) and Cragg and Uhler (1970)), travel mode (see, e.g., Domencich and McFadden (1975) and Hausman and Wise (1978)), and labor force participation (see, e.g., Gunderson (1974) and Parsons (1980)). (See Amemiya (1981) for a large list of empirical papers which use discrete models.) As a first step in modeling such examples, it is natural to consider a univariate random variable taking on two or more discrete values. For instance, we may consider a random variable taking on only two outcomes, those being whether to drive to work or take the bus; alternatively, we may describe a random variable whose two outcomes index an individual's decision whether or not to participate in the labor market. But now consider a more complicated problem in which an individual must decide not only whether to take the bus or drive to work but also whether to travel during rush hours or off-rush hours; or the decision of a husband and wife whether or not to each participate in the labor market. Certainly both of these problems involve joint decisions and should be modeled as multivariate (jointly dependent) qualitative random variables. Typically, we then specify a system of simultaneous equations in the form of a structural model. Such a system is simultaneously determined in the sense that the interaction of all variables as specified by the model determines simultaneously the entire set of jointly dependent qualitative variables.

At this point it may be useful to recall what we mean by a structural model.

> "In the fundamental economic sense, the property that characterizes a structural system is the (truth of the) assertion that it describes accurately (or adequately) the precise fashion in which all the current endogenous and predetermined variables mutually interact within the specified economic system. Thus, as of a given moment of time, observations on the system represent the result of such interaction."
>
> P.J. Dhrymes (1974)

If these observations on the system represent the result of interaction among individuals or firms, then it must be recognized that these interactions will most likely be characterized by strategic behavior. Therefore, it is incumbent that this strategic behavior be incorporated directly into econometric models of individual choice using the concepts of game theory. Until now, this has not been done and will be the focus of this dissertation. To be more precise, the major purpose of this dissertation is to present an alternative approach for formulating simultaneous equations models for qualitative endogenous variables. The endogenous variables will be generated successively as Nash and Stackelberg equilibria of a game played between two individuals, and the statistical model will be generated by invoking the random utility framework introduced by McFadden (1974, 1981). We then apply these models to a labor force participation problem. Finally, we propose a test to choose the model which most accurately generates the observations. Hence, the title of this dissertation.

Before developing the ideas in this dissertation, it is first necessary to review the literature on qualitative multivariate random variables. This literature has essentially developed in two separate directions by those who assume an underlying unobserved or latent random variable, and those who make no such assumption but instead follow the loglinear model approach to analyze contingency tables. As Fienberg (1975) points out, this controversy dates back to the beginning of the 20th century with Pearson (1900) advocating the former approach and Yule (1900) advocating the latter. We will discuss the latter approach first.

The distribution of a set of K dichotomous random variables is completely described by the 2^K joint cell probabilities $Pr(Y_1,Y_2,\ldots,Y_K)$ in the K dimensional contingency table. In the case of K = 2, for example with Y_1 and Y_2 taking on values one or zero, the four probabilities $Pr(Y_1 = y_1, Y_2 = y_2)$ may be described in the following table.

		Y_2: 1	0
Y_1	1	Pr(1,1)	Pr(1,0)
	0	Pr(0,1)	Pr(0,0)

The probabilities must sum to one and each must lie in the unit interval.

One representation of this problem is known as the loglinear model (LLM) (see, e.g., Bishop, Fienberg and Holland (1975), Goodman (1972) and Haberman (1974)). For the case of $K = 2$, with Y_1 and Y_2 each taking on the values 1 and 0, the loglinear model is given by

$$\log[\Pr(Y_1, Y_2)] = u_0 + u_1(2Y_1-1) + u_2(2Y_2-1) + u_{12}(2Y_1-1)(2Y_2-1). \quad (1)$$

Note that the $2^K = 2^2 = 4$ cell probabilities are fully represented by the four single and double subscripted u terms. In this formulation, the terms u_1 and u_2 are commonly known as "main effects" and the u_{12} term is known as the "second-order interaction effect". Since the four probabilites must sum to one, the standard normalization imposed on the overall mean, u_0, is

$$e^{-u_0} = \sum_{Y_1} \sum_{Y_2} \exp[u_1(2Y_1-1) + u_2(2Y_2-1) + u_{12}(2Y_1-1)(2Y_2-1)]. \quad (2)$$

While the loglinear model may be successful in describing the outcomes in a purely scientific experiment, it is of little use when one of the purposes of statistical estimation is to predict the effect of changes in economic policy or some other structural change. Indeed, the loglinear model does not make the distinction between endogenous and exogenous variables; as such, it has been described by some (see, e.g. Maddala (1983)) as more a correlation model than a causal model. It is probably for this reason that the loglinear model has been little used in economics. To this end, the conditional loglinear model (CLLM) has been put forth as a model which purports to more adequately explain causal relationships (see, e.g., Nerlove and Press (1973, 1976) and Vuong (1982b). Given that we may want to explain one variable given advance knowledge of another variable, a formulation in terms of conditional probabilities seems particularly appealing. With this in mind, the conditional probability of Y_1 given Y_2 is written as

$$\Pr(Y_1 \mid Y_2) = \frac{\Pr(Y_1,Y_2)}{\Pr(1,Y_2) + \Pr(0,Y_2)}$$

$$= \frac{\exp(u_1 Y_1 + u_{12} Y_1 Y_2)}{\exp(u_1 + u_{12} Y_2) + \exp(-u_1 - u_{12} Y_2)},$$

where we have used the loglinear specification of the joint probability of (Y_1,Y_2) as given in (1) and the normalization given in (2). A representation that is more commonly used is the log-odds ratio:

$$L_{1|2} = \log \left[\frac{\Pr(Y_1 = 1 | Y_2)}{\Pr(Y_1 = 0 | Y_2)} \right] = 2u_1 + 2u_{12} Y_2 = w_0^1 + w_2^1 Y_2. \qquad (3)$$

For completeness, we have a similar expression for the probability of Y_2 given Y_1:

$$L_{2|1} = \log \left[\frac{P(Y_2 = 1 | Y_1)}{P(Y_2 = 0 | Y_1)} \right] = 2u_2 + 2u_{12} Y_1 = w_0^2 + w_1^2 Y_1. \qquad (4)$$

Because of the similarity between equations (3) and (4) and the structural equations in a simultaneous model for continuous variables, some authors have been led to interpret these conditional loglinear equations as structural (see, e.g., Schmidt and Strauss (1975a,b)). To do this is misleading, for a couple of reasons (see, e.g., Nelson (1979) for an alternative discussion). First, and foremost, as opposed to the situation in classical simultaneous models where the endogenous variables are determined jointly, the derivation at equations (3) and (4) is based on conditional distributions. Second, note the implicit cross-equation parameter constraint in equations (3) and (4), viz, $w_2^1 = w_1^2$. Thus we are required to live with the unappealing and restrictive notion that the effect of Y_1 on Y_2 is equal to the reverse effect of Y_2 on Y_1. We are thus led to the conclusion that the conditional loglinear model has no more of a structural interpretation than the loglinear model; indeed, this should not be surprising given that the latter model is nothing more than an alternative representation of the former model. A further attempt to model causal relationships is given by the recursive logistic model (RLM), the topic to which we now turn.

Unlike the usual simultaneous equations model where it is not possible to interpret each equation as a conditional expectation except in the case of a fully recursive model, we find that the formulation of multivariate qualitative variables in a logistic framework allows us to do precisely this. But as we will shortly see, recursive logistic models suffer from the same drawbacks as conditional loglinear models. Let us first examine models that are only partially recursive. Consider the case of three binary variables Y_1, Y_2 and Y_3 where we posit that Y_1 is determined first, and that Y_1 then effects both Y_2 and Y_3 which are determined jointly. Then the model is given as $Pr(Y_1,Y_2,Y_3) = Pr(Y_1) \cdot Pr(Y_2,Y_3|Y_1)$. Since it is easily shown that all parameters appearing in $Pr(Y_2,Y_3|Y_1)$ also appear in $Pr(Y_2|Y_1,Y_3)$ and $Pr(Y_3|Y_1,Y_2)$, we will use these latter two conditional probability expressions. Then expressing each of the three probability components as log-odds, we have:

$$L_1 = \log\left[\frac{Pr(Y_1 = 1)}{Pr(Y_1 = 0)}\right] = 2u_1, \tag{5}$$

$$L_{2|13} = \log\left[\frac{Pr(Y_2 = 1|Y_1,Y_3)}{Pr(Y_2 = 0|Y_1,Y_3)}\right]$$
$$= 2u_2 + 2u_{12}Y_1 + 2u_{23}Y_3 + 2u_{123}Y_1Y_3, \tag{6}$$

and

$$L_{3|12} = \log\left[\frac{Pr(Y_3 = 1|Y_1,Y_2)}{Pr(Y_3 = 0|Y_1,Y_2)}\right]$$
$$= 2u_3 + 2u_{13}Y_1 + 2u_{23}Y_2 + 2u_{123}Y_1Y_2, \tag{7}$$

where the u terms appearing in the above three equations refer to the coefficients in the loglinear model. But now note from equation (6) that the coefficient attached to the Y_3 term is the same as the coefficient attached to the Y_2 term in equation (7), viz, $2u_{23}$. Therefore we see, just as in the conditional loglinear model, that the effect of Y_2 on Y_3 is identical to the opposite effect of Y_3 on Y_2.

We now turn to fully recursive logistic models.

Let us now alter the example used in the discussion of partially recursive models so that now Y_1 is determined first, that Y_1 determines Y_2 and that Y_1 and Y_2 now jointly determine Y_3. Thus we need to examine the marginal distribution of Y_1, the conditional distribution of Y_2 given Y_1, and finally the conditional distribution of Y_3 given the joint determination of Y_1 and Y_2. Again, expressing each of the three probability components as log-odds ratios, we have:

$$L_1 = \log\left[\frac{\Pr(Y_1 = 1)}{\Pr(Y_1 = 0)}\right] = 2u_1, \tag{8}$$

$$L_{2|1} = \log\left[\frac{\Pr(Y_2 = 1|Y_1)}{\Pr(Y_2 = 0|Y_1)}\right] = 2u_2 + 2u_{12}Y_1, \quad \text{and} \tag{9}$$

$$L_{3|12} = \log\left[\frac{\Pr(Y_3 = 1|Y_1,Y_2)}{\Pr(Y_3 = 0|Y_1,Y_2)}\right]$$

$$= 2u_3 + 2u_{13} + 2u_{23}Y_2 + 2u_{123}Y_1Y_2. \tag{10}$$

Although it is seen that there are no cross-equation constraints appearing in equations (8) through (10), in contrast to the previous two models, this model is inadequate unless the process generating the observations is truly recursive. Moreover, even if the true process is recursive, the model still does not have a structural interpretation in the sense that we are modeling the conditional determination of the outcomes rather than their joint determination. For this reason, a number of authors feel that a more useful approach in modeling multivariate qualitative variables is to assume the existence of underlying continuous response functions. Since the general model introduced by Heckman (1978) considers a number of other models as special cases, we will confine ourselves to a discussion of his model (see also Maddala and Lee (1976)).

Heckman considers the following two equation model

$$Y^*_{1t} = X_{1t}\alpha_1 + \beta_1 Y_{2t} + Y^*_{2t}\gamma_1 + \varepsilon_{1t} \tag{11}$$

$$Y^*_{2t} = X_{2t}\alpha_2 + \beta_2 Y_{1t} + Y^*_{1t}\gamma_2 + \varepsilon_{2t} \tag{12}$$

where Y^*_{1t} and Y^*_{2t} are continuous latent random variables and Y_{1t} and Y_{2t} are observed qualitative variables generated using the following dichotomization:

$$Y_{1t} = \begin{matrix} 1 \text{ iff } Y^*_{1t} > 0 \\ 0 \text{ otherwise} \end{matrix} \tag{13}$$

$$Y_{2t} = \begin{matrix} 1 \text{ iff } Y^*_{2t} > 0 \\ 0 \text{ otherwise} \end{matrix} .$$

By making specific assumptions on β_1 and β_2, and on Y^*_{1t} and Y^*_{2t}, a number of different models arise from the general model specified by equations (11) - (13). For example, if Y^*_{1t} and Y^*_{2t} are both observed and if β_1 and β_2 are both zero, we have the classical simultaneous equations model. Second, it may be the case that while we only observe whether the event $Y^*_{2t} \gtrless 0$, we may in fact observe Y^*_{1t}; Y^*_{1t} is therefore treated as a continuous random variable. Heckman calls this the "Hybrid Model with Structural Shift." Naturally, then we have the simple "Hybrid Model" when no structural shift is permitted; this is when $\beta_1 = \beta_2 = 0$. Finally, when $\gamma_1 = \gamma_2 = 0$ and only the events $Y^*_{1t} \gtrless 0$ and $Y^*_{2t} \gtrless 0$ are observed, we have a bivariate probit model with structural shift. When, in addition, $\beta_1 = \beta_2 = 0$, we have the simple bivariate probit model of Ashford and Sowden (1970). Since all the models discussed above exhibit similar characteristics, we will confine the discussion to the case of the bivariate probit model with structural shift.

As is well-known, all simultaneous equations models of dummy endogenous variables that have the realization of Y^*_{1t} and Y^*_{2t}, namely Y_{1t} and Y_{2t}, appearing as right-hand side variables must satisfy a constraint on the parameters so as to insure that the probabilities add up to one. This restriction on the parameters is commonly known as the logical consistency condition; for the case at hand, the restriction is that $\beta_1 \cdot \beta_2 = 0$. (See Heckman (1978) and Maddala (1983); see also Schmidt (1981) for a general discussion of logical consistency conditions required for various forms of simultaneous equations probit and tobit models.) As will be seen later, the implications of the logical consistency condition for bivariate probit models with structural shift are quite important and striking.

Just as we have previously argued that the loglinear model and its conditional loglinear and recursive logistic variants are not adequate for formulating structural econometric models, Heckman argues that the bivariate probit model with structural shift is superior to the loglinear model in the sense that the loglinear model is not sufficiently rich in parameters to distinguish structural association among discrete random variables from purely statistical association among discrete random variables. As we shall see, Heckman's argument differs slightly from the arguments put forth above.

To fix ideas, consider again the loglinear model for two dichotomous variables

$$\log[\Pr(Y_1,Y_2)] = u_0 + u_1(2Y_1-1) + u_2(2Y_2-1) + u_{12}(2Y_1-1)(2Y_2-1).$$

Then the four joint probabilities generated by this model give us the following set of equations:

$$\log[\Pr(1,1)] = u_0 + u_1 + u_2 + u_{12},$$

$$\log[\Pr(1,0)] = u_0 + u_1 - u_2 - u_{12},$$

$$\log[\Pr(0,1)] = u_0 - u_1 + u_2 - u_{12},$$

$$\log[\Pr(0,0)] = u_0 - u_1 - u_2 + u_{12},$$

where again $u_0 = -\log[\exp(u_1 + u_2 + u_{12}) +$
$\exp(u_1 - u_2 - u_{12}) + \exp(-u_1 + u_2 - u_{12}) + \exp(-u_1 - u_2 + u_{12})]$
so as to insure that the individual probabilities add up to one. Now note that in the loglinear model, the conditional probability that $Y_1 = 0$ given Y_2 may be written as

$$\Pr(Y_1 = 0|Y_2) = \frac{1}{(1 + e^{-2u_1})e^{(2u_{12}Y_2 - 2u_{12}(1-Y_2))}}.$$

Now let us consider comparable expressions for Heckman's bivariate probit model with structural shift. Repeating again the two equation model, and omitting the subscript t, we have

$$Y_1^* = X_1\alpha_1 + \beta_1 Y_2 + \varepsilon_1$$

$$Y_2^* = X_2\alpha_2 + \beta_2 Y_1 + \varepsilon_2$$

Now let us impose the logical consistency condition that $\beta_2 = 0$, say. Then, again, the four joint probabilities generated by this model give us the following set of equations

$$\Pr(0,0) = F(-X_1\alpha_1, -X_2\alpha_2, \rho),$$

$$\Pr(0,1) = F(-X_1\alpha_1 - \beta_1, X_2\alpha_2, -\rho),$$

$$\Pr(1,0) = F(X_1\alpha_1, -X_2\alpha_2, -\rho),$$

$$\Pr(1,1) = F(X_1\alpha_1 + \beta_1, X_2\alpha_2, \rho),$$

where $F(\cdot,\cdot, \rho)$ is the bivariate normal cumulative density function with correlation coefficient ρ. Then using the four equations above, we can write the conditional probability that $Y_1 = 0$ given Y_2 as

$$\Pr(Y_1 = 0|Y_2) = \left[\frac{F(-X_1\alpha_1 - \beta_1, X_2\alpha_2, -\rho)}{F(\infty, -X_2\alpha_2, -\rho)}\right]^{Y_2} \cdot \left[\frac{F(-X_1\alpha_1, -X_2\alpha_2, \rho)}{F(\infty, -X_2\alpha_2, \rho)}\right]^{1-Y_2}.$$

Heckman now wishes to point out that while u_1 and $X_1\alpha_1$, u_2 and $X_2\alpha_2$ play similar roles in the models in which they appear, it is important to note that u_{12} and ρ <u>and</u> β_1 also play similar roles. In the bivariate probit model with structural shift, the probability that $Y_1 = 0$ given Y_2 depends on Y_2 for two conceptually distinct reasons: one related to the true structure of the model ($\beta_1 \neq 0$) and the other due to covariance in latent errors ($\rho \neq 0$). In the loglinear model, these two effects are indistinguishable. Thus the loglinear parameter of association, u_{12}, corresponds to two distinct parameters in the bivariate probit model with structural shift, ρ and β_1. Although one must judge for himself whether or not the preceding argument is convincing, two further points deserve mention.

First, while both the loglinear model and the Heckman model each contain a parameter which attempts to capture the association between the random variables Y_1 and Y_2, both these parameters are subject to restrictions; these restrictions call into question the usefulness of both models. As recalled from above, both the conditional loglinear model and the partially recursive logistic model contain the implicit cross-equation parameter constraint which requires that the effect of Y_2 on Y_1 be equal to the reverse effect of Y_2 on Y_1. In the Heckman model, on the other hand, although the shift parameters are included as an attempt to explicitly model the structural association between the random variables Y_1 and Y_2, it will later be shown that the required logical consistency condition implies that the effect of Y_1, say, on Y_2 is structurally independent from the reverse effect; this point was alluded to by Schmidt (1981).

The second point concerns the Heckman model in particular. With rare exception (Waldman (1981)), the required logical consistency conditions do not have an economic interpretation. That is, there is usually no a priori reason to impose that either β_1 or β_2 must be zero.

With this in mind, the purpose of Chapter II of this dissertation is to present an alternative approach for formulating simulataneous equations models for qualitative endogenous variables. To be explicit, the endogenous variables will be generated as Nash equilibria of a game between two individuals, and the statistical model will be generated by invoking the random utility fromework introduced by McFadden (1974, 1981). (See Heckman (1978) p. 954, for a discussion of the random utility framework in bivariate probit models with structural shift.) Contrary to the earlier simultaneous equations models, the approach presented in Chapter II will not impose logical consistency constraints on the parameters.

Moreover, a distinctive feature of the model is that it extends the usual simultaneous model with structural shift to cases where the parameters need not satisfy the logical consistency conditions. Indeed, when the logical consistency conditions are imposed, the model coincides with the usual simultaneous equations model with structural shift. This will provide a structural interpretation to the usual dichotomization.

With this as a background, Chapter II is organized as follows.

Section 1 provides a short introduction to the Chapter. In Section 2, we show that the logical consistency condition implied by simultaneous equations models with structural shift requires these models to be recursive. Then in Section 3, our game theoretic approach will suggest a natural rule for generating the observed dichotomous variables, other than the dichotomization used in the literature up until now. The resulting statistic model is derived, and it is seen that no logical consistency conditions are implied on the parameters. In Section 4, identification of the model is discussed along with problems of estimation that relate particularly to identification. Section 5 concludes the paper. Proofs of all propositions are found in the text of the Chapter itself. The first partial derivatives of the expressions for the probabilities are found in the Appendix to Chapter II, as they will be needed both for identification and maximization of the likelihood function.

Following the game theoretic formulation set out in Chapter II where the discrete dichotomous random variables are generated as Nash equilibria of a game played by two individuals, Chapter III proposed an alternative model where the equilibrium concept used is that of Stackelberg. As in Chapter II, we will still assume that each player maximizes his own utility, with the statistical model again being derived using the random utility approach suggested by McFadden (1974, 1981). A distinctive feature of our model is that it contains as a special case the usual recursive model for discrete endogenous variables (see, e.g., Maddala and Lee (1976)). A structural interpretation of this latter model can then be given in terms of a Stackelberg game in which the leader is indifferent to the follower's actions.

With this as a brief background, Chapter III is organized as follows. Section 1 provides a brief introduction to the Chapter. In Section 2, we derive the statistical model where the outcomes are generated as Stackelberg equilibria of a game played between two players. Section 3 compares the usual formulation of the problem in terms of recursive models with our alternative formulation. In particular, it is shown that the usual recursive model is nested in our more general model. In Section 4, we discuss identification of the model and estimation issues as they relate to identification. Section 5 concludes the Chapter. Proofs of all propositions are found in the text of the Chapter. The first partial derivatives of the probabilities, which are needed both for identification and estimation of the model, are found in the Appendix to Chapter III.

With Chapters II and III as a theoretical background, the purpose of Chapter IV is to present an empirical study of the Nash and Stackelberg equilibrium models. The problem we will examine concerns a married couple's joint decision whether or not to participate in the labor market. As such, we will examine three competing specifications. First, we will assume that the joint work decision is the outcome of a Nash game played by the husband and wife. Second, we will assume that the married couple plays a Stackelberg game in which the husband plays the role of the leader, while his wife plays the role of the follower. Third, we will specify another Stackelberg game in which the roles of the leader and follower are reversed.

We feel that an examination of this problem will yield two useful results. First, it will demonstrate that the game theoretic models we have proposed are in fact empirically tractable. Second, we feel that the proposed study will make a contribution to the literature on labor force participation because we explicitly model the behavior of a married couple in a game theoretic framework, while previous work has either taken the husband's decision whether or not to work as exogenous (see, e.g., Heckman (1974), Heckman and Macurdy (1980)), or specified the labor supply of a husband and wife from the outcome of a joint utility function (see, e.g., Ashenfelter and Heckman (1974), Cotterman (1981), and Gronau (1973)).

After one reads Chapters II and III, it will hopefully be clear that the structures of the Nash and Stackelberg models contain some similarities. As such, the similarities in the structure of the models will naturally translate into similarities in the specification of the three applications. Thus, after a brief introduction in Section 1, these similarities will be discussed all at once in Section 2. Section 3 will discuss peculiar features of the structure of the Nash model along with the empirical results of the problem under study. Section 4 will discuss the Stackelberg model for the case in which the husband plays the role of the leader while his wife plays the role of the follower. Empirical results will also be presented. Section 5 discusses the third empirical example where the roles of the husband and wife are reversed. Section 6 provides a brief conclusion. A description of the data set used in the three empirical studies is included as an Appendix to Chapter IV. As such, the Appendix will discuss the source and description of the original data set, the selection criteria we used in choosing the appropriate set of observations, and the means and variances of the explanatory variables.

With Chapters II through IV as a background, it is natural to conclude this dissertation with a discussion of which of the three empirical models most accurately describes the joint labor force participation decision for the random sample of married couples. Chapter V proposes to do precisely this. As will be made explicit in Chapter V, none of the three models are completely nested in each other; as such, we will not be able to employ any of the classical tests, viz, the Wald test, the Lagrange Multiplier test or the likelihood ratio test. To get around this difficulty, two alternative methods have been developed over the last couple of decades, those being the techniques of model selection as proposed first by Akaike (1973, 1974), and tests of non-nested hypotheses first proposed by Cox (1961, 1962). Since neither of these two methods has been widely employed in the empirical econometrics literature, it is therefore necessary to present a brief review of the literature on both of them.

With this as a brief background to the problem of choosing the most adequate model, Chapter V is organized as follows. Section 1 presents a brief introduction. Sections 2 and 3 present a review of the literature on model selection and tests of non-nested hypotheses, respectively. As will be explicitly pointed out in both sections, major difficulties are encountered in carrying out either of the two methods. As such, an alternative method will be employed which deals both with model selection and tests of non-nested hypotheses. Although this method is developed fully in Vuong (1985), it will be discussed briefly in Section 4. Using the method suggested in Section 4, Section 5 will attempt to choose the most adequate of the three proposed models. Section 6 presents a brief conclusion.

Upon the completion of Chapter V, two separate appendices will be found. Appendix A presents the documentation and the computer program used to estimate the Nash model. Similarly, Appendix B presents the documentation and the computer program used to estimate both specifications of the Stackelberg model.

CHAPTER II: AN ECONOMETRIC MODEL OF A NASH GAME

1. INTRODUCTION

Over the last decades, economists have been increasingly interested in studying economic decisions involving choice among a finite number of alternatives. For instance, frequently analyzed behavioral phenomena are decisions on labor force participation, travel modes, and brands of commodity purchases. A reason for such a trend may lie in the increasing availability and quality of large microdata sets. Since the behavioral phenomena of interest were qualitative in nature, new statistical models such as the by-now well-known logit model were introduced in econometrics (McFadden (1974), Nerlove and Press (1973, 1976)).

Following the development of the standard linear simultaneous equations model (Koopmans and Hood (1953)), the literature on discrete variables models has rapidly evolved in simultaneous modeling. In particular, the first simultaneous models that were proposed were directly issued from the standard linear simultaneous equations model. Specifically, the models were formulated in terms of a linear simultaneous equations model in latent continuous variables from which the observed qualitative variables were generated using a dichotomization such as $Y_{1t} = 1$ if $Y^*_{1t} > 0$, and $Y_{1t} = 0$ if $Y^*_{1t} \leq 0$ (see e.g., Maddala and Lee (1976), Nelson and Olson (1978)).

More recently, starting with Amemiya (1974) and Heckman (1978), a new generation of simultaneous models for qualitative/truncated variables was introduced where the underlying simultaneous equations models were formulated in terms of both latent continuous variables and observed qualitative/truncated variables. These models are often called simultaneous models with structural shift. As is well-known, however, a major difficulty that arises with these latter models is that they require the parameters to satisfy some logical consistency conditions in order that the models be statistically meaningful (see, e.g., Heckman (1978), Gourieroux, Laffont, and Monfort (1980), and Schmidt (1981)).

With some rare exceptions (Waldman (1981)), the logical consistency conditions implied by the simultaneous models with structural shift do not have an economic interpretation. This fact explains the few applications of these models in economics. Moreover, as Schmidt (1981) has pointed out, when the exogenous variables are qualitative, the associated logical consistency conditions imply some recursivity in the simultaneous equation models. Although Maddala (1983, Sections 5.7 and 5.8) has warned against the mechanistic formulation of simultaneous models with latent continuous variables, and has argued that in many cases an alternative model without logical consistency conditions can be specified, the previous models are nevertheless the only ones available in the literature that have both latent continuous variables and observed qualitative variables in the equations.

The purpose of this chapter is to present an alternative approach for formulating simultaneous equations models for qualitative endogenous variables. For the simple model that we shall propose, both latent continuous variables and observed dichotomous variables will appear in the equations. To be explicit, the observed endogenous variables will be generated as Nash equilibria of a game between two individuals, and the statistical model will be generated by invoking the random utility framework introduced by McFadden (1974, 1981). Contrary to earlier simultaneous models, however, our approach will not impose logical consistency constraints on the parameters.

A distinctive feature of our model is that it extends the usual simultaneous model with structural shift to cases where the parameters need not satisfy the logical consistency conditions. Moreover, when the logical consistency conditions are imposed, our model coincides with the usual simultaneous model with structural shift. This provides a structural interpretation to the usual dichotomization.

The chapter is organized as follows. In Section 2, we shall show that the logical consistency conditions implied by simultaneous equation models with structural shift actually rule out simultaneity for the simple problem that we consider. Since simultaneity is an inherent feature of the problem, it follows that these models are inappropriate. Then in Section 3, our game-theoretic approach will suggest a natural rule for generating the observed dichotomous variables, other than the dichotomization rule used in the literature up to now. The resulting statistical model is derived, and it is seen that no logical consistency conditions are implied on the parameters. Section 4 discusses identification of the model and a few estimation problems as they relate to identification. Section 5 concludes the chapter. As will be discussed in Section 3, a difficulty arises when there do not exist unique pure strategy Nash equilibria. As a result, Appendix A of this chapter provides a justification of how we handle this difficulty. The first partial derivatives of the probabilities are found in Appendix B of this chapter, as they will be needed for both identification and estimation of the model.

2. SIMULTANEOUS EQUATIONS MODELS WITH STRUCTURAL SHIFT: A CRITIQUE

To simplify the discussion, we shall restrict our attention to the case where there are only two qualitative endogenous variables. To simplify further, we shall assume that these variables are dichotomous. As an illustration, it will be convenient to consider the decision to participate in the labor force by a husband and his wife:

$$\begin{aligned} Y_h &= 1 \quad \text{if the husband works,} \\ &= 0 \quad \text{otherwise,} \\ Y_w &= 1 \quad \text{if the wife works,} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

(The subscript indexing the observations is omitted in this section and the following one.)

Following the classical tradition in economics, we shall nevertheless postulate that each individual maximizes his or her own utility function. The decisions of the husband and wife need not, however, be independent since the utility derived by each individual naturally depends on the action taken by the other. Let $\tilde{U}_h(i,j)$ be the utility that the husband derives from taking action i if his wife takes action j where i = 1 if the husband works and 0 otherwise, and j = 1 if the wife works and 0 otherwise. The utility $\tilde{U}_w(j,i)$ derived by the wife from taking action j if her husband takes action i is defined similarly.

To generate the observed dichotomous variables Y_h and Y_w, the rule that has been used in the literature on simultaneous equations models for dummy endogenous variables is summarized in the following assumption.

ASSUMPTION A1:

$$
\begin{aligned}
Y_h &= 1 \quad \text{if } Y_h^* > 0, \\
&= 0 \quad \text{otherwise,} \\
Y_w &= 1 \quad \text{if } Y_w^* > 0, \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

where:

$$
\begin{aligned}
Y_h^* &= Y_w[\tilde{U}_h(1,1) - \tilde{U}_h(0,1)] + (1 - Y_w)[\tilde{U}_h(1,0) - \tilde{U}_h(0,0)], \\
Y_w^* &= Y_h[\tilde{U}_w(1,1) - \tilde{U}_w(0,1)] + (1 - Y_h)[\tilde{U}_w(1,0) - \tilde{U}_w(0,0)].
\end{aligned}
$$

The rationale for this model is the following. If the wife works ($Y_w = 1$), then the husband works ($Y_h = 1$) if and only if $\tilde{U}_h(1,1) \geq \tilde{U}_h(0,1)$.[1] On the other hand, if the wife does not work ($Y_w = 0$), the husband works if and only if $\tilde{U}_h(1,0) \geq \tilde{U}_h(0,0)$. Combining these two conditions, it follows that $Y_h = 1$ if and only if $Y_h^* > 0$ where Y_h^* is defined as above. The decision rule for Y_w is similarly derived.

Following McFadden (1974, 1981), the utilities $\tilde{U}_h(i,j)$ and $\tilde{U}_w(j,i)$ are treated as random, and decomposed into deterministic components and random components:

$$\tilde{U}_h(i,j) = U_h(i,j) + \eta_h(i,j),$$
$$\tilde{U}_w(j,i) = U_w(j,i) + \eta_w(j,i).$$

To complete the specification of the statistical model, assumptions must be made on the unobserved random components $\eta_h(i,j)$ and $\eta_w(j,i)$. From the expressions for Y_h^* and Y_w^*, it is clear that only the differences $\eta_h(1,1) - \eta_h(0,1)$, $\eta_h(1,0) - \eta_h(0,0)$, $\eta_w(1,1) - \eta_w(0,1)$, and $\eta_w(1,0) - \eta_w(0,0)$ are relevant. We make the following simplifying assumption:

ASSUMPTION A2:

$$\eta_h(1,1) - \eta_h(0,1) = \eta_h(1,0) - \eta_h(0,0) = \varepsilon_h,$$
$$\eta_w(1,1) - \eta_w(0,1) = \eta_w(1,0) - \eta_w(0,0) = \varepsilon_w.$$

where the pair $(\varepsilon_h, \varepsilon_w)$ is normally distributed with zero means, unit variances, and correlation ρ.

To simplify further, we shall assume that the difference in utility that the husband derives from working versus not working, when the wife works, differs only by a constant β_h from the utility he derives from working versus not working when the wife does not work. A similar simplifying assumption is made for the wife.[2] Formally, we have:

ASSUMPTION A3:

$$U_h(1,1) - U_h(0,1) = \beta_h + U_h(1,0) - U_h(0,0) \equiv \beta_h + \Delta_h,$$
$$U_w(1,1) - U_w(0,1) = \beta_w + U_w(1,0) - U_w(0,0) \equiv \beta_w + \Delta_w.$$

Using Assumptions A2 and A3 in the expressions for Y_h^* and Y_w^*, it follows that:

$$Y_h^* = \Delta_h + \beta_h Y_w + \varepsilon_h, \tag{1}$$
$$Y_w^* = \Delta_w + \beta_w Y_h + \varepsilon_w. \tag{2}$$

Given the previous assumptions, our model is a simultaneous equations model with structural shift (Heckman (1978), Schmidt (1981)). From Schmidt's condition 12.6, it follows that, for the model to be statistically meaningful, the following logical consistency condition must hold:

$$\beta_h \cdot \beta_w = 0, \tag{3}$$

i.e., either β_h or β_w must be zero.

In the simple problem considered here, the logical consistency condition can readily be interpreted. For instance, "$\beta_h = 0$" means that, apart from the statistical association between ε_h and ε_w, and hence between ε_h and Y_w, the husband's decision on labor force participation does not depend on the actual wife's decision.[3] Thus the logical consistency condition (3) implies that the decision of either one of the individuals must be structurally independent from the decision of the other. In other words, the logical consistency condition associated with the above simultaneous equation model with structural shift introduces structural recursivity in the model. Since there is no reason to impose a priori that β_h or β_w be zero, it follows that the usual approach for formulating simultaneous equation models with structural shift is inappropriate.

3. AN ALTERNATIVE FORMULATION

As argued by Heckman (1978), and made explicit in Chapter I, an important justification for the use of simultaneous equations models with structural shift is that these models can distinguish structural association from purely statistical association among discrete endogenous variables, while alternative models such as those developed by Goodman (1970) and Nerlove and Press (1976) cannot.[4] The previous section has shown, however, that the corresponding simultaneous model with structural shift is inappropriate in our case because of the implied logical consistency condition.

Although Assumption A1 defines the crucial dichotomization that generates the observed discrete variables Y_1 and Y_2, that assumption has not been questioned in the literature on discrete variables modeling. We shall argue that Assumption A1 is in fact the cause of the problems that are associated with the logical consistency conditions. Our approach relies instead on the following assumption:

ASSUMPTION A1': The observed dichotomous variables (Y_h, Y_w) are Nash Equilibrium (NE) outcomes of a game played between the two individuals.

Since the utility derived by each individual depends on the action of the other, the natural framework is that of a game between the two individuals. In situations other than the one considered here, the utility functions can obviously be replaced by the appropriate payoff functions. Assumption A1' considers the non-cooperative Nash Equilibrium concept, although alternative equilibrium concepts can be invoked as we shall soon see in the next chapter. (See Brown and Manser (1978), Manser and Brown (1980), McElroy and Horney (1981), and Kooreman and Kapteyn (1985) for related work using a cooperative game solution.)

Assumption A1' is not, however, sufficient to define how the observed dichotomous variables (Y_h, Y_w) are generated. This is so because in many games, and especially for the particular problem considered here, a Nash Equilibrium may not exist or multiple Nash Equilibria may arise. As seen below, this difficulty will be resolved

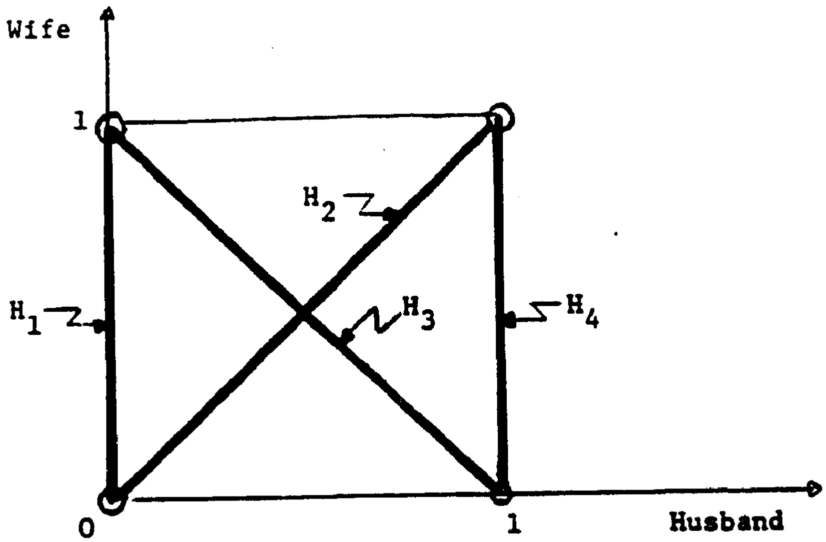

Figure 1: Husband's Reaction Functions

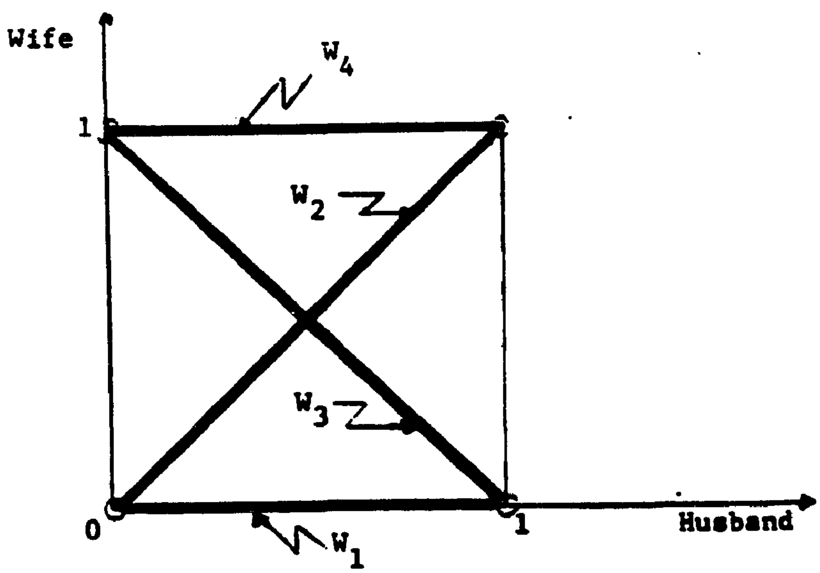

Figure 2: Wife's Reaction Functions

by increasing the number of parameters.

To determine the NE outcomes of the game, we derive the reaction functions of each individual.[5] Since the outcome space is simply {0,1} X {0,1}, there are only four possible reaction functions for each player. These are referred to as H_1, H_2, H_3, H_4 for the husband and W_1, W_2, W_3, W_4 for the wife, as displayed in Figures 1 and 2, respectively. For instance, reaction function H_1 for the husband says that, whether or not the wife works, the husband will always choose not to work.

Given that H_i and W_j are the husband and wife's reaction functions, we can readily find the Nash Equilibria. Table 1 indicates the Nash Equilibria (or lack thereof) for each of the 16 possible pairs of reaction functions

TABLE 1: Nash Equilibria

Husband/Wife	W_1	W_2	W_3	W_4
H_1	(0,0)	(0,0)	(0,1)	(0,1)
H_2	(0,0)	(0,0)&(1,1)	None	(1,1)
H_3	(1,0)	None	(1,0)&(0,1)	(0,1)
H_4	(1,0)	(1,1)	(1,0)	(1,1)

where the first number in each ordered pair refers to the husband and the second to the wife.

As mentioned earlier, a difficulty arises because of the non-existence or the multiplicity of Nash Equilibria for the pairs (H_2,W_2), (H_2,W_3), (H_3,W_2) and (H_3,W_3). As a result, we shall distribute the probability of occurrence of each of those pairs over the appropriate outcomes according to some weights.[6] The interpretation of the weights will be discussed more fully in Appendix A. Let Pr(i,j) be the probability that the random variables Y_h and Y_w take on the values i and j. Using Table 1, it follows that:

$$\begin{aligned} Pr(0,0) &= Pr(H_1 \,\&\, W_1) + Pr(H_1 \,\&\, W_2) + Pr(H_2 \,\&\, W_1) \\ &\quad + a_1 Pr(H_2 \,\&\, W_2) + c_1 Pr(H_2 \,\&\, W_3) + d_1 Pr(H_3 \,\&\, W_2) \end{aligned} \tag{4}$$

$$\begin{aligned} Pr(1,0) &= Pr(H_3 \,\&\, W_1) + Pr(H_4 \,\&\, W_1) + Pr(H_4 \,\&\, W_3) \\ &\quad + b_1 Pr(H_3 \,\&\, W_3) + c_2 Pr(H_2 \,\&\, W_3) + d_2 Pr(H_3 \,\&\, W_2) \end{aligned} \tag{5}$$

$$\begin{aligned} Pr(0,1) &= Pr(H_1 \,\&\, W_3) + Pr(H_1 \,\&\, W_4) + Pr(H_3 \,\&\, W_4) \\ &\quad + b_2 Pr(H_3 \,\&\, W_3) + c_3 Pr(H_2 \,\&\, W_3) + d_3 Pr(H_3 \,\&\, W_2) \end{aligned} \tag{6}$$

$$\begin{aligned} Pr(1,1) &= Pr(H_2 \,\&\, W_4) + Pr(H_4 \,\&\, W_2) + Pr(H_4 \,\&\, W_4) \\ &\quad + a_2 Pr(H_2 \,\&\, W_2) + c_4 Pr(H_2 \,\&\, W_3) + d_4 Pr(H_3 \,\&\, W_2) \end{aligned} \tag{7}$$

where

$$a_1 + a_2 = 1 \quad , \quad b_1 + b_2 = 1 \quad ,$$
$$c_1 + c_2 + c_3 + c_4 = 1 \quad , \quad d_1 + d_2 + d_3 + d_4 = 1 \quad ,$$

all these additional parameters being non-negative.

We are now in a position to derive the conditions under which the reaction functions for the husband and wife will occur. As expected, each reaction function will occur if certain conditions on the random components ε_h and ε_w are satisfied. For brevity, we derive only the conditions for the husband's reaction functions.

Using Figure 1, reaction function H_1 can be characterized by the following two conditions: $\tilde{U}_h(1,0) - \tilde{U}_h(0,0) < 0$ and $\tilde{U}_h(1,1) - \tilde{U}_h(0,1) < 0$. Using Assumptions 1 and 2 these conditions are equivalent to $\varepsilon_h < - \Delta_h$ and $\varepsilon_h < - \Delta_h - \beta_h$, respectively, which can be combined to give $\varepsilon_h < - \Delta_h - \max(0,\beta_h)$.

Reaction function H_2 is characterized by $\tilde{U}_h(1,0) - \tilde{U}_h(0,0) < 0$ and $\tilde{U}_h(1,1) - \tilde{U}_h(0,1) \geq 0$, which are equivalent to $\varepsilon_h < - \Delta_h$ and $\varepsilon_h \geq - \Delta_h - \beta_h$, respectively. When combined we have $-\Delta_h-\beta_h \leq \varepsilon_h < -\Delta_h$ if $\beta_h \geq 0$; otherwise, reaction function H_2 cannot occur.

Reaction function H_3 is characterized by $\tilde{U}_h(1,1) - \tilde{U}_h(0,1) < 0$ and $\tilde{U}_h(1,0) - \tilde{U}_h(0,0) \geq 0$. Using Assumptions 1 and 2 these conditions are equivalent to $\varepsilon_h < - \Delta_h - \beta_h$ and $\varepsilon_h \geq - \Delta_h$, respectively. When combined we get $-\Delta_h \leq \varepsilon_h < -\Delta_h-\beta_h$ if $\beta_h < 0$; otherwise reaction function H_3 cannot occur.

Reaction function H_4 is characterized by $\tilde{U}_h(1,0) - \tilde{U}_h(0,0) \geq 0$ and $\tilde{U}_h(1,1) - \tilde{U}_h(0,1) \geq 0$, which are equivalent to $\varepsilon_h \geq - \Delta_h$ and $\varepsilon_h \geq - \Delta_h - \beta_h$, respectively, which when combined give $\varepsilon_h \geq - \Delta_h - \min(0,\beta_h)$.

The following table thus gives the conditions on ε_h and ε_w

that must be satisfied for each of the husband's reaction functions to hold. Conditions for the wife's reaction functions are the same with the subscript h being replaced by w.

TABLE 2: Conditions for Husband's Reaction Functions

H_1: $\varepsilon_h < -\Delta_h - \max(0,\beta_h)$

H_2: $-\Delta_h - \beta_h \leq \varepsilon_h < -\Delta_h$ if $\beta_h \geq 0$; otherwise cannot occur

H_3: $-\Delta_h \leq \varepsilon_h < -\Delta_h - \beta_h$ if $\beta_h < 0$; otherwise cannot occur

H_4: $\varepsilon_h \geq -\Delta_h - \min(0,\beta_h)$

We can now derive the probabilities Pr(i,j) in terms of the unknown parameters. Let $F(a,b,\rho)$ be the c.d.f. evaluated at (a,b) of a bivariate normal distribution with zero means, unit variances, and correlation ρ. Let $I(a,b,c,d,\rho)$ be the integral of the corresponding bivariate density over the range $a \geq \varepsilon_h \geq c$, $b \geq \varepsilon_w \geq d$. Using Equations (4)-(7), we have:

PROPOSITION 1: The probabilities Pr(i,j) are:

$$\Pr(0,0) \begin{cases} = F(-\Delta_h, -\Delta_w, \rho) - a_2 I_{++} & \text{if } \beta_h \geq 0,\ \beta_w \geq 0 \\ = F(-\Delta_h, -\Delta_w, \rho) + c_1 I_{+-} & \text{if } \beta_h \geq 0,\ \beta_w < 0 \\ = F(-\Delta_h, -\Delta_w, \rho) + d_1 I_{-+} & \text{if } \beta_h < 0,\ \beta_w \geq 0 \\ = F(-\Delta_h, -\Delta_w, \rho) & \text{if } \beta_h < 0,\ \beta_w < 0 \end{cases} \tag{8}$$

$$\Pr(1,0) \begin{cases} = F(\Delta_h, -\Delta_w - \beta_w, -\rho) & \text{if } \beta_h \geq 0,\ \beta_w \geq 0 \\ = F(\Delta_h, -\Delta_w - \beta_w, -\rho) + c_2 I_{+-} & \text{if } \beta_h \geq 0,\ \beta_w < 0 \\ = F(\Delta_h, -\Delta_w - \beta_w, -\rho) + d_2 I_{-+} & \text{if } \beta_h < 0,\ \beta_w \geq 0 \\ = F(\Delta_h, -\Delta_w - \beta_w, -\rho) - b_2 I_{--} & \text{if } \beta_h < 0,\ \beta_w < 0 \end{cases} \tag{9}$$

$$\Pr(0,1) \begin{cases} = F(-\Delta_h - \beta_h, \Delta_w, -\rho) & \text{if } \beta_h \geq 0, \beta_w \geq 0 \\ = F(-\Delta_h - \beta_h, \Delta_w, -\rho) + c_3 I_{+-} & \text{if } \beta_h \geq 0, \beta_w < 0 \\ = F(-\Delta_h - \beta_h, \Delta_w, -\rho) + d_3 I_{-+} & \text{if } \beta_h < 0, \beta_w \geq 0 \\ = F(-\Delta_h - \beta_h, \Delta_w, -\rho) - b_1 I_{--} & \text{if } \beta_h < 0, \beta_w < 0 \end{cases} \quad (10)$$

$$\Pr(1,1) \begin{cases} = F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho) - a_1 I_{++} & \text{if } \beta_h \geq 0, \beta_w \geq 0 \\ = F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho) + c_4 I_{+-} & \text{if } \beta_h \geq 0, \beta_w < 0 \\ = F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho) + d_4 I_{-+} & \text{if } \beta_h < 0, \beta_w \geq 0 \\ = F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho) & \text{if } \beta_h < 0, \beta_w < 0 \end{cases} \quad (11)$$

where

$$\begin{aligned} I_{++} &= I(-\Delta_h, -\Delta_w, -\Delta_h - \beta_h, -\Delta_w - \beta_w, \rho), \\ I_{+-} &= I(-\Delta_h, -\Delta_w - \beta_w, -\Delta_h - \beta_h, -\Delta_w, \rho), \\ I_{-+} &= I(-\Delta_h - \beta_h, -\Delta_w, -\Delta_h, -\Delta_w - \beta_w, \rho), \\ I_{--} &= I(-\Delta_h - \beta_h, -\Delta_w - \beta_w, -\Delta_h, -\Delta_w, \rho). \end{aligned} \quad (12)$$

PROOF: From Table 2 and the conditions for the wife's reaction functions, it follows that some reaction functions cannot occur according to the signs of β_h and β_w. For instance, when $\beta_h \geq 0$ and $\beta_w \geq 0$, the reaction functions H_3 and W_3 cannot occur. From Equations (4)-(7) it follows that, when $\beta_h \geq 0$ and $\beta_w \geq 0$, we have:

$$\begin{aligned} \Pr(0,0) &= \Pr(H_1,W_1) + \Pr(H_1,W_2) + \Pr(H_2,W_1) + a_1\Pr(H_2,W_2), \\ \Pr(1,0) &= \Pr(H_4,W_1), \\ \Pr(0,1) &= \Pr(H_1,W_4), \\ \Pr(1,1) &= \Pr(H_2,W_4) + \Pr(H_4,W_2) + \Pr(H_4,W_4) + a_2\Pr(H_2,W_2). \end{aligned}$$

Similarly, when $\beta_h \geq 0$ and $\beta_w < 0$, the reaction functions H_3 and W_2 never occur so we have:

$$Pr(0,0) = Pr(H_1,W_1) + Pr(H_2,W_1) + c_1Pr(H_2,W_3),$$
$$Pr(1,0) = Pr(H_4,W_1) + Pr(H_4,W_3) + c_2Pr(H_2,W_3),$$
$$Pr(0,1) = Pr(H_1,W_3) + Pr(H_1,W_4) + c_3Pr(H_2,W_3),$$
$$Pr(1,1) = Pr(H_2,W_4) + Pr(H_4,W_2) + Pr(H_4,W_4) + c_4Pr(H_2,W_3).$$

When $\beta_h < 0$ and $\beta_w \geq 0$, the reaction functions H_2 and W_3 never occur. Thus:

$$Pr(0,0) = Pr(H_1,W_1) + Pr(H_1,W_2) + d_1Pr(H_3,W_2),$$
$$Pr(1,0) = Pr(H_3,W_1) + Pr(H_4,W_1) + d_2Pr(H_3,W_2),$$
$$Pr(0,1) = Pr(H_1,W_4) + Pr(H_3,W_4) + d_3Pr(H_3,W_2),$$
$$Pr(1,1) = Pr(H_4,W_2) + Pr(H_4,W_4) + d_4Pr(H_3,W_2).$$

Finally, when $\beta_h < 0$ and $\beta_w < 0$, the reaction functions H_2 and W_2 never occur. Thus:

$$Pr(0,0) = Pr(H_1,W_1),$$
$$Pr(1,0) = Pr(H_3,W_1) + Pr(H_4,W_1) + Pr(H_4,W_3) + b_1Pr(H_3,W_3),$$
$$Pr(0,1) = Pr(H_1,W_3) + Pr(H_1,W_4) + Pr(H_3,W_4) + b_2Pr(H_3,W_3),$$
$$Pr(1,1) = Pr(H_4,W_4).$$

It now suffices to use the conditions on s_h and s_w for obtaining particular reaction functions for the husband and wife. For each of the 4 possible pairs of signs for β_h and β_w, figures 3a-3d

Areas Defining Pr(0,0), Pr(1,0), Pr(0,1), Pr(1,1)

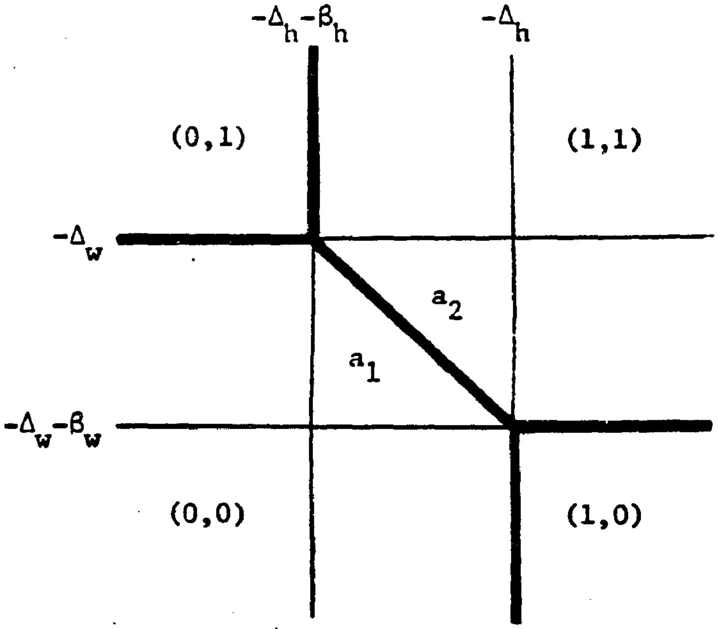

Fig. 3a : $\beta_h \geq 0$, $\beta_w \geq 0$

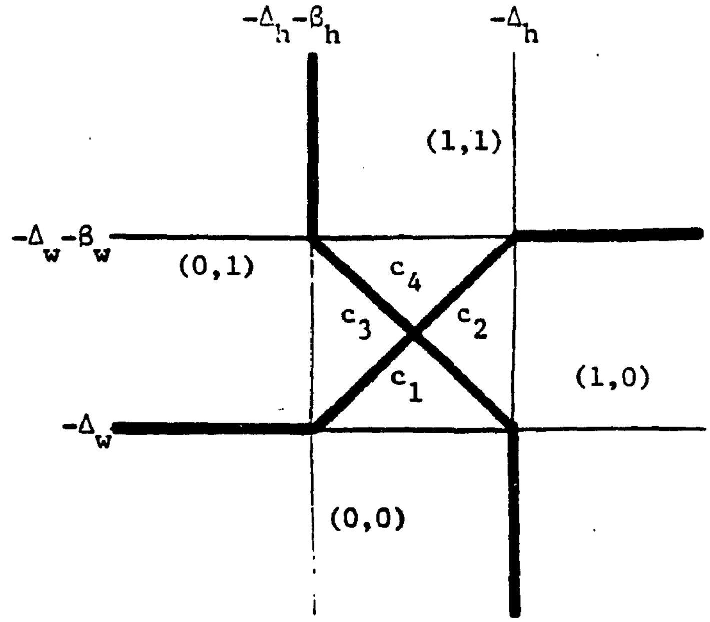

Fig. 3b : $\beta_h \geq 0$, $\beta_w < 0$

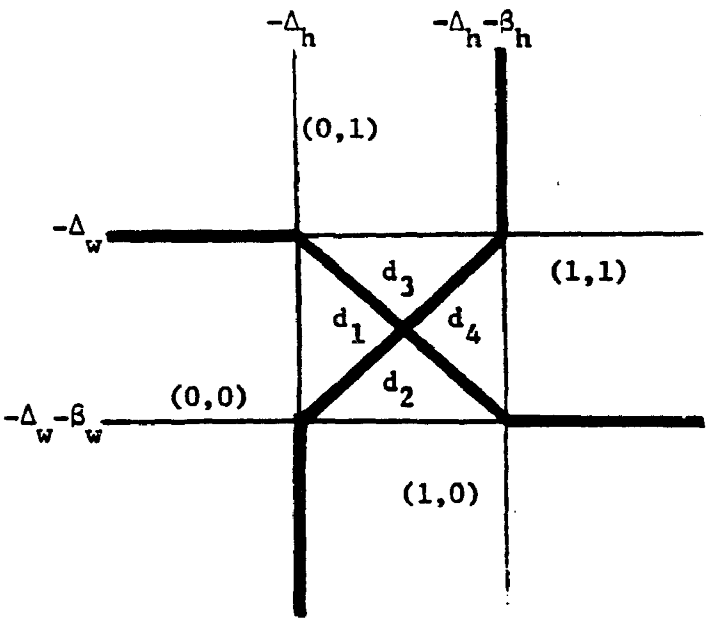

Fig. 3c : $\beta_h < 0$, $\beta_w \geq 0$

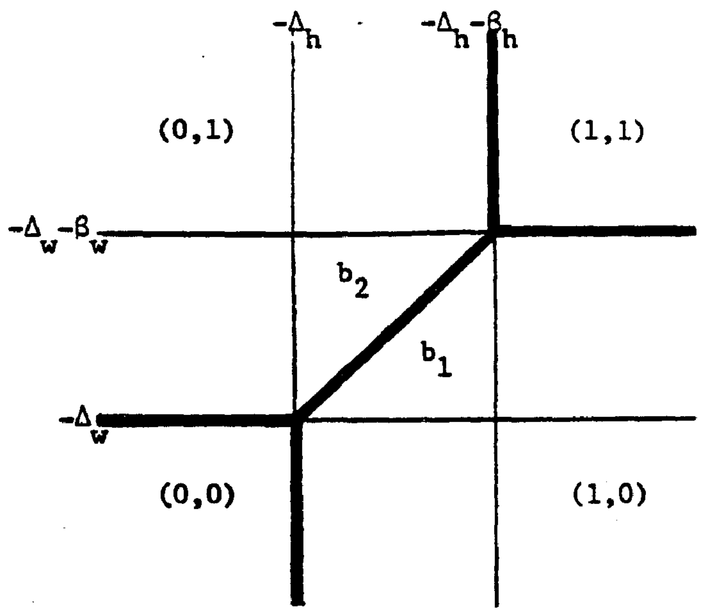

Fig. 3d : $\beta_h < 0$, $\beta_w < 0$

show the areas over which the bivariate normal density for $(\varepsilon_h, \varepsilon_w)$ must be integrated to obtain the 4 probabilities Pr(0,0), Pr(1,0), Pr(0,1), and Pr(1,1). (The areas are separated by heavy lines, while the lighter lines separate the areas corresponding to the realizations of the specific pairs of reaction functions.) It follows that the probabilities Pr(0,0), Pr(1,0), Pr(0,1) and Pr(1,1) are given by Equations (8)-(11).

Q.E.D.

One can readily check that the four probabilities Pr(i,j) add up to one irrespective of the signs of β_h and β_w, and are continuous with respect to all the parameters. It is then worth noting that the proposed model does not imply any logical consistency constraints on the structural parameters. In addition, Heckman's logical consistency condition (3) can be interpreted in our model.

PROPOSITION 2: Imposing $\beta_h \cdot \beta_w = 0$ is equivalent to imposing that the probability that each of the four pairs of reaction functions (H_2, W_2), (H_2, W_3), (H_3, W_2), (H_3, W_3) occurs is a priori zero.

PROOF: Since $(\varepsilon_h, \varepsilon_w)$ have a joint continuous distribution, it follows from the conditions for the husband's and wife's reaction functions given in Table 2 that if $\beta_h = 0$ or $\beta_w = 0$ then H_2, H_3, W_2, or W_3 occur with zero probability.

Conversely, if the pair (H_2, W_2) occurs with zero probability it follows from Table 2 that either $\beta_h \leq 0$ or $\beta_w \leq 0$, i.e., that

$\beta_h \cdot \beta_w \geq 0$. Similarly if the pair (H_2, W_3) occurs with zero probability it follows that $\beta_h \leq 0$ or $\beta_w \geq 0$, i.e., that $\beta_h \cdot \beta_w \leq 0$. Hence $\beta_h \cdot \beta_w = 0$.

Q.E.D.

From Figures 1 and 2, it follows that the logical consistency condition (3) requires that either the husband's or the wife's action be independent from the action of the other, as discussed in Section 2. More importantly, once the logical consistency condition is imposed, the standard rule used in the literature to generate the observed dichotomous variables is equivalent to implementing Nash outcomes. This is shown in the following proposition.

PROPOSITION 3: If the condition $\beta_h \cdot \beta_w = 0$ is imposed, Assumption A1' is equivalent to Assumption A1.

PROOF: Let us first assume that $\beta_h = 0$. From Table 2, it is clear that reaction functions H_2 and H_3 now occur with probability zero. Depending on the sign of β_w, we have two cases to consider. For brevity, we only present the case where $\beta_w < 0$.

Again from Table 2, imposing $\beta_w < 0$ prevents reaction function W_2 from occurring. Referring to Table 1, we see that outcome pair (0,0) occurs if and only if reaction functions H_1 and W_1 are used. But from Table 2 we see that the use of (H_1, W_1) is equivalent to the occurrence of $\varepsilon_h < -\Delta_h$ and $\varepsilon_w < -\Delta_w$ which is equivalent to $Y_h^* < 0$ and $Y_w^* < 0$ as seen from equations (1) and (2).

The occurrence of outcome pair (1,0) is equivalent to the use of (H_4,W_1) or (H_4,W_3). Now, (H_4,W_1) is used if and only if $\varepsilon_h \geq -\Delta_h$ and $\varepsilon_w < -\Delta_w$; (H_4,W_3) is used if and only if $\varepsilon_h \geq -\Delta_h$ and $-\Delta_w \leq \varepsilon_w < -\Delta_w - \beta_w$. Thus, the use of (H_4,W_1) or (H_4,W_3) is equivalent to $\varepsilon_h \geq -\Delta_h$ and $\varepsilon_w < -\Delta_w - \beta_w$, which is equivalent to $Y_h^* \geq 0$ and $Y_w^* < 0$.

As seen from Table 1, outcome pair (0,1) occurs if and only if reaction functions (H_1,W_3) or (H_1,W_4) are used. The occurrence of (H_1,W_3) is equivalent to $\varepsilon_h < -\Delta_h$ and $-\Delta_w \leq \varepsilon_w < -\Delta_w - \beta_w$; (H_1,W_4) is equivalent to $\varepsilon_h < -\Delta_n$ and $\varepsilon_w \geq -\Delta_w - \beta_w$. Thus, (H_1,W_3) or (H_1,W_4) occur if and only if $\varepsilon_h < -\Delta_h$ and $\varepsilon_w \geq -\Delta_w$, which is equivalent to $Y_h^* < 0$ and $Y_w^* \geq 0$.

Outcome pair (1,1) is realized if and only if reaction function pair (H_4,W_4) is used. The use of (H_4,W_4) is equivalent to $\varepsilon_h \geq -\Delta_h$ and $\varepsilon_w \geq -\Delta_w$, which is equivalent to $Y_h^* \geq 0$ and $Y_w^* \geq 0$.

To show that the proposition holds when $\beta_w = 0$, we proceed identically.

Q.E.D.

To illustrate Proposition 3, consider the case where $\beta_h = 0$ and $\beta_w < 0$. Under the usual dichotomization, outcome (1,1) is observed if and only if $Y_h^* \geq 0$ and $Y_w^* \geq 0$, while outcome (1,1) is a Nash Equilibrium if and only if the husband's reaction function is H_4 and the wife's reaction function is W_4. Proposition 3 says that $Y_h^* \geq 0$ and $Y_w^* \geq 0$ if and only if the husband and wife have reaction functions H_4 and W_4 respectively.

Second, Proposition 3 says that our interpretation of the observed dichotomous variables Y_h and Y_w as Nash Equilibrium outcomes reduces to the usual dichotomization when the logical consistency condition holds. It follows that our probability model defined by Equations (8)-(11) is identical to the usual simultaneous equations model with structural shift defined by equations (1)-(2) when $\beta_h \cdot \beta_w = 0$. For instance, if $\beta_h = 0$, the probability model becomes:

$$\Pr(0,0) = F(-\Delta_h, -\Delta_w, \rho)$$
$$\Pr(1,0) = F(\Delta_h, -\Delta_w - \beta_w, -\rho)$$
$$\Pr(0,1) = F(-\Delta_h, \Delta_w, -\rho)$$
$$\Pr(1,1) = F(\Delta_h, \Delta_w + \beta_w, \rho)$$

(see, e.g., Heckman (1978, p. 949)

Finally, while the usual dichotomization cannot generate a well-defined statistical model when $\beta_h \cdot \beta_w \neq 0$, our formulation based on Assumption A1' can still generate a simultaneous equations model for discrete endogenous variables whether or not the structural parameters satisfy the logical consistency condition. When one considers a game where only a unique pure strategy Nash Equilibrium is allowed, our game theoretic model is related to models based on Assumption A1. The following Corollary is a straightforward consequence of Propositions 2 and 3.

COROLLARY 1: If the condition $\beta_h \cdot \beta_w = 0$ is imposed, Assumption A1 is equivalent to a game in which only pure strategy Nash Equilibria are allowed.

It is of interest to know the direction of change in the probabilities that the husband and wife will work as $\Delta_h, \Delta_w, \beta_h, \beta_w$ vary. These are given by $Pr(1,\cdot) \equiv Pr(1,1) + Pr(1,0)$ for the husband and $Pr(\cdot,1) \equiv Pr(1,1) + Pr(0,1)$ for the wife.

PROPOSITION 4: (1) An increase in β_h or Δ_h

(i) always increases the probability that the husband will work $Pr(1,\cdot)$;

(ii) increases (decreases) the probability that the wife will work $Pr(\cdot,1)$ if $\beta_w \geq 0$ ($\beta_w < 0$).

(2) By symmetry, the effects of an increase in β_w or Δ_w can be deduced from (1).

As expected, an increase in β_h or Δ_h increases the probability that the husband will work, regardless of whether or not the wife works. Also when $\beta_w \geq 0$ the effect of an increase in β_h or Δ_h increases the probability that the wife will work, since it increases the likelihood that the husband will work. Similar remarks can be made for an increase in β_w and Δ_w. (Included with the proof of Proposition 4 is a table indicating the direction of change in the probabilities $Pr(i,j)$ as Δ_h, Δ_w, β_h, and β_w vary).

PROOF: Easily established by using either the areas defining the probabilities $Pr(i,j)$ or by using the first partial derivatives found in Appendix B. The table below can similarly be established.

increase in Δ_h

	Pr(1,1)	Pr(1,0)	Pr(0,1)	Pr(0,0)	Pr(1,·)	Pr(·,1)
$\beta_h \geq 0, \beta_w \geq 0$	+	+	−	−	+	+
$\beta_h \geq 0, \beta_w < 0$	?	+	−	?	+	−
$\beta_h < 0, \beta_w \geq 0$	+	?	?	−	+	+
$\beta_h < 0, \beta_w < 0$	+	+	−	−	+	−

Increase in β_h

	Pr(1,1)	Pr(1,0)	Pr(0,1)	Pr(0,0)	Pr(1,·)	Pr(·,1)
$\beta_h \geq 0, \beta_w \geq 0$	+	no change	−	−	+	+
$\beta_h < 0, \beta_w < 0$	+	+	−	+	+	−
$\beta_h < 0, \beta_w \geq 0$	+	−	−	−	+	+
$\beta_h < 0, \beta_w < 0$	+	+	−	no change	+	−

The effects of an increase Δ_w and β_w on the probabilities can be found by reversing rows 2 and 3, reversing columns 1 and 2, and reversing columns 5 and 6.

Q.E.D.

4. IDENTIFICATION AND ESTIMATION

Given the previous expressions for the probabilities Pr(i,j) of the observed dichotomous variables Y_h and Y_w, the log-likelihood function under random sampling can be written as:

$$L = \sum_t \log Pr_t(Y_{ht}, Y_{wt}) \tag{13}$$

$$= \sum_t [Y_{ht}Y_{wt} \log \Pr_t(1,1) + Y_{ht}(1 - Y_{wt}) \log \Pr_t(1,0)$$

$$+ (1 - Y_{ht})Y_{wt} \log \Pr_t(0,1) + (1 - Y_{ht})(1 - Y_{wt}) \log \Pr_t(0,0)],$$

where the subscript t indexes the observations. The probabilities are subscripted by t since Δ_h and Δ_w are in general functions of explanatory variables. We assume:

$$\Delta_{ht} = x'_{ht}\gamma_h \quad \text{and} \quad \Delta_{wt} = x'_{wt}\gamma_w, \tag{14}$$

where x_{ht} may include characteristics of the t-th household in addition to characteristics of the husband. A similar remark applies to x_{wt}.

The parameters of the model are $(\gamma_h,\beta_h,\gamma_w,\beta_w,\rho)$ together with the weights a's, b's, c's, and d's introduced in Equations (4)-(7). Although for the purposes of estimation we will constrain a priori these weights to satisfy $a_1 = a_2 = 1/2$, $b_1 = b_2 = 1/2$, $c_1 = c_2 = c_3 = c_4 = 1/4$, and $d_1 = d_2 = d_3 = d_4 = 1/4$, we will next show that the parameters $(\gamma_h,\beta_h\gamma_w,\beta_w,\rho,a,b)$ are identified. Notice that imposing these a priori weights is equivalent to distributing the probabilities of the four pairs of reaction functions (H_2,W_2), (H_2,W_3), (H_3,W_2), and (H_3,W_3) equally over the appropriate outcomes. We now turn to the problem of identification.

In order to discuss the conditions under which our model is identified, we must first introduce some notation. Define the following partitioned matrix $\tilde{A}$ as

$$\tilde{A} = \left[D_h\bar{X}_h \;\middle|\; D_w\bar{X}_w \;\middle|\; D_\rho\bar{X}_\rho \right],$$

where D_h , D_w and D_ρ are each block diagonal matrices of order 3T, the t-th blocks given as follows:

$$D_{ht} = \begin{bmatrix} (1-a_1)e_t^h & a_1 f_t^h & 0 \\ 0 & g_t^h & 0 \\ 0 & 0 & h_t^h \end{bmatrix}, \quad D_{wt} = \begin{bmatrix} (1-a_1)e_t^w & 0 & a_1 f_t^w \\ 0 & g_t^w & 0 \\ 0 & 0 & h_t^w \end{bmatrix},$$

$$D_{\rho t} = \begin{bmatrix} (1-a_1)r_t^1 & -a_1 r_t^2 & -I_{++} \\ 0 & -r_t^3 & 0 \\ -r_t^4 & 0 & 0 \end{bmatrix};$$

if $\beta_h < 0$ and $\beta_w < 0$

$$D_{ht} = \begin{bmatrix} e_t^h & 0 & 0 \\ 0 & b_1 g_t^h & -(1-b_1)h_t^h \\ 0 & f_t^h & 0 \end{bmatrix}, \quad D_{wt} = \begin{bmatrix} e_t^w & 0 & 0 \\ 0 & b_1 g_t^w & -(1-b_1)h_t^w \\ 0 & 0 & f_t^w \end{bmatrix}, \quad D_{\rho t} = \begin{bmatrix} r_t^1 & 0 & 0 \\ (1-b_1)r_t^4 & -b_1 r_t^3 & I_{--} \\ -r_t^2 & 0 & 0 \end{bmatrix};$$

if $\beta_h \geq 0$ and $\beta_w < 0$, or $\beta_h < 0$ and $\beta_w \geq 0$

$$D_{ht} = \begin{bmatrix} e_t^h & f_t^h & 0 \\ 0 & (g_t^h + f_t^h) & 0 \\ 0 & f_t^h & h_t^h \end{bmatrix}, \quad D_{wt} = \begin{bmatrix} e_t^w & f_t^w & 0 \\ 0 & (g_t^w + f_t^w) & 0 \\ 0 & f_t^w & h_t^w \end{bmatrix},$$

$$D_{\rho t} = \begin{bmatrix} r_t^1 & -r_t^2 & 0 \\ 0 & (-r_t^2 - r_t^3) & 0 \\ 0 & -r_t^4 & -r_t^2 \end{bmatrix},$$

with

$$
\begin{aligned}
e_t^h &= \varphi(\Delta_{ht} + \beta_h)\Phi(\Delta_{wt}^* + \beta_w^* - \rho(\Delta_{ht}^* + \beta_h^*)) \\
f_t^h &= \varphi(\Delta_{ht})\Phi(-\Delta_{wt}^* + \rho\Delta_{ht}^*) \\
g_t^h &= \varphi(\Delta_{ht})\Phi(-\Delta_{wt}^* - \beta_w^* + \rho\Delta_{ht}^*) \\
h_t^h &= -\,\varphi(\Delta_{ht} + \beta_h)\Phi(\Delta_{wt}^* - \rho(\Delta_{ht}^* + \beta_h^*)) \\
e_t^w &= \varphi(\Delta_{wt} + \beta_w)\Phi(\Delta_{ht}^* + \beta_h^* - \rho(\Delta_{wt}^* + \beta_w^*)) \\
f_t^w &= \varphi(\Delta_{wt})\Phi(-\Delta_{ht}^* + \rho\Delta_{wt}^*) \\
g_t^w &= -\,\varphi(\Delta_{wt} + \beta_w)\Phi(\Delta_{ht}^* - \rho(\Delta_{wt}^* + \beta_w^*)) \\
h_t^w &= \varphi(\Delta_{wt})\Phi(-\Delta_{ht}^* - \beta_h^* + \rho\Delta_{wt}^*),
\end{aligned}
$$

where Φ is the univariate normal c.d.f. and φ is the corresponding p.d.f. A quantity with a "*" means that quantity is divided by the square root of $1 - \rho^2$. In addition, let $f(x,y,\rho)$ be the p.d.f. corresponding to the bivariate normal p.d.f. $F(x,y,\rho)$. Now define:

$$
\begin{aligned}
r_t^1 &= f(\Delta_{ht} + \beta_h, \Delta_{wt} + \beta_w, \rho) \\
r_t^2 &= f(\Delta_{ht}, \Delta_{wt}, \rho) \\
r_t^3 &= f(\Delta_{ht}, \Delta_{wt} + \beta_w, \rho) \\
r_t^4 &= f(\Delta_{ht} + \beta_h, \Delta_{wt}, \rho).
\end{aligned}
$$

The matrices $\bar{X}_h$ and $\bar{X}_w$ are of dimension 3T by $K_h + 1$ and 3T by $K_w + 1$, the t-th blocks given respectively as:

$$
\begin{bmatrix} 1 & x_{ht}' \\ 0 & x_{ht}' \\ 1 & x_{ht}' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & x_{wt}' \\ 1 & x_{wt}' \\ 0 & x_{wt}' \end{bmatrix}.
$$

In addition, $\bar{X}_\rho$ is a unit vector of dimension 3T when $\beta_h > 0$ and $\beta_w < 0$ or $\beta_h < 0$ and $\beta_w > 0$; when β_h and β_w have the same sign, $\bar{X}\rho$ is given by the matrix

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

PROPOSITION 5: The parameters $(\beta_h, \beta_w, \gamma_h, \gamma_w, \rho, a, b)$ of the model are identified if and only if $\tilde{A}$ has full column rank.

PROOF:

Let $Z_t = (Y_{ht}, Y_{wt}, X'_{ht}, X'_{wt})$ and $\theta = (\beta_h, \beta_w, \gamma_h, \gamma_w, \rho, a, b)$. Define:

$$B = E\left[\sum_{t=1}^{T} \frac{\partial \log f(Z_t,\theta)}{\partial\theta} \cdot \frac{\partial \log f(Z_t,\theta)}{\partial\theta'}\right] = \sum_{t=1}^{T} B_t.$$

From equation (13), we have, omitting the subscript t, that

$$\frac{\partial \log f(Z,\theta)}{\partial\theta} = \frac{Y_h Y_w}{Pr(1,1)} \frac{\partial Pr(1,1)}{\partial\theta} + \frac{Y_h(1 - Y_w)}{Pr(1,0)} \frac{\partial Pr(1,0)}{\partial\theta}$$

$$+ \frac{(1 - Y_h)Y_w}{Pr(0,1)} \frac{\partial Pr(0,1)}{\partial\theta} + \frac{(1 - Y_h)(1 - Y_w)}{Pr(0,0)} \frac{\partial Pr(0,0)}{\partial\theta}.$$

Then, $\frac{\partial \log f}{\partial\beta_h} \frac{\partial \log f}{\partial\beta_h}$ is given by

$$\left[\frac{Y_h Y_w}{Pr(1,1)} \frac{\partial Pr(1,1)}{\partial\beta_h}\right]^2 + \left[\frac{Y_h(1 - Y_w)}{Pr(1,0)} \frac{\partial Pr(1,0)}{\partial\beta_h}\right]^2$$

$$+ \left[\frac{(1 - Y_h)Y_w}{Pr(0,1)} \frac{\partial Pr(0,1)}{\partial\beta_h}\right]^2 + \left[\frac{(1 - Y_h)(1 - Y_w)}{Pr(0,0)} \frac{\partial Pr(0,0)}{\partial\beta_h}\right]^2,$$

where we have used the fact that Y_h and Y_w take on only the values zero or one. Since Y_h and Y_w are random variables where $Y_h = i$, $Y_w = j$ with probability $Pr(i,j)$, $i,j \in \{0,1\}$, we have that

$$E\left[\frac{\partial \log f}{\partial \beta_h}\frac{\partial \log f}{\partial \beta_h}\right] = \frac{1}{Pr(1,1)}\left[\frac{\partial Pr(1,1)}{\partial \beta_h}\right]^2$$

$$+ \frac{1}{Pr(1,0)}\left[\frac{\partial Pr(1,0)}{\partial \beta_h}\right]^2 + \frac{1}{Pr(0,1)}\left[\frac{\partial Pr(0,1)}{\partial \beta_h}\right]^2 + \frac{1}{Pr(0,0)}\left[\frac{\partial Pr(0,0)}{\partial \beta_h}\right]^2 .$$

Proceeding analogously, the remaining terms in B are given by:

$$E\left[\frac{\partial \log f}{\partial \theta_k} \cdot \frac{\partial \log f}{\partial \theta_h}\right] = \sum_{i=0}^{1}\sum_{j=0}^{1}\frac{1}{Pr(i,j)}\,\frac{\partial Pr(i,j)}{\partial \theta_k}\,\frac{\partial Pr(i,j)}{\partial \theta_h}.$$

Notice that B can be decomposed into $B = A'DA$ where A is of dimension 4T by K. When β_h and β_w have the same sign, $K = K_h + K_w + 4$; otherwise, $K = K_h + K_w + 3$. For clarity if $\beta_h > 0$ and $\beta_w > 0$, the t-th block of matrix A, denoted by A_t is defined as:

$$\begin{matrix}
\frac{\partial Pr_t(1,1)}{\partial \beta_h} & \frac{\partial Pr_t(1,1)}{\partial \beta_w} & \frac{\partial Pr_t(1,1)}{\partial \gamma_h'} & \frac{\partial Pr_t(1,1)}{\partial \gamma_w'} & \frac{\partial Pr_t(1,1)}{\partial \rho} & \frac{\partial Pr_t(1.1)}{\partial a_1} \\
\frac{\partial Pr_t(1,0)}{\partial \beta_h} & \frac{\partial Pr_t(1,0)}{\partial \beta_w} & \frac{\partial Pr_t(1,0)}{\partial \gamma_h'} & \frac{\partial Pr_t(1,0)}{\partial \gamma_w'} & \frac{\partial Pr_t(1,0)}{\partial \rho} & \frac{\partial Pr_t(1,0)}{\partial a_1} \\
\frac{\partial Pr_t(0,1)}{\partial \beta_h} & \frac{\partial Pr_t(0,1)}{\partial \beta_w} & \frac{\partial Pr_t(0,1)}{\partial \gamma_h'} & \frac{\partial Pr_t(0,1)}{\partial \gamma_w'} & \frac{\partial Pr_t(0,1)}{\partial \rho} & \frac{\partial Pr_t(0,1)}{\partial a_1} \\
\frac{\partial Pr_t(0,0)}{\partial \beta_h} & \frac{\partial Pr_t(0,0)}{\partial \beta_w} & \frac{\partial Pr_t(0,0)}{\partial \gamma_h'} & \frac{\partial Pr_t(0,0)}{\partial \gamma_w'} & \frac{\partial Pr_t(0,0)}{\partial \rho} & \frac{\partial Pr(0,0)}{\partial a_1}
\end{matrix}$$

and D is a block diagonal matrix of order 4T, the t-th block given by

$$\begin{bmatrix} Pr_t(1,1) & 0 & 0 & 0 \\ 0 & Pr_t(1,0) & 0 & 0 \\ 0 & 0 & Pr_t(0,1) & 0 \\ 0 & 0 & 0 & Pr_t(0,0) \end{bmatrix}^{-1}.$$

As shown by Rothenberg (1971), the model will be (locally) identified if and only if B is nonsingular. Since D is of full rank and 4T > K, a necessary and sufficient condition is that A have full column rank. As can be seen from Appendix B, the partial derivatives of $Pr_t(i,j)$ with respect to the vector θ depend on the signs of β_h and β_w; we must therefore check that matrix A is nonsingular for all cases.

Case 1: $\beta_h > 0$, $\beta_w > 0$

Substituting into A_t the partial derivatives, using the notation $e_t^i, f_t^i, g_t^i, h_t^i$, $i = h,w$, and r_t^j, $j=1,2,3,4$, established earlier, we perform the following matrix algebra

(i) add Row 4 to Row 1, noting that $a_2 = 1 - a_1$

(ii) add Rows 2 and 3 to Row 1

(iii) add $(1 - a_1)$ (Row 2 and Row 3) to Row 4

(iv) switch Rows 2 and 4; swith Rows 3 and 4.

(v) switch Columns 2 and 3

Omitting Row 1 since it is identically null, we have

$$\begin{bmatrix} (1-a_1)e_t^h & (a_1f_t^h+(1-a_1)e_t^h)x'_{ht} & (1-a_1)d_t^w & (a_1f_t^w+(1-a_1)e_t^w)x'_{wt} & (1-a_1r_t^1-a_1r_t^2 & -I_{++} \\ 0 & g_t^h x'_{ht} & g_t^w & g_t^w x'_{wt} & -r_t^3 & 0 \\ h_t^h & h_t^h x'_{ht} & 0 & h_t^w x'_{wt}\ . & -r_t^4 & 0 \end{bmatrix}$$

We now decompose this matrix into a partitioned matrix

$$A = \left[D_h\bar{X}_h \;\middle|\; D_w\bar{X}_w \;\middle|\; D_\rho\bar{X}_\rho \right],$$

where D_h, D_w, and D_ρ are each block diagonal matrices of order 3T, the t-th blocks being D_{ht}, D_{wt}, and $D_{\rho t}$ respectively, as given previously.

Case 2: $\beta_h < 0$, $\beta_w < 0$

Substituting into A_t the partial derivatives found in Appendix B, again using e_t^i, f_t^i, g_t^i, h_t^i, $i = h,w$, and r_t^j, $t = 1,2,3,4$, as established in the text, now perform the following matrix algebra on matrix A

(i) add Rows 2 and 3 to Row 1, noting that $b_2 = 1 - b_1$

(ii) add Row 1 to Row 3

(iii) add Row 4 to Row 3

(iv) add $(1 - b_1)$ (Row 1 and Row 4) to Row 2

(v) reverse Columns 2 and 3

Again, omitting Row 3 since it is identically null, we have

$$\begin{bmatrix} e_t^h & e_t^h x'_{ht} & e_t^w & e_t^w x'_{wt} & r_t^1 & 0 \\ -(1-b_1)h_t^h & (b_1 g_t^h - (1-b_1)h_t^h)x'_{ht} & b_1 g_t^w & (b_1 g_t^w - (1-b_1)h_t^w)x'_{wt} & -b_1 r_t^3 + (1-b_{1-} r_t^4 & I_{--} \\ 0 & f_t^h x'_{ht} & 0 & f_t^w x'_{wt} & -r_t^2 & 0 \end{bmatrix}$$

which can be written as $\tilde{A}$.

Case 3: $\beta_h > 0$, $\beta_w < 0$

Proceeding as in Case 2, perform the following algebra

(i) subtract Row 4 from Rows 1, 2 and 3, noting that

$$d_1 = d_2 = d_3 = d_4 = \frac{1}{4}$$

(ii) add 1/4(Row 1 + Row 2 + Row 3) to Row 4

(iii) reverse Columns 2 and 3

Deleting Row 4 since it is null, we have

$$\bar{A}_t = \begin{bmatrix} e_t^h & (e_t^h + f_t^h)x_{ht}' & e_t^w & (e_t^w + f_t^w)x_{wt}' \\ 0 & (g_t^h + f_t^h)x_{ht}' & g_t^w & (g_t^w + f_t^w)x_{wt}' \\ h_t^h & (h_t^h + f_t^h)x_{ht}' & 0 & (h_t^w + f_t^w)x_{wt}' \end{bmatrix}$$

which can be written as $\tilde{A}$.

Case 4: $\beta_h < 0$, $\beta_w > 0$

Identical to Case 3; this can be easily seen by noting that $c_i = d_i = 1/4$, $i = 1,2,3,4$.

Q.E.D.

As can be easily seen, D_h, D_w, and D_ρ are nonsingular under all four cases since D_{ht}, D_{wt}, and $D_{\rho t}$ are either triangular matrices or can be made triangular by suitable permutations of rows and columns. By examining matrix $\tilde{A}$ above, it is clear that if $\tilde{A}$ does not have full column rank, it will occur extremely rarely for some specific values of the parameters as an artifact of certain explanatory variables (note the exception in Corollary 2 below).

When compared to the restrictions needed for the identification of linear models, our results are quite surprising. For example, even if x_{ht}' and x_{wt}' are the same, it will occur only rarely that our model is not identified, a sharp contrast to the order conditions which is necessary for the identification of linear models.

A corollary to Proposition 5 provides a necessary condition for identification as shown next. Suppose that constant terms are included in both equations of (14). Define $\Delta_{ht} = \gamma_h^0 + x_{ht}'\tilde{\gamma}_h$ and $\Delta_{wt} = \gamma_w^0 + x_{wt}'\tilde{\gamma}_w$ where γ_h^0 is the coefficient for the husband's constant term and x_{ht} is a vector of explanatory variables for husband t; γ_w^0 and x_{wt} are defined analogously for the wife. We then have:

COROLLARY 2: If constant terms (γ_h^0 and γ_w^0) are included in both equations, and if $\tilde{\gamma}_h = 0$ and $\tilde{\gamma}_w = 0$, the model is not identified.

PROOF:

Note that when only constant terms are included, $\bar{X}_h$ and $\bar{X}_w$ reduce respectively to T repetitions of

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Note also that the blocks comprising D_h are now identical, as are the blocks comprising D_w now identical. (This is because $\Delta_{ht}, \Delta_{ht}^*, \Delta_{wt}$, and Δ_{wt}^* no longer depend on t). As a result, the matrix $\tilde{A}$ repeats itself every three rows. Since $K = 4 > 3$, B is singular.

Q.E.D.

Therefore if one intends to estimate the model with both constant terms (γ_h^0 and γ_w^0) included, at least one equation must include further explanatory variables or the model will certainly be not identified.

As a practical implication of Corollary 2 for estimation, if one includes a constant term in each equation along with one or more additional explanatory variables in at least one of the equations, one or more of the initial values for the parameters associated with these non-constant explanatory variables must be nonzero. Otherwise, the information matrix will be singular at the first iteration, and the optimization cannot be carried out.

5. CONCLUSION

In this Chapter, we presented an alternative approach for formulating simultaneous equations models for qualitative endogenous variables. Contrary to earlier simultaneous models, our model does not require any logical consistency conditions on the parameters. In addition, a distinctive feature of our approach is that the simultaneous model is derived from optimizing behavior within the random utility framework.

Our approach also emphasizes the role of the equilibrium concept used in order to define the process generating the observed random variables Y_h and Y_w. Our proposed model extends the usual simultaneous model with structural shift to cases where the parameters need not satisfy the logical consistency conditions. Moreover, when the logical consistency conditions are imposed, our model relates to the usual simultaneous equations model with structural shift in a number of important ways. First, imposing the logical consistency condition is equivalent to imposing that four of the sixteen possible pairs of reaction functions occur with probability zero. As a result, it is shown in Section 3 that either the husband's or the wife's actions are required to be structurally independent of the action of the other. Second, imposing the logical consistency condition on our game theoretic model insures that only pure strategy Nash Equilibria are allowed as solutions. Finally, and most importantly, once the logical consistency condition is imposed, our model coincides with the usual simultaneous equations model with structural shift. Since our model is still well defined in the absence of this restrictive condition, it follows that the simultaneous model with structural shift is nested in our model. Although the model studied in this Chapter is relatively simple, we believe that our approach is a first step in introducing strategic behavior directly into econometric models of structural shift. That the model can be extended to other game theoretic equilibrium concepts is shown in the following chapter.

FOOTNOTES

1. When an individual is indifferent between working and not working, we arbitrarily consider that he will work, hence the use of the weak inequality.

2. The assumption is that β_h and β_w do not depend on individuals. By allowing β_h or β_w not to be zero, one is allowing interactions in the constant terms of Δ_h and Δ_w defined below.

3. The condition $\beta_h = 0$ holds if one assumes that the husband's utility function is additively separable so that $U_h(Y_h, Y_w) = U_h(Y_h) + U_w(Y_w)$. Then $\Delta_h = U_h(1) - U_h(0)$. We are grateful to Donald Lien for pointing this out. Note, however, that assuming the husband's utility function to be additively separable essentially removes simultaneity as is argued in the text.

4. See, however, Vuong (1980, 1981, 1982b).

5. Although it is not necessary to use reactions functions in order to derive the probabilities Pr(i,j) given in Proposition 1, it is much easier to do so. For example, take the case where $\beta_h < 0$, $\beta_w < 0$ and we do not use reaction functions.

   Using Assumptions A1', A2 and A3, we find that outcome (0,0) is a pure strategy Nash Equilibrium (P.S.N.E.) if $\tilde{U}_h(0,0) \geq \tilde{U}_h(1,0)$ and $\tilde{U}_w(0,0) \geq \tilde{U}_w(1,0)$ which is equivalent to $\varepsilon_h \leq -\Delta_h$ and $\varepsilon_w \leq \Delta_w$. But since $\beta_h < 0$ and $\beta_w < 0$, outcome (1,1) cannot be a P.S.N.E. for that would require $\varepsilon_h < -\beta_h - \Delta_h$ and $\varepsilon_w < -\beta_w - \Delta_w$, which is inconsistent with the requirements for outcome (0,0) being a P.S.N.E. Therefore, if outcome (0,0) is a P.S.N.E., it must be unique P.S.N.E. when $\beta_h < 0$ and $\beta_w < 0$. Thus, $Pr(0,0) = F(-\Delta_h, -\Delta_w, \rho)$ as given in Proposition 1. A similar argument shows that if outcome (1,1) is a P.S.N.E., it is a unique P.S.N.E. when $\beta_h < 0$ and $\beta_w < 0$.

Alternatively, outcome (1,0) is a P.S.N.E if $\tilde{U}_h(1,0) \geq \tilde{U}_h(0,0)$ and $\tilde{U}_w(0,1) \geq \tilde{U}_w(1,1)$, which is equivalent to $\varepsilon_h \geq -\Delta_h$ and $\varepsilon_w < -\beta_w - \Delta_w$. Moreover, outcome (0,1) is a P.S.N.E. if $\tilde{U}_h(0,1) \geq \tilde{U}_h(1,1)$ and $\tilde{U}_w(1,0) \geq \tilde{U}_w(0,0)$, which is equivalent to $\varepsilon_h < -\beta_h - \Delta_h$ and $\varepsilon_w \geq -\Delta_w$. But now note that there exists the area $\{-\Delta_h \leq \varepsilon_h < -\beta_h - \Delta_h \times -\Delta_w \leq \varepsilon_w < -\beta_w - \Delta_w\}$ where both outcomes (1,0) and (0,1) are P.S.N.E. We therefore divide this area into two parts and assign the parts to outcomes (1,0) and (0,1) according to the weights b_2 and b_1, respectively, where $b_2 + b_1 = 1$.

6. An obvious generalization of our model is to specify the weights (a_h, b_i, c_j, d_k) as functions of explanatory variables. For example, a_h, h = 1, 2, can be specified in a logistic functional form as

$$\log a_{1t} = \mu_t^a + Z_t' \delta^a 1 \quad \text{and}$$
$$\log a_{2t} = \mu_t^a + Z_t' \delta^a 2,$$

where Z_t is a vector of explanatory variables for the t-th household and μ_t^a is a normalizing parameter so as to insure $a_{1t} + a_{2t} = 1$, viz. $\mu_t^a = -\log[\exp(Z_t'\delta^a 1) + \exp(Z_t' \delta^a 2)]$. Notice then that a_{1t} and a_{2t} can be written in familiar form as

$$a_{jt} = \frac{\exp(Z_t' \delta^a j)}{\exp(Z_t' \delta^a 1) + \exp(Z_t' \delta^a 2)}, \; j = 1, 2.$$ Notice also that

a_{1t} and 1_{2t} are constrained to lie in the open unit interval. This unconstrained model thus contains our model which assumes that the weight parameters are constant across observations, or equivalently that the parameters δ^a, δ^b, δ^c, and δ^d are zero, with the exception of those parameters associated with the constant term. It then follows that this assumption can be tested. Another interesting test would be to check whether $\delta^a = \delta^b = \delta^c = \delta^d = 0$, including the constant term. If this hypothesis is not rejected, the data would support the idea that, whenever there is no Nash Equilibrium, each of the four outcomes are equally likely to occur; alternatively, when there are two Nash Equilibria, a rejection of the hypothesis would support distributing the probability over the two outcomes with equal weights. (See Appendix A for a more complete discussion.)

APPENDIX A

As discussed in Section 3 of this chapter, a difficulty arises because of the non-existence of Nash Equilibria for the two pairs of reaction functions (H_2,W_3) and (H_3,W_2) and the multiplicity of Nash Equilibria for the pairs (H_2,W_2) and (H_3,W_3). As will be recalled, we handled this problem by distributing the probability of occurrence of each of these four pairs of reactions functions over the appropriate outcomes according to some fixed weights that were independent of the observations. For example, when the husband used reaction function H_2 and the wife used reaction function W_3, we distributed the probability of occurrence equally over each of the four outcomes. When the pair (H_2,W_2) was used, we found that two Nash Equilibria, (0,0) and (1,1), occur; we therefore distributed equally the probability over each of these two outcomes. The purpose of this Appendix is to discuss more fully these weights.

In the previous developments of Section 3, we have restricted ourselves to Nash Equilibria in pure strategies. A broader class of strategies is in general considered in game theory, namely the class of mixed strategies (see, e.g., Owen (1982)). In this more general approach, randomness enters into the statistical model now for two reasons; they are (i) the random nature of the sampling, or equivalently, the ignorance of the econometrician, and (ii) the randomization of the strategies. How can the simultaneous model proposed in the text of this chapter be interpreted when mixed strategies are allowed?

First, it is worth noting from Table 1 that in 12 out of the 16 cases, a unique pure strategy Nash Equilibrium exists. Moreover, as the following Lemma shows, when there is a unique Nash Equilibrium in pure strategies, then there are no Nash Equilibria in mixed strategies. It therefore follows that for those 12 cases, restricting ourselves to pure strategies is irrelevant.

LEMMA 1: For a two-person, two-strategy, normal form random payoff game, if there is a unique Nash Equilibrium in pure strategies, there is almost surely no mixed strategy Nash Equilibria.

PROOF: (by contradiction) Let the payoff matrix be given by:

		player B	
		1'	2'
player A	1	a_{11}, b_{11}	a_{12}, b_{12}
	2	a_{21}, b_{21}	a_{22}, b_{22}.

Without loss of generality, assume (1,1') is the unique Nash Equilibrium; we therefore require $a_{11} > a_{21}$ and $b_{11} > b_{12}$. Note that we use strong inequalities since $a_{11} = a_{21}$ or $b_{11} = b_{12}$ occur with probability zero. Assume a mixed strategy N.E. exists where player A plays strategy 1 with probability p and strategy 2 with probability (1-p), $0 < p < 1$. Similarly, player B plays strategy 1' with probability q and strategy 2' with probability (1-q), $0 < q < 1$. In choosing mixed strategies, players A and B must solve respectively:

$$\max_p E\pi_A = pqa_{11} + p(1 - q)a_{12} + (1 - p)qa_{21} + (1 - p)(1 - q)a_{22} \quad (1)$$

$$\max_q E\pi_B = pqb_{11} + p(1 - q)b_{12} + (1 - p)qb_{21} + (1 - p)(1 - q)b_{22}. \quad (2)$$

For a mixed strategy N.E. to exist, it must be the case that:

$$\frac{\partial E\pi_A}{\partial p} = 0 \Rightarrow a_{11}q + a_{12}(1 - q) = a_{21}q + a_{22}(1 - q) \quad (3)$$

and

$$\frac{\partial E\pi_B}{\partial q} = 0 \Rightarrow b_{11}p + b_{21}(1 - p) = b_{12}p + b_{22}(1 - p). \quad (4)$$

But when $a_{11} > a_{21}$ we have by (3) that $a_{12} < a_{22}$. Also, when $b_{11} > b_{12}$ we have by (4) that $b_{21} < b_{22}$. Therefore both (1,1') and (2,2') are pure strategy Nash Equilibria. Contradiction. We now need to show there do not exist any Nash Equilibria when one player plays a pure strategy while the other player plays a mixed strategy. Assume player A plays a mixed strategy while player B plays strategy 1' as a pure strategy. We therefore have $0 < p < 1$ and $q = 1$. But from (3) we then have $a_{11} = a_{21}$, which occurs with probability zero. Contradiction. The other three cases follow exactly and yield similar contradictions.

Q.E.D.

On the other hand, when there are two pure strategy Nash Equilibria, which is the case when reaction function pairs (H_2,W_2) or (H_3,W_3) are used, we estimate the model by imposing weights which distribute the probability of occurrence over the appropriate pair of pure strategy Nash Equilibria. Specifically, when reaction function pair (H_2,W_2) is used, the two pure strategy Nash Equilibria are (0,0) and (1,1); we set the weights a_1 and a_2 each at one half. Similarly, when reaction function pair (H_3,W_3) is used, the two pure strategy Nash Eqiulibria are (1,0) and (0,1); we set these weights b_1 and b_2 also at one half each. It follows that these weights could be estimated since Proposition 5 demonstrates that a and b are identified.

The third possibility is when there are no Nash Equilibria in pure strategies. This occurs when the husband and wife use reaction pair (H_2,W_3) or (H_3,W_2). In this case, as the next Lemma shows, there exists a unique Nash Equilibrium in mixed strategies.

LEMMA 2: For a two-person, two strategy, normal form game, if there are no Nash Equilibria in pure strategies, then there exists a unique Nash Equilibrium in mixed strategies.

PROOF: Existence is well known (see e.g., Owen (1982), p. 126). From Lemma 1 we know that for a mixed strategy N.E. to exist, we must have

$$\frac{\partial E\pi_A}{\partial p} = 0 \Rightarrow q = \frac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}},\ a_{11} - a_{12} - a_{21} + a_{22} \neq 0; \quad (5)$$

and

$$\frac{\partial E\pi_B}{\partial q} = 0 \Rightarrow p = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}},\ b_{11} - b_{12} - b_{21} + b_{22} \neq 0. \quad (6)$$

or

$$(a_{11} - a_{12} - a_{21} + a_{22})q - (a_{22} - a_{12}) = 0 \quad (7)$$

and

$$(b_{11} - b_{12} - b_{21} + b_{22})p - (b_{22} - b_{21}) = 0. \tag{8}$$

Now assume that there exists multiple mixed strategy Nash Equilibria for player A. Then from (7) we must have $(a_{11} - a_{12} - a_{21} + a_{22}) = 0$ and $(a_{22} - a_{12}) = 0$. Therefore, $a_{22} = a_{12}$ and $a_{11} = a_{21}$. Then when $b_{11} \geq b_{12}$ we have that (1,1') is a pure strategy Nash Equilibrium. Contradiction. Alternatively, when $b_{11} < b_{12}$ we have that (1,2') is a pure strategy Nash Equilibrium. Contradiction. The case where there exists multiple mixed strategy Nash Equilibria for player B proceeds similarly.

Let us now consider the case where one player plays a pure strategy while the other player plays a mixed strategy. Let player A play strategy 1 as a pure strategy while player B plays a mixed strategy. We then have $p = 1$ and $0 < q < 1$. Then from (6) we have that

$$1 = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}}.$$

Therefore $b_{11} = b_{12}$. Now, for (1,1') not to be a pure strategy N.E. we require $a_{21} > a_{11}$; for (2,1') not to be a pure strategy N.E. we require $b_{22} > b_{21}$; for (2,2') not to be pure strategy N.E. we require $a_{12} > a_{22}$. But these conditions then require (1,2') to be a pure strategy N.E. Contradiction. Similarly, let us examine the case where player A plays strategy 2 as a pure strategy while player B continues to play a mixed strategy. We then have $p = 0$ and $0 < q < 1$.

From equation (6) we then get $b_{22} = b_{21}$. Now, for (2,2') not to be a pure strategy N.E. we require $a_{22} < a_{12}$; for (1,2') not to be a pure strategy N.E. we require $b_{12} < b_{11}$; for (1,1') not to be a pure strategy N.E. we require $a_{11} < a_{21}$. But the above conditions then imply that (2,1') is a pure strategy N.E. Contradiction. The cases where player A plays a mixed strategy while player B plays a pure strategy lead to similar contradictions.

Q.E.D.

Let (q_{h0},q_{h1}), where $q_{h0} + q_{h1} = 1$, characterize the randomization of strategies between not working and working for the husband; similarly, let (q_{w0},q_{w1}) characterize the randomization between not working and working for the wife. Then the weights c_1, c_2, c_3, and c_4 for the reaction function pair (H_2,W_3) defined in Section 3 can be interpreted as:

$$c_1 = q_{h0}q_{w0} \quad , \qquad c_2 = q_{h1}q_{w0} \quad ,$$

$$c_3 = q_{h0}q_{w1} \quad , \qquad c_4 = q_{h1}q_{w1} \quad . \tag{9}$$

But we know from equations (3) and (4) of Lemma 1 that for a mixed strategy Nash Equilibrium to exist we must have

$$U_h(1,1)q_{w1} + U_h(1,0)(1-q_{w1}) = U_h(0,1)q_{w1} + U_h(0,0)(1-q_{w1}) \tag{10}$$

and

$$U_w(1,1)q_{h1} + U_w(1,0)(1-q_{h1}) = U_w(0,1)q_{h1} + U_w(0,0)(1-q_{h1}). \tag{11}$$

Solving (10) and (11) for q_{w1} and q_{h1} we get respectively:

$$q_{w1} = \frac{U_h(0,0) - U_h(1,0)}{U_h(1,1) - U_h(1,0) - U_h(0,1) + U_h(0,0)}$$

$$q_{h1} = \frac{U_w(0,0) - U_w(1,0)}{U_w(1,1) - U_w(1,0) - U_w(0,1) + U_w(0,0)}.$$

By substituting from ASSUMPTION A3 of the current chapter, we get $q_{w1} = -\Delta_h/\beta_h$ and $q_{h1} = -\Delta_w/\beta_w$. Noting that $q_{w0} = (1-q_{w1})$ and $q_{h0} = (1-q_{h1})$ and using (9), we get that

$$C_1 = \left(\frac{\beta_w + \Delta_w}{\beta_w}\right)\left(\frac{\beta_h + \Delta_h}{\beta_h}\right), \quad C_2 = \frac{-\Delta_w}{\beta_w}\left(\frac{\beta_h + \Delta_h}{\beta_h}\right),$$

$$C_3 = -\left(\frac{\beta_w + \Delta_w}{\beta_w}\right)\frac{\Delta_h}{\beta_h}, \quad C_4 = \frac{\Delta_w \cdot \Delta_h}{\beta_w \cdot \beta_h}.$$

Moreover, we know from Lemma 2 that the mixed strategy Nash Equilibrium is unique. As a result, $c_i = d_i$, $i = 1, \ldots, 4$. It therefore follows that the model proposed in Section 3 of this chapter can be interpreted as a model in which the weights c's and d's remain constrained as $c_i = d_i = \frac{1}{4}$, $i = 1, \ldots, 4$.

APPENDIX B

First Partial Derivatives of the Probabilities Pr(i,j)

For brevity, let us rewrite the four probabilities Pr(i,j) listed in Proposition 1 using the indicator variables I_h and I_w defined as:

$$I_h = \begin{cases} 1 & \text{if} \quad \beta_h \geq 0 \\ 0 & \text{if} \quad \beta_h < 0 \end{cases} \qquad \text{and } I_w = \begin{cases} 1 & \text{if} \quad \beta_w \geq 0 \\ 0 & \text{if} \quad \beta_w < 0 \end{cases}$$

Then we have:

$$\Pr(0,0) = F(-\Delta_h, -\Delta_w, \rho) - I_h I_w a_2 I_{++} + I_h(1-I_w)c_1 I_{+-} + (1-I_h)I_w d_1 I_{-+}$$

$$\Pr(1,0) = F(\Delta_h, -\Delta_w - \beta_w, -\rho) + I_h(1-I_w)c_2 I_{+-} + (1-I_h)I_w d_2 I_{-+} - (1-I_h)(1-I_w)b_2 I_{--}$$

$$\Pr(0,1) = F(-\Delta_h - \beta_h, \Delta_w, -\rho) + I_h(1-I_w)c_3 I_{+-} + (1-I_h)I_w d_3 I_{-+} - (1-I_h)(1-I_w)b_1 I_{--}$$

$$\Pr(1,1) = F(\Delta_h + \beta_h, \Delta_w + \beta, \rho) - I_h I_w a_1 I_{++} + I_h(1 - I_w)c_4 I_{+-} + (1-I_h)I_w d_4 I_{-+}.$$

In order to determine the first partial derivatives of the Probabilities Pr(i,j), we need the following two Lemmas.

LEMMA 4:

$\frac{\partial F(a,b,\rho)}{\partial a} = \varphi(a)\Phi(\frac{b - \rho a}{\sqrt{1 - \rho^2}})$ and $\frac{\partial F(a,b,\rho)}{\partial b} = \varphi(b)\ \Phi\ (\frac{a - \rho b}{\sqrt{1 - \rho^2}})$ where

Φ is the univariate normal c.d.f. and φ is the corresponding p.d.f.

PROOF: We have that $F(a,b,\rho) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y,\rho)\, dy\, dx$ where f is the bivariate normal p.d.f. with zero means, unit variances and correlation ρ. In addition, let a quantity with a "*" indicate a quantity divided by the square root of $1 - \rho^2$. We then have

$$f(x,y,\rho) = \varphi^*(x) \cdot \varphi(y^* - \rho x^*) = \varphi^*(y) \cdot \varphi(x^* - \rho y^*)$$

which gives

$$F(a,b,\rho) = \int_{-\infty}^{a} \varphi(x)\left[\int_{-\infty}^{b} \varphi^*(y^* - \rho x^*)dy\right]dx$$

$$= \int_{-\infty}^{a} \varphi(x)\left[\int_{-\infty}^{b} \frac{d\Phi(y^* - \rho x^*)}{dy}\, dy\right]dx = \int_{-\infty}^{a} \varphi(x)\Phi(b^* - \rho x^*)dx.$$

Then $\frac{\partial F(a,b,\rho)}{\partial a} = \varphi(a)\Phi(b^* - \rho a^*)$.

Similarly, $\frac{\partial F(a,b,\rho)}{\partial b} = \varphi(a)\Phi(a^* - \rho b^*)$.

Q.E.D.

LEMMA 5: Let f be the p.d.f. corresponding to the bivariate normal c.d.f. F. Then if ε_h and ε_w each have zero means and unit variances with correlation coefficient ρ, we have $\frac{\partial F(a,b,\rho)}{\partial \rho} = f(a,b,\rho)$.

PROOF: From Lemma 4, $f(a,b,\rho) = \int_{-\infty}^{a} \varphi(x)\Phi(b^* - \rho x^*)dx$.

$$\text{Then } \frac{\partial F}{\partial \rho} = \int_{-\infty}^{a} \varphi(x) \cdot \varphi(b^* - \rho x^*) \cdot \frac{\partial(b^* - \rho x^*)}{\partial \rho} dx$$

$$= \int_{-\infty}^{a} \varphi(x) \cdot \varphi(b^* - \rho x^*) \cdot \frac{\rho b - x}{(1-\rho^2)^{3/2}} \, dx.$$

But $\varphi(x)\varphi(b^* - \rho x^*) = (2\pi)^{-1/2} \varphi(x^* - \rho b^*) \cdot e^{-b^2/2}$. Thus,

$$\frac{\partial F}{\partial \rho} = -\int_{-\infty}^{a} (2\pi)^{-1/2} \varphi(x^* - \rho b^*) \cdot e^{-b^2/2} \cdot \frac{x - \rho b}{(1 - \rho^2)^{3/2}} \, dx.$$

Letting $z = x^* - \rho b^*$, we have

$$\frac{\partial F}{\partial \rho} = -\int_{-\infty}^{a^* - \rho b^*} (2\pi)^{-1/2} \cdot z \cdot \varphi^*(z) \cdot e^{-b^2/2} dz$$

$$= \varphi(a^* - \rho b^*) \cdot \varphi^*(b) = f(a,b,\rho).$$

Q.E.D.

Again, let a quantity with a "*" indicate a quantity divided by the square root of $1 - \rho^2$. Then using Lemmas 4 and 5, the first partial derivatives of the probabilities Pr(i,j) are as follows:

$$\frac{\partial F(-\Delta_h, -\Delta_w, \rho)}{\partial \beta_h} = 0 \, ,$$

$$\frac{\partial F(-\Delta_h, -\Delta_w, \rho)}{\partial \beta_w} = 0 \, ,$$

$$\frac{\partial F(-\Delta_h, -\Delta_w, \rho)}{\partial \gamma_h} = -\varphi(\Delta_h)\Phi(-\Delta_w^* + \rho\Delta_h^*)x_h \, ,$$

$$\frac{\partial F(-\Delta_h, -\Delta_w, \rho)}{\partial \gamma_w} = -\varphi(\Delta_w)\Phi(-\Delta_h^* + \rho\Delta_w^*)x_w \, ;$$

$$\frac{\partial F(-\Delta_h, -\Delta_w, \rho)}{\partial \rho} = f(-\Delta_h, -\Delta_w, \rho)$$

$$\frac{\partial F(\Delta_h, -\Delta_w - \beta_w, -\rho)}{\partial \beta_h} = 0,$$

$$\frac{\partial F(\Delta_h, -\Delta_w - \beta_w, -\rho)}{\partial \beta_w} = -\varphi(\Delta_w + \beta_w)\Phi(\Delta_h^* - \rho(\Delta_w^* + \beta_w^*)),$$

$$\frac{\partial F(\Delta_h, -\Delta_w - \beta_w, -\rho)}{\partial \gamma_h} = \varphi(\Delta_h)\Phi(-\Delta_w^* - \beta_w^* + \rho\Delta_h^*)x_h,$$

$$\frac{\partial F(\Delta_h, -\Delta_w - \beta_w, -\rho)}{\partial \gamma_w} = -\varphi(\Delta_w + \beta_w)\Phi(\Delta_h^* - \rho(\Delta_w^* + \beta_w^*))x_w \ ;$$

$$\frac{\partial F(\Delta_h, -\Delta_w - \beta_w, -\rho)}{\partial \rho} = -f(\Delta_h, -\Delta_w - \beta_w, -\rho)$$

$$\frac{\partial F(-\Delta_h - \beta_h, \Delta_w, -\rho)}{\partial \beta_h} = -\varphi(\Delta_h + \beta_h)\Phi(\Delta_w^* - \rho(\Delta_h^* + \beta_h^*)) \ ,$$

$$\frac{\partial F(-\Delta_h - \beta_h, \Delta_w, -\rho)}{\partial \beta_w} = 0,$$

$$\frac{\partial F(-\Delta_h - \beta_h, \Delta_w, -\rho)}{\partial \gamma_h} = -\varphi(\Delta_h + \beta_h)\Phi(\Delta_w^* - \rho(\Delta_h^* + \beta_h^*))x_h,$$

$$\frac{\partial F(-\Delta_h - \beta_h, \Delta_w, -\rho)}{\partial \gamma_w} = \varphi(\Delta_w)\Phi(-\Delta_h^* - \beta_h^* + \rho\Delta_w^*)x_w \ ;$$

$$\frac{\partial F(-\Delta_h - \beta_h, \Delta_w, -\rho)}{\partial \rho} = -f(-\Delta_h - \beta_h, \Delta_w, -\rho)$$

$$\frac{\partial F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)}{\partial \beta_h} = \varphi(\Delta_h + \beta_h)\Phi(\Delta_w^* + \beta_w^* - \rho(\Delta_h^* + \beta_h^*)) \ ,$$

$$\frac{\partial F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)}{\partial \beta_w} = \varphi(\Delta_w + \beta_w)\Phi(\Delta_h^* + \beta_h^* - \rho(\Delta_w^* + \beta_w^*)) \ ,$$

$$\frac{\partial F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)}{\partial \gamma_h} = \varphi(\Delta_h + \beta_h)\Phi(\Delta_w^* + \beta_w^* - \rho(\Delta_h^* + \beta_h^*))x_h,$$

$$\frac{\partial F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)}{\partial \gamma_w} = \varphi(\Delta_w + \beta_w)\Phi(\Delta_h^* + \beta_h^* - \rho(\Delta_w^* + \beta_w^*))x_w \ ;$$

$$\frac{\partial F(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)}{\partial \rho} = f(\Delta_h + \beta_h, \Delta_w + \beta_w, \rho)$$

and

$$\frac{\partial I_{++}}{\partial \beta_h} = -\varphi(\Delta_h + \beta_h)\Phi(-\Delta_w^* - \beta_w^* + \rho(\Delta_h^* + \beta_h^*))$$

$$+ \varphi(\Delta_h + \beta_h)\Phi(- \Delta_w^* + \rho(\Delta_h^* + \beta_h^*)) \ ,$$

$$\frac{\partial I_{++}}{\partial \beta_w} = - \varphi(\Delta_w + \beta_w)\Phi(- \Delta_h^* - \beta_h^* + \rho(\Delta_w^* + \beta_w^*))$$

$$+ \varphi(\Delta_w + \beta_w)\Phi(- \Delta_h^* + \rho(\Delta_w^* + \beta_w^*)) \ ,$$

$$\frac{\partial I_{++}}{\partial \gamma_h} = [-\varphi(\Delta_h)\Phi(- \Delta_w^* + \rho\Delta_h^*)$$

$$- \varphi(\Delta_h + \beta_h)\Phi(- \Delta_w^* - \beta_w^* + \rho(\Delta_h^* + \beta_h^*))$$

$$+ \varphi(\Delta_h)\Phi(- \Delta_w^* - \beta_w^* + \rho\Delta_h^*)$$

$$+ \varphi(\Delta_h + \beta_h)\Phi(- \Delta_w^* + \rho(\Delta_h^* + \beta_h^*))] \ x_h \ ,$$

$$\frac{\partial I_{++}}{\partial \gamma_w} = [- \varphi(\Delta_w)\Phi(-\Delta_h^* + \rho\Delta_w^*) - \varphi(\Delta_w + \beta_w)\Phi(- \Delta_h^* - \beta_h^* + \rho(\Delta_w^* + \beta_w^*))$$

$$+ \varphi(\Delta_w)\Phi(- \Delta_h^* - \beta_h^* + \rho\Delta_w^*)$$

$$+ \varphi(\Delta_w + \beta_w)\Phi(-\Delta_h^* + \rho(\Delta_w^* + \beta_w^*))] \ x_w \ ,$$

$$\frac{\partial I_{++}}{\partial \rho} = f(\Delta_h,\Delta_w,\rho) + f(\Delta_h + \beta_h,\Delta_w + \beta_w,\rho) - f(\Delta_h,\Delta_w + \beta_w,\rho)$$

$$- f(\Delta_h + \beta_h,\Delta_w,\rho)$$

with

$$\frac{\partial I_{+-}}{\partial(\cdot)} = \frac{\partial I_{-+}}{\partial(\cdot)} = - \frac{\partial I_{++}}{\partial(\cdot)} \quad \text{and} \quad \frac{\partial I_{--}}{\partial(\cdot)} = \frac{\partial I_{++}}{\partial(\cdot)} \ .$$

CHAPTER III: AN ECONOMETRIC MODEL OF A STACKELBERG GAME

1. INTRODUCTION

In the previous chapter, we proposed an alternative simultaneous model for discrete endogenous variables. A distinctive feature of that model is that no logical consistency constraints on the parameters need to be imposed. In addition, our simultaneous model was derived from optimizing behavior as an outcome of a game between two players. The equilibrium concept used was that of Nash. In the current chapter, we shall propose an econometric model of an alternative equilirium concept from noncooperative game theory.

In this alternative game theoretic formulation, we shall still assume that each player maximizes his own utility. The model proposed in the current chapter is, however, different from the simultaneous model of the previous chapter since the equilibrium concept used here will be that of Stackelberg. Although it may appear that the model is recursive, it will be seen that the model in fact generalizes recursive models for discrete endogenous variables that have been considered up to now in the literature (see, e.g., Maddala and Lee (1976)). As before, our model becomes stochastic by adopting the random utility framework introduced by McFadden (1974, 1981). For notational convenience, we derive the statistical model by assuming the husband is the Stackelberg leader and his wife the follower; that is, we assume the husband knows what action his wife will take conditional upon his action and he thus optimizes accordingly.

This chapter is organized as follows. In Section 2, we derive the statistical model where the outcomes are generated as Stackelberg equilibria of a game played between two players. Section 3 compares the usual formulation of the problem in terms of recursive models with our alternative formulation. In particular, it is shown that the usual recursive model is nested in our more general model. In Section 4, we discuss identification of the model and estimation issues as they relate to identification. Section 5 concludes the chapter. The first partial derivatives of the probabilities, which are needed both for identification and estimation, are found in the Appendix to this chapter.

2. THE MODEL

For ease of exposition, assume that the husband is the Stackelberg leader and the wife is the follower. Let $\tilde{U}_h(i,j)$ be the payoff to the husband when he takes action i and his wife takes action j, $i,j \in \{0,1\}$. Analogously, let $\tilde{U}_w(j,i)$ be the payoff to the wife. Then we have the extensive form:

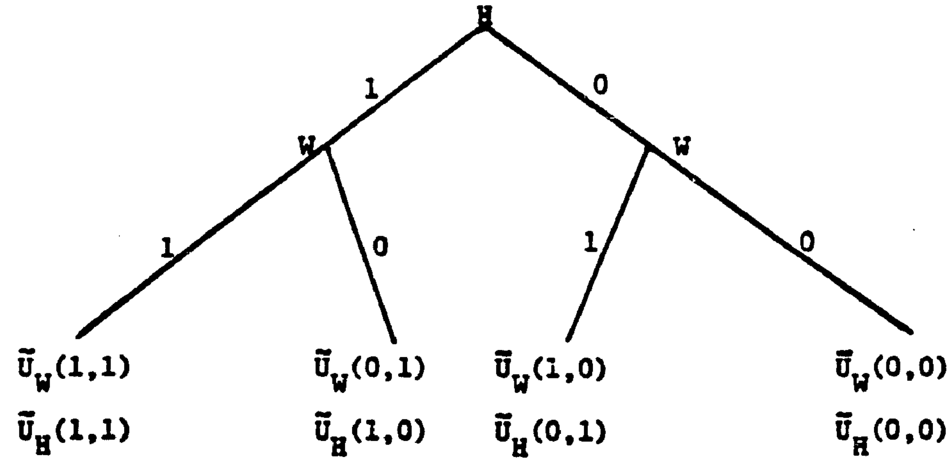

Figure 1

The husband, in making his decision whether to take action 1 or 0 must take the wife's payoffs into account. That is, the husband must take action i such that when the wife takes action j, conditional on i, $\tilde{U}_h(i,j)$ gives the husband the greatest possible payoff. There are four possible cases, W_1, W_2, W_3, and W_4 for the husband to consider before taking his action i:[1]

$$W_1 : \tilde{U}_w(1,0) \geq \tilde{U}_w(0,0) \ \& \ \tilde{U}_w(1,1) \geq \tilde{U}_w(0,1)$$
$$W_2 : \tilde{U}_w(1,0) < \tilde{U}_w(0,0) \ \& \ \tilde{U}_w(1,1) \geq \tilde{U}_w(0,1)$$
$$W_3 : \tilde{U}_w(1,0) < \tilde{U}_w(0,0) \ \& \ \tilde{U}_w(1,1) < \tilde{U}_w(0,1)$$
$$W_4 : \tilde{U}_w(1,0) \geq \tilde{U}_w(0,0) \ \& \ \tilde{U}_w(1,1) < \tilde{U}_w(0,1).$$

The four cases W_1, W_2, W_3, and W_4 are the wife's reaction functions as given in Figure 2. For example, reaction function W_1 says that whether the husband chooses action 1 or 0, the wife always chooses action 1. Conditional on the reaction function

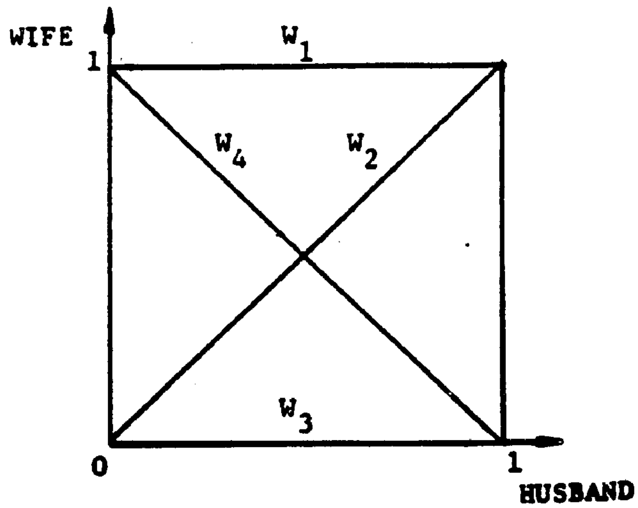

Figure 2: Wife's Reaction Functions

chosen by the wife, the husband then takes that action which maximizes his payoff. For example, if the wife follows reaction function W_1, the husband will choose action 1 when $\tilde{U}_h(1,1) \geq \tilde{U}_h(0,1)$, while choosing action 0 when the inequality is reversed. Thus, each reaction function W_i for the wife calls for a payoff comparison C_i for the husband. Therefore we define:

$$C_1 : \tilde{U}_h(1,1) \geq \tilde{U}_h(0,1)$$
$$C_2 : \tilde{U}_h(1,1) \geq \tilde{U}_h(0,0)$$
$$C_3 : \tilde{U}_h(1,0) \geq \tilde{U}_h(0,0)$$
$$C_4 : \tilde{U}_h(1,0) \geq \tilde{U}_h(0,1).$$

Let $\overline{C}_i$ indicate the negation of C_i.

Now that the reaction functions for the wife W_i and the payoff comparisons for the husband C_i have been defined, we can readily find the Stackelberg outcomes of this game, as indicated in Table 1.[2] Note that for each outcome, the first number in each ordered pair refers to the husband while the second number refers to the wife.

Table 1: Stackelberg Equilibria

W_1 & C_1	(1,1)	W_3 & C_3	(1,0)
W_1 & $\bar{C}_1$	(0,1)	W_3 & $\bar{C}_3$	(0,0)
W_2 & C_2	(1,1)	W_4 & C_4	(1,0)
W_2 & $\bar{C}_2$	(0,0)	W_4 & $\bar{C}_4$	(0,1)

To introduce a stochastic structure, we shall follow McFadden (1974, 1981). The utilities $\tilde{U}_h(i,j)$ and $\tilde{U}_w(j,i)$ are then treated as random, and decomposed into deterministic components and random components. Further, we shall allow for the possibility that the utility the husband receives depends on the wife's decision whether or not to work. We make a similar allowance for the wife. Then formally we have the following set of four equations:[3]

$$\tilde{U}_h(1,Y_w) = U_h^1 + \alpha_h^1 Y_w + \eta_h^1 \tag{1}$$
$$\tilde{U}_h(0,Y_w) = U_h^0 + \alpha_h^0 Y_w + \eta_h^0 \tag{2}$$
$$\tilde{U}_w(1,Y_h) = U_w^1 + \alpha_w^1 Y_h + \eta_w^1 \tag{3}$$
$$\tilde{U}_w(0,Y_h) = U_w^0 + \alpha_w^0 Y_h + \eta_w^0. \tag{4}$$

where

$$Y_h = \begin{cases} 1 & \text{if the husband works} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_w = \begin{cases} 1 & \text{if the wife works} \\ 0 & \text{otherwise} \end{cases}$$

To illustrate, the utility that the husband receives from working when his wife also works ($Y_w = 1$) is given by $\tilde{U}_{ht}(1,1) = U_{ht}^1 + \alpha_h^1 + \eta_h^1$. As can be seen from the wife's reaction functions W_i and the husband's utility (payoff) comparisons C_i, $i = 1,2,3,4$, only differences in utilities are relevant in the husband's and wife's respective decisions whether or not to work. As a result, we define $\varepsilon_h \equiv \eta_h^1 - \eta_h^0$ and $\varepsilon_w \equiv \eta_w^1 - \eta_w^0$. It is assumed from now on that the pair $(\varepsilon_h, \varepsilon_w)$ is normally distributed with zero means, unit variances and correlation ρ.

The distribution of the random components $(\varepsilon_h, \varepsilon_w)$ then induces a probabilistic structure on the observed decisions (Y_h, Y_w). Indeed, each reaction function W_i for the wife will occur if certain conditions on the random component ε_w are satisfied. Let us now derive these conditions.

Using Figure 2, reaction function W_1 is characterized by the following two conditions: $\tilde{U}_w(1,0) \geq \tilde{U}_w(0,0)$ and $\tilde{U}_w(1,1) \geq \tilde{U}_w(0,1)$. From (3) and (4) these conditions are equivalent to $\varepsilon_w \geq -(U_w^1 - U_w^0)$ and $\varepsilon_w \geq -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$, respectively, which can be combined to give $\varepsilon_w \geq -(U_w^1 - U_w^0) - \min(0, \alpha_w^1 - \alpha_w^0)$.

Reaction function W_2 is characterized by $\tilde{U}_w(1,0) < \tilde{U}_w(0,0)$ and $\tilde{U}_w(1,1) \geq \tilde{U}_w(0,1)$, which are equivalent to $\varepsilon_w \leq -(U_w^1 - U_w^0)$ and $\varepsilon_w > -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$, respectively. When combined, we get $-(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0) < \varepsilon_w < -(U_w^1 - U_w^0)$ if $\alpha_w^1 - \alpha_w^0 \geq 0$; otherwise, reaction function W_2 cannot occur.

Reaction function W_3 is characterized by $\tilde{U}_w(1,0) < \tilde{U}_w(0,0)$ and

$\tilde{U}_w(1,1) < \tilde{U}_w(0,1)$. Using (3) and (4), these conditions are equivalent to $\varepsilon_w < -(U_w^1 - U_w^0)$ and $\varepsilon_w < -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$, respectively. When combined, we get $\varepsilon_w < -(U_w^1 - U_w^0) - \max(0,\ \alpha_w^1 - \alpha_w^0)$.

Reaction function W_4 is characterized by $\tilde{U}_w(1,0) \geq \tilde{U}_w(0,0)$ and $\tilde{U}_w(1,1) < \tilde{U}_w(0,1)$, which are equivalent to $\varepsilon_w \geq -(U_w^1 - U_w^0)$ and $\varepsilon_w < -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$, respectively, which when combined give $-(U_w^1 - U_w^0) < \varepsilon_w < -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$ if $\alpha_w^1 - \alpha_w^0 < 0$; otherwise reaction function W_4 cannot occur. We thus have the following table.

Table 2: Conditions for Wife's Reaction Functions

W_1 : $\varepsilon_w > -(U_w^1 - U_w^0) - \min(0, \alpha_w^1 - \alpha_w^0)$

W_2 : $-(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0) < \varepsilon_w < -(U_w^1 - U_w^0)$ if $\alpha_w^1 - \alpha_w^0 \geq 0$

otherwise cannot occur

W_3 : $\varepsilon_w < -(U_w^1 - U_w^0) - \max(0, \alpha_w^1 - \alpha_w^0)$

W_4 : $-(U_w^1 - U_w^0) < \varepsilon_w < -(U_w^1 - U_w^0 + \alpha_w^1 - \alpha_w^0)$ if $\alpha_w^1 - \alpha_w^0 < 0$;

otherwise cannot occur

Once a reaction function for the wife is determined, a utility comparison for the husband is also determined; that is, if the wife's reaction function is given by W_i, the husband makes utility comparison C_i, $i = 1,\ldots,4$. As with the wife, each utility comparison C_i will occur if a certain condition on the random component ε_h is satisfied. These are now derived.

We see from Figure 1 that when the wife follows reaction function W_1, the husband compares $\tilde{U}_h(1,1)$ and $\tilde{U}_h(0,1)$. If $\tilde{U}_h(1,1) \geq \tilde{U}_h(0,1)$, then from (1) and (2) we have that

$\varepsilon_h \geq -(U_h^1 - U_h^0 + \alpha_h^1 - \alpha_h^0)$.

When the wife follows reaction function W_2, the husband compares $\tilde{U}_h(1,1)$ and $\tilde{U}_h(0,0)$. When $\tilde{U}_h(1,1) \geq \tilde{U}_h(0,0)$, we have that $\varepsilon_h \geq -(U_h^1 - U_h^0 + \alpha_h^1)$.

When reaction function W_3 is used, the husband compares $\tilde{U}_h(1,1)$ and $\tilde{U}_h(0,1)$. When $\tilde{U}_h(1,1) \geq \tilde{U}_h(0,1)$, we have that $\varepsilon_h \geq -(U_h^1 - U_h^0)$.

Finally, Figure 1 shows that when the wife uses W_4, the husband makes a comparison between $\tilde{U}_h(1,0)$ and $\tilde{U}_h(0,1)$. If $\tilde{U}_h(1,0) \geq \tilde{U}_h(0,1)$, we have from (1) and (2) that $\varepsilon_h \geq -(U_h^1 - U_h^0 - \alpha_h^0)$. We therefore have the following table.

Table 3: Conditions for Husband's Utility Comparisons

$$
\begin{aligned}
C_1 &: \varepsilon_h \geq -(U_h^1 - U_h^0 + \alpha_h^1 - \alpha_h^0) \\
C_2 &: \varepsilon_h \geq -(U_h^1 - U_h^0 + \alpha_h^1) \\
C_3 &: \varepsilon_h \geq -(U_h^1 - U_h^0) \\
C_4 &: \varepsilon_h \geq -(U_h^1 - U_h^0 - \alpha_h^0)
\end{aligned}
$$

Now that randomness has been introduced into the model, we can derive the joint probabilities on the part of both the husband and wife whether or not to work. Let $Pr(i,j)$ be the probability that the random variables Y_h and Y_w take on the values i and j, $i,j \in \{0,1\}$. From Table 1, we have

$$Pr(0,0) = Pr(W_2 \;\&\; \bar{C}_2) + Pr(W_3 \;\&\; \bar{C}_3) \tag{5}$$

$$Pr(1,0) = Pr(W_3 \;\&\; C_3) + Pr(W_4 \;\&\; C_4) \tag{6}$$

$$\Pr(0,1) = \Pr(W_1 \,\&\, \bar{C}_1) + \Pr(W_4 \,\&\, \bar{C}_4) \tag{7}$$

$$\Pr(1,1) = \Pr(W_1 \,\&\, C_1) + \Pr(W_2 \,\&\, C_2). \tag{8}$$

Using Tables 2 and 3 and Equations (5)-(8) we can derive the probabilities in terms of the unknown parameters. Let $F(a,b,\rho)$ be the c.d.f. evaluated at (a,b) of a bivariate normal distribution with zero means, unit variances, and correlation ρ. Moreover, let $I(a,b,c,d,\rho)$ be the integral corresponding to a bivariate density over the range $a \geq \varepsilon_h \geq c,\ b \geq \varepsilon_w \geq d$.[4] As can be seen from Table 2, the probabilites $\Pr(i,j)$ will depend on the sign of $\Delta\alpha_w \equiv (\alpha_w^1 - \alpha_w^0)$. We then have:

PROPOSITION 1:

$$\begin{aligned}
\Pr(0,0) &= F(-\Delta U_h, -\Delta U_w, \rho) - I_+^B && \text{if } \Delta\alpha_w \geq 0 && (9)\\
&= F(-\Delta U_h, -\Delta U_w, \rho) && \text{otherwise} &&\\
\Pr(1,0) &= F(\Delta U_h, -\Delta U_w - \Delta\alpha_w, -\rho) && \text{if } \Delta\alpha_w \geq 0 && (10)\\
&= F(\Delta U_h, -\Delta U_w - \Delta\alpha_w, -\rho) + I_-^B && \text{otherwise} &&\\
\Pr(0,1) &= F(-\Delta U_h - \alpha_h^1 + \alpha_h^0, \Delta U_w, -\rho) && \text{if } \Delta\alpha_w \geq 0 && (11)\\
&= F(-\Delta U_h - \alpha_h^1 + \alpha_h^0, \Delta U_w, -\rho) + I_-^A && \text{otherwise} &&\\
\Pr(1,1) &= F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho) - I_+^A && \text{if } \Delta\alpha_w \geq 0, && (12)\\
&= F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho) && \text{otherwise} &&
\end{aligned}$$

where

$$\begin{aligned}
I_+^A &= I(-\Delta U_h - \alpha_h^1, -\Delta U_w, -\Delta U_h - \alpha_h^1 + \alpha_h^0, -\Delta U_w - \Delta\alpha_w, \rho)\\
I_+^B &= I(-\Delta U_h, -\Delta U_w, -\Delta U_h - \alpha_h^1, -\Delta U_w - \Delta\alpha_w, \rho)\\
I_-^A &= I(-\Delta U_h + \alpha_h^0, -\Delta U_w - \Delta\alpha_w, -\Delta U_h - \alpha_h^1 + \alpha_h^0, -\Delta U_w, \rho) && (13)
\end{aligned}$$

$$I_{-}^{B} = I(-\Delta U_h, -\Delta U_w - \Delta\alpha_w, -\Delta U_h + \alpha_h^0, -\Delta U_w, \rho)$$

$$\Delta U_h \equiv U_h^1 - U_h^0 \text{ and } \Delta U_w \equiv U_w^1 - U_w^0.$$

PROOF:

From Table 2, it is clear that reaction function W_4 for the wife cannot occur when $(\alpha_w^1 - \alpha_w^0) \geq 0$, while reaction function W_2 cannot occur when $(\alpha_w^1 - \alpha_w^0) < 0$. Thus when $(\alpha_w^1 - \alpha_w^0) \geq 0$ it follows from equations (5) - (8) that

$$\Pr(0,0) = \Pr(W_2 \,\&\, \bar{C}_2) + \Pr(W_3 \,\&\, \bar{C}_3),$$
$$\Pr(1,0) = \Pr(W_3 \,\&\, C_3),$$
$$\Pr(0,1) + \Pr(W_1 \,\&\, \bar{C}_1),$$
$$\Pr(1,1) = \Pr(W_1 \,\&\, C_1) + \Pr(W_2 \,\&\, C_2).$$

Similarly, when $(\alpha_w^1 - \alpha_w^0) < 0$, we have that

$$\Pr(0,0) = \Pr(W_3 \,\&\, \bar{C}_3),$$
$$\Pr(1,0) = \Pr(W_3 \,\&\, C_3) + \Pr(W_4 \,\&\, C_4),$$
$$\Pr(0,1) = \Pr(W_1 \,\&\, \bar{C}_1) + \Pr(W_4 \,\&\, \bar{C}_4),$$
$$\Pr(1,1) = \Pr(W_1 \,\&\, C_1).$$

Now, using the conditions on ε_w and ε_h given in Tables 2 and 3, respectively, we can derive the needed comparisons between particular W_i, C_i, and $\bar{C}_i$, $i = 1,\ldots,4$. For the cases $\Delta\alpha_w \equiv (\alpha_w^1 - \alpha_w^0) \geq 0$ and $\Delta\alpha_w \equiv (\alpha_w^1 - \alpha_w^0) < 0$, figures 3a and 3b respectively show the areas over which the bivariate normal density must be integrated to obtain the four probabilities

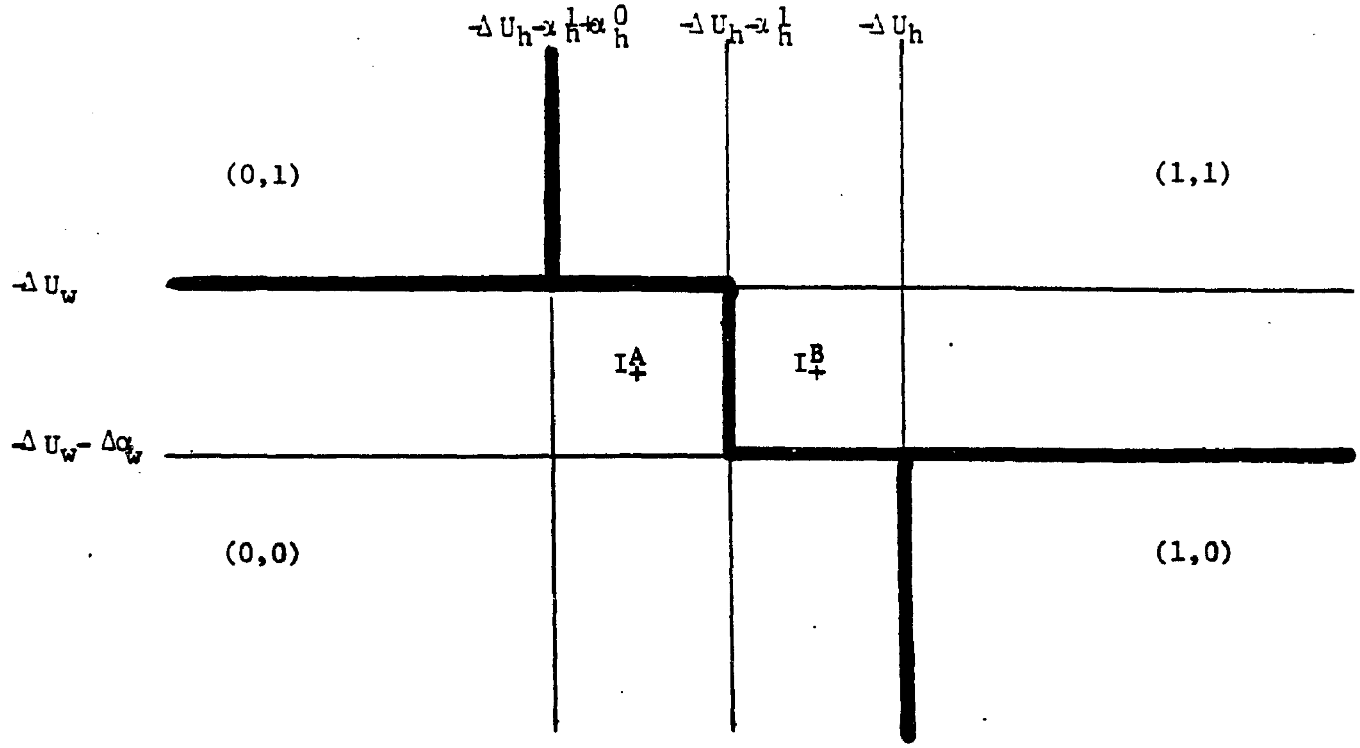

figure 3a

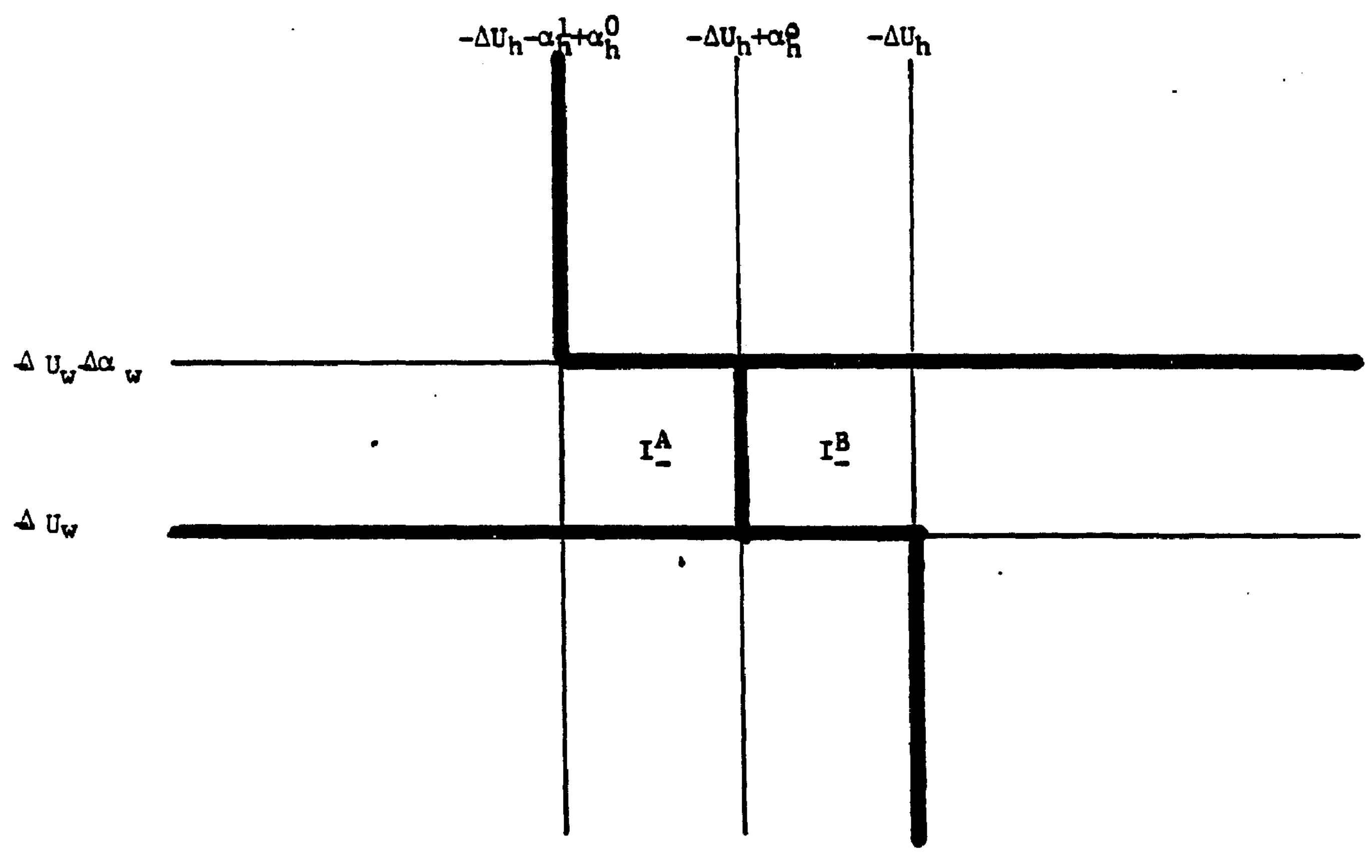

figure 3b

Pr(0,0), Pr(1,0), Pr(0,1), and Pr(1,1). Without loss of generality, figures 3a and 3b are drawn for the case $\alpha_h^0 < 0 < \alpha_h^1$. It can be seen from figures 3a and 3b that I_+^A, I_+^B, I_-^A and I_-^B correspond to the areas over the bivariate normal density given by (13). It follows that the probabilities Pr(0,0), Pr(1,0), Pr(0,1), and Pr(1,1) are given by equations (9) – (13) in Proposition 1.

Q.E.D.

Let us now examine how the probabilities $Pr(1,\cdot)$ and $Pr(\cdot,1)$ will change as we allow the parameters to vary. These are given as $Pr(1,\cdot) \equiv Pr(1,1) + Pr(1,0)$ and $Pr(\cdot,1) \equiv Pr(1,1) + Pr(0,1)$, respectively. We then have:

PROPOSITION 2:

(i) An increase in α_h^1 or ΔU_h always increases the probability that the husband will work, $Pr(1,\cdot)$;

(ii) an increase in α_h^0 always decreases the probability that the husband will work;

(iii) an increase in $\Delta\alpha_w$ or ΔU_w always increase the probability that the wife will work, $Pr(\cdot,1)$.

PROOF:

Easily established by using either the areas defining the probabilities Pr(i,j), as seen in figures 3a and 3b or by differentiating the probabilities found in Proposition 1. Included below is a table indicating the direction of change in the probabilities as all parameters are allowed to vary.

Case 1: $\Delta\alpha_w \equiv (\alpha_w^1 - \alpha_w^0) \geq 0$

	Pr(0,0)	Pr(1,0)	Pr(0,1)	Pr(1,1)	Pr(1,·)	Pr(·,1)
α_h^1	-	no change	-	+	+	+
α_h^0	no change	no change	+	-	-	no change
$\Delta\alpha_w$	?	-	no change	+	?	+
ΔU_h	-	+	-	+	+	+
ΔU_w	?	-	+	?	?	+

Case 2 : $\Delta\alpha_w \equiv (\alpha_w^1 - \alpha_w^0) < 0$

	Pr(0,0)	Pr(1,0)	Pr(0,1)	Pr(1,1)	Pr(1,·)	Pr(·,1)
α_h^1	no change	no change	-	+	+	no change
α_h^0	no change	-	+	-	-	+
$\Delta\alpha_w$	no change	-	?	+	?	+
ΔU_h	-	+	-	+	+	-
ΔU_w	-	?	?	+	?	+

Q.E.D.

As expected, an increase in ΔU_h increases the probability that the husband will work, whether or not the wife chooses to work; a similar remark holds for an increase in ΔU_w. Also, as can be seen from equation (1), an increase in α_h^1 increases the probability that the husband will work when he knows his wife wishes to work, while having no effect on his propensity to work when he knows his wife chooses not to work. From equation (2), it is clear that an increase in α_h^0 increases the husband's utility of not working. Finally using equations (3) and (4), it is seen that an increase in ΔU_w increases the wife's utility of joining the labor market.

3. A COMPARISON OF MODELS

Now that we have developed a model in which the outcomes of the sequential decision-making problem are generated as Stackelberg equilibria of a game between two players, we are in a position to compare it to the usual recursive probability model for dichotomous variables (see e.g., Maddala and Lee (1976)). According to the usual formulation, a recursive equation system is described in terms of latent continuous variables, where the observed dichotomous variables are generated using a dichotomization. In our case, the corresponding recursive probability model is

$$Y_w^* = \Delta_w + \beta_w Y_h + \varepsilon_w \qquad (14)$$

$$Y_h^* = \Delta_h + \varepsilon_h \qquad (15)$$

for some Δ_h and Δ_w, and

$$Y_h = \begin{cases} 1 & \text{if } Y_h^* > 0, \\ 0 & \text{otherwise,} \end{cases} \qquad Y_w = \begin{cases} 1 & \text{if } Y_w^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The purpose of this section is to show that this recursive probability model is nested in our model of Section 2.

Suppose that

$$\alpha_h^1 = \alpha_h^0 = 0; \tag{16}$$

then from equations (1) and (2) defining the husband's utilities, we have:

$$\tilde{U}_h(1,Y_w) = U_h^1 + \eta_h^1$$
$$\tilde{U}_h(0,Y_w) = U_h^0 + \eta_h^0.$$

Thus the restrictions (16) can be interpreted as imposing that the utilities derived by the husband from working or not working do not depend on the wife's decision whether or not to work.

But now note that if the restrictions (16) hold, then from Table 3, the four conditions C_1, C_2, C_3, and C_4 are identical; that is, $\varepsilon_h \geq -\Delta U_h$. Looking now at the conditions for the wife's reaction functions, we have to distinguish two cases according to the sign of $\Delta\alpha_w$. Suppose first that $\Delta\alpha_w \geq 0$. Then it is readily seen from Tables 1 and 2 that the pairs (1,1), (1,0), (0,1), and (0,0) occur under the following conditions:

(1,1) if and only if $\Delta U_h + \varepsilon_h \geq 0$ and $\Delta U_w + \Delta\alpha_w + \varepsilon_w \geq 0$,

$(1,0)$ if and only if $\Delta U_h + \varepsilon_h \geq 0$ and $\Delta U_w + \Delta\alpha_w + \varepsilon_w < 0$,

$(0,1)$ if and only if $\Delta U_h + \varepsilon_h < 0$ and $\Delta U_w + \varepsilon_w \geq 0$,

$(0,0)$ if and only if $\Delta U_h + \varepsilon_h < 0$ and $\Delta U_w + \varepsilon_w < 0$.

It suffices now to note that these conditions are exactly identical to the ones that are obtained from the recursive probability model (14)-(15) with the usual dichotomization where $\Delta_h = \Delta U_h$, $\Delta_w = \Delta U_w$ and $\beta_w = \Delta\alpha_w$. The case $\Delta\alpha_w < 0$ is similarly studied, and gives the same conditions as above on the errors ε_h and ε_w. We have therefore established the following proposition.

PROPOSITION 3: If the restrictions $\alpha_h^1 = \alpha_h^0 = 0$ hold, then the usual recursive probability model using the dichotomization rule is identical to our model in which the observed outcomes are generated as Stackelberg equilibria.

The import of Proposition 3 is that it gives a structural interpretation to the usual recursive probability model in terms of a Stackelberg game. In addition , since the restrictions (16) on the parameters of our model must hold in order for the result in Proposition 3 to hold, it follows that the usual recursive probability model is nested in our proposed model. As an empirical consequence, it is then possible to test the specification of the usual recursive model by testing $\alpha_h^1 = \alpha_h^0 = 0$. Finally, given the above interpretation of these restrictions, it can be seen that these restrictions are unrealistic since they impose that the utility the husband derives from working or not working does not depend on whether the wife is working. Thus, the usual recursive formulation is inappropriate since it implicitly assumes that the leader is indifferent to the follower's action. Let us also note that although the husband is moving first and in principle should take into account his wife's conditional action when making his decision, the restrictions (16), when imposed, lead the husband to ignore his wife's action.

4. IDENTIFICATION AND ESTIMATION

Given the previous expressions for the probabilities Pr(i,j) of the observed dichotomous variables Y_h and Y_w, the log-likelihood function under random sampling is written as:

$$\begin{aligned} L &= \sum_t \log Pr_t(Y_{ht}, Y_{wt}) \qquad (17) \\ &= \sum_t [Y_{ht}Y_{wt} \log Pr_t(1,1) + Y_{ht}(1 - Y_{wt}) \log Pr_t(1,0) \\ &+ (1 - Y_{ht})Y_{wt} \log Pr_t(0,1) + (1 - Y_{ht})(1 - Y_{wt}) \log Pr_t(0,0)], \end{aligned}$$

where the subscript t indexes the observations. The probabilities are subscripted by t since ΔU_h and ΔU_w are in general functions of explanatory variables. We assume as in Chapter II:

$$\Delta U_{ht} = x_{ht}'\gamma_h \quad \text{and} \quad \Delta U_{wt} = x_{wt}'\gamma_w, \qquad (18)$$

where x_{ht} may include characteristics of the t-th household and characteristics of the husband. A similar remark applies to x_{wt}. We now turn to the conditions under which the parameters $(\rho, \Delta\alpha_w, \alpha_h^0, \alpha_h^1, \gamma_h, \gamma_w)$ of our model are identified.

In order to discuss identification, we first need to introduce some notation. Define the following partitioned matrix $\tilde{A}$ as

$$\tilde{A} = \left[D_\rho \bar{X}_\rho \;\middle|\; D_h \bar{X}_h \;\middle|\; D_w \bar{X}_w \right],$$

where D_ρ , D_h and D_w are block diagonal matrices of order 3T, the t-th blocks given as follows:

if $\Delta\alpha_w \geq 0$

$$D_{\rho t} = \begin{bmatrix} r_t^{1+} & 0 & 0 \\ 0 & r_t^{2+} & r_t^{3+} \\ 0 & 0 & r_t^{4+} \end{bmatrix}, \; D_{ht} = \begin{bmatrix} 0 & 0 & a_t^{h+} \\ 0 & (b_t^{h+}+c_t^{h+}) & 0 \\ -d_t^{h+} & 0 & 0 \end{bmatrix}, \text{ and}$$

$$D_{wt} = \begin{bmatrix} 0 & 0 & a_t^{w+} \\ b_t^{w+} & c_t^{w+} & 0 \\ d_t^{w+} & 0 & 0 \end{bmatrix};$$

if $\Delta\alpha_w < 0$

$$D_{\rho t} = \begin{bmatrix} r_t^{1-} & 0 & 0 \\ 0 & r_t^{2-} & 0 \\ 0 & r_t^{3-} & r_t^{4-} \end{bmatrix}, \; D_{ht} = \begin{bmatrix} 0 & 0 & a_t^{h-} \\ 0 & b_t^{h-} & 0 \\ -(c_t^{h-}+d_t^{h-}) & 0 & 0 \end{bmatrix}, \text{ and}$$

$$D_{wt} = \begin{bmatrix} a_t^{w-} & 0 & 0 \\ 0 & b_t^{w-} & 0 \\ c_t^{w-} & 0 & d_t^{w-} \end{bmatrix}.$$

The elements of the above matrices are described in the Appendix to this chapter. The matrices $\bar{X}_h$ and $\bar{X}_w$ are of dimension 3T by $K_h + 2$ and 3T by $K_w + 1$, the t-th blocks given respectively as:

$$\begin{bmatrix} -1 & 0 & x_{ht}' \\ 0 & 1 & x_{ht}' \\ 0 & 0 & x_{ht}' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & x_{wt}' \\ 1 & x_{wt}' \\ 1 & x_{wt}' \end{bmatrix}.$$

In addition, $\bar{X}_\rho$ is a unit vector of dimension 3T.

PROPOSITION 4: The parameters $(\rho, \Delta\alpha_w, \alpha_h^1, \alpha_h^0, \gamma_h, \gamma_w)$ of the model are identified if and only if $\tilde{A}$ has full column rank.

PROOF:

Let $Z_t = (Y_{ht}, Y_{wt}, X_{ht}', X_{wt}')$ and $\theta = (\rho, \Delta\alpha_w, \alpha_h^0, \alpha_h^1, \gamma_h, \gamma_w)$. Define:

$$B = E\left[\sum_{t=1}^{T} \frac{\partial \log f(Z_t,\theta)}{\partial\theta} \cdot \frac{\partial \log f(Z_t,\theta)}{\partial\theta'}\right] = \sum_{t=1}^{T} B_t.$$

From expression (17), we have, omitting the subscript t, that

$$\frac{\partial \log f(Z,\theta)}{\partial\theta} = \frac{Y_h Y_w}{\Pr(1,1)} \frac{\partial \Pr(1,1)}{\partial\theta} + \frac{Y_h(1 - Y_w)}{\Pr(1,0)} \frac{\partial \Pr(1,0)}{\partial\theta}$$

$$+ \frac{(1 - Y_h)Y_w}{\Pr(0,1)} \frac{\partial \Pr(0,1)}{\partial\theta} + \frac{(1 - Y_h)(1 - Y_w)}{\Pr(0,0)} \frac{\partial \Pr(0,0)}{\partial\theta}.$$

Then, $\frac{\partial \log f}{\partial\rho} \frac{\partial \log f}{\partial\rho}$, say, is given by

$$\left[\frac{Y_h Y_w}{\Pr(1,1)} \frac{\partial \Pr(1,1)}{\partial\rho}\right]^2 + \left[\frac{Y_h(1 - Y_w)}{\Pr(1,0)} \frac{\partial \Pr(1,0)}{\partial\rho}\right]^2$$

$$+ \left[\frac{(1 - Y_h)Y_w}{Pr(0,1)} \frac{\partial Pr(0,1)}{\partial \rho}\right]^2 + \left[\frac{(1 - Y_h)(1 - Y_w)}{Pr(0,0)} \frac{\partial Pr(0,0)}{\partial \rho}\right]^2,$$

where we have used the fact that Y_h and Y_w take on only the values zero or one. Since Y_h and Y_w are random variables where $Y_h = i$, $Y_w = j$ with probability $Pr(i,j)$, $i,j \in \{0,1\}$, we have that

$$E\left[\frac{\partial \log f}{\partial \rho} \frac{\partial \log f}{\partial \rho}\right] = \frac{1}{Pr(1,1)}\left[\frac{\partial Pr(1,1)}{\partial \rho}\right]^2$$

$$+ \frac{1}{Pr(1,0)}\left[\frac{\partial Pr(1,0)}{\partial \rho}\right]^2 + \frac{1}{Pr(0,1)}\left[\frac{\partial Pr(0,1)}{\partial \rho}\right]^2 + \frac{1}{Pr(0,0)}\left[\frac{\partial Pr(0,0)}{\partial \rho}\right]^2.$$

Proceeding analogously, the remaining terms in B are given by:

$$E\left[\frac{\partial \log f}{\partial \theta_k} \cdot \frac{\partial \log f}{\partial \theta_h}\right] = \sum_{i=0}^{1}\sum_{j=0}^{1}\frac{1}{Pr(i,j)} \frac{\partial Pr(i,j)}{\partial \theta_k} \frac{\partial Pr(i,j)}{\partial \theta_h}.$$

Notice that B can be decomposed into B = A'DA where A is of dimension 4T by K, $K = K_h + K_w + 4$, that has as its t-th block A_t defined as:

$$\begin{array}{cccccc}
\frac{\partial Pr_t(1,1)}{\partial \rho} & \frac{\partial Pr_t(1,1)}{\partial \Delta\alpha_w} & \frac{\partial Pr_t(1,1)}{\partial \alpha_h^0} & \frac{\partial Pr_t(1,1)}{\partial \alpha_h^1} & \frac{\partial Pr_t(1,1)}{\partial \gamma_h'} & \frac{\partial Pr_t(1,1)}{\partial \gamma_w'} \\[2ex]
\frac{\partial Pr_t(1,0)}{\partial \rho} & \frac{\partial Pr_t(1,0)}{\partial \Delta\alpha_w} & \frac{\partial Pr_t(1,0)}{\partial \alpha_h^0} & \frac{\partial Pr_t(1,0)}{\partial \alpha_h^1} & \frac{\partial Pr_t(1,0)}{\partial \gamma_h'} & \frac{\partial Pr_t(1,0)}{\partial \gamma_w'} \\[2ex]
\frac{\partial Pr_t(0,1)}{\partial \rho} & \frac{\partial Pr_t(0,1)}{\partial \Delta\alpha_w} & \frac{\partial Pr_t(0,1)}{\partial \alpha_h^0} & \frac{\partial Pr_t(0,1)}{\partial \alpha_h^1} & \frac{\partial Pr_t(0,1)}{\partial \gamma_h'} & \frac{\partial Pr_t(0,1)}{\partial \gamma_w'} \\[2ex]
\frac{\partial Pr_t(0,0)}{\partial \rho} & \frac{\partial Pr_t(0,0)}{\partial \Delta\alpha_w} & \frac{\partial Pr_t(0,0)}{\partial \alpha_h^0} & \frac{\partial Pr_t(0,0)}{\partial \alpha_h^1} & \frac{\partial Pr_t(0,0)}{\partial \gamma_h'} & \frac{\partial Pr_t(0,0)}{\partial \gamma_w'}
\end{array}$$

and D is a block diagonal matrix of order 4T, the t-th block given by

$$\begin{bmatrix} Pr_t(1,1) & 0 & 0 & 0 \\ 0 & Pr_t(1,0) & 0 & 0 \\ 0 & 0 & Pr_t(0,1) & 0 \\ 0 & 0 & 0 & Pr_t(0,0) \end{bmatrix}^{-1} .$$

The model will be locally identified if and only if B is nonsingular (see, e.g., Rothenberg (1971)). Since D is of full rank and $4T > K$, a necessary and sufficient condition is that A have full column rank. From the Appendix to this chapter, it is seen that the partial derivatives of $Pr_t(i,j)$ with respect to the vector θ depend on the sign of $\Delta\alpha_w$; we must therefore check that matrix A has full column rank in both cases.

Case 1: $\Delta\alpha_w \geq 0$

Substituting into A_t the partial derivatives, using the notation $a_t^{i+}, b_t^{i+}, c_t^{i+}, d_t^{i+}, i = h,w,$ and $r_t^{j+}, j = 1,2,3,4,$ found in the Appendix, we perform the following matrix algebra:

(i) add rows (2+3+4) to row 1

(ii) add row 2 to row 4

(iii) add column 3 to column 4

(iv) multiply columns 1, 2 and 6 by -1

(v) Switch rows 3 and 4.

Rearranging columns and omitting row 1 since it is identically null, we have

$$\bar{A}_t = \begin{bmatrix} r_t^{1+} & 0 & 0 & a_t^{h+}x_h' & a_t^{w+} & a_t^{w+}x_w' \\ (r_t^{2+} + r_t^{3+}) & 0 & (b_t^{h+} + c_t^{h+}) & (b_t^{h+} + c_t^{h+})x_h' & c_t^{w+} & (b_t^{w+} + c_t^{w+})x_w' \\ r_t^{4+} & d_t^{h+} & 0 & -d_t^{h+}x_h' & 0 & d_t^{w+}x_w' \end{bmatrix}.$$

We now decompose the resulting matrix $\tilde{A}$ into a partitioned matrix

$$\bar{A} = \left[D_\rho \bar{X}_\rho \;\middle|\; D_h \bar{X}_h \;\middle|\; D_w \bar{X}_w \right],$$

where D_ρ, D_h, and D_w are each block diagonal matrices of order 3T, the t-th blocks being $D_{\rho t}$, D_{ht}, and D_{wt} respectively, as given above.

Case 2: $\Delta\alpha_w < 0$

Substitute into A_t the partial derivatives found in the Appendix, again using $a_t^{i-}, b_t^{i-}, c_t^{i-}, d_t^{i-}, i = h, w$, and $r_t^{j-}, j = 1,2,3,4$. Now perform the following matrix algebra on matrix A:

(i) add rows (1+2+4) to row 3

(ii) add row 4 to row 2

(iii) add column 4 to column 3

(iv) multiply column 6 by −1

(v) switch rows 2 and 4.

Rearranging columns and omitting row 3 since it is identically null, we have

$$\bar{A}_t = \begin{bmatrix} r_t^{1-} & 0 & 0 & a_t^{h-}x_h^1 & 0 & a_t^{w-}x_w^1 \\ r_t^{2-} & 0 & b_t^{h-} & b_t^{h-}x_h^1 & b_t^{w-} & b_t^{w-}x_w^1 \\ (r_t^{3-} + r_t^{4-}) & c_t^{h-} & 0 & -(c_t^{h-} + d_t^{h-})x_h^1 & c_t^{w-} & (c_t^{w-} + d_t^{w-})x_w^1 \end{bmatrix}$$

which can be written as $\tilde{A}$.

Q.E.D.

As seen in the Appendix, the elements of the matrices D_ρ, D_h, and D_w are all nonzero. Moreover, these matrices are nonsingular in both cases since they are either triangular matrices or can be made triangular by suitable permutations of rows and columns. By examining matrix $\tilde{A}$ above, it is clear that if $\tilde{A}$ does not have full column rank, it will occur only extremely rarely for some specific values of the parameters as an artifact of certain explanatory variables. We have, although, the following ne'cessary condition for identification.

COROLLARY 1: If $\Delta\alpha_w \equiv \alpha_w^1 - \alpha_w^0 = 0$, the model is not identified.

PROOF:

When $\Delta\alpha_w = 0$, it is seen from above that $b_t^{h+} + c_t^{ht} = 0$. Therefore matrix D_{ht} is singular for all t which implies that matrix $\tilde{A}$ no longer has full column rank.

Q.E.D.

As a practical implication of the corollary for estimation, it must be the case that the initial values chosen for α_w^1 and α_w^0 not be the same. Otherwise, the information matrix will be nonsingular at the first iteration, and the optimization cannot be carried out.

5. CONCLUSION

In this chapter, we presented a new approach for formulating simultaneous equations models for qualitative endogenous variables. As we have seen, this approach integrates results from noncooperative game theory and discrete choice modeling. In contrast to the model proposed in the previous chapter, we assume a model where two individuals play a Stackelberg game in which each player maximizes his own utility. As in the model of the previous chapter, the current model is also made stochastic by adopting the random utility framework.

Just as the previous chapter proposed a generalization of simultaneous equations models with structural shift in which the discrete endogenous variables were generated as Nash equilibria of a game between two players, the model proposed in the current chapter generalizes the recursive models for discrete endogenous variables that have been proposed up to now in the literature. As we have seen in Section 3, the usual recursive model is nested in our game theoretic model. Although recursive models have been used in the formulation of many econometric problems in which sequential decision making is a distinct feature, these models implicitly assume that the leader is indifferent to the follower's action. If this is not the case, then the usual recursive models are misspecified since they ignore the optimizing

behavior of the leader who is taking into account the conditional action of the second agent when choosing his action. As such, the usual recursive model of a sequential decision making problem is inadequate in many problems. In contrast, our formulation in terms of a Stackelberg model allows for optimizing behavior on the part of both agents.

FOOTNOTES

1. Where an individual is indifferent, we arbitrarily assume that he or she will take action 1.

2. Let us note that the husband is fully informed about the utility function of the wife; that is, he not only knows the deterministic components in the utilities (3)-(4) given below, but also the random components. An interesting generalization, which will be pursued in future work, arises when the huband knows only the deterministic components, in which case one has a Stackelberg game under uncertainty (see also Vuong (1982)).

3. Let us note that we allow the utilities $\tilde{U}_h(1,Y_w)$ and $\tilde{U}_w(1,Y_h)$ to depend on Y_w and Y_h respectively. This contrasts with the formulation adopted in Chapter II.

4. If $a < c$, $I(a,b,c,d,\rho)$ is by convention the negative of the integral of the bivariate density over the range $[a,c] \times [d,b]$. A similar remark applies if $b < d$. If both $a < c$ and $b < d$, then $I(a,b,c,d,\rho)$ is by convention the integral of the bivariate density over $[a,c] \times [b,d]$.

APPENDIX

First partial derivatives of the Probabilities Pr(i,j): Let Φ be the univariate normal c.d.f. and let φ be the corresponding p.d.f. We then use the relations $\frac{\partial F(x,y,\rho)}{\partial x} = \varphi(x)\Phi(y^* - \rho x^*)$, $\frac{\partial F(x,y,\rho)}{\partial y} = \varphi(y)\Phi(x^* - \rho y^*)$, and $\frac{\partial F(x,y,\rho)}{\partial \rho} = f(x,y,\rho)$ where a quantity with a "*" means that quantity is divided by the square root of $(1 - \rho^2)$. Proofs of these relations are found in Appendix B to Chapter II. In addition, let $f(x,y,\rho)$ be the p.d.f. corresponding to the bivariate normal c.d.f. $F(x,y,\rho)$. Then from equations (9)-(13) found in Section 2 of this chapter, the first partial derivatives of the probabilities Pr(i,j) use the following:

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \gamma_h} = -\varphi(\Delta U_h)\Phi(-\Delta U_w^* + \rho\Delta U_h^*)x_h,$$

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \gamma_w} = -\varphi(\Delta U_w)\Phi(-\Delta U_h^* + \rho\Delta U_w^*)x_w,$$

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \Delta\alpha_w} = 0,$$

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \alpha_h^0} = 0,$$

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \alpha_h^1} = 0,$$

$$\frac{\partial F(-\Delta U_h,-\Delta U_w,\rho)}{\partial \rho} = f(-\Delta U_h,-\Delta U_w,\rho);$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \gamma_h} = \varphi(\Delta_h)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho\Delta U_h^*)x_h,$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \gamma_w} = -\varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* - \rho(\Delta U_w^* + \Delta\alpha_w^*))x_w,$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \Delta\alpha_w} = -\varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* - \rho(\Delta U_w^* + \Delta\alpha_w^*)),$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \alpha_h^0} = 0,$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \alpha_h^1} = 0,$$

$$\frac{\partial F(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho)}{\partial \rho} = -f(\Delta U_h,-\Delta U_w - \Delta\alpha_w,-\rho);$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \gamma_h} = -\varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\ \Phi(\Delta U_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))x_h,$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \gamma_w} = \varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho\Delta U_w^*)x_w,$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \Delta\alpha_w} = 0,$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \alpha_h^0} = \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\ \Phi(\Delta U_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \alpha_h^1} = -\varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\ \Phi(\Delta U_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho)}{\partial \rho} = -f(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\Delta U_w,-\rho);$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0,\Delta U_w + \Delta\alpha_w,\rho)}{\partial \gamma_h} = \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\ \Phi(\Delta U_w^* + \Delta\alpha_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))x_h,$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0,\Delta U_w + \Delta\alpha_w,\rho)}{\partial \gamma_w} = \varphi(\Delta U_w + \Delta\alpha_w)$$

$$\times\ \Phi(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*} - \rho(\Delta U_w^* + \Delta\alpha_w^*))x_w,$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0,\Delta U_w + \Delta\alpha_w,\rho)}{\partial \Delta\alpha_w} = \varphi(\Delta U_w + \Delta\alpha_w)$$

$$\times\, \Phi(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*} - \rho(\Delta U_w^* + \Delta\alpha_w^*)),$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho)}{\partial \alpha_h^0} = -\varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\, \Phi(\Delta U_w^* + \Delta\alpha_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho)}{\partial \alpha_h^1} = \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)$$

$$\times\, \Phi(\Delta U_w^* + \Delta\alpha_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho)}{\partial \rho} = f(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho);$$

and

$$\begin{aligned}
\frac{\partial I_+^A}{\partial \gamma_h} = [&-\varphi(\Delta U_h + \alpha_h^1)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*})) \\
&+ \varphi(\Delta U_h + \alpha_h^1)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*})) \\
&+ \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})) \\
&- \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))]x_h,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial I_+^A}{\partial \gamma_w} = [&-\varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \rho\Delta U_w^*) \\
&+ \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \rho(\Delta U_w^* + \Delta\alpha_w^*)) \\
&+ \varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho\Delta U_w^*) \\
&- \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*))]x_w,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial I_+^A}{\partial \Delta\alpha_w} = &\ \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \rho(\Delta U_w^* + \Delta\alpha_w^*)) \\
&- \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*)),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial I_+^A}{\partial \alpha_h^0} = &-\varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})) \\
&+ \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{*0})),
\end{aligned}$$

$$\frac{\partial I_+^A}{\partial a_h^1} = -\varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + a_h^{1*}))$$

$$+ \varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho(\Delta U_h^* + a_h^{1*}))$$

$$+ \varphi(\Delta U_h + a_h^1 - a_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + a_h^{1*} - a_h^{0*}))$$

$$- \varphi(\Delta U_h + a_h^1 - a_h^0)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho(\Delta U_h^* + a_h^{1*} - a_h^{0*})),$$

$$\frac{\partial I_+^A}{\partial \rho} = f(-\Delta U_h - a_h^1, -\Delta U_w, \rho) - f(-\Delta U_h - a_h^1, -\Delta U_w - \Delta a_w, \rho)$$

$$- f(-\Delta U_h - a_h^1 + a_h^0, -\Delta U_w, \rho) + f(-\Delta U_h - a_h^1 + a_h^0, -\Delta U_w - \Delta a_w, \rho);$$

$$\frac{\partial I_+^B}{\partial \gamma_h} = [-\varphi(\Delta U_h)\Phi(-\Delta U_w^* + \rho\Delta U_h^*) + \varphi(\Delta U_h)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho\Delta U_h^*)$$

$$+ \varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + a_h^{1*}))$$

$$- \varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho(\Delta U_h^* + a_h^{1*}))]x_n,$$

$$\frac{\partial I_+^B}{\partial \gamma_w} = [-\varphi(\Delta U_w)\Phi(-\Delta U_h^* + \rho\Delta U_w^*) + \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + \rho(\Delta U_w^* + \Delta a_w^*))$$

$$+ \varphi(\Delta U_w)\Phi(-\Delta U_h^* - a_h^{1*} + \rho\Delta U_w^*)$$

$$- \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* - a_h^{1*} + \rho(\Delta U_w^* + \Delta a_w^*))]x_w,$$

$$\frac{\partial I_+^B}{\partial \Delta a_w} = \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + \rho(\Delta U_w^* + \Delta a_w^*))$$

$$- \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* - a_h^{1*} + \rho(\Delta U_w^* + \Delta a_w^*)),$$

$$\frac{\partial I_+^B}{\partial a_h^0} = 0,$$

$$\frac{\partial I_+^B}{\partial a_h^1} = \varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + a_h^{1*}))$$

$$- \varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho(\Delta U_h^* + a_h^{1*})),$$

$$\frac{\partial I_+^B}{\partial \rho} = f(-\Delta U_h, -\Delta U_w, \rho) - f(-\Delta U_h, -\Delta U_w - \Delta a_w, \rho) - f(-\Delta U_h - a_h^1, -\Delta U_w, \rho)$$

$$+ f(-\Delta U_h - a_h^1, -\Delta U_w - \Delta a_w, \rho);$$

$$\frac{\partial I_{-}^{A}}{\partial \gamma_h} = [-\varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* - \alpha_h^{*0}))$$

$$+ \varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$

$$+ \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))$$

$$- \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))]x_h,$$

$$\frac{\partial I_{-}^{A}}{\partial \gamma_w} = [-\varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*))$$

$$+ \varphi(\Delta U_w)\Phi(-\Delta U_h^* + \alpha_h^{0*} + \rho\Delta U_w^*)$$

$$+ \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*))$$

$$- \varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho\Delta U_w^*))]x_w,$$

$$\frac{\partial I_{-}^{A}}{\partial \Delta\alpha_w} = -\varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*))$$

$$+ \varphi(\Delta U_w + \Delta\alpha_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho(\Delta U_w^* + \Delta\alpha_w^*)),$$

$$\frac{\partial I_{-}^{A}}{\partial \alpha_h^0} = \varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$

$$- \varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$

$$- \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))$$

$$+ \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial I_{-}^{A}}{\partial \alpha_h^1} = \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))$$

$$- \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*})),$$

$$\frac{\partial I_{-}^{A}}{\partial \rho} = f(-\Delta U_h + \alpha_h^0, -\Delta U_w - \Delta\alpha_w, \rho) - f(-\Delta U_h + \alpha_h^0, -\Delta U_w, \rho)$$

$$- f(-\Delta U_h - \alpha_h^1 + \alpha_h^0, -\Delta U_w - \Delta\alpha_w, \rho) + f(-\Delta U_h - \alpha_h^1 + \alpha_h^0, -\Delta U_w, \rho);$$

$$\frac{\partial I_{-}^{B}}{\partial \gamma_h} = [-\varphi(\Delta U_h)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho\Delta U_h^*) + \varphi(\Delta U_h)\Phi(-\Delta U_w^* + \rho\Delta U_h^*)$$

$$+ \varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$

$$- \varphi(\Delta U_h - a_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* - a_h^{0*}))]x_n,$$

$$\frac{\partial I_-^B}{\partial \gamma_w} = [-\varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + \rho(\Delta U_w^* + \Delta a_w^*)) + \varphi(\Delta U_w)\Phi(-\Delta U_h^* + \rho\Delta U_w^*)$$

$$+ \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + a_h^{0*} + \rho(\Delta U_w^* + \Delta a_w^*))$$

$$- \varphi(\Delta U_w)\Phi(-\Delta U_h^* + a_h^{0*} + \rho\Delta U_w^*)]x_w,$$

$$\frac{\partial I_-^B}{\partial \Delta a_w} = -\varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + \rho(\Delta U_w^* + \Delta a_w^*))$$

$$+ \varphi(\Delta U_w + \Delta a_w)\Phi(-\Delta U_h^* + a_h^{0*} + \rho(\Delta U_w^* + \Delta a_w^*))$$

$$\frac{\partial I_-^B}{\partial a_h^0} = -\varphi(\Delta U_h - a_h^0)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho(\Delta U_h^* - a_h^{0*}))$$

$$+ \varphi(\Delta U_h - a_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* - a_h^{0*})),$$

$$\frac{\partial I_-^B}{\partial a_h^1} = 0,$$

$$\frac{\partial I_-^B}{\partial \rho} = f(-\Delta U_h, -\Delta U_w - \Delta a_w, \rho) - f(-\Delta U_h, -\Delta U_w, \rho)$$

$$- f(-\Delta U_h + a_h^0, -\Delta U_w - \Delta a_w, \rho) + f(-\Delta U_h + a_h^0, -\Delta U_w, \rho)$$

Elements of the matrices $D_{\rho t}$, D_{ht}, and D_{wt}.

For simplicity, we drop the subscript t in the following expressions.

$$r^{1+} \equiv f(\Delta U_h, \Delta U_w + \Delta a_w, \rho)$$

$$r^{2+} \equiv -f(\Delta U_h + a_h^1, \Delta U_w, \rho)$$

$$r^{3+} \equiv f(\Delta U_h + a_h^1, \Delta U_w + \Delta a_w, \rho)$$

$$r^{4+} \equiv f(\Delta u_h + a_h^1 - a_h^0, \Delta U_w, \rho)$$

$$a^{h+} \equiv \varphi(\Delta U_h)\Phi(-\Delta U_w^* - \Delta a_w^* + \rho\Delta U_h^*)$$

$$b^{h+} \equiv -\varphi(\Delta U_h + a_h^1)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* + a_h^{1*}))$$

$$c^{h+} \equiv \varphi(\Delta U_h + \alpha_h^1)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* + \alpha_h^{1*}))$$
$$d^{h+} \equiv \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(\Delta U_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))$$
$$a^{w+} \equiv \varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* - \rho(\Delta U_w^* + \Delta\alpha_w^*))$$
$$b^{w+} \equiv \varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \rho\Delta U_w^*)$$
$$c^{w+} \equiv \varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* + \alpha_h^{1*} - \rho(\Delta U_w^* + \Delta\alpha_w^*))$$
$$d^{w+} \equiv -\varphi(\Delta U_w)\Phi(-\Delta U_h^* - \alpha_h^{1*} + \alpha_h^{0*} + \rho\Delta U_w^*)$$
$$r^{1-} \equiv f(\Delta U_h, \Delta U_w, \rho)$$
$$r^{2-} \equiv f(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho)$$
$$r^{3-} \equiv -f(\Delta U_h - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho)$$
$$r^{4-} \equiv f(\Delta U_h - \alpha_h^0, \Delta U_w, \rho)$$
$$a^{h-} \equiv -\varphi(\Delta U_h)\Phi(-\Delta U_w^* + \rho\Delta U_h^*)$$
$$b^{h-} \equiv \varphi(\Delta U_h + \alpha_h^1 - \alpha_h^0)\Phi(\Delta U_w^* + \Delta\alpha_w^* - \rho(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*}))$$
$$c^{h-} \equiv -\varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* - \Delta\alpha_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$
$$d^{h-} \equiv \varphi(\Delta U_h - \alpha_h^0)\Phi(-\Delta U_w^* + \rho(\Delta U_h^* - \alpha_h^{0*}))$$
$$a^{w-} \equiv \varphi(\Delta U_w)\Phi(-\Delta U_h^* + \rho\Delta U_w^*)$$
$$b^{w-} \equiv \varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* + \alpha_h^{1*} - \alpha_h^{0*} - \rho(\Delta U_w^* + \Delta\alpha_w^*))$$
$$c^{w-} \equiv -\varphi(\Delta U_w + \Delta\alpha_w)\Phi(\Delta U_h^* - \alpha_h^{0*} - \rho(\Delta U_w^* + \Delta\alpha_w^*))$$
$$d^{w-} \equiv \varphi(\Delta U_w)\Phi(-\Delta U_h^* + \alpha_h^{0*} + \rho\Delta U_w^*).$$

CHAPTER IV: AN EMPIRICAL STUDY OF THE GAME THEORETIC MODELS

1. INTRODUCTION

The purpose of the current chapter is to present an econometric study of the Nash and Stackelberg equilibrium models proposed in the previous two chapters. The problem we will examine concerns the joint decision of a married couple whether or not to participate in the labor force. We feel that an examination of this problem will yield two useful results. First, it will demonstrate that the game theoretic models we have proposed are in fact empirically tractable. Second, we feel that the proposed study will make a contribution to the literature on labor force participation because we explicitly model the behavior of a married couple in a game theoretic framework, while previous empirical work has either taken the husband's decision whether or not to work as exogenous (see, e.g., Heckman (1974), Heckman and McCurdy (1980)), or specified the labor supply of a husband and wife from the outcome of a joint utility function (see, e.g., Ashenfelter and Heckman (1974), and Gronau (1973)). Although Brown and Manser (1978), Manser and Brown (1980) and McElroy and Horney (1981) have done related work using a cooperative game solution, none have provided an empirical application.

By now it should be clear after reading Chapters II and III that the structure of the Nash and Stackelberg models contain many similarities. These common elements arise for essentially four reasons. First, the dimensionality of both models is the same; in both models, attention is restricted to the case of two qualitative endogenous variables where each variable takes on only two outcomes. Second, both models follow the classical tradition in economics, postulating that each member of the married couple maximizes his or her own utility function. The decisions made by the husband and wife need not be independent, however, since the utility derived by each individual naturally depends on the action taken by the other. In this sense, both the Nash and Stackelberg models allow as arguments in each utility function the action taken by both members of the married couple. Third, since the ultimate goal of the previous two chapters was to specify an econometric model of the two game theoretic notions, it was necessary to introduce a stochastic structure into the two models. This was done by treating the utilities as random, decomposing the utility of each of the four outcomes into a deterministic component and an additive random component; the deterministic components were in turn decomposed into continuous explanatory variables and shift parameters. Finally, since the observed dichotomous variables were generated as equilibrium outcomes of game theoretic notions, it was necessary in both models to derive the reaction functions of each individual. Given the probabilistic structure, each reaction function occured if certain relations were satisfied between the disturbance terms, the shift parameters, and differences in the continuous explanatory variables.

Given the similarities in the structure of both the Nash and Stackelberg models, it should not be surprising that the specification of the three empirical models will in turn yield many similarities. As to economize on space, these similarities are discussed in the following section. Section 3 will discuss peculiar features of the structure of the Nash model along with the empirical results of the problem under study. Sections 4 will discuss the Stackelberg model for the case in which the husband plays the role of the leader while his wife plays the role of the follower. Empirical results will also be presented. Section 5 discusses the third empirical example where the roles of the husband and wife are reversed. Section 6 provides a brief conclusion. A description of the data set used in the three empirical studies is included as the Appendix to this chapter. As such, it discusses the source and description of the original data set, the selection criteria we used in choosing the appropriate set of observations, and the means and variances of the explanatory variables.

2. ON THE COMMON STRUCTURE OF THE MODELS

In each of the three models, the following four equations will be used to describe the joint behavior of a representative married couple:

$$W_h^r = Z_h'\gamma_h^r + d_h^r Y_w + \eta_h^r \tag{1}$$
$$W_w^r = Z_w'\gamma_w^r + d_w^r Y_h + \eta_w^r \tag{2}$$
$$W_h^m = X_h'\gamma_h^m + d_h^m Y_w + \eta_h^m \tag{3}$$
$$W_w^m = X_w'\gamma_w^m + d_w^m Y_h + \eta_w^m, \tag{4}$$

where

$$Y_h = \begin{cases} 1 \text{ if the husband works,} \\ 0 \text{ otherwise,} \end{cases}$$

$$Y_w = \begin{cases} 1 \text{ if the wife works,} \\ 0 \text{ otherwise.} \end{cases}$$

Equations (1) and (2) describe the reservation wages, or equivalently, the shadow price of time for the husband and wife, respectively. Note that the wife's decision of whether or not to work, given by the dichotomous variable Y_w, affects the husband's reservation wage in (1). Analogously, the husband's decision of whether or not to work, given by Y_h, affects the wife's reservation wage in (2). Equations (3) and (4) describe the market wages for the husband and the wife, respectively. Note that in the general specification of the three models, we also allow for the possibility that one of the determinants of the husband's market wage is whether or not he has a working wife; we make a similar allowance for the wife.

Now let the husband's (wife's) reservation wage play the role of the payoff he (she) derives from not working, given below by equations (5) and (6) respectively. Similarly, let the husband's (wife's) market wage play the role of the payoff he (she) derives from working, given by equations (7) and (8) respectively. We then have

$$U_h(0,Y_w) = Z_h'\gamma_h^r + d_h^r Y_w + \eta_h^r \tag{5}$$

$$U_w(0,Y_h) = Z_w'\gamma_w^r + d_w^r Y_h + \eta_w^r \tag{6}$$

$$U_h(1,Y_w) = X_h'\gamma_h^m + d_h^m Y_w + \eta_h^m \tag{7}$$

$$U_w(1,Y_h) = X_w'\gamma_w^m + d_w^m Y_h + \eta_w^m. \tag{8}$$

Since we are describing the empirical structure of the three models only in general terms in this section, it should be noted that the structural assumptions of both the Nash and Stackelberg models will place zero restrictions on certain of the shift parameters d_i^j, $i = h,w$, $j = m,r$.

In each of the three models, we must now specify the set of explanatory variables used to estimate the market wage equations and the reservation wage equations for the husband and wife.[1] Market wages for the husband and wife are specified in (9) and (10) respectively. Reservation wages for the husband and wife are specified in (11) and (12) respectively.

$$U_h(1,Y_w) = W_h^m = \gamma_h^0 + \gamma_h^1 \mathrm{AGEH} + \gamma_h^2 \mathrm{EDUCH} + \gamma_h^3 \mathrm{UNEM} + \gamma_h^4 \mathrm{RACE} \tag{9}$$

$$+ d_h^m Y_w + \eta_h^m$$

$$U_w(1,Y_h) = W_w^m = \gamma_w^0 + \gamma_w^1 \mathrm{AGEW} + \gamma_w^2 \mathrm{AGEW{*}{*}2} + \gamma_w^3 \mathrm{EDUCW} \tag{10}$$

$$+ \gamma_w^4 \mathrm{UNEM} + \gamma_w^5 \mathrm{RACE} + d_w^m Y_h + \eta_w^m$$

$$\tilde{U}_h(0,Y_w) = W_h^r = \tilde{\gamma}_h^0 + \tilde{\gamma}_h^1 AGEH + \tilde{\gamma}_h^2 EDUCH + \tilde{\gamma}_h^3 UNEM + \tilde{\gamma}_h^4 RACE + \tilde{\gamma}_h^5 ASSETS$$

$$+ \tilde{\gamma}_h^6 KIDS{<}13 + \tilde{\gamma}_h^7 KIDS{>}14 + d_h^r Y_w + \eta_h^r \qquad (11)$$

$$\tilde{U}_w(0,Y_h) = W_w^r = \tilde{\gamma}_w^0 + \tilde{\gamma}_w^1 AGEW + \tilde{\gamma}_w^2 AGEW{**}2 + \tilde{\gamma}_w^3 EDUCW + \tilde{\gamma}_w^4 UNEM$$

$$+ \tilde{\gamma}_w^5 RACE + \tilde{\gamma}_w^6 ASSETS + \tilde{\gamma}_w^7 KIDS1\text{-}2 + \tilde{\gamma}_w^8 KIDS3\text{-}5$$

$$+ \tilde{\gamma}_w^9 KIDS6\text{-}13 + \tilde{\gamma}_w^{10} KIDS{>}14 - d_w^r Y_h + \eta_w^r, \qquad (12)$$

where

AGEH	Age of husband
AGEW	Age of wife
AGEW**2	Squared age of wife
EDUCH	Number of years of formal schooling of husband
EDUCW	Number of years of formal schooling of wife
UNEM	Local unemployment rate
RACE	Dummy variable indicating race of married couple; 1 = Black or Hispanic, Ø otherwise[2]
ASSETS	Family's annual income other than from wages or salaries [3]
KIDS1-2	Number of children in family unit ages 1 and 2.
KIDS3-5	Number of children between ages 3 and 5.
KIDS6-13	Number of children between 6 and 13.
KIDS<13	Number of children in family unit 13 years or younger
KIDS>14	Number of children in family unit 14 years or older

The plus and minus signs under the explanatory variables in equations

(9)-(12) indicate the expected impact upon estimation. Note that not only does the wife's decision of whether or not to work effect the husband's market wage, it also affects the husband's reservation wage. Analogously, the husband's decision to work or not affects the wife's market wage and also her reservation wage.

The data used in this study on married couples is from the 1982 wave of the University of Michigan Survey Research Center's Panel Study on Income Dynamics, 1968-1982. The data was restricted to 2012 records for married couples living in the U.S., where both the husband and the wife were able-bodied, neither older than 64 years of age with no nonrelative living in the family unit. A more detailed description of the Panel Study on the selection criteria we used is found in the Appendix to this chapter. Let us now briefly discuss the estimation technique used as it will apply to all three empirical problems.

Although many of the nonlinear optimization methods developed over the years could be used to estimate the Nash and Stackelberg models, we choose to use a version of the iterative procedure suggested by Berndt, Hall, Hall, and Hausman (1974). A major advantage of the BHHH algorithm is that it makes use of the information matrix equality when the method of maximum likelihood is applied to correctly specified models. As such, only the first partial derivatives of the expressions for the probabilities need to be derived.

For completeness, a brief digression on the derivation of the BHHH algorithm, a variation on the method of scoring, is in order.

The method of scoring is as follows. Given a sample of n i.i.d. observations $(y_1,\ldots,y_n)$ from a density function $f(y,\theta)$, the log-likelihood function is given as

$$\log L(\theta) = \sum_{i=1}^{n} \log f(y_i,\theta).$$

Let θ_0 be an initial or trial value of the parameter vector θ. Then by taking a Taylor expansion of $\frac{\partial \log L}{\partial \theta}$ around θ_0 we get $\frac{\partial \log L}{\partial \theta} \simeq \frac{\partial \log L}{\partial \theta}\Big|_{\theta_0} + \frac{\partial^2 \log L}{\partial\theta\partial\theta'}\Big|_{\theta_0} \cdot (\theta - \theta_0)$. But since

$$n^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\Big|_{\theta_0} = n^{-1}\sum_{i=1}^{n}\frac{\partial^2 \log f(y_i,\theta)}{\partial\theta\partial\theta'}\Big|_{\theta_0}$$

$$\xrightarrow{a.s.} -n^{-1} I(\theta_0) \equiv E\left[n^{-1}\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]_{\theta_0},$$

we have $\frac{\partial \log L}{\partial \theta} \simeq S(\theta_0) - (I(\theta_0) \cdot (\theta - \theta_0)$ where $S(\theta_0) = \frac{\partial \log L}{\partial \theta}\Big|_{\theta_0}$. But at a maximum we must have $\frac{\partial \log L}{\partial \theta}\Big|_{\theta} = 0$ or $\theta = \theta_0 + I(\theta_0)^{-1}S(\theta_0)$. The method of scoring now works as follows: For any initial θ_0 construct $S(\theta_0)$ and $I(\theta_0)$. The new estimate θ_1 is given by $\theta_1 = \theta_0 + I(\theta_0)^{-1}S(\theta_0)$. Since the scoring method involves second partial derivatives, we exploit the fact that under correct model specification the information matrix may also be written as

$$n^{-1} \cdot I(\theta_0) = n^{-2} E\left[\frac{\partial \log L}{\partial \theta} \cdot \frac{\partial \log L}{\partial \theta'}\right].$$ Since

$$n^{-1}Q(\theta) \equiv n^{-1}\sum_{i=1}^{n}\frac{\partial \log f(y_i,\theta)}{\partial\theta} \cdot \frac{\partial \log f(y_i,\theta)}{\partial\theta'}$$

$$\xrightarrow{a.s.} E\, n^{-2}\left[\frac{\partial \log L}{\partial \theta} \cdot \frac{\partial \log L}{\partial \theta'}\right],$$

Berndt, et. al. suggest using $\theta_1 = \theta_0 + Q(\theta_0)^{-1}S(\theta_0)$ as an iterative procedure.

3. THE NASH MODEL

A. SPECIAL FEATURES OF THE NASH SPECIFICATION

Recall from Assumption A3 of Chapter II that only two shift parameters, β_h and β_w, appear in the equations specifying the joint determination of the decision of the husband and wife whether or not to participate in the labor force. As a result, equations (1)-(4) must be suitably modified. We thus assume that the wife's decision of whether or not to work, given by the dichotomous variable Y_w, affects the husband's reservation wage but not his market wage. Similarly, we assume that the husband's decision of whether or not to work, given by Y_h, affects the wife's reservation wage but not her market wage. We then have:

$$W_h^r = Z_h'\gamma_h^r - \beta_h Y_w + \eta_h^r \tag{13}$$
$$W_w^r = Z_w'\gamma_w^r - \beta_w Y_h + \eta_w^r \tag{14}$$
$$W_h^m = X_h'\gamma_h^m + \eta_h^m \tag{15}$$
$$W_w^m = X_w'\gamma_w^m + \eta_w^m. \tag{16}$$

Letting the husband's (wife's) reservation wage play the role of the payoff he (she) derives from not working, and letting the husband's (wife's) market wage play the role he (she) derives from working, we have the following four equations

$$U_h(0,Y_w) = Z_h'\gamma_h^r - \beta_h Y_w + \eta_h^r \tag{17}$$

$$\tilde{U}_w(0,Y_h) = Z_w'\gamma_w^r - \beta_w Y_h + \eta_w^r \tag{18}$$

$$\tilde{U}_h(1,Y_w) = X_h'\gamma_h^m + \eta_h^m \tag{19}$$

$$\tilde{U}_w(1,Y_h) = X_w'\gamma_w^m + \eta_w^m. \tag{20}$$

Notice from Chapter II that Assumption A2 on the error terms is naturally satisfied. Moreover, we have:

$$U_h(1,1) - U_h(0,1) = \beta_h + U^h(1,0) - U^h(0,0) = X_h'\gamma_h^m - Z_h'\gamma_h^r + \beta_h \tag{21}$$

$$U_w(1,1) - U_w(0,1) = \beta_w + U^w(1,0) - U^w(0,0) = X_w'\gamma_w^m - Z_w'\gamma_w^r + \beta_w. \tag{22}$$

Thus Assumption A3 of Chapter II is also satisfied. In addition, note that in specifying the husband's reservation wage and market wage equations, given by (17) and (19) respectively, it may be the case that certain explanatory variables appear in both equations, implying that the associated coefficient in (21) will be measuring the difference between market and reservation wages. A similar comment holds for the wife.

Applying (21) and (22) to the general specification given in the previous section, we now have

$$\begin{aligned} U_h(1,1) - U_h(0,1) &= \beta_h + U_h(1,0) - U_h(0,0) \\ &= (\gamma_h^0 - \tilde{\gamma}_h^0) + (\gamma_h^1 - \tilde{\gamma}_h^1)\text{AGEH} + (\gamma_h^2 - \tilde{\gamma}_h^2)\text{EDUCH} + (\gamma_h^3 - \tilde{\gamma}_h^3)\text{UNEM} \\ &\quad + (\gamma_h^4 - \tilde{\gamma}_h^4)\text{RACE} - \tilde{\gamma}_h^5\text{ASSETS} - \tilde{\gamma}_h^6\text{KIDS13} - \tilde{\gamma}_H^7\text{KIDS14} + \beta_h \end{aligned} \tag{23}$$

and

$$\begin{aligned} U_w(1,1) - U_w(0,1) &= \beta_w + U_w(1,0) - U_w(0,0) \\ &= (\gamma_w^0 - \tilde{\gamma}_w^0) + (\gamma_w^1 - \tilde{\gamma}_w^1)\text{AGEW} + (\gamma_w^2 - \tilde{\gamma}_w^2)\text{AGEW**2} + (\gamma_w^3 - \tilde{\gamma}_w^3)\text{EDUCW} \\ &\quad + (\gamma_h^4 - \tilde{\gamma}_h^4)\text{UNEM} + (\gamma_w^5 - \tilde{\gamma}_w^5)\text{RACE} - \tilde{\gamma}_w^6\text{ASSETS} - \tilde{\gamma}_w^7\text{KIDS1-2} \end{aligned} \tag{24}$$

$$- \tilde{\gamma}_w^8 \text{KIDS3-5} - \tilde{\gamma}_w^9 \text{KIDS6-13} - \tilde{\gamma}_w^{10} \text{KIDS>14} + \beta_w.$$

B. EMPIRICAL RESULTS FOR THE NASH MODEL

Assuming that the parameters are identified, maximization is straightforward although the log-likelihood takes four different functional forms according to the signs of β_h and β_w. To complicate matters, we do not have a set of initial consistent estimates of the parameters (γ_h, β_h, γ_w, β_w, ρ) nor do we believe that the log-likelihood function is globally concave. Although this situation is common in problems of nonlinear estimation, one can never be totally certain that he has found the global maximum. The best that one can do is experiment with various sets of initial parameter values until becoming reasonably certain that the global maxmum has been reached. To this end, we provided various initial values for (γ_w, β_h, γ_w, β_w) with a grid search over possible values of ρ. Having no strong priors over ρ, we estimated all values from -.9 to +.9 in increments of 1. We then iterated until convergence, convergence being reached when, on average, each element of the score vector was no greater than .01. Different trials of the parameter vector (γ_h, β_h, γ_w, β_w) included (i) the zero vector; (ii) our a priori best guess of the parameter values, making use of equations (23) and (24); (iii) the previous trial with the signs of the coefficients reversed; (iv) dividing be ten the final estimates from each of the trials (i), (ii) and (iii). Upon estimation, it was found that the parameter estimates and log-likelihood values were almost the same for each of the give trials listed above; we took this as sufficient evidence that we had found the global maximum of the log-likelihood function.

As can be seen from Table 1, the value of ρ that maximizes the log-likelihood function appears to lie in the interval (-1.0, -.8). Therefore we tried all values within this interval in increments of .01. The maximizing value of ρ is -.91, which also appears in Table 1. It may appear surprising that the maximizing value of ρ is close to minus one. One possible explanation for this result may be due to the fact that the contingency table is very unbalanced. As we see from Figure 1 of the Appendix to this chapter, there are very few observations for which the husband does not work; for only 77 of the 2020 observations on married couples does the husband choose not to work. If we now look at Figure 3c of Chapter II (since $\hat{\beta}_h < 0$ and $\hat{\beta}_w > 0$), we see that an ellipsoid in $(\varepsilon_h, \varepsilon_w)$ space will be centered, most likely, in the (1,1) region since this corresponds to the largest element in the contingency table; for 1204 observations out of 2020, both the husband and wife work. Moreover, an ellipsoid whose major acis has slope near minus one and centered in the (1,1) region will cover an area associated with the largest number of observations. Since the method of estimation employed is maximum likelihood, this explanation seems quite plausible. In addition, a likelihood ratio test of $\rho = 0$ vs. $\rho = -.91$ can be easily constructed from Table 1. Since -2(log-likelihood value for $\rho = 0$ - log-likelihood value for $\rho = -.91) \sim \chi^2_1$, we can reject the hypothesis that $\rho = 0$ at the 95 percent level.

Table 3 lists the estimated coefficients and the associated

TABLE 1

Correlation Coefficient (ρ)	Log-likelihood Value
-.99	-1514.95
-.91	-1514.57
-.9	-1514.59
-.8	-1515.05
-.7	-1515.72
-.6	-1516.41
-.5	-1517.07
-.4	-1516.56
-.3	-1516.93
-.2	-1517.35
-.1	-1517.79
0.	-1518.23
.1	-1518.67
.2	-1519.10
.3	-1519.51
.4	-1519.90
.5	-1520.27
.6	-1520.65
.7	-1521.01
.8	-1521.58
.9	-1521.92

TABLE 2
$\rho = 0$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	β_h	0.384	1.00			
				β_w	-0.681	-1.29
ONSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	0.835	1.22	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	-0.161	-0.24
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.014	1.89*			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	-0.084	-3.46**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.001	-4.24**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.058	1.78*			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.039	3.18**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.043	-2.41**	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.015	-1.59
CE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.262	-1.97**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.399	5.31**
SSET	$-\tilde{\gamma}_h^5$	0.381	1.27	$-\tilde{\gamma}_w^6$	-0.012	-2.06**
DS1-2				$-\tilde{\gamma}_w^7$	-0.700	-11.20**
DS3-5				$-\tilde{\gamma}_w^8$	-0.451	-7.37**
DS6-13				$-\tilde{\gamma}_w^9$	-0.200	-5.21**
DS < 13	$-\tilde{\gamma}_h^6$	-0.046	-0.72			
DS > 14	$-\tilde{\gamma}_h^7$	0.088	0.79	$-\tilde{\gamma}_w^{10}$	-0.136	-2.74**

g-likelihood value = -1518.23

* significant at the 10% level
** significant at the 5% level

TABLE 3
$\rho = -.91$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	β_h	-.972	-2.44**	β_w	2.63	11.80**
ISTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	1.04	1.55	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	-3.02	-6.16**
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.009	1.10			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.067	2.84**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.001	-3.73**
JCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.142	4.71**			
JCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.032	1.85*
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.055	-2.67**	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.001	-0.08
CE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.128	-0.89	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.442	5.87**
SET	$-\tilde{\gamma}_h^5$	0.427	1.48	$-\tilde{\gamma}_w^6$	-0.012	-2.22**
DS1-2				$-\tilde{\gamma}_w^7$	-0.650	-11.00**
DS3-5				$-\tilde{\gamma}_w^8$	-0.426	-7.30**
DS6-13				$-\tilde{\gamma}_w^9$	-0.153	-4.01**
DS < 13	$-\tilde{\gamma}_h^6$	-0.074	-1.26			
DS > 14	$-\tilde{\gamma}_h^7$	0.211	1.68*	$-\tilde{\gamma}_w^{10}$	-0.148	-3.00**

g-likelihood value = -1514.57

* significant at the 10% level
** significant at the 5% level

t-statistics for the case $\rho = -.91$. Several points deserve mention. First, if we compare Table 2, the estimated coefficients and t-statistics for the case of $\rho = 0$, with Table 3 we see that the signs of the estimated coefficients are quite robust with respect to ρ. Second, we see that both β_h and β_w are significantly different from zero, providing evidence that the husband's decision whether or not to work depends on the wife's decision and vice versa. From (17) we see that the negative estimated coefficient of β_h from Table 3 implies that if the wife works, the husband's reservation wage increases as expected. A priori, we would also expect the estimate of β_w to be negative also; we offer no explanation for the disconcerting result that β_w is positive.

Finally, and most importantly, we can provide a test of whether or not the logical consistency condition, $\beta_h \cdot \beta_w = 0$, holds for the empirical problem presently under study. Since the results presented in Table 3 provide unrestricted estimates of the two shift parameters β_h and β_w, we can easily perform a Wald test of the logical consistency condition. As a reminder, let us define the Wald statistic. If θ is a k vector of parameters, $r(\theta)$ a vector of g restrictions ($g < k$) imposed by the null hypothesis on θ, and R the g x k matrix of partial derivatives $\partial r(\theta)/\partial\theta$, the Wald statistic is given by

$$W = r(\hat{\theta})'[RI(\hat{\theta})^{-1}R']^{-1}r(\hat{\theta}).$$

Asymptotically, W is distributed $\chi^2(g)$. For the case at hand,

$\theta = (\beta_h, \beta_w, \gamma_h, \gamma_w, \rho)$ and $r(\theta) = \beta_h \cdot \beta_w = 0$ so that $g = 1$. Since the value of the Wald statistic in this case is 4.63, we can reject the hypothesis that $\beta_h \cdot \beta_w = 0$ at the 95 percent level. If we recall from Chapter II that the usual simultaneous equations model for dummy endogenous variables developed by Heckman (1978) is not well-defined when $\beta_h \cdot \beta_w \neq 0$, but that our formulation in which the dummy endogenous variables are generated as outcomes of a Nash game always yields a well-defined statistical model, we are led to conclude that the Heckman formulation is inappropriate for modeling the empirical problem at hand.

Looking again at Table 3, we see that most of the coefficients explaining the wife's decision whether or not to work are in agreement with our expectations and are highly significant. For example, family income from sources other than wages and salaries (ASSET) has the expected effect of increasing the wage at which a wife is willing to accept work outside the home. Concerning children, one would certainly expect that mothers would be least likely to leave the home when children are very young and be more inclined to seek outside employment as children become older and more self-sufficient. Indeed, this is what we see from Table 3. Children between the ages of one and two (KIDS1-2) raise the mother's reservation wage more than do children between three and five (KIDS3-5); her reservation wage is higher for children between three and five than for children six to thirteen (KIDS6-13); finally, the mother is more likely to stay home when her children are between six and thirteen than when they are fourteen years or older (KIDS14).[4] The estimated positive coeff ient on the female race dummy (RACE) seems to suggest that women of racial minorities, on average, can command a higher market wage than they are willing to accept to enter the labor market; that is, minority women are on average worth more in the marketplace than they think they are worth. While one may interpret this result as saying that racial discrimination in the marketplace is not as widespread as minority women are led to believe, the discrepancy between reservation wage and market wage may be a reflection of past discrimination. The coefficient on the wives' education (EDUCW) is also consistent with our prior expectation; although an increase in education should increase the wife's market wage, it should also increase her reservation wage. Turning finally to the effect of age on a wife's decision whether or not to work, a life-cycle model of employment would suggest that women are more likely to work during middle age than either early or late in their life times. That is, the probability of working as a function of age first increases, then reaches a maximum, and then decreases. As can be seen from Table 3, the combined effects of a linear term on age (AGEW) and a quadratic term (AGEW**2) does indeed impart the expected shape.

Turning next to the variables used to explain the husband's decision of whether or not to work, we see that while a number of the coefficients are insignificant, some of the variables to which we attached strong priors appear to be significant. For example, the coefficients attached to both the husband's age (AGEH) and the local unemployment rate (UNEM) measure the difference between the husband's market wage and his reservation wage; if the husband is behaving logically, both of these estimated coefficients should be close to zero, which they are. The positive estimated coefficient on the husband's level of education (EDUCH) is surprising; one would a priori expect that an increase in education would raise equally the husband's market wage and reservation wage. One possible explanation for this surprising result is that additional education is viewed by many as a consumption good rather than an investment good.[5]

4. THE STACKELBERG MODEL: HUSBAND LEADER

A. SPECIAL FEATURES OF THE STACKELBERG SPECIFICATION

Recall from the specification of the Stackelberg model given in Chapter III that we allow for the inclusion of four shift parameters. Thus, equations (1)-(4) need only be altered to conform to the notation used in Chapter III. We thus have:

$$W_h^r = Z_h'\gamma_h^r + \alpha_h^0 Y_w + \eta_h^0 \tag{25}$$

$$W_w^r = Z_w'\gamma_w^r + \alpha_w^0 Y_h + \eta_w^0 \tag{26}$$

$$W_h^m = X_h'\gamma_h^m + \alpha_h^1 Y_w + \eta_h^1 \tag{27}$$

$$W_w^m = X_w'\gamma_w^m + \alpha_w^1 Y_h + \eta_w^1. \tag{28}$$

Equations (25) and (26) describe the reservation wages, or equivalently, the shadow price of time for the husband and wife, respectively. Note that the wife's decision of whether or not to work, given by the dichotomous variable Y_w, affects the husband's reservation wage in (25). Analogously, the husband's decision of whether or not to work, given by Y_h, affects the wife's reservation wage in (26). Equations (27) and (28) describe the market wages for the husband and the wife, respectively. Note that we allow the possibility that one of the determinants of the husband's market wage is whether or not he has a working wife; we make a similar allowance for the wife.

As in the Nash model, let the husband's (wife's) reservation wage play the role of the payoff he (she) derives from not working. Therefore we have $W_h^r = \tilde{U}_h(0,Y_w)$ and $W_w^r = \tilde{U}_w(0,Y_h)$. Similarly, let the husband's (wife's) market wage play the role of the payoff he (she) derives from working. We thus have $W_h^m = \tilde{U}_h(1,Y_w)$ and $W_w^m = \tilde{U}_w(1,Y_h)$.

We then see from equations (25) through (28) that $U_h^1 - U_h^0 \equiv \Delta U_h = X_h'\gamma_h^m - Z_h'\gamma_h^r$ and $U_w^1 - U_w^0 \equiv \Delta U_w = X_w'\gamma_w^m - Z_w'\gamma_w^r$. Moreover, note that in specifying the husband's reservation wage and market wage equations, given by (25) and (27) respectively, it may be that certain explanatory variables appear in both equations, implying that the associated coefficient in ΔU_h will be measuring the difference between the market and reservation wage coefficients. A similar remark holds for the wife. In addition, note that the assumptions on the error terms are also satisfied, namely $\varepsilon_h \equiv \eta_h^1 - \eta_h^0$ and $\varepsilon_w = \eta_w^1 - \eta_w^0$.

Using the general specification set out in Section 2 of this Chapter, we then have the following two expressions for ΔU_h and ΔU_w

$$\Delta U_h = (\gamma_h^0 - \tilde{\gamma}_h^0) + (\gamma_h^1 - \tilde{\gamma}_h^1)\mathrm{AGEH} + (\gamma_h^2 - \tilde{\gamma}_h^2)\mathrm{EDUCH} \tag{29}$$

$$+ (\gamma_h^3 - \tilde{\gamma}_h^3)\text{UNEM} + (\gamma_h^4 - \tilde{\gamma}_h^4)\text{RACE} - \tilde{\gamma}_h^5\text{ASSETS} - \tilde{\gamma}_h^6\text{KIDS13} - \tilde{\gamma}_h^7\text{KIDS14}$$

and

$$\begin{aligned} \Delta U_w = {} & (\gamma_w^0 - \tilde{\gamma}_w^0) + (\gamma_w^1 - \tilde{\gamma}_w^1)\text{AGEW} + (\gamma_w^2 - \tilde{\gamma}_w^2)\text{AGEW**2} \\ & + (\gamma_w^3 - \tilde{\gamma}_w^3)\text{EDUCW} + (\gamma_h^4 - \tilde{\gamma}_h^4)\text{UNEM} + (\gamma_w^5 - \tilde{\gamma}_w^5)\text{RACE} - \tilde{\gamma}_w^6\text{ASSETS} \\ & - \tilde{\gamma}_w^7\text{KIDS1-2} - \tilde{\gamma}_w^8\text{KIDS3-5} - \tilde{\gamma}_w^9\text{KIDS6-13} - \tilde{\gamma}_w^{10}\text{KIDS>14}. \end{aligned} \tag{30}$$

B. EMPIRICAL RESULTS

From equations (25) and (27) it will be recalled that not only does the model allow for the possibility that one of the determinants of the husband's reservation wage is whether or not his wife chooses to work, the model also allows for the possibility that the husband's market wage is affected by his wife's decision. Although economic theory suggests that only the former effect should be meaningful, we can test that hypothesis in our model by allowing for the presence of both effects; that is, both α_h^0 and α_h^1 are included.

As with the Nash model, we currently do not have a set of initial consistent estimates of the parameters nor do we believe that the log-likelihood function is globally concave. As before, we performed a grid search using various sets of initial values of the parameter vector $(\alpha_h^0, \alpha_h^1, \gamma_h, \Delta\alpha_w, \gamma_w)$ with a grid search over possible values of ρ. Again, having no strong priors over the most likely value of ρ, we estimated all values from -.9 to +.9 in increments of .1. The convergence criterion we used was the same one previously used in the Nash model, viz., the mean value of each

TABLE 4

a_h^1 Included

Correlation Coefficient (ρ)	Log-likelihood Value
-.9	-1516.82
-.8	-1517.03
-.7	-1515.72
-.6	-1515.20
-.5	-1514.99
-.4	-1514.93
-.3	-1515.00
-.2	-1515.14
-.1	-1515.32
0.	-1515.58
.1	-1515.89
.2	-1516.24
.3	-1516.64
.4	-1517.06
.5	-1517.50
.6	-1517.98
.7	-1518.50
.8	-1519.19
.9	-1520.53

TABLE 5
α_h^1 Included
$\rho = 0$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	α_h^0	-1.76	-1.83*			
	α_h^1	-0.510	-0.68	$\Delta\alpha_w$	-1.44	-3.55**
NSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	-0.361	-0.32	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	0.580	0.90
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.017	2.01*			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.087	3.57**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.134	-4.39**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.067	1.62			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.039	3.22**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.043	-1.98**	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.015	-1.63*
CE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.300	-2.16**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.406	5.49**
SET	$-\tilde{\gamma}_h^5$	0.454	1.36	$-\tilde{\gamma}_w^6$	-0.012	-2.13**
DS1-2				$-\tilde{\gamma}_w^7$	-0.695	-11.20**
DS3-5				$-\tilde{\gamma}_w^8$	-0.444	-7.30**
DS6-13				$-\tilde{\gamma}_w^9$	-0.211	-5.63**
DS < 13	$-\tilde{\gamma}_h^6$	-0.023	-0.29			
DS > 14	$-\tilde{\gamma}_h^7$	0.074	0.62	$-\tilde{\gamma}_w^{10}$	-0.133	-2.71**

g-likelihood value = -1515.58

* significant at the 10% level
** significant at the 5% level

TABLE 6
α_h^1 Included
$\rho = -.40$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	α_h^0	-1.98	-1.72*			
	α_h^1	-0.256	-0.30	$\Delta\alpha_w$	-1.15	-2.02**
ONSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	-0.736	-0.60	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	0.330	0.43
GEH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.014	1.75*			
GEW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.084	3.48**
GEW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.131	-4.32**
DUCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.071	1.81*			
DUCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.039	3.27**
NEM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.040	-1.98**	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.014	-1.48
ACE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.330	-2.55**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.420	5.66**
SSET	$-\tilde{\gamma}_h^5$	0.410	1.35	$-\tilde{\gamma}_w^6$	-0.012	-2.13**
IDS1-2				$-\tilde{\gamma}_w^7$	-0.685	-11.10**
IDS3-5				$-\tilde{\gamma}_w^8$	-0.444	-7.30**
IDS6-13				$-\tilde{\gamma}_w^9$	-0.212	-5.66**
IDS < 13	$-\tilde{\gamma}_h^6$	0.021	0.29			
IDS > 14	$-\tilde{\gamma}_h^7$	0.104	0.93	$-\tilde{\gamma}_w^{10}$	-0.132	-2.70**

og-likelihood value = -1514.93

* significant at the 10% level
** significant at the 5% level

TABLE 7

α_h^1 Not Included

Correlation Coefficient (ρ)	Log-likelihood Value
-.9	-1519.15
-.8	-1517.52
-.7	-1515.85
-.6	-1515.21
-.45	-1514.93
-.5	-1515.00
-.4	-1515.03
-.3	-1515.20
-.2	-1515.44
-.1	-1515.74
0.	-1516.06
.1	-1516.41
.2	-1516.78
.3	-1517.16
.4	-1517.55
.5	-1517.97
.6	-1518.46
.7	-1519.08
.8	-1520.08
.9	-1522.29

TABLE 8
α_h^1 Not Included
$\rho = 0$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	α_h^0	-1.47	-2.01**			
				$\Delta\alpha_w$	-1.49	-3.71**
STANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	-0.300	-0.27	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	0.61	0.95
H	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.016	1.92*			
W				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.086	3.56**
W**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.133	-4.37**
CH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.052	1.28*			
CW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.042	3.45**
M	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.039	-1.91*	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.016	-1.77*
E	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.339	-2.56**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.403	5.43**
ET	$-\tilde{\gamma}_h^5$	0.460	1.41	$-\tilde{\gamma}_w^6$	-0.012	-2.10**
S1-2				$-\tilde{\gamma}_w^7$	-0.697	-11.30**
S3-5				$-\tilde{\gamma}_w^8$	-0.450	-7.45**
S6-13				$-\tilde{\gamma}_w^9$	-0.212	-5.76**
S < 13	$-\tilde{\gamma}_h^6$	0.011	0.20			
S > 14	$-\tilde{\gamma}_h^7$	0.081	0.69	$-\tilde{\gamma}_w^{10}$	-0.132	-2.68**

-likelihood value = -1516.06

* significant at the 10% level
** significant at the 5% level

TABLE 9
α_h^1 Not Included
$\rho = -.45$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	α_h^0	-1.82	-2.16**			
				$\Delta\alpha_w$	-1.12	-1.92*
NSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	-0.784	-0.63	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	0.31	0.40
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.013	1.68*			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.083	3.47**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.130	-4.31**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.069	1.79*			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.040	3.34**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.038	-1.97**	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.014	-1.50
ACE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.335	-2.64**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.442	5.69**
SSET	$-\tilde{\gamma}_h^5$	0.410	1.35	$-\tilde{\gamma}_w^6$	-0.012	-2.13**
DS1-2				$-\tilde{\gamma}_w^7$	-0.684	-11.20**
DS3-5				$-\tilde{\gamma}_w^8$	-0.447	-7.45**
DS6-13				$-\tilde{\gamma}_w^9$	-0.214	-5.88**
DS < 13	$-\tilde{\gamma}_h^6$	0.034	0.63			
DS > 14	$-\tilde{\gamma}_h^7$	0.109	0.99	$-\tilde{\gamma}_w^{10}$	-0.132	-2.70**

og-likelihood value = -1514.93

* significant at the 10% level
** significant at the 5% level

element of the score vector was allowed to be no greater than .01. Different trials of the parameter vector (α_h^0, α_h^1, γ_h, $\Delta\alpha_w$, γ_w) included (i) the zero vector; (ii) our best a priori guess of the parameter vector; (iii) the previous trial with the signs of the coefficients reversed; (iv) dividing by ten the final estimates from each of trials (i), (ii) and (iii); and (v) dividing by ten and multiplying by minus one the final estimates from each of trials (i), (ii) and (iii).

As can be seen from Table 4, the value of ρ that maximizes the likelihood function is $-.40$. The maximum likelihood estimates of the parameters of the full model for case $\rho = -.40$ are thus presented in Table 6. For completeness, Table 5 lists the maximum likelihood estimates for the case $\rho = 0$. It is seen in a comparison of Tables 5 and 6 that the estimates are quite robust with respect to ρ. Looking at the t-statistic associated with α_h^1 on Table 6, it follows that the hypothesis that $\alpha_h^1 = 0$ cannot be rejected at any reasonable level of significance, as theory suggests. Although we see that most of the explanatory variables, especially for the wife, have the a priori correct sign and are highly significant, we have therefore reestimated the model without α_h^1. Companion estimates for this case include the grid search over possible values of ρ, given in Table 7; the maximum likelihood estimates for the case $\rho = 0$, given in Table 8; and the coefficient estimates corresponding to the value of ρ which maximizes the likelihood function, given in Table 9.

As can be seen from Table 7, the value of ρ that maximizes the log-likelihood function is -.45. Although it may at first appear surprising that this maximizing value of ρ is not positive, it must be remembered that ρ is not simply the correlation between omitted variables in the husband's and wife's equations, but arises from a more complicated relationship between the disturbance terms ε_h and ε_w, viz, $\varepsilon_h \equiv \eta_h^1 - \eta_h^0$ and $\varepsilon_w \equiv \eta_w^1 - \eta_w^0$ as seen in Section 2 of Chapter III. From Table 9 we see that both $\Delta\alpha_w$ and α_h^0 are significantly different from zero, providing evidence that the wife's decision whether or not to work depends on the husband's decision and vice versa. Although it will be recalled from Section 4 of Chapter III that only the difference $\Delta\alpha_w \equiv \alpha_w^1 - \alpha_w^0$ can be identified in our model, economic theory again suggests that α_w^1 should be a priori zero since α_w^1 measures the effect of the husband's decision whether or not to work on the wife's market wage. Therefore the estimate -1.12 of $\Delta\alpha_w$ is actually an estimate of $-\alpha_w^0$. With this in mind then, we see from equation (26) that if the husband works, the wife's reservation wage increases as expected since $\hat{\alpha}_w^0$ is positive. It should also be noticed from Table 9 that we can provide a test of Proposition 3 of Chapter III. Since α_h^1 is restricted to be a priori zero and α_h^0 is significantly different from zero at the 5 percent level, we can reject the hypothesis that the data are generated from the usual recursive probability model using the dichotomization rule in favor of our model in which the observed outcomes are generated as Stackelberg equilibria. In other words, we must accept the hypothesis that the husband takes his wife's conditional action into account when making his decision whether or not to work.

A priori, we would expect the estimate of α_h^0 to be positive; that is, we expect that the wife's decision to work should increase the husband's reservation wage. In contrast, we find that the estimate of α_h^0 is negative and significant at the 5% level. One possible explanation for this result is that no husband wishes to suffer the embarrassment of staying at home when his wife chooses to work; that is, the husband lowers his reservation wage when his wife is working.

Looking again at Table 9, we see that most of the coefficients explaining the wife's decision whether or not to work are in agreement with our expectations and are highly significant. In reading Table 9 it should be noted that all estimated coefficients represent either differences between market and reservation wages or minus the reservation wage coefficients, as seen in Equations (29) and (30). For example, family income from souces other than wages and salaries (ASSET) has the expected effect of increasing the wage at which the wife is willing to accept work outside the home ($\tilde{\gamma}_w^6 = +0.012$). Concerning children, one would certainly expect that mothers would be least likely to leave the home when children are very young and be more inclined to seek outside employment as children become older and more self-sufficient. That is, younger children should have the effect of increasing mother's reservation wage more than do older children. Indeed, this is what we see from Table 9. Children between the ages of one and two (KIDS1-2) raise the mother's reservation wage more than do children between three and five (KIDS3-5); her reservation wage is higher for children between three and five than for children six to thirteen (KIDS6-13); finally, mothers are more likely to stay at home when their children are between six and thirteen than when they are fourteen years or older (KIDS>14). The coefficient on the wives' education (EDUCW) is also consistent with our a priori expectations; although an increase in education should increase the wife's market wage, it should also have the effect of increasing her reservation wage. The estimated positive coefficient on the female race dummy (RACE) seems to suggest again that women of racial minorities, on average, can command a higher market wage than the wage necessary to entice them into the labor market; that is, minority women are on average worth more in the market place than they think they are worth. Turning finally to the effect of age on a wife's decision whether or not to work, a life-cycle model of employment would suggest that women are more likely to work during middle age than either early or late in their lifetimes. That is, the probability that an individual will work exhibits a concave shape. As can be seen from Table 9, the combined effects of a positive linear term on age (AGEW) and a negative quadratic term (AGE**2) does indeed impart an increasing then a decreasing shape with respect to age.

Turning next to the variables used to explain the husband's decision whether or not to work, we see that while some of the coefficients are insignificant, many of the variables to which we attach strong priors are indeed significant. For example, the coefficients attached to the husband's age (AGEH), his education (EDUCH) and the local unemployment rate (UNEM) are each significant. Since each of these three coefficients measure the difference between the husband's market wage and his reservation wage, it is not surprising that they all should be close to zero if the husband is behaving rationally; for example, the effect of an increase in education should not only increase an individual's market wage but should also increase his reservation wage. Finally we see that the effects of racial discrimination on minorities have the effect of lowering their market wages relative to those of nonminorities.

5. THE STACKELBERG MODEL: WIFE LEADER

A. SPECIAL FEATURES OF THE STACKELBERG SPECIFICATION

In this third empirical example we shall reverse the roles of the husband and wife in the Stackelberg model so that the wife now plays the role of the leader while her husband plays the role of the follower. The current specification differs from the specification of the previous section in only one respect. Recall from Chapter III that only the <u>difference</u> in the shift parameters associated with the market wage and reservation wage can be identified for the individual playing the role of the follower, while both shift parameters can be identified for the leader. Since the husband is now playing the role of the follower, only $\Delta\alpha_h = \alpha_h^1 - \alpha_h^0$ can be identified, although both α_w^1 and α_w^0 can now be identified separately for the wife. The method we employ to estimate the parameter vector $(\Delta\alpha_h, \gamma_h, \alpha_w^0, \alpha_w^1, \gamma_w, \rho)$ is the same one used to estimate the Stackelberg model in which the husband plays the role of the leader; it is fully stated in Section 5B.

B. EMPIRICAL RESULTS

Recall from the previous section that while it is only possible to estimate the difference $\Delta\alpha_h = \alpha_h^1 - \alpha_h^0$, it is possible to estimate both α_w^0 and α_w^1; that is, we can separately test the effect of the husband's decision of whether or not to work on both the wife's reservation wage and also on her market wage. Although economic theory suggests that the latter effect should not be significant, we test this hypothesis in the model by allowing for the presence of both effects; that is, we estimate the model where both α_w^0 and α_w^1 are included. The empirical results are given in Tables 10 through 12. As can be seen from Table 10, the value of the correlation coefficient, ρ, that maximizes the log-likelihood function is .93. Constructing a likelihood ratio test of $\rho = 0$ vs. $\rho = .93$, we see that the hypothesis of $\rho = 0$ can be rejected at any reasonable level of significance. The maximum likelihood estimates of the parameters along with the associated t-statistics are found in Table 12 for the case $\rho = .93$. For comparison, companion estimates for the case $\rho = 0$ are included in Table 11.

From Table 12, we see that economic theory is confirmed in the sense that we cannot reject at the 10% level the hypothesis that $\alpha_w^1 = 0$. As a result, we have therefore reestimated the model without α_w^1; that is, we a priori constrain the effect of the husband's

TABLE 10

α_W^1 Included

Correlation Coefficient (ρ)	Log-likelihood Value
-.9	-1518.16
-.8	-1516.18
-.7	-1515.59
-.6	-1515.62
-.5	-1516.18
-.4	-1516.56
-.3	-1515.91
-.2	-1516.13
-.1	-1514.96
0.	-1514.37
.1	-1513.74
.2	-1513.10
.3	-1512.45
.4	-1511.70
.5	-1510.91
.6	-1510.02
.7	-1509.07
.8	-1507.71
.9	-1506.57
.93	-1506.26
.99	-1507.29

TABLE 11
α_w^1 Included
$\rho = 0$

	Husband Coefficient	Husband Estimate	Husband t-Statistic	Wife Coefficient	Wife Estimate	Wife t-Statistic
	$\Delta\alpha_h$	1.46	8.12**	α_w^0	1.83	7.05**
				α_w^1	-2.52	-0.22
NSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	1.15	1.94*	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	2.96	0.26
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.008	1.19			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.101	3.40**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.150	-4.06**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	-0.050	-1.70*			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.052	3.58**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.028	-1.68*	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.030	-2.53**
ACE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.354	-3.22**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.325	3.15**
SSET	$-\tilde{\gamma}_h^5$	0.344	1.68*	$-\tilde{\gamma}_w^6$	-0.004	-0.57
DS1-2				$-\tilde{\gamma}_w^7$	-0.864	-8.21**
DS3-5				$-\tilde{\gamma}_w^8$	-0.552	-6.57**
DS6-13				$-\tilde{\gamma}_w^9$	-0.260	-5.32**
DS < 13	$-\tilde{\gamma}_h^6$	0.020	0.39			
DS > 14	$-\tilde{\gamma}_h^7$	0.052	0.50	$-\tilde{\gamma}_w^{10}$	-0.138	-2.13**

g-likelihood value = -1514.37

* significant at the 10% level
** significant at the 5% level

TABLE 12

α_w^1 Included

$\rho = .93$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	$\Delta\alpha_h$	1.03	5.94**	α_w^0	2.05	15.70**
				α_w^1	-2.24	-1.54
NSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	2.02	3.50**	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	3.24	2.25
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.015	2.28**			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.084	3.89**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.118	-4.43**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	-0.099	-3.82**			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.024	2.07**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.030	-1.64	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.029	-2.55**
ACE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.378	-3.16**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.110	1.10
SSET	$-\tilde{\gamma}_h^5$	0.482	2.01**	$-\tilde{\gamma}_w^6$	-0.002	-0.29
DS1-2				$-\tilde{\gamma}_w^7$	-0.595	-9.90**
DS3-5				$-\tilde{\gamma}_w^8$	-0.374	-6.32**
DS6-13				$-\tilde{\gamma}_w^9$	-0.162	-4.02**
DS < 13	$-\tilde{\gamma}_h^6$	0.026	0.46			
DS > 14	$-\tilde{\gamma}_h^7$	0.010	0.01	$-\tilde{\gamma}_w^{10}$	-0.087	-1.47

g-likelihood value = -1506.26

* significant at the 10% level

** significant at the 5% level

TABLE 13
α_w^1 Not Included

Correlation Coefficient (ρ)	Log-likelihood Value
-.9	-1519.16
-.8	-1516.73
-.7	-1516.13
-.6	-1516.10
-.5	-1516.29
-.4	-1516.81
-.3	-1516.81
-.2	-1516.86
-.1	-1516.17
0.	-1515.77
.1	-1515.37
.2	-1514.97
.3	-1514.58
.4	-1514.19
.5	-1513.83
.6	-1513.52
.7	-1513.34
.71	-1513.33
.8	-1521.03
.9	-1522.60

TABLE 14
α_w^1 Not Included
$\rho = 0$

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	$\Delta\alpha_h$	1.33	6.82**	α_w^0	1.79	6.79**
NSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	1.10	1.88*	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	0.450	0.77
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.009	1.22			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.101	3.52**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.150	-4.20**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	0.033	-1.13			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.050	3.49**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.030	-1.79*	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.028	-2.46**
CE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.340	-3.01**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.328	3.28**
SET	$-\tilde{\gamma}_h^5$	0.365	1.56	$-\tilde{\gamma}_w^6$	-0.005	-0.74
DS1-2				$-\tilde{\gamma}_w^7$	-0.835	-8.36**
DS3-5				$-\tilde{\gamma}_w^8$	-0.526	-6.53**
DS6-13				$-\tilde{\gamma}_w^9$	-0.262	-5.39**
DS < 13	$-\tilde{\gamma}_h^6$	-0.013	-0.28			
DS > 14	$-\tilde{\gamma}_h^7$	0.035	0.34	$-\tilde{\gamma}_w^{10}$	-0.138	-2.22**

g-likelihood value = -1515.77

* significant at the 10% level
** significant at the 5% level

TABLE 15
α_w^1 Not Included
ρ = .71

	Husband			Wife		
	Coefficient	Estimate	t-Statistic	Coefficient	Estimate	t-Statistic
	$\Delta\alpha_h$	1.04	5.85**	α_w^0	2.46	13.8**
ONSTANT	$(\gamma_h^0 - \tilde{\gamma}_h^0)$	1.80	3.26**	$(\gamma_w^0 - \tilde{\gamma}_w^0)$	1.27	2.37**
EH	$(\gamma_h^1 - \tilde{\gamma}_h^1)$	0.012	1.71*			
EW				$(\gamma_w^1 - \tilde{\gamma}_w^1)$	0.094	3.74**
EW**2				$(\gamma_w^2 - \tilde{\gamma}_w^2)$	-0.134	-4.33**
UCH	$(\gamma_h^2 - \tilde{\gamma}_h^2)$	-0.072	-2.68**			
UCW				$(\gamma_w^3 - \tilde{\gamma}_w^3)$	0.032	2.53**
EM	$(\gamma_h^3 - \tilde{\gamma}_h^3)$	-0.033	-1.90*	$(\gamma_w^4 - \tilde{\gamma}_w^4)$	-0.032	-2.76**
CE	$(\gamma_h^4 - \tilde{\gamma}_h^4)$	-0.342	-2.95**	$(\gamma_w^5 - \tilde{\gamma}_w^5)$	0.174	1.66*
SET	$-\tilde{\gamma}_h^5$	0.324	1.81*	$-\tilde{\gamma}_w^6$	-0.001	-0.16
DS1-2				$-\tilde{\gamma}_w^7$	-0.733	-9.40**
DS3-5				$-\tilde{\gamma}_w^8$	-0.459	-6.62**
DS6-13				$-\tilde{\gamma}_w^9$	-0.244	-5.41**
DS < 13	$-\tilde{\gamma}_h^6$	-0.067	-1.44			
DS > 14	$-\tilde{\gamma}_h^7$	-0.005	-0.05	$-\tilde{\gamma}_w^{10}$	-0.112	-1.74*

g-likelihood value = -1513.33

* significant at the 10% level
** significant at the 5% level

decision of whether or not to work on the wife's market wage to be zero. Companion estimates for this case include the grid search over possible values of ρ, given in Table 13; the maximum likelihood estimates and the associated t-statistics for the case $\rho = 0$, given in Table 14; similar estimates for the value of ρ which maximizes the value of the likelihood function are given in Table 15.

As can be seen from Table 13, the value of ρ that maximizes the likelihood function is $\rho = .71$. Again, a likelihood ratio test at any reasonable level of significance would reject the hypothesis that $\rho = 0$. We also see from Table 15 that both $\Delta\alpha_h$ and α_w^0 are significantly different from zero, providing evidence that the husband's decision of whether or not to work depends on the wife's decision and vice versa. Although it will be recalled from the previous section that only the difference $\Delta\alpha_h \equiv \alpha_h^1 - \alpha_h^0$ can be identified in the model where the husband plays the role of the follower, economic theory again suggests that α_h^1 should be a priori zero. Thus, the estimate of $\Delta\alpha_h = 1.04$ is actually an estimate of $-\alpha_h^0$. With this in mind, we see from equation (25) of Section 4-A that if the wife works, the husband's reservation wage in fact declines. The only explanation we offer for this disconcerting result is that no husband wishes to suffer the embarrassment of staying out of the labor force when his wife chooses to work; that is, the husband chooses to lower his reservation wage when he has a working wife.

Notice also from Table 15 that we can again provide a test of Proposition 3 of Chapter III. Since we a priori restrict α_h^1 to be zero and α_h^0 is significantly different from zero at the 5 percent level, we can reject the hypothesis that the usual recursive model is identical to our proposed model in which the outcomes are generated as Stackelberg equilibria of a game in which the wife plays the role of the leader while her husband plays the role of the follower. In this model, in other words, we cannot reject the hypothesis that the wife takes her husband's conditional action into account when making her decision of whether or not to participate in the labor market. Note also that the estimate of α_w^0 is positive and highly significant, providing evidence that the husband's decision to work has the expected effect of increasing the wife's reservation wage.

Examining again Table 15, we see that most of the coefficients explaining the wife's decision of whether or not to work have the a priori correct sign and are highly significant. As is clear from looking at Table 15, many of the estimated coefficients represent the difference between market and reservation wages. As such, our a priori expectations would suggest that these estimated coefficients should be zero since any variable which has the effect of increasing the wife's market wage should have the balancing effect of increasing her reservation wage. Indeed, this is what we see when we examine the effects of an increase in the education of the wife (EDUCW) or an increase in the local unemployment rate (UNEM). The estimated positive coefficient on the female race dummy (RACE), however, does not meet with our a priori expectations. One possible explanation for this unexpected result is that women of racial minorities, on average, can command a wage in the market higher than the wage necessary to entice them into the labor market; in other words, minority women are on average worth more in the market than they think they are worth.

Turning next to the effect of age on a wife's decision of whether or not to work, a life cycle model of employment is again confirmed. As seen from Table 15, the combined effects of a positive linear term on age (AGEW) and a negative quadratic term (AGE**2) does indeed impart an increasing than a decreasing shape with respect to age. That is, the probability that a wife will work exhibits a concave shape. We need now only examine the effects of children on the wife's decision whether or not to work. As we have seen in the previous two studies, our a priori expectations are confirmed in the sense that younger children have the effect of increasing the mother's reservation wage more than do other children. Indeed, again we see that an increase in reservation wages for the wife exhibits a monotonically decreasing shape with respect to the age of an additional child.

Turning next to the variables used to explain the husband's decision whether or not to work, we see that all estimated coefficients are significant at the 10 percent level except those associated with the two age categories of children, viz, (KIDS<13) and (KIDS>14). Again, it should be noted in reading Table 15 that the coefficients associated with the husband's age (AGEH), his education (EDUCH), his race (RACE) and the local unemployment rate (UNEM) represent the difference between his market wage and his reservation wage. As such, we would again expect that if the husband is behaving logically, these estimated coefficients should be close to zero. Indeed, this is what we see for all variables except race (RACE). The negative and significant estimate associated with race seems to sugest that the effects of racial discrimination tend to lower the market wages of minorities relative to the wages of nonminorities. Finally, we see that an increase in assets has the effect of lowering the husband's reservation wage; we offer no explanation for this disconcerting result.

6. CONCLUSION

The purpose of the current chapter was to present an econometric study of the Nash and Stackelberg equilibrium models as set forth in Chapters II and III, respectively. The problem examined concerned the joint decision of a married couple whether or not to participate in the labor force. Two useful results should be apparent from this study. First, we have demonstrated that the game theoretic models proposed are in fact empirically tractable. Second, we feel that the proposed study has made a contribution to the literature on labor force participation because we explicitly modeled the behavior of a married couple in a game theoretic framework, while previous work has either taken the husband's decision to work as exogenous or specified the labor supply of a husband and wife from the outcome of a joint utility function.

In addition, we were able to test the hypothesis that the usual recursive probability model using the dichotomization rule is identical to our model in which the observed outcomes are generated as Stackelberg equilibria. In both specifications of the Stackelberg model, we were able to safely reject this hypothesis. That is, in both specifications, we were required to accept the hypothesis that the individual playing the role of the leader takes the conditional action of the follower into account when making his or her decision whether or not to work. Moreover, we were heartened in all three empirical examples to get strong results, both in terms of correct signs on coefficients for which we had strong priors and significant t-statistics.

APPENDIX: DESCRIPTION OF THE DATA

This Appendix describes the source of the data, the selection criteria used in choosing the approriate observations, and the means and variances of the explanatory variables. The data used in all three studies comprising the empirical part of this dissertation is from the 1982 wave of the Panel Study of Income Dynamics, 1968-1982. The sample is a combination of a representative cross-section of nearly 3,000 families selected by the Survey Research Center at the Uniersity of Michigan, and a subsample of about 1,900 low income families previously interviewed by the Census Bureau for the Office of Economic Opportunity. This data was then restricted to 2,012 records for married couples living in the United States, where both the husband and wife were able-bodied, neither older that 64 years of age with no nonrelative living with the family; in addition, the sample was further restricted to couples whose marital status had not changed within the previous twelve months.

Below we list the selection criteria used. Of the original 6,742 observations, we were left with 2,012 observations after selection.

(1) Family composition change in 1982 was restricted to children moving in or out of the home; husband and wife remained married and in the home. Loss: 2,295 observations.

(2) Family was restricted to husband, wife and children. Loss: 1,349 observations.

(3) both the husband and the wife were restricted to be 64 years or

less in age. Loss: 219 observations.

(4) Husbands who stated they were retired, permanently disabled, temporarily laid-off, or students were excluded. Those who stated they were "working now" were classified by us as working; those who answered "looking for work, unemployed," were considered to be not working. Loss: 331 observations.

(5) Wives who stated they were retired, permanently disables, temporarily laid-off, unemployed but looking for work, or students were excluded from the sample. Those who stated they were "working now" were classified by us as working; those who answered "housewife" were considered to be not working. Loss: 174 observations.

(6) If either the husband or the wife had a physical or nervous condition that limited the type or the amount of work they could do, they were excluded from the sample. Loss: 287 observations.

(7) If any record contained missing data for the 10 explanatory variables used in the analysis, that record was dropped. Loss: 67 observations.

(8) Eight observations were dropped because of incorrectly reported unemployment data.

Of the 2,012 observations remaining after selection, the numerical breakdown based upon husband/wife employment status is described in the following table.

		HUSBAND	
		Working	Not Working
Wife	Working	1204	48
	Not Working	739	29

FIGURE 1

MEANS AND VARIANCES OF THE EXPLANATORY VARIABLES

		MEAN	VARIANCE
AGEH	Age of Husband	36.7	109.0
AGEW	Age of Wife	34.3	97.4
AGEW**2	Squared Age of Wife	1270	592000
EDUCH	Husband's Education (years)	12.6	4.32
EDUCW	Wife's Education (years)	12.7	6.65
UNEM	Local unemployment rate	9.49	13.8
ASSETS	Family non-wage income (in thousands $)	.986	22.6
KIDS1-2	Number of children ages 1 and 2	.325	.299
KIDS3-5	Number of children between ages 3 and 5	.296	.270
KIDS6-13	Number of children between ages 6 and 13	.602	.735
KIDS<13	Number of children 13 or younger	1.22	1.28
KIDS>14	Number of children 14 or older	.296	.458
RACE	Race dummy (1 if Black or Hispanic, 0 otherwise)	.252	—

FOOTNOTES

1. The set of explanatory variables used in our empirical analysis will present no surprises. Indeed, most empirical studies of labor force particpation using cross-section data use a fairly common set of explanatory variables (see, e.g., Ashenfelter and Heckman (1974), Gronau (1973), Heckman (1974, 1976), and Nakamura and Nakamura (1981)).

2. The Panel Study of Income Dynamics asked only the race of the head of household; if married, we assumed the spouse to be of the same race.

3. For a discussion of the appropriateness of including current assets in a labor supply equation, see Cotterman (1981).

4. Note from (24) that $\tilde{\gamma}_w^6$ through $\tilde{\gamma}_w^{10}$ enter with negative signs.

5. See, e.g., Lazear (1977) for a discussion of this particular hypothesis.

CHAPTER V: A CHOICE OF THE MOST ADEQUATE MODEL

1. INTRODUCTION

In Chapters II and III, respectively, we proposed econometric models of two different game theoretic equilibrium notions, those being Nash Equilibrium and Stackelberg Equilibrium. In Chapter IV, three different empirical models were proposed and estimated concerning the joint labor force participation decision of a married couple. The first model assumed that the husband and wife both played a Nash game. The second model assumed that the married couple played a Stackelberg game where the husband played the role of the leader and his wife played the role of the follower. Model three, while also a Stackelberg game, assumed that the roles of the two players were reversed; that is, it was assumed that the wife played the role of the leader while the husband played the role of the follower. The purpose of the present chapter is to determine the most adequate model among the three proposed for explaining the joint labor force participation decision over a large sample of married couples.

The problem at hand can be considered to take the following form. Let us define two sequences of random vectors $y_1,\dots,y_n$ and $x_1,\dots,x_n$. The y_i's will be thought of as the endogenous variables and the x_i's as the exogenous variables, and we are interested in testing various hypotheses about the conditional distribution of the sequence $y_1,\ldots,y_n$ with respect to $x_1,\ldots,x_n$. The hypotheses to be tested are two functional forms of the conditional density of y_i given x_i denoted by

$$H_f = \{f(y_i|x_i,\alpha),\ \alpha \in A \subset \mathbb{R}^H\} \quad \text{and}$$

$$H_g = \{g(y_i|x_i,\beta),\ \beta \in B \subset \mathbb{R}^K\}$$

where $\mathbb{R}^H$ and $\mathbb{R}^K$ represent H dimensional and K dimensional real parameter spaces, respectively. The functions f and g will depend on parameter vectors which are given respectively as α and β.

If these models are nested in the sense that for any parameter value $\tilde{\alpha}$, the p.d.f. $f(y|x,\tilde{\alpha})$ can be approximated arbitrarily closely by $g(y|x,\tilde{\beta})$ or if for any parameter vector $\tilde{\beta}$, the p.d.f. $g(y|x,\tilde{\beta})$ can be approximated arbitrarily closely by $f(y|x,\tilde{\alpha})$, the problem of choosing the more adequate model is one in which the classical tests may be applied, viz, the Wald test, the Lagrange multiplier test, and the likelihood ratio test. Alternatively, the models may belong to separate families in the sense that for any parameter value $\tilde{\alpha}$, the p.d.f. $f(y|x,\tilde{\alpha})$ cannot be approximated arbitrarily closely by $g(y|x,\tilde{\beta})$ and for any parameter value $\tilde{\beta}$, the p.d.f. $g(y|x,\tilde{\beta})$ cannot be approximated arbitrarily closely by $f(y|x,\tilde{\alpha})$. When the two models belong to separate families we say they are non-nested. As we shall soon see, two different approaches have been suggested in the literature for choosing the most adequate model when the competing models belong to separate families.

The situation encountered, however, when we try to examine any two of the three proposed game theoretic models is neither one of comparing two nested models nor of testing two non-nested models. As will be shown later, the models are pairwise overlapping. As such, the traditional methods for choosing the most adequate model are, strictly speaking, inappropriate.[1] They will nevertheless be discussed for completeness. Given the drawbacks of the traditional methods which we shall point out, we therefore rely on a new technique developed by Vuong (1985) which handles as separate cases those situations in which the models may be nested, non-nested or overlapping. We also contrast the traditional approaches with the approach we adopt for choosing the game theoretic model which most adequately describes the labor force participation decision of a random sample of married couples.

In an attempt to discriminate between alternative models that arise from separate families, two different approaches have been suggested in the literature over the last two decades. The first approach, originating with Akaike (1973, 1974), has come to be known as model selection. As that name implies, one simply seeks to choose one model from a specified pair of models which minimizes an appropriately defined loss function; in the Akaike setting, the loss function is defined as two times the Kullback-Leibler (1951) measure of information. Although work subsequent to Akaike has been done on model selection, it has all concentrated on linear models; as such, it is inappropriate for the task at hand. (See Kinal and Lahiri (1983) for a review of this literature.)

The second approach has come to be known as non-nested hypothesis testing. Here, two main principles have been proposed in the literature: the Cox (1961, 1962) principle which is based on a modified likelihood ratio and the Atkinson (1970) principle which consists of nesting the non-nested models in a more general model. Although the Cox formulation does not explicitly allow the presence of exogenous variables, subsequent work by Aguirre-Torres and Gallant (1983) explicitly incorporates explanatory variables. Recent work by Gourieroux, Monfort and Trognon (1983) takes a different approach and rests on the notion of a pseudo-true parameter value and its associated pseudo-true maximum likelihood estimator (see, e.g., Sawa (1978) and Gourieroux, Monfort and Trognon (1984)). This work also allows for the presence of explanatory variables.

With this as a brief background, the present chapter is organized as follows. In Section 2, the Akaike principle of model selection will be developed more fully. As will be seen, a certain amount of controversy still exists as to whether or not the formula proposed by Akaike for the selection among competing models is correct. Moreover, it will be pointed out that the criteria proposed by Akaike is not probabilistic. That is, one simply chooses that model, among a group of competing models, with the largest model selection criteria. As such, we are required to choose a "best" model even though the "best" model may be statistically indistinguishable from one or more of the competing models. Section 3 will discuss in some detail, starting with Cox, the papers mentioned above on non-nested hypothesis testing. As will be pointed out, all tests of non-nested hypotheses are difficult to implement. In addition, as Vuong (1985) has argued, the Cox-type tests are inadequate for choosing the best model among a group of competing models. Since the techniques to be proposed in Sections 2 and 3 are inappropriate for the task at hand, Section 4 will discuss the new approach developed by Vuong (1985). As will be seen, this approach places the problem of model selection in a hypothesis testing framework. Section 5 will actually apply this test to the three models estimated in Chapter IV. Section 6 concludes the chapter.

2. MODEL SELECTION

Any discussion of model selection criteria for non-nested models must begin with Akaike (1973, 1974), both because he originated the subject and also because his framework is quite general. The basic attitude taken by Akaike toward the subject of model selection is to recognize the fact that, in general, a certain amount of discrepancy exists between the true, but unknown, distribution of a random variable and any proposed model. The best that can be done in trying to cope with this sort of situation is to identify the most adequate model relatively among a given set of alternative models. The adequacy of a proposed model thus needs to be quantified by defining a suitable measure of the distance of the model from the unknown true distribution. The measure of distance used by Akaike is based upon the Kullback-Leibler (1951) Information Criterion (KLIC);

as such, Akaike's statistic for measuring this distance is called the Akaike Information Criterion and is abbreviated as the AIC.

Suppose that one is concerned with the probabilistic structure of a set of independent random variables $y = (y_1,\dots,y_n)$. To simplify the discussion, let us further suppose that there are no exogenous variables. Let $G^0(y)$ be the true distribution of y with associated probability density function $g^0(y)$. We now postulate a distribution $F(y,\alpha)$ to approximate the unknown distribution $G^0(y)$. Then the Kullback-Leibler Information Criterion (KLIC), which defines a measure of distance between the true distribution $G^0(y)$ and the proposed distribution $F(y,\alpha)$, is defined as

$$I(G,F) = E_G^0\left[\log\frac{g^0(y)}{f(y,\alpha)}\right] = \int \log\frac{g^0(y)}{f(y,\alpha)}\, g^0(y)dy$$

where f (resp. g^0) is the density function of F (resp. G^0) and $E_G^0(\cdot)$ stands for the expectation with respect to the true distribution G^0. Since it can be shown that the KLIC is nonnegative, $I(G^0,F) \geq 0$, and equals zero only when $F(y,\alpha) = G^0(y)$, it is therefore natural to choose that distribution, among a group of distributions, which minimizes the KLIC. Actually, Akaike proposes a slightly different measure of the discrepancy between the true distribution $G^0(y)$ and a proposed distribution:

$$W(G^0,F) = -2\int \log\frac{f(y,\alpha)}{g^0(y)}\, g^0(y)dy.$$

Then the distance between the true distribution $G^0(y)$ and a proposed model $\mathcal{F} = \{f(y,\alpha);\ \alpha \in A\}$ is $W(G^0, \mathcal{F}) = -2\int \log\frac{f(y,\alpha^*)}{g^0(y)}\, g^0(y)dy$ where α^* is the pseudo-true value of α; that is, α^* is the value of α which minimizes $W(G^0,\mathcal{F})$. As can be seen, $W(G^0,\mathcal{F})$ is two times the Kullback-Leibler Information measure given previously. Why Akaike uses a modified KLIC will be seen later. Associated with the above loss function is the following function

$$R(G^0,\mathcal{F}) = E_{\hat{\alpha}}\, W(G^0,\mathcal{F})$$

where the expectation is taken with respect to the distribution of $\hat{\alpha}$, the maximum likelihood estimator of α. Note that we take the expectation with respect to $\hat{\alpha}$ so that $R(G^0,\mathcal{F})$ is no longer a random variable.

Since $W(G^0,\mathcal{F})$ is unknown because G^0 and α^* are unknown the next step then is to propose a criterion which provides an estimate of $R(G^0,F)$. The approach relies on the fact that α^* can be replaced with the maximum likelihood estimator $\hat{\alpha}$ which converges almost surely to $\hat{\alpha}$ (see, e.g. White (1982)). Although it is not necessary to go through the complete derivation of the criterion, we need to examine the first step in order to make clear the major drawback of the AIC. Assume that α is of dimension L. Now consider the problem of the selection of $g(y,{}_K\alpha)$, $K = 0,1,\ldots,L$ based on the independent observations $y_1,\ldots, y_n$ where ${}_K\alpha$ is restricted to the space with ${}_K\alpha_{K+1} = {}_K\alpha_{K+2} = \cdots = {}_K\alpha_L = 0$. If we let ${}_K\hat{\alpha}$ be the maximum likelihood estimator in ${}_K\alpha$ space, then

$${}_K W_L = \frac{-2}{N}\sum_{i=1}^{N} \log \frac{g(y_i, {}_K\hat{\alpha})}{g(y_i, {}_L\hat{\alpha})}$$ will be a consistent estimate of

$W(\tilde{\alpha}, {}_K\hat{\alpha})$ since we are assuming the model with L parameters, $g(y,\tilde{\alpha})$, is the true model. Let us now treat $W(\tilde{\alpha}, {}_K\hat{\alpha})$ as a function of ${}_K\hat{\alpha}$ and take a second order Taylor expansion of $W(\tilde{\alpha},\hat{\alpha})$ around $\tilde{\alpha}$ assuming that ${}_K\hat{\alpha}$ is close to $\hat{\alpha}$. We get

$$W(\tilde{\alpha}, {}_K\hat{\alpha}) \simeq W(\tilde{\alpha},\tilde{\alpha}) + \sum_{\ell=1}^{L} ({}_K\hat{\alpha}_\ell - \tilde{\alpha}_\ell) \frac{\partial W(\tilde{\alpha},\alpha)}{\partial \alpha_\ell} \Big|_{\alpha=\tilde{\alpha}}$$

$$+ \frac{1}{2}\sum_{\ell=1}^{L}\sum_{m=1}^{L} ({}_K\hat{\alpha}_\ell - \tilde{\alpha}_\ell)({}_K\hat{\alpha}_m - \tilde{\alpha}_m) \frac{\partial^2 W(\tilde{\alpha},\alpha)}{\partial\alpha_\ell\, \partial\alpha_m}\Big|_{\alpha=\tilde{\alpha}} + 0_p(n^{-3/2})$$

$$= W(\tilde{\alpha},\tilde{\alpha}) + W_1(\tilde{\alpha}, {}_K\hat{\alpha}) + W_2(\tilde{\alpha}, {}_K\hat{\alpha}) + 0_p(n^{-3/2})$$

But $W(\tilde{\alpha},\tilde{\alpha}) = 0$ and $\frac{\partial W(\tilde{\alpha},\alpha)}{\partial\alpha}\Big|_{\alpha=\tilde{\alpha}} = -2\int \frac{\partial g(y,\tilde{\alpha})}{\partial\alpha} dy = 0$ since

$\int g(y,\alpha)dy = 1$ for every α. Now, $\frac{\partial^2 W(\tilde{\alpha},\alpha)}{\partial\alpha\partial\alpha'}\Big|_{\alpha=\tilde{\alpha}}$

$$= 2\int \left(\frac{\partial g(y,\tilde{\alpha})}{\partial\alpha}\, \frac{1}{g(y,\alpha)}\right)\left(\frac{\partial g(y,\tilde{\alpha})}{\partial\alpha}\, \frac{1}{g(y,\alpha)}\right) g(y,\tilde{\alpha})dy$$

$$= 2\int \frac{\partial^2 \log g(y,\tilde{\alpha})}{\partial\alpha\partial\alpha'} g(y,\tilde{\alpha})dy.$$ We therefore have that

$W(\tilde{\alpha}, {}_K\hat{\alpha}) \simeq W_2(\tilde{\alpha}, {}_K\hat{\alpha})$. Akaike next shows that

$R({}_L\hat{\alpha}, {}_K\hat{\alpha}) = \frac{1}{N}({}_K\eta_L + 2K - L)$ is a good estimate of $EW_2(\tilde{\alpha}, {}_K\hat{\alpha})$ where ${}_K\eta_L = N \cdot {}_KW_L = -2\sum_{i=1}^{N} \log\left[\frac{g(y_i, {}_K\hat{\alpha})}{g(y_i, {}_L\hat{\alpha})}\right]$. Since we are only concerned with finding the ${}_K\hat{\alpha}$ which will give the minimum of $R({}_L\hat{\alpha}, {}_K\hat{\alpha})$, we have only to compute

$${}_K\lambda_L = -2\sum_{i=1}^{N} \log\, g(y_i, {}_K\hat{\alpha}) + 2K$$ and adopt the ${}_K\hat{\alpha}$ which

gives the minimum of ${}_K\lambda_L$, $0 \leq K \leq L$. ${}_K\lambda_L$ is called the Akaike Information Criterion, AIC. The first term in the AIC measures the goodness of fit of the model to a given set of data since $g(y, {}_K\hat{\alpha})$ is the maximized likelihood function. The second term is interpreted as a penalty that is paid for increasing the number of parameters (see Leamer (1979) for a further discussion of this point).

Three points are now worth noting. First, although the above derivation required that $g(y, {}_K\alpha)$ be nested in $g(y, {}_L\alpha)$, the use of the AIC requires no such assumption. That is, the AIC can be used to select between non-nested models since each model can be thought of as being nested in a larger model that contains both. Second, as pointed out by Sawa (1978), the above requirement that ${}_K\hat{\alpha}$ be sufficiently close to the true parameter value $\tilde{\alpha}$ lessens to some extent the plausability of the AIC. To see this, consider the problem where we must choose between f_1 and f_2, say. The AIC for f_1 is evaluated assuming that f_1 is sufficiently close to the true model, while the AIC for f_2 is evaluated assuming that f_2 is sufficiently close to the true model. Third, as noted above, the AIC trades off goodness of fit for the number of parameters. As such, one can consider it to be analogous to the multiple correlation coefficient adjusted for degrees of freedom, $\bar{R}$, suitably generalized so that one may compare nonlinear and non-nested models. Moreover, just as the use of $\bar{R}^2$ can be misleading when choosing among nested regression models, so can the use of the AIC when choosing among nonlinear and/or non-nested models. To see this, recall from Pesaran (1974) the following example. Consider a linear model with K-1 regressors and the same model with an extra regressor. He then shows that the two models are related as

$$\bar{R}^2_K - \bar{R}^2_{K-1} = \frac{(1 + \bar{R}^2_K)}{n - K + 1} (\hat{t}_K - 1) \text{ where } \bar{R}_K \text{ and } \bar{R}_{K-1} \text{ denote}$$

the adjusted multiple correlation coefficients for the models with K and K-1 explanatory variables respectively, $\hat{t}_K$ is the t-statistic of the added variable and n is the sample size. It is then clear from this relation that an estimated value of the t-ratio which is slightly greater than one will increase the adjusted multiple correlation coefficient but will only be significant at the 25 percent level, which is much larger than the normally adhered to significance levels of 5 percent or 10 percent. A similar problem exists with the Akaike Information Criterion: we are required to choose among alternative models simply on the basis of which one gives the minimum AIC, ignoring the fact that the competing models may be statistically indistinguishable. As we shall see in the next section, this problem is overcome when we study those techniques collectively known as non-nested selection criteria. We will also see in Section 4 that the recent work by Vuong (1985) places in a hypothesis testing framework the problem of selecting among competing models whether they be nested, non-nested or overlapping.

3. NON-NESTED HYPOTHESES TESTING

Although important strides have been made in the last few years in developing tests for separate families of hypotheses, all of this work draws heavily from Cox (1961, 1962). As such, it is only fitting that we discuss the work of Cox first. As we shall soon see, the test proposed by Cox is a modification of the likelihood ratio test.

In keeping with the notation of the previous section, suppose that the observed value of a random vector $y = (y_1,\ldots,y_n)$ is to be used to test the null hypothesis, H_f, that the true probability density function is $f(y,\alpha^0)$, where α^0 is an unknown parameter vector. Let H_g be the alternative hypothesis with p.d.f. $g(y,\beta)$, where β is an unknown vector parameter, and where $f(y,\alpha)$ and $g(y,\beta)$ are separate families. In addition, let $L_f(\hat{\alpha}) = \sum \log f(y_i,\hat{\alpha})$ be the maximized log likelihood under H_f where $\hat{\alpha}$ is the maximum likelihood estimator. $L_g(\hat{\beta}) = \sum \log g(y_i,\hat{\beta})$ is similarly defined. Cox (1961) proposes the following test

$T_f = \{L_f(\hat{\alpha}) - L_g(\hat{\beta})\} - E_{\hat{\alpha}}\{L_f(\hat{\alpha}) - L_g(\hat{\beta})\}$ where $E\{\cdot\}$ means that you evaluate the expectation with respect to $f(y;\hat{\alpha})$. Thus, the test

compares the observed difference in log-likelihood with an estimate of that to be expected under H_f. As such, an estimate of T_f close to zero would indicate the acceptance of H_f; a positive estimate of T_f would indicate a departure from H_f away from H_g; and a negative estimate of T_f also indicates a departure from H_f but in the direction of H_g. Examining the latter case more carefully, although a significantly negative value of T_f indicates rejection of H_f, it obviously does not indicate acceptance of H_g, since a model cannot be accepted (or rejected) until it has been put to a test. It is therefore for this reason that we must reverse the roles of the two hypotheses and repeat the test procedure. If the roles of H_f and H_g as null and alternative hypotheses are interchanged, a test statistic T_g is obtained, where

$T_g = [L_g(\hat{\beta}) - L_f(\hat{\alpha})] - E_{\hat{\beta}}[L_g(\hat{\beta}) - L_f(\hat{\alpha})]$. Cox (1962) and White (1982) then show

that if the components $(y_1,\cdots,y_n)$ are independently and identically distributed, $n^{-1/2}T_f$ is under H_f distributed asymptotically normal with mean zero and asymptotic covariance matrix

$$\left[V_\alpha(F - G) - \frac{C_\alpha^2(F - G, F_\alpha)}{V_\alpha(F_\alpha)}\right]$$

where $F = \log f(y,\alpha)$, $F_\alpha = \frac{\partial \log f(y,\alpha)}{\partial \alpha}$, $G = \log g(y,\beta)$, and the subscript on the variance and covariance operators, V and C, indicates that the expectation is taken with respect to $f(y;alph^0)$. The major difficulty with the test proposed by Cox is the computation of the expectation of $L_f(\hat{\alpha}) - L_g(\hat{\beta})$ with respect to the P.D.F. $f(y;\hat{\alpha})$. Two methods have been proposed to get around this problem by numerically simulating the distribution of the test statistic. These methods are known as parametric and nonparametric bootstrapping; we will discuss the former first.

The idea of parametric bootstrapping (see Williams (1970) and Loh (1985)) is quite simple. Assume two hypotheses, H_0: $f(y,\alpha)$ and H_1: $g(y,\beta)$. Under H_0, we assume $\alpha = \hat{\alpha}$ and simulate for sufficiently large K, K sets of artificial data $(y^*_{1k},\cdots,y^*_{nk})$, $k = 1,\ldots,K$, drawn from the population with density $f(y^*,\hat{\alpha})$. From the k-th set of artificial data, the pseudo maximum likelihood estimates $\hat{\alpha}^*_k$, $\hat{\beta}^*_k$ are calculated along with the test statistic

$T^*_{nk} = n^{-1} \sum \log[f(y^*_{ik},\hat{\beta}^*_k)/g(y^*_{ik},\hat{\alpha}^*_k)]$. This is done for each k, $k = 1,\ldots,K$. Calculating the comparable test statistic $T_n = n^{-1} \sum \log[f(y_i,\hat{\beta})/g(y_i,\hat{\alpha})]$ from the real data, hypothesis H_0 is rejected if T_n exceeds the largest element of the set $\{T^*_{nk}\}$. Turning the problem around, we can again by simulation obtain observations on the distribution of the statistic assuming model $g(y,\beta)$ is true with parameters equal to its maximum likelihood estimates, $\hat{\beta}$, from the original data. Once these steps are completed, the problem is then one of determining whether the observed data arise from just one, both or neither of the two models $f(y,\alpha)$ and $g(y,\beta)$. We now turn to a discussion of nonparametric bootstrapping.

The work of Aguirre-Torres and Gallant (AGT) is similar to the work previously described in the sense that the goal is to provide a bootstrap estimate of the expectation under the null of the Cox difference, viz

$$(1/n)\sum_{t=1}^{n} \int [\log f(y|x_t,\hat{\alpha}) - \log g\ (y|x_t,\hat{\beta})] f(y|x_t,\hat{\alpha})dy. \qquad (1)$$

This method differs from the previous bootstrap method, however, in that the error structure of both models is free of any distributional assumptions. It is necessary to adopt the notation of AGT for they consider the Cox test for choosing between two nonlinear, non-nested models

$$q_1(y_t,x_t,\alpha) = e_{1t} \quad \text{and} \quad q_2(y_t,x_t,\beta) = e_{2t,}\ \ t = 1,\cdots,n\ ,$$

based on the observed data $\{y_t,x_t\}_{t=1}^{n}$. (When distributions on the error terms are specified, each model determines a conditional density: $f(y|x,\alpha)$ corresponding to $q_1(y,x,\alpha) = e_1$ and $g(y|x,\beta)$ corresponding to $q_2(y,x,\beta) = e_2$.)

The distribution free estimate of the expectation of the Cox difference under the null is then calculated as follows. For <u>each</u> <u>t</u>, compute a random sample of size n denoted as $\hat{Y}_{tj}$, $j = 1,\ldots,n$, by

(i) finding the maximum likelihood estimate $\hat{\alpha}$,

(ii) for each x_t, generate the n dimensional vector of residuals $\hat{e}_t$ from $\hat{e}_t = q_1(y_t,x_t,\hat{\alpha})$ by varying

y_t, $t = 1,\ldots,n$,

(iii) find $\hat{y}_{tj}$ by solving the equations $\hat{e}_j = q_1(\hat{y}_{tj}, x_t, \hat{\alpha})$

This procedure thus specifies a way of generating a random sample of dependent variables, y_t, conditional upon x_t under the conditonal density $f(y|x,\alpha)$ without making any distributional assumptions on the errors, e_{1t}. The distribution free estimate of the expression given in (1) is then

$$(1/n^2) \sum_{t=1}^{n} \sum_{j=1}^{n} [\log f(\hat{y}_{tj}|x_t,\hat{\alpha}) - \log g(\hat{y}_{tj}|x_t,\hat{\beta})]. \qquad (2)$$

Two further points deserve mention.[2] First, we must assume that the errors, e_{1t} or e_{2t}, are not only independent but also identically distributed in order that the estimate given in (2) converges almost surely to (1). Therefore, the bootstrapping method proposed by AGT is not valid for models with a heteroscedastic error structure. Second, in order to calculate the bootstrap estimate it is necessary to be able to express the conditional density under the null, $f(y_t|x_t,\alpha)$, in terms of the expression $q_1(y_t,x_t,\alpha) = e_{1t}$; certainly the latter expression contains more structure than the former.

Let us now discuss an alternative approach for testing non-nested hypotheses as proposed by Gourieroux, Monfort and Trognon (1983). They propose a test based on the difference between the pseudo-maximum likelihood estimator of the parameter of the alternative model and an estimator of the pseudo-true value obtained from the maximum likelihood estimator for the parameter of the null model. An added feature of their approach is that it takes explicitly into account the presence of explanatory variables.

As before, the hypotheses to be tested are two functional forms of the conditional density of y_t given x_t denoted as

$$H_f = \{f(y_t|x_t,\alpha),\ \alpha \in A \subset \mathbb{R}^H\},$$

$$H_g = \{g(y_t|x_t,\beta),\ \beta \in B \subset \mathbb{R}^K\}.$$

For convenience $f(y_t|x_t,\alpha)$ and $g(y_t|x_t,\beta)$ will also be denoted by $f_t(\alpha)$ and $g_t(\beta)$ and the conditional expectations with respect to the densities will be written as $\underset{\alpha}{E}$ and $\underset{\beta}{E}$. Therefore, the conditional log-likelihood functions associated with H_f and H_g are respectively given by

$$L_f^T(\alpha) = \sum_{t=1}^{T} \log f_t(\alpha), \quad L_g^T(\beta) = \sum_{t=1}^{T} \log g_t(\beta).$$

We now allow for the possibility that the true distribution generating the observations may not be associated with either of the families H_f and H_g. As such, these estimators have been termed pseudo-maximum likelihood (PML) estimators (see, e.g., Gourieroux, Monfort and Trognon (1984)) and are defined to be the solutions of $\max_{\alpha} L_f^T(\alpha)$ and $\max_{\beta} L_g^T(\beta)$. These estimators are denoted as $\hat{\alpha}$ and $\hat{\beta}$, respectively.

We now define the finite sample pseudo-true value of β for a given α as the solution to the problem

$$\max_{\beta} \sum_{t=1}^{T} \underset{\alpha}{\overset{t}{E}} \log g_t(\beta) \;;$$

this solution is denoted as $b^T(\alpha)$, where

$$\underset{\alpha}{\overset{t}{E}} \log g_t(\beta) = \int \log g(y|x_t,\beta) \cdot f(y|x_t,\alpha)dy.$$

Since this problem is equivalent to

$$\min_{\beta} \sum_{t=1}^{T} \underset{\alpha}{\overset{t}{E}} \log \frac{f_t(\alpha)}{g_t(\beta)}$$

where

$$\underset{\alpha}{\overset{t}{E}} \log \frac{f_t(\alpha)}{g_t(\beta)} = \int \log \frac{f(y|x_t,\alpha)}{g(y|x_t,\beta)} f(y|x_t,\alpha)dy$$

is the Kullback-Leibler information criterion, it is clear that the sample pseudo-true value of β, $b^T(\alpha)$, is the β which minimizes the distance between H_f and H_g when we assume that $f(y|x,\alpha)$ is true. In the same way, the finite sample pseudo-true value of α for a given β is denoted as $a^T(\beta)$ and is defined as the solution to

$$\max_{\alpha} \sum_{t=1}^{T} \underset{\beta}{\overset{t}{E}} \log f_t(\alpha).$$

Moreover, Gourieroux, Monfort and Trognon (1984) show that the pseudo-maximum likelihood estimator, $\hat{\beta}$, is a consistent estimator of the asymptotic pseudo-true value, $b(\alpha)$. It therefore seems natural to compare under H_f the pseudo-maximum likelihood estimator $\hat{\beta}$ of β with $b^T(\hat{\alpha})$, the finite sample pseudo-true value of β. Since, under H_f,

$\hat{\beta} - b^T(\hat{\alpha})$ converges to zero, and since this is not in general the case under H_g, a significant departure from zero of $\hat{\beta} - b^T(\hat{\alpha})$ will be in favor of H_g. The authors then show that the distribution of $\sqrt{T}\,[\hat{\beta} - b^T(\hat{\alpha})]$ is asymptotically normal.

We now discuss another method of testing non-nested hypotheses which yields a test statistic that is asymptotically equivalent to the Cox test. Although this method was originally suggested by Cox (1961, 1962), it was formalized by Atkinson (1970) and consists of nesting two non-nested models into a general model in which the two smaller models would both be special cases. Explicit account is not taken for exogenous variables.

Again, letting the two component p.d.f.'s be $f(y,\alpha)$ and $g(y,\beta)$, the combined p.d.f. is of the form

$$f_\lambda(y) = \frac{[f(y,\alpha)]^\lambda \; [g(y,\beta)]^{1-\lambda}}{\int [f(z,\alpha)]^\lambda \; [g(z,\beta)]^{1-\lambda} dz}$$

when we assume that the observations are identically and independently distributed. Note that the denominator has been introduced so that $f_\lambda(y)$ has the properties of a density. Since we are ultimately concerned with an estimate of λ, Atkinson follows a suggestion by Bartlett (1953) and uses an asymptotically normal statistic for testing hypotheses about the value of single parameter, λ, in the presence of nuisance parameters, α and β. For ease of exposition, we assume that L, the log-likelihood of the observations, contains only the nuisance scalar parameter Θ.

Let

$$I_{11} = -E(\frac{\partial^2 L}{\partial \lambda^2}),\ I_{12} = -E(\frac{\partial^2 L}{\partial \theta \partial \lambda}),\ I_{22} = -E(\frac{\partial^2 L}{\partial \theta^2})$$

where the partial derivatives are evaluated and the expectations calculated under the null hypothesis. The test statistic proposed by Bartlett is

$$T = \frac{(\partial L/\partial \lambda) - (I_{12}/I_{22})(\partial L/\partial \theta)}{(I_{11} - (I_{12}^2/I_{22}))^{1/2}}\ ,$$

which has an asymptotic normal distribution. If we use as estimates of the nuisance parameters the maximum likelihood estimates under the null hypothesis, we see that the second term in the numerator in the above expression drops out.

Let us now assume that the null hypothesis is $\lambda = 1$; that is, we assume $f(y,\alpha)$ is the true model and we test for departures from $f(y,\alpha)$ in the direction of $g(y,\beta)$. It is then only necessary to differentiate the log-likelihood function associated with $f_\lambda(y)$ with respect to α, β and λ and evaluate these partial derivatives at $\lambda = 1$. Since the test statistic T requires that we estimate λ and β under the null hypothesis $\lambda = 1$, we use the maximum likelihood estimator $\hat{\alpha}$ and the pseudo-maximum likelihood estimator $b(\hat{\alpha})$.

As with all the non-nested hypotheses tests discussed so far, we also need to reverse the roles of the two distributions and test the hypothesis $\lambda = 0$; that is, we must test for departures from $g(y,\beta)$ in the direction of $f(y,\alpha)$.

4. LIKELIHOOD RATIO TESTS FOR MODEL SELECTION AND

The purpose of the current section is to provide a brief overview of the method proposed by Vuong (1985) for the selection of a "best" model from a set of competing models in a hypothesis testing framework. As in Akaike (1973, 1974) the selection criteria used is based on the Kullback-Leibler (1951) Information Criterion (KLIC) which measures the difference between a given distribution and the true distribution. The work discussed here differs, however, from Akaike in three important respects. First, conditional models are considered so as to allow for explanatory variables. Second, Vuong (1985) examines the cases in which the competing models may be non-nested, overlapping or nested; in contrast, Akaike (1973, 1974) limits himself to a discussion of non-nested models. Moreover, in contrast to Akaike, Vuong does not require one of the models to be correctly specified. Finally, the test statistic proposed is based on the log-likelihood ratio uncorrected for the number of estimated parameters.

Using the notation established earlier, let us define two conditional models $F_\alpha = \{f(y|x,\alpha), \alpha \in A\}$ and $G_\beta = \{g(y|x,\beta), \beta \in B\}$. Then the distance of each of these models from the true conditional distribution $h^0(y|x)$, as measures by the KLIC, is defined as
$E^0[\log h^0(y|x)] - E^0[\log f(y|x,\alpha^*)]$ and
$E^0[\log h^0(y|x)] - E^0[\log g(y|x,\beta^*)]$, respectively, where $E^0[\cdot]$ denotes the expectation with respect to the true joint distribution of (y,x) and α^* and β^* are the pseudo-true values of α and β (see e.g., Sawa (1978) and Gourieroux, Monfort and Trognon (1984)). As noted before,

both of these expressions are nonnegative and equal zero, say, only when $h^0(y|x) = f(y|x,\alpha^*)$. As such, an equivalent selection criterion is based on $E^0[\log f(y|x,\alpha^*)]$ and $E^0[\log g(y|x,\beta^*)]$, the better of the two models being the one which gives the larger value. Although both of these quantitites are unknown, they can each be consistently estimated by (1/n) times the log-likelihood evalated at the respective pseudo-maximum likelihood estimators. With this as a background, Vuong (1985) suggests tests for model selection where the null hypothesis is that $E^0[\log g(y|x,\beta^*)] = E^0[\log g(y|x,\beta^*)]$, indicating that the two models are equivalent, against $E^0[\log f(y|x,\alpha^*)] > E^0[\log g(y|x,\beta^*)]$ indicating that F_α is better than G_β, or against $E^0[\log f(y|x,\alpha^*)] < E^0[\log g(y|x,\beta^*)]$ indicating that G_β is better than F_α. As such, the test will be based on the likelihood ratio statistic

$$
\begin{aligned}
LR_n(\hat{\alpha}_n,\hat{\beta}_n) &\equiv L_n^f(\hat{\alpha}_n) - L_n^g(\hat{\beta}_n) \\
&= \sum_{t=1}^{n} \log \frac{f(y_t|x_t,\hat{\alpha}_n)}{g(y_t|x_t,\hat{\beta}_n)}.
\end{aligned}
$$

where $\hat{\alpha}_n$ and $\hat{\beta}_n$ are the pseudo-maximum likelihood estimators of α^* and β^*.

As noted in the Introduction to this chapter, the model selection criterion proposed by Akaike (1973, 1974) is not probabilistic. Since Vuong (1985) derives tests for model selection it is therefore necessary to examine the asymptotic distribution of the likelihood ratio statistic. A complication arises, although, because the asymptotic distribution of the likelihood ratio statistic depends on whether or not $f(\cdot|\cdot,\alpha_*) = g(\cdot|\cdot,\beta_*)$. As is shown in Lemma 3.1 of Vuong (1985), if $f(\cdot|\cdot,\alpha_*) = g(\cdot|\cdot,\beta_*)$, then $2LR_n(\hat{\alpha}_n,\hat{\beta}_n)$ has a limiting distribution which is a weighted sum of chi square random variables, the weights being the eigenvalues of a matrix to be discussed later.

On the other hand, if $f(\cdot|\cdot,\alpha_*) \neq g(\cdot|\cdot,\beta_*)$, then $n^{-1/2}LR_n(\hat{\alpha},\hat{\beta})$ is a consistent estimator of $LR(\alpha_*,\beta_*)$ $\equiv n^{-1/2}E^0[\log f(y|x,\alpha^*) - \log g(y|x,\beta^*)]$. Indeed, as Theorem 3.4 of Vuong (1985) shows, if $f(\cdot|\cdot, \alpha_*) \neq g(\cdot|\cdot,\beta_*)$ then

$$n^{-1/2}LR_n(\hat{\alpha}_n,\hat{\beta}_n) - n^{-1/2}E^0\left[\log\frac{f(y|x,\ \alpha_*)}{g(y|x,\ \beta_*)}\right] \xrightarrow{D} N(0,\ \omega_*^2)$$

where ω_*^2 is the variance of $\log[f(y)|x,\ \alpha^*)/g(y|x,\ \beta^*)]$ where the variance is computed with respect to the joint distribution H^0 of (y,x).

Then under H^0: $E^0\left[\log\frac{f(y|x,\ \alpha^*)}{g(y|x,\ \beta^*)}\right] = 0$,

Theorem 5.2 of Vuong (1985) shows that $n^{-1/2}LR_n(\hat{\alpha}_n,\ \hat{\beta}_n)/\hat{\omega}_n \xrightarrow{D} N(0,1)$ where $\hat{\omega}_n^2$ is a consistent estimator of ω_*^2. This consistent estimator will be defined shortly. As should be clear, if the models are non-nested, one must have $f(\cdot|\cdot,\alpha^*) \neq g(\cdot|\cdot,\beta^*)$; alternatively, when the models are either nested or overlapping, one may also have $f(\cdot|\cdot,\alpha^*) = g(\cdot|\cdot,\beta^*)$.

As will be shown in the next section, the Nash model and the two Stackelberg models are all pairwise overlapping. In a pairwise test of the three proposed models, it is therefore necessary to test whether or not $f(\cdot|\cdot,\alpha^*) = g(\cdot|\cdot,\beta^*)$. As shown in Lemma 4.1 of Vuong (1985) a necessary and sufficient condition for $f(\cdot|\cdot,\alpha^*) = g(\cdot|\cdot,\beta^*)$ is $\omega_*^2 = 0$, ω_*^2 denoting the variance of $\log[f(y|x,\alpha^*)/g(y|x,\beta^*)]$. Since

$$\omega_*^2 = \mathrm{var}^0\left[\log \frac{f(y|x,\alpha^*)}{g(y|x,\beta^*)}\right]$$

$$= E^0\left[\log \frac{f(y|x,\alpha^*)}{g(y|x,\beta^*)}\right]^2 - \left[E^0\left[\log \frac{f(y|x,\alpha^*)}{g(y|x,\beta^*)}\right]\right]^2, \text{ the}$$

expectation taken with respect to the true joint distribution H^0 of (y,x), a consistent estimator, $\hat{\omega}_n^2$, of ω_*^2 is given by the sample analog:

$$\hat{\omega}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\left[\log \frac{f(y_t|x_t,\hat{\alpha}_n)}{g(y_t|x_t,\hat{\beta}_n)}\right]^2 - \left[\frac{1}{n}\sum_{t=1}^{n}\log \frac{f(y_t|x_t,\hat{\alpha}_n)}{g(y_t|x_t,\hat{\beta}_n)}\right]^2.$$

As mentioned above, the asymptotic distribution of $\hat{\omega}_n$ is seen to be a weighted sum of independent chi-square random variables, the weights being the eigenvalues of the matrix product $V\sum$ where

$$V = \left[\begin{array}{cc} B_f(\alpha^*) & -B_{fg}(\alpha^*,\beta^*) \\ \hline -B_{gf}(\beta^*,\alpha^*) & B_g(\beta^*) \end{array}\right]$$

and

$$\sum = \left[\begin{array}{c c c} A_f^{-1}(\alpha^*)B_f(\alpha^*)A_f^{-1}(\alpha^*) & ; & A_f^{-1}(\alpha^*)B_{fg}(\alpha^*,\beta^*)A_g^{-1}(\beta^*) \\ \hline A_g^{-1}(\alpha^*)B_{gf}(\alpha^*,\beta^*)A_g^{-1}(\beta^*) & ; & A_g^{-1}(\beta^*)B_g(\beta^*)A_g^{-1}(\beta^*) \end{array}\right].$$

Submatrices within V and $\sum$ are defined as, for example

$$A_f(\alpha) \equiv E^0\left[\frac{\partial^2 \log f(y|x,\alpha)}{\partial\alpha\partial\alpha'}\right]$$

and

$$B_f(\alpha) \equiv E^0\left[\frac{\partial \log f(y|x,\alpha)}{\partial\alpha} \cdot \frac{\partial \log f(y|x,\alpha)}{\partial\alpha'}\right].$$

Consistent estimates of the matrices V and $\sum$ can be given by their syntax error file chap5i, between lines 919 and 919 sample analogs, $\hat{V}$ and $\sum$.

With this in mind, Vuong (1985) proposes a two-step sequential procedure for choosing the better model of two competing but overlapping models. If we let $H_0^\omega : \omega_*^2 = 0$ and $H_A^\omega : \omega_*^2 \neq 0$, the sequential procedure is

(i) Test H_0^ω against H_A^ω using the test based on $\hat{\omega}_n^2$. This is termed the variance test by Vuong (1985). If H_0^ω cannot be rejected, then conclude that the models F_α and G_β cannot be discriminated given the data. Alternatively, if H_0^ω is rejected, then

(ii) Test H_0 against H_f or H_g using the model selection test based on the statistic $n^{-1/2}LR_n(\hat{\alpha}_n,\hat{\beta}_n)$ which has a limiting standard normal distribution.

In the next section we then show that there exists a set of restrictions on the parameters of the Nash and the two Stackelberg models such that the three models pairwise overlap. This being the case, we would normally then implement the above sequential procedure. Since the first step of this sequential procedure is difficult to implement, we will instead use an alternative procedure. This alternative procedure is based on whether we can reject the hypothesis $H_0^{\omega}:\omega_*^2 = 0$ or equivalently $f(\cdot|\cdot,\alpha^*) = g(\cdot|\cdot,\beta^*)$ if we can statistically show that the estimated conditional distributions for the competing but overlapping models are not pairwise identical. To do this, we need to characterize the intersection of any two overlapping models. We therefore propose in the next section a necessary and sufficient condition for the models to be pairwise identical. If we can then reject this condition based upon the estimated distributions of Chapter IV, we need then only implement the second step of the above sequential procedure.

5. IMPLEMENTING THE TEST FOR THE CHOICE OF THE MOST ADEQUATE MODEL

As stated in the previous section, we will now present necessary and sufficient conditions for the three game theoretic models to be pairwise identical. Let us first start with the conditions that relate the two Stackelberg models, those being the models in which the husband and the wife successively play the role of the leader. The conditional distribution for the Stackelberg game in which the husband plays the role of the leader is given as Proposition

1 of Chapter III. In order to avoid confusion, we now state explicitly the conditional distribution of the Stackelberg game in which the wife plays the leading role. The probabilities relating the four possible discrete outcomes are as follows:

$$\tilde{P}r(0,0) = F(-\Delta\tilde{U}_w,-\Delta\tilde{U}_h,\tilde{\rho}) - \tilde{I}_+^B \qquad \text{if } \Delta\tilde{a}_h \geq 0$$

$$F(-\Delta\tilde{U}_w,-\Delta\tilde{U}_h,\tilde{\rho}) \qquad \text{otherwise}$$

$$\tilde{P}r(1,0) = F(\Delta\tilde{U}_w,-\Delta\tilde{U}_h-\Delta\tilde{a}_h,-\tilde{\rho}) \qquad \text{if } \Delta\tilde{a}_h \geq 0$$

$$F(\Delta\tilde{U}_w,-\Delta\tilde{U}_h-\Delta\tilde{a}_h,-\tilde{\rho}) + \tilde{I}_-^B \qquad \text{otherwise}$$

$$\tilde{P}r(0,1) = F(-\Delta\tilde{U}_w-\tilde{a}_w^1+\tilde{a}_w^0,\Delta\tilde{U}_h,-\tilde{\rho}) \qquad \text{if } \Delta\tilde{a}_h \geq 0$$

$$F(-\Delta\tilde{U}_w-\tilde{a}_w^1+\tilde{a}_w^0,\Delta\tilde{U}_h,-\tilde{\rho}) + \tilde{I}_-^A \qquad \text{otherwise}$$

$$\tilde{P}r(1,1) = F(\Delta\tilde{U}_w+\tilde{a}_w^1-\tilde{a}_w^0,\Delta\tilde{U}_h+\Delta\tilde{a}_h,\tilde{\rho}) - \tilde{I}_+^A \qquad \text{if } \Delta\tilde{a}_h \geq 0$$

$$F(\Delta\tilde{U}_w+\tilde{a}_w^1-\tilde{a}_w^0,\Delta\tilde{U}_h+\Delta\tilde{a}_h,\tilde{\rho}) \qquad \text{otherwise}$$

where

$$\tilde{I}_+^A = I(-\Delta\tilde{U}_w-\tilde{a}_w^1,-\Delta\tilde{U}_h,-\Delta\tilde{U}_w-\tilde{a}_w^1+\tilde{a}_w^0,-\Delta\tilde{U}_h-\Delta\tilde{a}_h,\tilde{\rho})$$

$$\tilde{I}_+^B = I(-\Delta\tilde{U}_w, -\Delta\tilde{U}_h, -\Delta\tilde{U}_w - \tilde{\alpha}_w^1, -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h, \tilde{\rho})$$

$$\tilde{I}_-^A = I(-\Delta\tilde{U}_w + \tilde{\alpha}_w^0, \Delta\tilde{U}_h - \Delta\tilde{\alpha}_h, -\Delta\tilde{U}_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0, -\Delta\tilde{U}_h, \tilde{\rho})$$

$$\tilde{I}_-^B = I(-\Delta\tilde{U}_w, -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h, -\Delta\tilde{U}_w + \tilde{\alpha}_w^0, -\Delta\tilde{U}_h, \tilde{\rho})$$

As in Chapter III, let us decompose $\Delta\tilde{U}_h$ and $\Delta\tilde{U}_w$ into linear combinations of coefficients and explanatory variables. We then have

$$\Delta\tilde{U}_h = \tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z$$

and

$$\Delta\tilde{U}_w = \tilde{\gamma}_w X_w + \tilde{\delta}_w^0 + \tilde{\delta}_w^1 Z$$

where Z contains those explanatory variables that are common to both the husband and the wife while X_h and X_w contain those variables that are specific to the husband and wife, respectively. Without loss of generality, let us also assume that X_h, X_w and Z are one-dimensional. We first present the following Lemma.

LEMMA 1: For every a and b, $I(a+x,b+y,a,b,\rho) = I(a+\tilde{x},b+\tilde{y},a,b,\rho)$ if and only if $x=\tilde{x}$ and $y=\tilde{y}$, provided $\widetilde{xy} \neq 0$.

PROOF:[3] (<==) Obvious.

(==>) Define $G(a,b;x,y,\tilde{x},\tilde{y}) = I(a+x,b+y,a,b) - I(a+\tilde{x},b+\tilde{y},a,b)$

$$= \int_a^{a+x}\int_b^{b+y} f(u,v;\rho)\,dv\,du - \int_a^{a+\tilde{x}}\int_b^{b+\tilde{y}} \Phi(u,v;\rho)\,dv\,du.$$

Since $G(a,b;x,y,\tilde{x},\tilde{y})$ is identically zero for all a and b, we have

$$\frac{\partial G(a,b;x,y,\tilde{x},\tilde{y})}{\partial a} = \int_b^{b+y} [f(a+x,v;\rho) - f(a,v;\rho)]dv$$

$$- \int_b^{b+\tilde{y}} [f(a+\tilde{x},v;\rho) - f(a,v;\rho)]dv = 0 \qquad \forall a,b.$$

Also, $\dfrac{\partial^2 G(a,b;x,y,\tilde{x},\tilde{y})}{\partial a\partial b} = 0$, $\forall a,b$ implies that

$$f(a+x,b+y;\rho) - f(a+x,b;\rho) - f(a,b+y;\rho)$$

$$= f(a+\tilde{x},b+\tilde{y};\rho) - f(a+\tilde{x},b;\rho) - f(a,b+\tilde{y};\rho), \qquad \forall a,b. \tag{1}$$

Now let

$$f(a\ \tilde{x},b\ \tilde{y};\rho) = f(a+x,b+y;\rho) + (\varepsilon_1+\varepsilon_2)$$

$$\text{and } f(a+\tilde{x},b;\rho) = f(a+x,b;\rho) + \varepsilon_1 . \tag{2}$$

Then $f(a,b+\tilde{y};\rho) = f(a,b+y;\rho) + \varepsilon_2$.

Moreover, $\dfrac{\partial^3 G(a,b;x,y,\tilde{x},\tilde{y})}{\partial a^2\partial b} = 0$ implies that

$$[(a+x)-\rho(b+y)]f(a+x,b+y;\rho)-[(a+x)-\rho b]f(a+x,b;\rho)-[a-\rho(b+y)]f(a,b+y;\rho)$$

$$=[(a+\tilde{x})-\rho(b+\tilde{y})][f(a+x,b+y;\rho)+\varepsilon_1+\varepsilon_2]-[(a+\tilde{x})-\rho b][f(a+x,b;\rho)+\varepsilon_1]$$

$$-[a-\rho(b+\tilde{y})][f(a,b+y;\rho)+\varepsilon_2]\ . \tag{3}$$

Upon simplification, we get

$$-\rho\tilde{y}\varepsilon_1+\tilde{x}\varepsilon_2=[x-\tilde{x}-\rho(y-\tilde{y})]f(a+x,b+y;\rho)-[x-\tilde{x}]f(a+x,b;\rho)+\rho(y-\tilde{y})f(a,b+y;\rho)\ .$$

Similarly, $\dfrac{\partial^3 G(a,b;x,y,\tilde{x},\tilde{y})}{\partial a\partial b^2}=0$ implies that

$$[(b+y)-\rho(a+x)]f(a+x,b+y;\rho)-[b-\rho(a+x)f(a+x,b;\rho)-[b+y-\rho a]f(a,b+y;\rho)$$

$$=[(b+\tilde{y})-\rho(a+\tilde{x})][f(a+x,b+y;\rho)+\varepsilon_1+\varepsilon_2]-[b-\rho(a+\tilde{x})][f(a+x,b;\rho)+\varepsilon_1]$$

$$-[b+\tilde{y}-\rho a][f(a,b+y;\rho)+\varepsilon_2]\ .$$

Again, upon simplification we get

$$\tilde{y}\varepsilon_1-\rho\tilde{x}\varepsilon_2=[y-\tilde{y}-\rho(x-\tilde{x})]f(a+x,b+y;\rho)+\rho[x-\tilde{x}]f(a+x,b;\rho)-(y-\tilde{y})f(a,b+y;\rho) \tag{4}$$

Combining equations (3) and (4), we have

$$\begin{bmatrix}-\rho\tilde{y} & \tilde{x}\\ \tilde{y} & -\rho\tilde{x}\end{bmatrix}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix}$$

$$=\begin{bmatrix}(x-\tilde{x}-\rho(y-\tilde{y}))f(a+x,b+y;\rho)-(x-\tilde{x})f(a+x,b;\rho)+\rho(y-\tilde{y})f(a,b+y;\rho)\\ (y-\tilde{y}-\rho(x-\tilde{x}))f(a+x,b+y;\rho)+\rho(x-\tilde{x})f(a+x,b;\rho)-(y-\tilde{y})f(a,b+y;\rho)\end{bmatrix}.$$

Therefore, since $\widetilde{xy} \neq 0$ and $|\rho| \neq 1$, we have

$$\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix} = \frac{1}{(\rho^2 - 1)\widetilde{xy}} \begin{bmatrix}-\rho\tilde{x} & -\tilde{x}\\ -\tilde{y} & -\rho\tilde{y}\end{bmatrix} \times$$

$$\begin{bmatrix}(x-\tilde{x})[f(a+x,b+y;\rho)-f(a+x,b;\rho)]-\rho(y-\tilde{y})[f(a+x,b+y;\rho)-f(a,b+y;\rho)]\\ -\rho(x-\tilde{x})[f(a+x,b+y;\rho)-\Phi(a+x,b;\rho)]+(y-\tilde{y})[f(a+x,b+y;\rho)-f(a,b+y;\rho)]\end{bmatrix}$$

$$= \frac{1}{(\rho^2-1)\widetilde{xy}} \begin{bmatrix}(\rho^2-1)\tilde{x}(y-\tilde{y})[f(a+x,b+y;\rho)-f(a,b+y;\rho)]\\ (\rho^2-1)\tilde{y}(x-\tilde{x})[f(a+x,b+y;\rho)-f(a+x,b;\rho)]\end{bmatrix}$$

$$= \begin{bmatrix}\dfrac{y-\tilde{y}}{\tilde{y}}[f(a+x,b+y;\rho)-f(a,b+y;\rho)]\\ \\ \dfrac{x-\tilde{x}}{\tilde{x}}[f(a+x,b+y;\rho)-f(a+x,b;\rho)]\end{bmatrix}. \tag{5}$$

Now, we also have $\dfrac{\partial^4 G(a,b;x,y,\tilde{x},\tilde{y})}{\partial a^3 \partial b} = 0$

which implies that

$$[(a+x)-\rho(b+y)]^2 f(a+x,b+y;\rho)-[(a+x)-\rho b]^2 f(a+x,b;\rho)$$
$$-[a-\rho(b+y)]^2 f(a,b+y;\rho)$$

$$= [(a+\tilde{x})-\rho(b+\tilde{y})]^2[f(a+x,b+y;\rho)+\varepsilon_1+\varepsilon_2]-[(a+\tilde{x}-\rho b]^2[f(a+x,b;\rho)+\varepsilon_1]$$

$$-[a-\rho(b+\tilde{y})]^2[f(a,b+y;\rho)+\varepsilon_2].$$

After rearranging, we have

$$\varepsilon_1(\rho^2\tilde{y}^2 - 2\rho\tilde{y}(a+\tilde{x}-\rho b)) + \varepsilon_2(\tilde{x}^2 + 2\tilde{x}(a-\rho b-\rho\tilde{y}))$$

$$= [x-\tilde{x}-\rho(y-\tilde{y})][2a+x+\tilde{x}-\rho(2b+y+\tilde{y})]f(a+x,b+y;\rho)$$

$$- [x-\tilde{x}][2a+x+\tilde{x}-2\rho b]f(a+x,b;\rho) + \rho(y-\tilde{y})(2a-2\rho b-\rho y-\rho\tilde{y})f(a,b+y;\rho).$$

Substituting (5) into the above equation, we have

$$\{(x-\tilde{x})[2a+x+\tilde{x}-2\rho b-\rho y-\rho\tilde{y}-\tilde{x}-2a+2\rho b+2\rho\tilde{y}]$$

$$+ \rho(y-\tilde{y})[-2a-x-\tilde{x}+2\rho b+\rho y+\rho\tilde{y}-\rho\tilde{y}+2a+2\tilde{x}-2\rho b]\}f(a+x,b+y;\rho)$$

$$=(x-\tilde{x})(-\tilde{x}-2a+2\rho b+2\rho\tilde{y}+2a+x+\tilde{x}-2\rho b)f(a+x,b;\rho)$$

$$+ \rho(y-\tilde{y})(-\rho\tilde{y}+2a+2\tilde{x}-2\rho b-2a+2\rho b+\rho y+\rho\tilde{y})f(a,b+y;\rho)$$

Upon simplification, we then have

$$(x-\tilde{x})(x+2\rho\tilde{y})f(a+x,b;\rho) + \rho(y-\tilde{y})(\rho y-2\tilde{x})f(a,b+y;\rho)$$

$$= [(x-\tilde{x})(x+\rho\tilde{y}-\rho y) + \rho(y-\tilde{y})(\rho y+\tilde{x}-x)]f(a+x,b+y;\rho), \quad \forall a,b. \tag{6}$$

Note that $f(\cdot,\cdot;\rho)$ is of the exponential form. Therefore (6) is possible only when all of the coefficients are zero. Note also that the coefficients are independent of a and b. We now have four cases to consider, depending on whether or not $x-\tilde{x}$ and $y-\tilde{y}$ are each zero.

Case 1: $x-\tilde{x} = 0$, $y-\tilde{y} = 0$. The proof is complete.

Case 2: $y-\tilde{y} = 0$, $x-\tilde{x} \neq 0$.

We have from (6) that $x+2\rho\tilde{y} = 0$ and $x+\rho\tilde{y}-\rho y = 0$ which implies $x = 0$ and $\tilde{y} = 0$ or $x = 0$ and $\rho = 0$. If $\tilde{y} = 0$, we have a contradiction since we have assumed that $\widetilde{xy} \neq 0$. If $\rho = 0$, $\tilde{y} \neq 0$, then

$I(a+x,b+y,a,b;\rho) = I(a+\tilde{x},b+\tilde{y},a,b;\rho)$ implies that

$$\int_a^{a+x} \varphi(u)du \int_b^{b+y} \varphi(v)dv = \int_a^{a+\tilde{x}} \varphi(u)du \int_b^{b+\tilde{y}} \varphi(v)dv.$$

Since $y = \tilde{y}$, we must have $x = \tilde{x}$. Contradiction.

Case 3: $x-\tilde{x} = 0$, $y-\tilde{y} \neq 0$.

From (6) we have that $2\tilde{x}+\rho y = 0$ and $\rho y+\tilde{x}-x = 0$ which implies $\rho y = 0$ and $\tilde{x}=0$ which is a contradiction since $\widetilde{xy} = 0$ by assumption.

Case 4: $x-\tilde{x} \neq 0$, $y-\tilde{y} \neq 0$. Then $x+2\rho\tilde{y} = 0$ and $\rho(\rho y+2\tilde{x}) = 0$. If $\rho = 0$, then $x = 0$ which implies $\tilde{x} = 0$ or $\tilde{y} = 0$. Contradiction. If $\rho \neq 0$, then $x+2\rho\tilde{y} = 0$ and $\rho y+2\tilde{x} = 0$ which implies $xy = 4\widetilde{xy}$. We also have from (6) that

$(x-\tilde{x})(x-\rho y+\rho\tilde{y}) + \rho(y-\tilde{y})(\rho y+\tilde{x}-x) = 0$ which implies

$(x-\tilde{x})(-\rho y-\rho\tilde{y}) + \rho(y-\tilde{y})(-x-\tilde{x}) = 0$ since $x = -2\rho\tilde{y}$ and $\rho y = -2\tilde{x}$.

We then have $-\rho[(x-\tilde{x})(y+\tilde{y}) + (y-\tilde{y})(x+\tilde{x})] = 0$ which implies

$-2\rho(xy - \widetilde{xy}) = 0$ which in turn implies $xy = \widetilde{xy}$ since we have assumed that $\rho \neq 0$.

Thus $\widetilde{\widetilde{xy}} = 4\widetilde{\widetilde{xy}}$ implies $\tilde{x} = 0$ or $\tilde{y} = 0$. Contradiction.

Q.E.D.

We are now in a position to prove the following necessary and sufficient condition such that the two Stackelberg models are identical.

PROPOSITION 1: Suppose that either $\gamma_h \neq 0$ and $\gamma_w \neq 0$ or $\tilde{\gamma}_h \neq 0$ and $\tilde{\gamma}_w \neq 0$. Then $Pr(i,j) = \widetilde{Pr}(j,i)$, $i,j = 0,1$, for any X_h, X_w, Z if and only if

$$\rho = \tilde{\rho},\ \gamma_h = \tilde{\gamma}_h,\ \gamma_w = \tilde{\gamma}_w;$$

$$\alpha_h^0 = \tilde{\alpha}_w^0 = 0,\ \alpha_h^1 = \tilde{\alpha}_w^1 = 0,\ \Delta\alpha_w = \Delta\tilde{\alpha}_h = 0;$$

$$\delta_w = \tilde{\delta}_w \text{ where } \delta_w = (\delta_w^0,\ \delta_w^1),\ \tilde{\delta}_w = (\tilde{\delta}_w^0,\ \tilde{\delta}_w^1); \text{ and}$$

$$\delta_h = \tilde{\delta}_h \text{ where } \delta_h = (\delta_h^0,\ \delta_h^1),\ \tilde{\delta}_h = (\tilde{\delta}_h^0,\ \tilde{\delta}_h^1).$$

PROOF: We have four cases to consider, depending upon the signs of $\Delta\tilde{\alpha}_w$ and $\Delta\tilde{\alpha}_h$.

Case 1: $\Delta\alpha_w \geq 0$, $\Delta\tilde{\alpha}_h \geq 0$

($\Leftarrow$) Obvious.

($\Rightarrow$) Without loss of generality, let us suppose $\gamma_h \neq 0$ and $\gamma_w \neq 0$. We must have $Pr(1,0) = \widetilde{Pr}(0,1)$ which gives

$$F(\Delta U_h,\ -\Delta U_w - \Delta\alpha_w,\ -\rho) = F(-\Delta\tilde{U}_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0,\ \Delta\tilde{U}_h,\ -\tilde{\rho}) \quad \text{or}$$

$$F(\gamma_h X_h + \delta_h^0 + \delta_h^1 Z, -\gamma_w X_w - \delta_w^0 - \delta_w^1 Z - \Delta\alpha_w, -\rho) =$$

$$F(-\tilde{\gamma}_w X_w - \tilde{\delta}_w^0 - \tilde{\delta}_w^1 Z - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0, \tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z, -\tilde{\rho})$$

First note that we must have $\tilde{\gamma}_h \neq 0$ and $\tilde{\gamma}_w \neq 0$. Otherwise, the right hand side would be independent of X_h and X_w, which contradicts the assumption that $\gamma_h \neq 0$ and $\gamma_w \neq 0$. Moreover, note that γ_h and $\tilde{\gamma}_h$ must have the same sign; otherwise, one side would be decreasing in X_h while the other side would be increasing in X_h. A similar requirment holds for γ_w and $\tilde{\gamma}_w$. Without loss of generality, let us suppose that $\gamma_h > 0$, $\tilde{\gamma}_h > 0$, $\gamma_w > 0$, $\tilde{\gamma}_w > 0$. Since the results must hold for all X_h, let $X_h \to +\infty$. This implies

$$\Phi(-\gamma_w X_w - \delta_w^0 - \delta_w^1 Z - \Delta\alpha_w) = \Phi(-\tilde{\gamma}_w X_w - \tilde{\delta}_w^0 - \tilde{\delta}_w^1 Z - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0) \quad \text{or}$$

$$\gamma_w X_w + \delta_w^1 Z + (\delta_w^0 + \Delta\alpha_w) = \tilde{\gamma}_w X_w + \tilde{\delta}_w^1 Z + (\tilde{\delta}_w^0 + \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0) \text{ for any } X_w, Z.$$

Hence, we must have

$$\gamma_w = \tilde{\gamma}_w, \tag{7}$$

$$\delta_w^1 = \tilde{\delta}_w^1, \text{ and} \tag{8}$$

$$\delta_w^0 + \Delta\alpha_w = \tilde{\delta}_w^0 + \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0. \tag{9}$$

Similarly, let $X_w \to -\infty$. This implies

$$\Phi(\gamma_h X_h + \delta_h^0 + \delta_h^1 Z) = \Phi(\tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z) \quad \text{or}$$

$$\gamma_h X_h + \delta_h^0 + \delta_h^1 Z = \tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z \text{ for any } X_h,\ Z.$$

This in turn implies that

$$\gamma_h = \tilde{\gamma}_h, \tag{10}$$

$$\delta_h^0 = \tilde{\delta}_h^0 \text{ and } \delta_h^1 = \tilde{\delta}_h^1, \text{ or, for notational convenience, } \delta_h = \tilde{\delta}_h. \tag{11}$$

In addition, we must have $Pr(0,1) = \tilde{P}r(1,0)$ or

$$F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\ \Delta U_w,\ -\rho) = F(\Delta\tilde{U}_w,\ -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h,\ -\tilde{\rho}) \text{ or} \tag{12}$$

$$F(-\gamma_h X_h - \delta_h^1 Z - (\delta_h^0 + \alpha_h^1 - \alpha_h^0),\ \gamma_w X_w + \delta_w^1 Z + \delta_w^0,\ -\rho)$$

$$= F(\tilde{\gamma}_w X_w + \tilde{\delta}_w^1 Z + \tilde{\delta}_w^0,\ -\tilde{\gamma}_h X_h - \tilde{\delta}_h^1 Z - (\tilde{\delta}_h^0 + \Delta\tilde{\alpha}_h),\ -\tilde{\rho}).$$

By the same argument as above, if we let $X_h \to -\infty$ we obtain

$$\gamma_w = \tilde{\gamma}_w, \tag{13}$$

$$\delta_w^0 = \tilde{\delta}_w^0 \text{ and } \delta_w^1 = \tilde{\delta}_w^1 \text{ which gives } \delta_w = \tilde{\delta}_w. \tag{14}$$

In addition, we get from (9) that

$$\Delta\alpha_w = \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0. \tag{15}$$

Similarly, letting $X_w \to +\infty$, we obtain

$$\gamma_h = \tilde{\gamma}_h, \tag{16}$$

$$\delta_h^1 = \tilde{\delta}_h^1 \quad \text{and} \tag{17}$$

$$\delta_h^0 + \alpha_h^1 - \alpha_h^0 = \tilde{\delta}_h^0 + \Delta\tilde{\alpha}_h. \tag{18}$$

But since we have from (11) that $\delta_h^0 = \tilde{\delta}_h^0$ equation (18) implies

$$\alpha_h^1 - \alpha_h^0 = \Delta\tilde{\alpha}_h. \tag{19}$$

Now note that equations (16) and (11) imply that $\Delta U_h = \Delta\tilde{U}_h$ while equations (13) and (14) imply $\Delta U_w = \Delta\tilde{U}_w$.

If we now use the fact that $F(x,y,\rho) = F(y,x,\rho)$, equation (12) can be rewritten as

$F(-\Delta U_h - \alpha_h^1 + \alpha_h^0, \Delta U_w, -\rho) = F(-\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h, \Delta\tilde{U}_w, -\tilde{\rho})$. But

$\Delta U_h = \Delta\tilde{U}_h$, $\Delta U_w = \Delta\tilde{U}_w$ and $\alpha_h^1 - \alpha_h^0 = \Delta\tilde{\alpha}_h$ from (19).

Moreover, the preceding equation must hold for all X_h, X_w, Z, and hence for all ΔU_h and ΔU_w. This implies $\rho = \tilde{\rho}$.

Third, setting $Pr(1,1) = \tilde{P}r(1,1)$ we get

$$F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \Delta U_w + \Delta\alpha_w, \rho) - I_+^A$$

$$= F(\Delta\tilde{U}_w + \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0, \Delta\tilde{U}_h + \Delta\tilde{\alpha}_h, \tilde{\rho}) - \tilde{I}_+^A.$$

But since we have from above that

$$\Delta U_h = \Delta\tilde{U}_h,\ \Delta U_w = \Delta\tilde{U}_w,\ \alpha_h^1 - \alpha_h^0 = \Delta\tilde{\alpha}_h,\ \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0 = \Delta\alpha_w,\ \text{and}\ \rho=\tilde{\rho}$$

we must have $I_+^A = \tilde{I}_+^A$ other words,

$I(-\Delta U_h - \alpha_h^1,\ -\Delta U_w,\ -\Delta U_h - \alpha_h^1 + \alpha_h^0,\ -\Delta U_w - \Delta\alpha_w,\ \rho)$

$= I(-\Delta\tilde{U}_w - \tilde{\alpha}_w^1,\ -\Delta\tilde{U}_h,\ \Delta\tilde{U}_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0,\ -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h,\ \tilde{\rho})$

$= I(-\Delta\tilde{U}_h,\ -\Delta\tilde{U}_w - \tilde{\alpha}_w^1,\ -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h,\ -\Delta\tilde{U}_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_h^0,\ \tilde{\rho})$

This last inequality comes about by switching the order of integration. Noting again that

$\Delta U_h = \Delta\tilde{U}_h,\ \Delta U_w = \Delta\tilde{U}_w,\ \rho = \tilde{\rho},\ \alpha_h^1 - \alpha_h^o, = \Delta\tilde{\alpha}_h$ and $\Delta\alpha_w = \tilde{\alpha}_w^1 - \alpha_w^0$, we have

$I(-\Delta U_h - \alpha_h^1,\ -\Delta U_w,\ -\Delta U_h -(\alpha_h^1 - \alpha_h^o),\ -\Delta U_w - \Delta\alpha_w,\ \rho)$

$= I(-\Delta U_h,\ -\Delta U_w - \tilde{\alpha}_w^1,\ -\Delta U_h - (\alpha_h^1 - \alpha_h^0),\ - \Delta U_w - \Delta\alpha_w,\ \rho)$,

which must hold for any ΔU_h and ΔU_w. Employing Lemma 1 on the above equation, we get $\alpha_h^1 = 0$ and $\tilde{\alpha}_w^1 = 0$. Finally, setting $Pr(0,0) = \tilde{P}r(0,0)$, we get $F(-\Delta U_h, -\Delta U_w, \rho) - I_+^B = F(-\Delta\tilde{U}_w, -\Delta\tilde{U}_h, \tilde{\rho}) - \tilde{I}_+^B$.

Using the same argument as above, we get $I_+^B = \tilde{I}_+^B$. In other words,

$I(-\Delta U_h,\ -\Delta U_w,\ -\Delta U_h - \alpha_h^1,\ -\Delta U_w - \Delta\alpha_w,\ \rho)$

$= I(-\Delta U_h,\ -\Delta U_w,\ -\Delta U_h - \Delta\tilde{\alpha}_h,\ -\Delta U_w - \tilde{\alpha}_w^1,\ \rho)$. Employing Lemma 1 again, we get $\alpha_h^1 = \Delta\tilde{\alpha}_h$ and $\Delta\alpha_w = \tilde{\alpha}_w^1$. But we have from above that $\alpha_h^1 = 0$ and

$\tilde{a}_w^1 = 0$. Therefore, $\Delta\tilde{a}_h = 0$ and $\Delta a_w = 0$. Now, using equations (15) and (19) respectively , we get $\tilde{a}_w^0 = 0$ and $a_h^0 = 0$.

Case 2: $\Delta a_w \geq 0$ and $\Delta\tilde{a}_h < 0$.

Without loss of generality, let us assume $\gamma_h \neq 0$ and $\gamma_w \neq 0$. If we first set $Pr(0,0) = \tilde{P}r(0,0)$, we get

$F(-\Delta U_h, -\Delta U_w, \rho) - I_+^B = F(-\Delta\tilde{U}_w, -\Delta\tilde{U}_h, \tilde{\rho})$, or upon substitution

$$F(-\gamma_h X_h - \delta_h^0 - \delta_h^1 Z, -\gamma_w X_w - \delta_w^0 - \delta_w^1 Z, \rho) -$$

$$I(-\gamma_h X_h - \delta_h^0 - {}_h^1 Z, - \gamma_w X_w - \delta_w^0 - \delta_w^1 Z, - \gamma_h X_h - \delta_h^0 - \delta_h^1 Z - a_h^1,$$

$$-\gamma_w X_w - \delta_w^0 - \delta_w^1 Z - \Delta a_w, \rho)$$

$$= F(-\tilde{\gamma}_w X_w - \tilde{\delta}_w^0 - \tilde{\delta}_w^1 Z, -\tilde{\gamma}_h X_h - \tilde{\delta}_h^0 - \tilde{\delta}_h^1 Z, \tilde{\rho}). \tag{20}$$

Since we assume that $\gamma_h \neq 0$ and $\gamma_w \neq 0$, it must also be the case that $\tilde{\gamma}_h \neq 0$ and $\tilde{\gamma}_w \neq 0$; otherwise, the right hand side of the above equation would not depend on X_h and X_w while the left hand side would. In addition, γ_h must have the same sign as $\tilde{\gamma}_h$; similarly γ_w must have the same sign as $\tilde{\gamma}_w$. Without loss of generality, let us assume $\gamma_w > 0$, $\gamma_w > 0$, $\tilde{\gamma}_h > 0$, $\tilde{\gamma}_w > 0$.

Now let $X_h \to -\infty$. We see then that $I_+^B \to 0$. Equation (14) then becomes

$$\Phi(-\gamma_w X_w - \delta_w^0 - \delta_w^1 Z) = \Phi(-\tilde{\gamma}_w X_w - \tilde{\delta}_w^0 - \tilde{\delta}_w^1 Z) \text{ or}$$

$$\gamma_w X_w + \delta_w^0 + \delta_w^1 Z = \tilde{\gamma}_w X_w + \tilde{\delta}_w^0 + \tilde{\delta}_w^1 Z \quad \forall \ X_w, \ Z.$$

We then have

$$\gamma_w = \tilde{\gamma}_w, \tag{21}$$

$$\delta_w^0 = \tilde{\delta}_w^0 \text{ and } \delta_w^1 = \tilde{\delta}_w^1 \text{ which imply } \delta_w = \tilde{\delta}_w \tag{22}$$

Similarly, if we let $X_w \rightarrow -\infty$ we get

$$\Phi(-\gamma_h X_h - \delta_h^0 - \delta_h^1 Z) = \Phi(-\tilde{\gamma}_h X_h - \tilde{\delta}_h^0 - \tilde{\delta}_h^1 Z) \text{ which implies}$$

$$\gamma_h X_h + \delta_h^0 + \gamma_h^1 Z = \tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z \quad \forall \ X_h, \ Z. \text{ We then have}$$

$$\gamma_h = \gamma_h, \tag{23}$$

$$\delta_h^0 = \tilde{\gamma}_h^0 \text{ and } \delta_h^1 = \tilde{\delta}_h^1 \text{ or } \delta_h = \tilde{\delta}_h. \tag{24}$$

Note also that from (21) and (22) we get $\Delta U_w = \Delta\tilde{U}_w$; from (23) and (24) we get $\Delta U_h = \Delta\tilde{U}_h$. let us next look at the conditions under which $Pr(1,1) = \tilde{P}r(1,1)$. This equality implies

$$F(\Delta U_h + \alpha_h^1 - \alpha_h^0, \ \Delta U_w + \Delta\alpha_w, \ \rho) - I_+^A = F(\Delta\tilde{U}_w = \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0, \ \Delta\tilde{U}_h + \Delta\tilde{\alpha}_h, \ \tilde{\rho})$$

which in turn implies

$$F(\gamma_h X_h + \delta_h^0 + \delta_h^1 Z + \alpha_h^1 - \alpha_h^0,\ \gamma_w X_w + \delta_w^0 + \delta_w^1 Z + \Delta\alpha_w,\ \rho) -$$
$$I(-\gamma_h X_h - \delta_h^0 - \delta_h^1 Z - \alpha_h^1,\ -\gamma_w X_w - \delta_w^0 - \delta_w^1 Z,\ -\gamma_h X_h - \delta_h^0 - \delta_h^1 Z - \alpha_h^1 + \alpha_h^0,\ -\gamma_w X_w - \delta_w^0 - \delta_w^1 Z - \Delta\alpha_w,\ \rho)$$
$$= F(\tilde{\gamma}_w X_w + \tilde{\delta}_w^0 + \tilde{\delta}_w^1 Z + \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0,\ \tilde{\gamma}_h X_h + \tilde{\delta}_h^0 + \tilde{\delta}_h^1 Z + \Delta\tilde{\alpha}_h,\ \tilde{\rho}). \qquad (25)$$

Now let $X_h \to +\infty$. Since $I_+^A \to 0$, we get

$\Phi(\Delta U_w + \Delta\alpha_w) = \Phi(\Delta\tilde{U}_w + \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0)$. Since $\Delta U_w = \Delta\tilde{U}_w$, we get

$$\Delta\alpha_w = \tilde{\alpha}_w^1 - \tilde{\alpha}_w^0. \qquad (26)$$

Similarly, if we let $X_w \to +\infty$, we get

$$\Delta\tilde{\alpha}_h = \alpha_h^1 - \alpha_h^0. \qquad (27)$$

The conditions for which $Pr(1,0) = \tilde{P}r(0,1)$ are

$$F(\Delta U_h,\ -\Delta U_w - \Delta\alpha_w,\ -\rho) = F(-\Delta\tilde{U}_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0,\ \Delta\tilde{U}_h,\ -\tilde{\rho}) + \tilde{I}_-^A,$$

which yield upon substitution

$$F(\Delta U_h,\ -\Delta U_w - \Delta\alpha_w,\ -\rho) = F(-\Delta U_w - \Delta\alpha_w,\ \Delta U_h,\ -\tilde{\rho}) \qquad (28)$$
$$+ I(-\Delta U_w + \tilde{\alpha}_w^0,\ -\Delta U_h - \Delta\tilde{\alpha}_h,\ -\Delta U_w - \tilde{\alpha}_w^1 + \tilde{\alpha}_w^0,\ -\Delta U_h,\ \tilde{\rho}).$$

Finally, $Pr(0,1) = \tilde{P}r(1,0)$ gives

$$F(-\Delta U_h - \alpha_h^1 + \alpha_h^0,\ \Delta U_w,\ -\rho) = F(\Delta\tilde{U}_w,\ -\Delta\tilde{U}_h - \Delta\tilde{\alpha}_h,\ -\tilde{\rho}) + \tilde{I}_-^B,$$

or upon substitution

$$F(-\Delta U_h - \alpha_h^1 + \alpha_h^0, \Delta U_w, -\rho) = F(\Delta U_w, -\Delta U_h - \alpha_h^1 + \alpha_h^0, -\tilde{\rho}) \tag{29}$$

$$+ I(-\Delta U_w, -\Delta U_h - \Delta\tilde{\alpha}_h, -\Delta U_w + \tilde{\alpha}_w^0, -\Delta U_h, \tilde{\rho}).$$

We now need to show that $\rho = \tilde{\rho}$; to do this, we will use conditions (20), (25), (28) and (29) and the facts that $\Delta\alpha_w \geq 0$ and $\Delta\tilde{\alpha}_h \leq 0$. In addition, we will note the relation $\frac{\partial F(x,y,\rho)}{\partial \rho} = f(x,y,\rho) > 0$ where $f(x,y,\rho)$ is the bivariate normal p.d.f. Note that the signs of I_+^B and I_+^A from equations (20) and (25), respectively, will depend upon the signs of α_h^1 and α_h^0. This is so because we have assumed that $\Delta\alpha_w \geq 0$. Note also that the signs of I_-^A and I_-^B from equations (28) and (29), respectively, will not depend upon the individual signs of α_h^1 and α_h^0. If we can thus determine the signs of I_+^B and I_+^A, we can determine the relationship betweend ρ and $\tilde{\rho}$ since we know that $\partial F(x,y,\rho)/\partial\rho = f(x,y,\rho) > 0$. A similar argument holds when we vary the signs of $\tilde{\alpha}_w^1$ and $\tilde{\alpha}_w^0$. We thus have eight subcases to consider.

Subcase 1: If $\alpha_h^1 \geq 0$ and $\alpha_h^0 \geq 0$, then $I_+^B \geq 0$ which implies that $\rho \geq \tilde{\rho}$. On the other hand, $I_+^A \leq 0$ which gives $\rho \leq \tilde{\rho}$. Thus $\rho = \tilde{\rho}$.

Subcase 2: If $\alpha_h^1 \leq 0$ and $\alpha_h^0 \leq 0$, then $I_+^B \leq 0$ giving $\rho \leq \tilde{\rho}$. Similarly, $I_+^A \geq 0$ which gives $\rho \geq \tilde{\rho}$. Therefore, $\rho = \tilde{\rho}$.

Subcase 3: If $\alpha_h^1 \geq 0$ and $\alpha_h^0 \leq 0$, it must be the case that both are zero since we require $\Delta\tilde{\alpha}_h \leq 0$. As a result, $I_+^B = I_+^A = 0$

which gives $\rho = \tilde{\rho}$.

Subcase 4: If $\alpha_h^1 \leq 0$ and $\alpha_h^0 \geq 0$, then $I_+^B \leq 0$ which implies $\rho \leq \tilde{\rho}$. On the other hand, $I_+^A \geq 0$, giving us $\rho \geq \tilde{\rho}$. Combining these results, we get $\rho = \tilde{\rho}$.

Subcase 5: If $\tilde{\alpha}_w^1 \geq 0$ and $\tilde{\alpha}_w^0 \geq 0$, then $I_-^A \geq 0$ which implies that $-\rho \geq -\tilde{\rho}$ or $\rho \leq \tilde{\rho}$. Similarly, $I_-^B \leq 0$ which gives $\rho \geq \tilde{\rho}$. We thus have $\rho = \tilde{\rho}$.

Subcase 6: If $\tilde{\alpha}_w^1 \leq 0$ and $\tilde{\alpha}_w^0 \leq 0$, then $I_-^A \leq 0$ and $I_-^B \geq 0$. These two conditions require that $\rho = \tilde{\rho}$.

Subcase 7: If $\tilde{\alpha}_w^1 \geq 0$ and $\tilde{\alpha}_w^0 \leq 0$, then $I_-^A \leq 0$ and $I_-^B \geq 0$. We thus respectively require $\rho \geq \tilde{\rho}$ and $\rho \leq \tilde{\rho}$ which again gives $\rho = \tilde{\rho}$.

Subcase 8: If $\tilde{\alpha}_w^1 \leq 0$ and $\tilde{\alpha}_w^0 \geq 0$, it must be the case that both are zero since we require $\Delta\alpha_w \geq 0$. As a result, $I_-^A = I_-^B = 0$ which gives $\rho = \tilde{\rho}$.

Since $\rho = \tilde{\rho}$, we have from (20) and (25) that
$I(-\Delta U_h, -\Delta U_w, -\Delta U_h - \alpha_h^1, -\Delta U_w - \Delta\alpha_w, \rho) =$
$I(-\Delta U_h, -\Delta U_w, -\Delta U_h - \Delta\tilde{\alpha}_h, -\Delta U_w - \Delta\alpha_w, \rho)$. But from Lemma 1, we get

$\Delta\tilde{\alpha}_h = 0$. Similarly, we have from (28) and (29) that $I_-^A = I_-^B$. Applying Lemma 1 again, we get that $\Delta\alpha_w = 0$. Finally, we can equate I_+^B to I_-^A from equations (20) and (28), respectively, which give

$\alpha_h^1 = \Delta\tilde{\alpha}_h$ and $\tilde{\alpha}_w^0 = 0$. But since $\Delta\tilde{\alpha}_h = 0$ from above, we get that

$\alpha_h^1 = \alpha_h^0 = 0$. Moreover, since $\Delta\alpha_w = \tilde{\alpha}_w^0 = 0$, we get $\tilde{\alpha}_w^1 = 0$.

Case 3: $\Delta\alpha_w \leq 0$ and $\Delta\tilde{\alpha}_h \leq 0$. This case proceeds identically as Case 1.

Case 4: $\Delta\alpha_w \leq 0$ and $\Delta\tilde{\alpha}_h \geq 0$. This case proceeds identically as Case 2.

Q.E.D.

We now state necessary and sufficient conditions such that the Nash model and the two Stackelberg models are overlapping. Again, we need to state another Lemma.

LEMMA 2: For every a and b,

$F(a,b,\rho) - F(a+x,b+y,\rho) = F(a,b,\tilde{\rho}) - F(a+x,b+y,\tilde{\rho})$ if and only if $\rho = \tilde{\rho}$, provided $x \cdot y \neq 0$, $|\rho| \neq 1$ and $|\tilde{\rho}| \neq 1$.

PROOF[4]: ($\Longleftarrow$) Obvious

($\Longrightarrow$) Let us define

$$H(a,b;x,y,\rho,\tilde{\rho}) = F(a,b,\rho) - F(a + x,b + y,\rho)$$
$$-F(a,b,\tilde{\rho}) + F(a + x,b + y,\tilde{\rho}) = 0 \quad \forall\ a,b.$$

Now $\frac{\partial^2 H}{\partial a \partial b} = f(a,b,\rho) - f(a + x, b + y, \rho)$

$$-f(a,b,\tilde{\rho}) + f(a + x, b + y, \tilde{\rho}) = 0 \quad \forall\ a,b.$$

Then if $f(a,b,\tilde{\rho}) = f(a,b,\rho) + \varepsilon_1$, we also must have

$f(a + x, b + y, \tilde{\rho}) = f(a + x, b + y, \rho) + \varepsilon_1$.

Differentiating again, we get $\frac{\partial^3 H}{\partial a^2 \partial b} = 0$ which implies

$$\frac{a - \rho b}{1 - \rho^2} f(a,b,\rho) - \frac{a + x - \rho(b + y)}{1 - \rho^2} f(a + x, b + y, \rho)$$

$$= \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2} f(a,b,\tilde{\rho}) - \frac{a + x - \tilde{\rho}(b + y)}{1 - \tilde{\rho}^2} f(a + x, b + y, \tilde{\rho})$$

$$= \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2}[f(a,b,\rho)+\varepsilon_1] - \frac{a + x - \tilde{\rho}(b + y)}{1 - \tilde{\rho}^2}[f(a + x, b + y, \rho)+\varepsilon_1].$$

This last term comes about by substitution from above. If we rearrange terms, the above equation gives us

$$\frac{\tilde{\rho} y - x}{1 - \tilde{\rho}^2} \cdot \varepsilon_1 = \left[\frac{a - \rho b}{1 - \rho^2} - \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2}\right] f(a,b,\rho)$$

$$- \left[\frac{a + x - \rho(b + y)}{1 - \rho^2} - \frac{a + x - \tilde{\rho}(b + y)}{1 - \tilde{\rho}^2}\right] f(a + x, b + y, \rho). \qquad (30)$$

Moreover, $\frac{\partial^3 H}{\partial a \partial b^2} = 0$ yields

$$\frac{b - \rho a}{1 - \rho^2} f(a,b,\rho) - \frac{(b + y) - \rho(a + x)}{1 - \rho^2} f(a + x, b + y, \rho)$$

$$= \left[\frac{b - \tilde{\rho} a}{1 - \tilde{\rho}^2}\right] [f(a,b,\rho) + \varepsilon_1] - \frac{b + y - \tilde{\rho}(a + x)}{1 - \tilde{\rho}^2} [f(a + x, b + y, \rho) + \varepsilon_1].$$

If we rearrange terms, the above equation can be written as

$$\frac{\tilde{\rho} x - y}{1 - \tilde{\rho}^2} \cdot \varepsilon_1 = \left[\frac{b - \rho a}{1 - \rho 2} - \frac{b - \tilde{\rho} a}{1 - \tilde{\rho}^2}\right] f(a,b,\rho) \tag{31}$$

$$- \left[\frac{b + y - \rho(a + x)}{1 - \rho^2} - \frac{b + y - \tilde{\rho}(a + x)}{1 - \tilde{\rho}^2}\right] f(a + x, b + y, \rho).$$

Now, if we combine equations (30) and (31) and assume that $\tilde{\rho} y - x \neq 0$, we get

$$\frac{\tilde{\rho} x - y}{\tilde{\rho} y - x} \left[\frac{a - \rho b}{1 - \rho^2} - \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2}\right] f(a,b,\rho)$$

$$- \left[\frac{a + x - \rho(b + y)}{1 - \rho^2} - \frac{a + x - \tilde{\rho}(b + y)}{1 - \tilde{\rho}^2}\right] f(a + x, b + y, \rho)$$

$$= \left[\frac{b - \rho a}{1 - \rho^2} - \frac{b - \tilde{\rho} a}{1 - \tilde{\rho}^2}\right] f(a,b,\rho) -$$

$$\left[\frac{b + y - \rho(a + x)}{1 - \rho^2} - \frac{b + y - \tilde{\rho}(a + x}{1 - \tilde{\rho}^2)}\right] f(a + x, b + y, \rho)$$

But this is possible only when

$$\frac{\tilde{\rho} x - y}{\tilde{\rho} y - x} \left[\frac{a - \rho b}{1 - \rho^2} - \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2}\right] = \left[\frac{b - \rho a}{1 - \rho^2} - \frac{b - \tilde{\rho} a}{1 - \tilde{\rho}^2}\right] = 0, \quad \forall\ a,b.$$

Rewriting this last equality, $\frac{b - \rho a}{1 - \rho^2} - \frac{b - \tilde{\rho} a}{1 - \tilde{\rho}^2} = 0$, gives us

$$(\rho^2 - \tilde{\rho}^2)b - [\rho(1-\tilde{\rho}^2) - \tilde{\rho}(1 - \rho^2)]a = 0, \quad \forall\ a,b.$$

This in turn implies that

$\rho^2 = \tilde{\rho}^2$ and $\rho(1-\tilde{\rho}^2) = \tilde{\rho}(1-\rho^2)$. Now either $\rho = \tilde{\rho}$ and we are done or

$\rho = -\tilde{\rho}$ which implies

$\rho(1-\tilde{\rho}^2) = \tilde{\rho}(1-\tilde{\rho}^2)$. But this implies

$\tilde{\rho} = 0 = \rho$. Alternatively, if $\tilde{\rho} y - x = 0$, we have from (30) that

$$\left[\frac{a - \rho b}{1 - \rho^2} - \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2}\right] f(a,b,\rho) -$$

$$\left[\frac{a + x - \rho(b + y)}{1 - \rho^2} - \frac{a + x - \tilde{\rho}(b + y)}{1 - \tilde{\rho}^2}\right] f(a + x, b + y, \rho) = 0 , \quad \forall\ a,b.$$

But this implies that

$$\frac{a - \rho b}{1 - \rho^2} - \frac{a - \tilde{\rho} b}{1 - \tilde{\rho}^2} = 0 \quad \text{which in turn implies}$$

$(\rho^2 - \tilde{\rho}^2)a + b(\tilde{\rho}(1 - \rho^2) - \rho(1 - \tilde{\rho}^2)) = 0$, $\forall$ a, b. We therefore require

$\rho^2 = \tilde{\rho}^2$ and $\tilde{\rho}(1 - \rho^2) = \rho(1 - \tilde{\rho}^2)$. Now, either $\rho = \tilde{\rho}$ and we are done,

or $\rho = -\tilde{\rho}$ which implies

$\tilde{\rho}(1 - \tilde{\rho}^2) = -\tilde{\rho}(1 - \tilde{\rho}^2)$ which in turn gives $\tilde{\rho} = 0 = \rho$.

Q.E.D.

We are now in a position to provide necessary and sufficient conditions such that the Nash model and each of the two Stackelberg models are overlapping. Since the conditions under which the Nash model overlaps with the husband leader Stackelberg model are identical to the conditions under which the Nash model overlaps with the wife leader Stackelberg model, we will only prove the former so as to economize on space. Let those terms with a tilde superscript be associated with the husband leader Stackelberg model and those without a tilde be asasociated with the Nash model. We then have:

PROPOSITION 2: Suppose that $\beta_h \leq 0$, $\beta_w \geq 0$, and $\Delta\tilde{\alpha}_w \leq 0$. In addition, suppose that either $\gamma_h \neq 0$ and $\gamma_w \neq 0$ or $\tilde{\gamma}_h \neq 0$ and $\tilde{\gamma}_w \neq 0$.

Then $Pr(i,j) = \widetilde{Pr}(i,j)$, i, j = 0, 1, if and only if

$\gamma_h = \tilde{\gamma}_h$, $\delta_h = \tilde{\delta}_h$, $\gamma_w = \tilde{\gamma}_w$, $\delta_w = \tilde{\delta}_w$;

$\beta_w = \Delta\tilde{\alpha}_w$, $\beta_h = \tilde{\alpha}_h^1 - \tilde{\alpha}_h^0$, $\rho = \tilde{\rho}$, $\beta_h \cdot \beta_w = 0$; and

$\tilde{\alpha}_h^1 = \tilde{\alpha}_h^0 = 0$ or $\Delta\tilde{\alpha}_w = 0$.

PROOF: ($\Leftarrow$) Obvious.

($\Rightarrow$) Without loss of generality, assume that $\gamma_h \neq 0$ and $\gamma_w \neq 0$. Then, when we set $Pr(0,0) = \widetilde{Pr}(0,0)$, it is clear that $\tilde{\gamma}_w \neq 0$ and $\tilde{\gamma}_w \neq 0$; otherwise $Pr(0,0)$ would depend on X_h and X_w while $\widetilde{Pr}(0,0)$ would not. Moreover, γ_h and $\tilde{\gamma}_h$ must have the same sign. Similarly, γ_w and $\tilde{\gamma}_w$ must have the same sign. Without loss of generality, let us then set

$\gamma_h > 0$, $\gamma_w > 0$, $\tilde{\gamma}_h > 0$, $\tilde{\gamma}_w > 0$. Then from the equality $Pr(0,0) = \widetilde{Pr}(0,0)$, if we let $X_h \rightarrow -\infty$ we get $\Phi(-\Delta U_w) = \Phi(-\Delta\tilde{U}_w)$ which implies that $\Delta U_w = \Delta\tilde{U}_w$. Written out more fully, we get

$\gamma_w X_w + \delta_w Z = \tilde{\gamma}_w X_w + \tilde{\delta}_w Z \quad \forall X_w, Z$. This implies that

$$\gamma_w = \tilde{\gamma}_w \text{ and } \delta_w = \tilde{\delta}_w. \tag{32}$$

Similarly, if we let $X_w \rightarrow -\infty$, we have

$$\gamma_h = \tilde{\gamma}_h \text{ and } \delta_h = \tilde{\delta}_h. \tag{33}$$

Next set $Pr(1,1) = \widetilde{Pr}(1,1)$. If we let $X_h \rightarrow +\infty$ we get

$$\beta_h = \tilde{\alpha}_h^1 - \tilde{\alpha}_h^0. \tag{34}$$

Similarly, if we let $X_w \to +\infty$ we get

$$\beta_w = \Delta\tilde{\alpha}_w. \tag{35}$$

We next show by contradiction that $\beta_h \cdot \beta_w = 0$. Assume not; that is, assume $\beta_h \neq 0$ and $\beta_w \neq 0$. If we take the difference

$\Pr(0,0) - \Pr(1,1) = \widetilde{\Pr}(0,0) - \widetilde{\Pr}(1,1)$ we get

$F(-\Delta U_h, -\Delta U_w, \rho) - F(\Delta U_h + \beta_h, \Delta U_w + \beta_w, \rho)$

$= F(-\Delta\tilde{U}_h, -\Delta\tilde{U}_w, \tilde{\rho}) - F(\Delta\tilde{U}_h + \tilde{\alpha}_h^1 - \tilde{\alpha}_h^0, \Delta\tilde{U}_w + \Delta\tilde{\alpha}_w, \tilde{\rho})$.

From (32) we have that $\Delta U_w = \Delta\tilde{U}_w$; from (33) we have that $\Delta U_h = \Delta\tilde{U}_h$. Using (34) and (35) the above difference implies

$$F(-\Delta U_h, -\Delta U_w, \rho) - F(\Delta U_h + \beta_h, \Delta U_w + \beta_w, \rho)$$
$$= F(-\Delta U_h, -\Delta U_w, \rho) - F(\Delta U_h + \beta_h, \Delta U_w + \beta_w, \tilde{\rho}).$$

But from Lemma 2 we see that we must have $\rho = \tilde{\rho}$. Using again the equality $\Pr(0,0) = \widetilde{\Pr}(0,0)$ we have

$F(-\Delta U_h, -\Delta U_w, \rho) + d_1 \cdot I_{-+} = F(-\Delta\tilde{U}_h, -\Delta\tilde{U}_w, \tilde{\rho}) = F(-\Delta U_h, -\Delta U_w, \rho)$

which implies $I_{-+} = 0$. As a result, we must have $\beta_h \cdot \beta_w = 0$, which is a contradiction. Hence, we must have $\beta_h \cdot \beta_w = 0$. Thus $I_{-+} = 0$, and $\Pr(0,0) = \widetilde{\Pr}(0,0)$ gives $\rho = \tilde{\rho}$. It now remains to show that $\tilde{\alpha}_h^1 = \tilde{\alpha}_h^0 = 0$ or $\Delta\tilde{\alpha}_w = 0$. We use the previous result that $\beta_h \cdot \beta_w = 0$.

First, suppose that $\beta_w \neq 0$. Equation (35) then implies that $\Delta\tilde{\alpha}_w \neq 0$. Second, suppose that $\beta_h \neq 0$. Equation (34) implies that $\tilde{\alpha}_h^1 - \tilde{\alpha}_h^0 = 0$. Third, suppose that $\Delta\tilde{\alpha}_w = 0$; then we have what we wanted to show. Finally, suppose that $\Delta\tilde{\alpha}_w \neq 0$. The required equality $Pr(1,0) = \tilde{P}r(1,0)$ gives

$$F(\Delta U_h, -\Delta U_w - \beta_w, -\rho) + d_2 \cdot I_{-+} = F(\Delta\tilde{U}_h, -\Delta\tilde{U}_w - \Delta\tilde{\alpha}_w, -\tilde{\rho}) + I_-^B.$$

Since $\beta_h \cdot \beta_w = 0$, we see that $I_{-+} = 0$. Moreover, $I_-^B = 0$ since $\Delta U_h = \Delta\tilde{U}_h$, $\Delta U_w = \Delta\tilde{U}_w$, $\beta_w = \Delta\tilde{\alpha}_w$, and $\rho = \tilde{\rho}$. $I_-^B = 0$ implies that $\tilde{\alpha}_h^0 = 0$ since we assume $\Delta\tilde{\alpha}_w \neq 0$. But if $\alpha_h^0 = 0$ we also see that $\tilde{\alpha}_h^1 = 0$ by (34). The conditions under which $Pr(0,1) = \tilde{P}(0,1)$ have already been established.

Q.E.D.

It is now important to notice that the necessary and sufficient conditions for the three game theoretic models to be pairwise identical fail to hold. Looking at Tables 6 and 12 of Chapter IV along with the statement of Proposition 1, we see that the necessary conditions for the two Stackelberg models to be overlapping fail to hold when we examine the estimated coefficients given in these two tables. For example, we see from Table 6 that $\hat{\alpha}_h^0$ is significantly different from zero at the 90 percent level while $\Delta\hat{\alpha}_w$ differs from zero at the 95 percent level. Correspondingly, Table 12 indicates that both $\hat{\alpha}_w^0$ and $\Delta\hat{\alpha}_w$ are both significantly different from zero at the

95 percent level.

We also see from Tables 6 and 12 and the statement of Proposition 2 of this chapter that neither Stackelberg model is pairwise to the Nash model. First, Table 6 indicates that the husband-leader Stackelberg model is not identical to the Nash model because both $\hat{\alpha}_h^0$ and $\Delta\hat{\alpha}_w$ are significantly different from zero at the 90 and 95 percent levels, respectively. Second, we see from Table 12 that the Nash model and the wife-leader Stackelberg models are not identical because both $\hat{\alpha}_w^0$ and $\Delta\hat{\alpha}_h$ differ significantly from zero at the 95 percent level.

We are now in a position where we can appeal to Theorem 5.2 of Vuong (1985) which under

$$H_0:\ E^0\left[\log \frac{f(y|x;\ \alpha^*)}{g(y|x;\ \beta^*)}\right]$$

the statistic

$$\frac{n^{1/2}LR_n(\hat{\alpha},\hat{\beta})}{\hat{\omega}_n} = \frac{LR_n(\hat{\alpha}_n,\hat{\beta}_n)}{\sum_{t=1}^{n}\left[\log\frac{f(y_t|x_t;\ \hat{\alpha}_n)}{g(y_t|x_t;\ \hat{\beta})}\right]^2 - \frac{1}{n}[LR_n(\hat{\alpha}_n,\hat{\beta}_n)]^2\ \ ^{1/2}}$$

is asymptotically distributed as standard normal. The values of the three statistics are presented in the following table, where N stands for the Nash model and SH and SW stand for the husband leader Stackelberg model and wife leader Stackelberg, respectively.

Model G

Model F		N	SW
	SH	-.056	-1.24
	SW	.855	

As we see from the above table, we are unable at the commonly adhered to levels of confidence able to pairwise discriminate between any of the three game theoretic models. That is, we are required to accept the hypothesis the Nash model is as equally good (or bad) as the husband leader Stackelberg model for explaining the joint labor participation decision of a random sample of married couples. The other two hypotheses we must similarly accept.

6. CONCLUSION

In Chapters II and III respectively, we proposed econometric models of two different game theoretic equilibrium notions, those being Nash Equilibrium and Stackelberg Equilibrium. In Chapter IV, three different empirical models were proposed and estimated concerning the joint labor force participation decision of a married couple. The first model assumed that the husband and wife both played a Nash game. The second model assumed that the married couple played a Stackelberg game where the husband played the role of the leader and his wife played the role of the follower. Model three, while also a Stackelberg game, assumed that the roles of the two players were reversed; that is, it was assumed that the wife played the role of the leader while the husband played the role of the follower. The purpose of the present Chapter was to determine the most adequate model among the three proposed for explaining the joint labor force participation decision over a large random sample of married couples.

The situation encountered, however, when we attempted to examine any two of the three proposed game theoretic models was neither one of comparing two nested models nor of comparing two non-nested models. As was shown, the models are pairwise overlapping. As such the traditional methods developed for choosing the most adequate model were inappropriate for the task. We therefore relied on a new technique developed by Vuong (1985) which handles as separate cases those situations in which the models may be nested, non-nested or overlapping. The results of the previous section indicated that although it was not possible to reject the hypotheses that any model was pairwise better than the remaining models in explaining the joint labor force participation decision for a random sample of married couples, the log-likelihood values may suggest that the Stackelberg model in which the wife plays the role of the leader may be the best. An alternative specification of the three competing models may yield more definitive results; this project will be tackled in future work.

FOONOTES

1. It should be pointed out that applied econometricians commonly confuse non-nested models with overlapping models, and then proceed to apply the techniques reserved for non-nested models to overlapping models. The reason for this confusion, as pointed out by McAleer and Bera (1983), appears to be that definition of non-nested models is usually not well-stated.
2. These observations arose from a discussion with Quang Vuong.
3. I owe this proof to Donald Lien.
4. I owe this proof to Donald Lien.

APPENDIX A: DOCUMENTATION AND COMPUTER PROGRAM FOR THE NASH MODEL

The computational procedures required for the estimates obtained in Chapter IV for the Nash model involve the formulation of the log-likelihood function in the parameters involved and the maximization of this log-likelihood function given observations on the two jointly dependent dichotomous endogenous variables, and whatever exogenous explanatory variables are thought to affect the occurrence or nonoccurrence of the qualitative dependent variables. The program, consistent with the model described in Chapter II, assumes that the disturbance pair $(\varepsilon_h, \varepsilon_w)$ is bivariate normally distributed with zero means, unit variances and correlation coefficient ρ. For computational ease, the program provides for a grid search over possible values of ρ. In addition, to economize on the number of lines of code, three IMSL routines are used: MDNOR and MDBNOR, the univariate and bivariate normal cumulative distribution functions, respectively, and LINV3P which inverts a positive definite matrix. To control input and output, three files are required by the program. First, an input file called PARAM.DAT consists of a number of lines of control parameters; these parameters will be discussed later. Second, another input file, INPUT.DAT, is the user's data set. It consists of two dichotomous dependent variables and the exogenous explanatory variables. Finally, an output file, OUTPUT.DAT, is used for the printing of summary statistics. Both dichotomous variables should be coded as either zero or one. For each observation, the explanatory variables follow immediately. Note also that if one wishes to provide for a constant term among the explanatory variables, a vector of "ones" should be included in the data file; there is no built-in option for a constant term in the program. The program uses a modified version of the algorithm suggested by Berndt, Hall, Hall and Hausman (1974). We now discuss briefly the program flow of control.

After reading in the control cards from file PARAM.DAT (SUBROUTINE RDFILE1), the program reads in the user's data set from INPUT.DAT (SUBROUTINE RDFILE2); as a check, the first four observations are printed out. The following steps are now performed for each value of the correlation coefficient, ρ, used in the grid search. SUBROUTINE INIT now calculates and prints the initial values of the coefficients and the initial value of the log-likelihood function. Control is now transferred to SUBROUTINE BHHH, the routine which performs the iterations for each value of the correlation coefficient. Now, for each iteration, SUBROUTINE QSCORE performs two tasks. First, the total score is calculated; that is, the individual score vectors are summed over the number of observations. This is performed using subroutines ISCORE00, ISCORE01, ISCORE10 and ISCORE11 since the functional form for the score will obviously depend on the values taken jointly by the dichotomous endogenous variables. Second, the outer product of each individual score vector is computed and these outer product matrices are then summed over the number of observations producing a new matrix Q. This is done with the aid of SUBROUTINE IQ. Next, IMSL routine LINV3P calculates the inverse of the Q matrix and postmultiplies it by the score vector. Control is now transferred to SUBROUTINE OPTSTP where the optimal step size is calculated; once the step size if found, the new parameter values are calculated. SUBROUTINE QSCORE is now called again and SUBROUTINE TEST checks for convergence. If convergence is not attained, the preceding sequence will be repeated either until convergence is attained or, failing that, until the maximum number of iterations has been completed. If convergence is attained, routine LINV3P is again called to calculate the inverse of the Q matrix, the asymptotic covariance matrix. Control is then transferred to SUBROUTINE WRIE which prints out the number of iterations, the final score vector, the log-likelihood value, and the estimated coefficients along with the associated t-statistics.

The program described here is subject to a number of limitations, most of which may be relaxed easily. The program has been written for data sets with up to 3000 observations, where the behavior of each player in the Nash game can each be estimated by up to 40 explanatory variables. Note also, as mentioned earlier, that the algorithm has no built-in constant term; if one wishes to include a constant term as an explanatory variable, a vector of "ones" should be included in the input file INPUT.DAT.

We now describe the control cards required by the program. Each control card should be entered on a separate line in PARAM.DAT.

1. First control card (5I2)

(a) INDS: One or zero, indicating whether or not the score is to be written after each iteration on output file OUTPUT.DAT.

(b) IE1G: One or zero, indicating whether or not the eigenvalues of the Q matrix are to be printed out at the end of each iteration.

(c) ILOG: One or zero, indicating whether or not the log-likelihood value is to be written out after each iteration; if ILOG = 0, the log-likelihood value is written only after the final iteration, whether or not convergence is achieved.

(d) IOPT: The value of this parameter, either zero or one, determines whether or not a fixed stepsize is automatically taken at each iteration. Although in theory a fixed stepsize (equal to one) is asymptotically efficient, in many numerical problems a variable stepsize is often required.

(e) IPARAM: One or zero, indicating whether or not the estimated parameters are to be written out after each iteration. IF IPARAM = 0, the parameter estimates are written out only at the end of the final iteration.

2. Second control card (3I2)

(a) IVAR: One or zero, depending on whether or not the asymptotic covariance matrix is to be written after the final iteration.

(b) IFIT1: One or zero, depending on whether or not the predicted probabilities for each observation are to written after the final iteration.

(c) IFIT2: One or zero, depending on whether or not the observed and predicted two by two contingency tables are to be written after the final iteration.

3. Third control card (I6, 5I3, F4.3)

(a) NOBS: Number of observations in data set INPUT.DAT.

(b) NHW: Total number of unique explanatory variables in data set INPUT.DAT.

(c) NH: Number of explanatory variables used to describe the behavior of the leader in the Stackelberg model.

(d) NW: Number of explanatory variables used to describe the behavior of the follower in the Stackelberg model.

(e) NRHO: Number of values of the correlation coefficient, ρ, used in the grid search.

(f) MITER: Limit on the number of iterations for each value of the correlation coefficient.

(g) EPS: The convergence criterion. The mean taken over the number of observations, for each element of the score must be less then or equal to EPS for convergence to be attained.

4. Fourth control card (20F4.2)

VRHO(I), I = 1, NRHO: The values of the correlation coefficient, ρ, which comprise the grid search. Up to 20 values are allowed.

5. Fifth control card (8E10.3)

XO(I), I = 1, NH+NW+3: Initial values of the parameters. Eight values are allowed per line.

6. Sixth control card (40I2)

KH(I), I = 1,NH: Positive integers corresponding to the column locations in INPUT.DAT of the explanatory variables for the first player in the Nash game.

7. Seventh control card (40I2)

KW(I), I=1,NW: Positive integers corresponding to the column locations in INPUT.DAT of the explanatory variables for the second player in the Nash game.

8. Eighth control card

Format in which data file INPUT.DAT is to be read. The format must be placed in parentheses, (·).

It consists of two dichotomous dependent variables and the exogenous explanatory variables. Finally, an output file, OUTPUT.DAT, is used for the printing of summary statistics. Both dichotomous variables should be coded as either zero or one. In addition, the first element of each observation should correspond to the action taken by the leader in the Stackelberg model, while the second element should correspond to the follower. For each observation, the explanatory variables follow immediately. Note also that if one wishes to provide for a constant term among the explanatory variables, a vector of "ones" should be included in the data file; there is no built-in option for a constant term in the program.

It should also be noted that through internal control of the program, two convergence algorithms are provided. In normal use, it will be the case that the sum of the outer products of the individual score vector will be a nonsingular matrix; in this case, the matrix will be invertible and the program uses the algorithm suggested by Berndt, Hall, Hall and Hausman (1974). Occasionally, though, it may be the case, especially within the first few iterations, that the model is algorithmically not identified. In this situation, the program employs the method of steepest ascent, a routine that does not require the inverse of the matrix discussed above. (See Mickle and Sze (1972, pp. 126-128) for a discussion of this algorithm.) We now discuss briefly the program flow of control.

After reading in the control cards from file PARAM.DAT (SUBROUTINE RDFILE1), the program reads in the user's data set from

```
      DIMENSION VRHO(100),X(40),Q(820),SCORE(40),XO(40)
      DIMENSION XH(3000,40),XW(3000,40)
      INTEGER NYH(3000),NYW(3000),KH(40),KW(40),RDFT(20)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      OPEN(UNIT=1,RECORDTYPE='VARIABLE',FORM='FORMATTED',
     1FILE='PARAM.DAT',STATUS='OLD')
      OPEN(UNIT=2,RECORDTYPE='VARIABLE',FORM='FORMATTED',
     1FILE='INPUT.DAT',STATUS='OLD')
      OPEN(UNIT=3,RECORDTYPE='VARIABLE',FORM='FORMATTED',
     1FILE='OUTPUT.DAT',STATUS='OLD')
      CALL RDFILE1(N,NOBS,NHW,NH,NW,KH,KW,VRHO,NRHO,XO,RDFT,MITER,EPS,
     1  INDS,IEIG,ILOG,IOPT,IPARAM,IVAR,IFIT1,IFIT2)
      CALL RDFILE2(NOBS,NHW,NH,NW,KH,KW,RDFT,XH,XW,NYH,NYW)
      DO 10 K=1,NRHO
        WRITE(6,1000) K
1000    FORMAT(' CASE ',I2,' STARTED')
        CALL INIT(K,N,VRHO,XO,RHO,X,VLIKE,ILOG,IOPT)
        CALL BHHH(N,MITER,EPS,RHO,X,Q,SCORE,VLIKE,STPSIZE,NITER,INDS,
     1    IEIG,ILOG,IOPT,IPARAM,IER)
        IF(IER.EQ.130) GO TO 10
          CALL WRTE(N,NH,NW,RHO,VLIKE,STPSIZE,NITER,X,Q,SCORE,IVAR)
          IF(IFIT1.EQ.1.OR.IFIT2.EQ.1) CALL FIT(RHO,X,VLIKE,IFIT1,IFIT2)
10    CONTINUE
      STOP
      END

      SUBROUTINE RDFILE1(N,NOBS,NHW,NH,NW,KH,KW,VRHO,NRHO,XO,RDFT,
     1MITER,EPS,INDS,IEIG,ILOG,IOPT,IPARAM,IVAR,IFIT1,IFIT2)
      DIMENSION VRHO(100),XO(40)
      INTEGER KH(40),KW(40),RDFT(20)
      READ(1,1003) INDS,IEIG,ILOG,IOPT,IPARAM
1003  FORMAT(5I2)
      READ(1,1004) IVAR,IFIT1,IFIT2
1004  FORMAT(3I2)
      READ(1,1000) NOBS,NHW,NH,NW,NRHO,MITER,EPS
1000  FORMAT(I6,5I3,F4.3)
      READ(1,2000) (VRHO(I),I=1,NRHO)
2000  FORMAT(20F4.2)
      N=NH+NW+2
      M=N/8
      IF(M.EQ.0) GO TO 11
        DO 15 IM=1,M
          M1=(IM-1)*8+1
          M2=IM*8
          READ(1,5000) (XO(J),J=M1,M2)
                                                    5000   FORMAT(8E10.3)
15      CONTINUE
11    M3=M*8+1
      IF(M3.GT.N) GO TO 12
        READ(1,5000) (XO(J),J=M3,N)
```

```
12    IF(NH.EQ.0) GO TO 10
        READ(1,3000) (KH(J),J=1,NH)
10    IF(NW.EQ.0) GO TO 20
        READ(1,3000) (KW(J),J=1,NW)
3000    FORMAT(40I2)
20    CONTINUE
      READ(1,4000) RDFT
4000  FORMAT(20A4)
      CALL WRTRD1(NOBS,NHW,NH,NW,NRHO,VRHO,MITER,EPS,RDFT)
      RETURN
      END

      SUBROUTINE WRTRD1(NOBS,NHW,NH,NW,NRHO,VRHO,MITER,EPS,RDFT)
      DIMENSION VRHO(100)
      INTEGER RDFT(20)
      WRITE(3,1001)
1001  FORMAT(' NOBS ','NHW',' NH',' NW',' NRHO',' MITER',' EPS')
      WRITE(3,1002) NOBS,NHW,NH,NW,NRHO,MITER,EPS
1002  FORMAT(I6,3I3,2X,I3,3X,I3,F4.3)
      WRITE(3,2001) (VRHO(I),I=1,NRHO)
2001  FORMAT(/,'RHO VALUES :',20F4.2)
      WRITE(3,4001) RDFT
4001  FORMAT(/,'READING FORMAT :',20A4)
      RETURN
      END

      SUBROUTINE RDFILE2(NOBS,NHW,NH,NW,KH,KW,RDFT,XH,XW,
     1                      NYH,NYW)
      DIMENSION XX(40),XH(3000,40),XW(3000,40)
      INTEGER KH(40),KW(40),RDFT(20),NYH(3000),NYW(3000)
      DO 100 I=1,NOBS
        READ(2,RDFT) NYH(I),NYW(I),(XX(J),J=1,NHW)
        DO 110 J=1,NH
          KJ=KH(J)
          XH(I,J)=XX(KJ)
110     CONTINUE
        DO 120 J=1,NW
          KJ=KW(J)
          XW(I,J)=XX(KJ)
120     CONTINUE
100   CONTINUE
      CALL WRTOBS(NH,NW,NYH,NYW,XH,XW)
      RETURN
      END

      SUBROUTINE WRTOBS(NH,NW,NYH,NYW,XH,XW)
      DIMENSION NYH(3000),NYW(3000),XH(3000,40),XW(3000,40)
      DIMENSION XXH(40),XXW(40)
      WRITE(3,1003)
1003  FORMAT(/,'FIRST 4 OBSERVATIONS :')
```

```
      DO 100 I=1,4
        IYH=NYH(I)
        IYW=NYW(I)
        WRITE(3,1000) I,IYH,IYW
1000    FORMAT(/,'OBSERVATION ',I4,' :',' IYH =',I2,' IYW =',I2)
        DO 110 J=1,NH
          XXH(J)=XH(I,J)
110     CONTINUE
        DO 120 J=1,NW
          XXW(J)=XW(I,J)
120     CONTINUE
        WRITE(3,1001)
1001    FORMAT(/,'  XH :')
        CALL WRTVEC(NH,XXH)
        WRITE(3,1002)
1002    FORMAT(/,'  XW :')
        CALL WRTVEC(NW,XXW)
 100  CONTINUE
      RETURN
      END

      SUBROUTINE INIT(K,N,VRHO,XO,RHO,X,VLIKE,ILOG,IOPT)
      DIMENSION XH(3000,40),XW(3000,40),XO(40),X(40),VRHO(100)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      RHO=VRHO(K)
      WRITE(3,1000) K,RHO
1000  FORMAT(///,'CASE ',I3,3X,'RHO =',F4.2)
      DO 20 J=1,N
        X(J)=XO(J)
20    CONTINUE
      WRITE(3,2001)
2001  FORMAT(/,'STARTING VALUES :')
      CALL WRTVEC(N,X)
      IF(ILOG.EQ.0.AND.IOPT.EQ.0) GO TO 13
        CALL VALUE(N,RHO,X,VLIKE)
13    IF(ILOG.EQ.0) GO TO 14
        WRITE(3,3000) VLIKE
3000    FORMAT(/,'INITIAL LOG-LIKELIHOOD VALUE = ',E13.6)
14    RETURN
      END

      SUBROUTINE VALUE(N,RHO,X,VLIKE)
      DIMENSION XH(3000,40),XW(3000,40),X(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      VLIKE=0.
      DO 100 I=1,NOBS
        INYH=NYH(I)
        INYW=NYW(I)
```

```
      CALL DELTA(I,X,DELTAH,DELTAW)
      IF (INYH.EQ.1.OR.INYW.EQ.1) GO TO 11
        CALL IPROB00(RHO,X,DELTAH,DELTAW,PR00)
        CALL TOL(PR00)
        VLIKE=VLIKE+ALOG(PR00)
        GO TO 100
11    IF (INYH.EQ.0.OR.INYW.EQ.1) GO TO 12
        CALL IPROB10(RHO,X,DELTAH,DELTAW,PR10)
        CALL TOL(PR10)
        VLIKE=VLIKE+ALOG(PR10)
        GO TO 100
12    IF (INYH.EQ.1.OR.INYW.EQ.0) GO TO 13
        CALL IPROB01(RHO,X,DELTAH,DELTAW,PR01)
        CALL TOL(PR01)
        VLIKE=VLIKE+ALOG(PR01)
        GO TO 100
13    CALL IPROB11(RHO,X,DELTAH,DELTAW,PR11)
      CALL TOL(PR11)
      VLIKE=VLIKE+ALOG(PR11)
100 CONTINUE
    RETURN
    END

    SUBROUTINE DELTA(I,X,DELTAH,DELTAW)
    DIMENSION X(1),XH(3000,40),XW(3000,40)
    INTEGER NYH(3000),NYW(3000)
    COMMON NOBS,NH,NW,XH,XW,NYH,NYW
    DELTAH=0.0
    DELTAW=0.0
    DO 100 J=1,NH
      JH=J+2
      DELTAH=DELTAH+X(JH)*XH(I,J)
100 CONTINUE
    DO 200 J=1,NW
      JW=J+NH+2
      DELTAW=DELTAW+X(JW)*XW(I,J)
200 CONTINUE
    RETURN
    END

    SUBROUTINE BHHH(N,MITER,EPS,RHO,X,Q,SCORE,VLIKE,STPSIZE,NITER,
   1INDS,IEIG,ILOG,IOPT,IPARAM,IER)
    DIMENSION XH(3000,40),XW(3000,40),X(40),Q(820),SCORE(40)
    INTEGER NYH(3000),NYW(3000)
    COMMON NOBS,NH,NW,XH,XW,NYH,NYW
    NITER=999
    NIN=0
    NOUT=3
    CALL UGETIO(3,NIN,NOUT)
    CALL QSCORE(N,RHO,X,Q,SCORE)
```

```
      IF(INDS.EQ.0.AND.IEIG.EQ.0) GO TO 1002
        CALL WRTBH1(N,Q,SCORE,INDS,IEIG)
1002  CONTINUE
      DO 10 ITER=1,MITER
        CALL LINV3P(Q,SCORE,3,N,IER)
        IF(IER.EQ.130) GO TO 999
        CALL OPTSTP(N,RHO,X,SCORE,STPSIZE,VLIKE,ILOG,IOPT)
        CALL QSCORE(N,RHO,X,Q,SCORE)
        CALL TEST(NOBS,N,SCORE,EPS,IEND)
        IF (IEND.EQ.1) GO TO 20
        IF(ITER.EQ.MITER) GO TO 11
          CALL WRTBHH(N,ITER,STPSIZE,X,Q,SCORE,VLIKE,INDS,IEIG,ILOG,
     1                 IPARAM)
11      CONTINUE
10    CONTINUE
      GO TO 50
20    NITER=ITER
50    IF(ILOG.EQ.1.OR.IOPT.EQ.1) GO TO 40
        CALL VALUE(N,RHO,X,VLIKE)
40    CALL LINV3P(Q,SCORE,1,N,IER)
999   RETURN
      END

      SUBROUTINE QSCORE(N,RHO,X,Q,SCORE)
      DIMENSION XH(3000,40),XW(3000,40),X(40),Q(820),SCORE(40)
      DIMENSION QI(820),SCOREI(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      NN=N*(N+1)/2
      DO 400 J=1,N
        SCORE(J)=0.
400   CONTINUE
      DO 500 K=1,NN
        Q(K)=0.
500   CONTINUE
      DO 100 I=1,NOBS
        CALL ISCORE(I,N,RHO,X,SCOREI)
        CALL IQ(N,SCOREI,QI)
        DO 200 J=1,N
          SCORE(J)=SCORE(J)+SCOREI(J)
200     CONTINUE
        DO 300 K=1,NN
          Q(K)=Q(K)+QI(K)
300     CONTINUE
100   CONTINUE
      RETURN
      END

      SUBROUTINE WRTBHH(N,ITER,STPSIZE,X,Q,SCORE,VLIKE,INDS,IEIG,ILOG,
     1                 IPARAM)
```

```
      IF(INDS.EQ.0.AND.IEIG.EQ.0.AND.ILOG.EQ.0.AND.IPARAM.EQ.0)
     1  GO TO 11
        CALL WRTBH2(N,ITER,X,STPSIZE,VLIKE,ILOG,IPARAM)
        IF(INDS.EQ.0.AND.IEIG.EQ.0) GO TO 12
          CALL WRTBH1(N,Q,SCORE,INDS,IEIG)
12      CONTINUE
11    RETURN
      END

      SUBROUTINE WRTBH1(N,Q,SCORE,INDS,IEIG)
      DIMENSION Q(820),SCORE(40)
      DIMENSION D(40),Z(40,40),WK(860)
      IF(INDS.EQ.0) GO TO 1001
        WRITE(3,4000)
4000    FORMAT(/,'    SCORE    : ')
        CALL WRTVEC(N,SCORE)
1001  CONTINUE
      IF(IEIG.EQ.0) GO TO 1002
        CALL EIGRS (Q,6,2,D,Z,6,WK,IER)
        WRITE(3,5000)
5000    FORMAT(/,'    EIGEN    : ')
        CALL WRTVEC(N,D)
1002  CONTINUE
      RETURN
      END

      SUBROUTINE OPTSTP(N,RHO,X,SCORE,STPSIZE,VLIKE,ILOG,IOPT)
      DIMENSION XH(3000,40),XW(3000,40),X(40),Q(820),SCORE(40)
      DIMENSION X1(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      VLIKE0=VLIKE
      STPSIZE=1.
      DO 100 NTRY=1,9
        DO 200 J=1,N
          X1(J)=X(J)+SCORE(J)
200     CONTINUE
        IF(ILOG.EQ.0.AND.IOPT.EQ.0) GO TO 300
          CALL VALUE(N,RHO,X1,VLIKE)
          IF(IOPT.EQ.0) GO TO 300
            DVLIKE=VLIKE-VLIKE0
            IF(DVLIKE.GE.0.) GO TO 300
              STPSIZE=STPSIZE-.1
              STPSIZ1=STPSIZE+.1
              DO 500 J=1,N
                SCORE(J)=SCORE(J)*STPSIZE/STPSIZ1
500           CONTINUE
            CONTINUE
          CONTINUE
        CONTINUE
```

```
100  CONTINUE
300  DO 400 J=1,N
       X(J)=X1(J)
400  CONTINUE
     RETURN
     END

     SUBROUTINE TEST(NOBS,N,SCORE,EPS,IEND)
     DIMENSION SCORE(1)
     IEND=0
     DO 10 J=1,N
       A=ABS(SCORE(J))/NOBS
       IF (A.GT.EPS) GO TO 20
10   CONTINUE
     IEND=1
20   RETURN
     END

     SUBROUTINE WRTBH2(N,ITER,X,STPSIZE,VLIKE,ILOG,IPARAM)
     DIMENSION X(40)
     WRITE(3,1000) ITER
1000 FORMAT(//,'ITERATION ',I3)
     IF(ILOG.EQ.0) GO TO 10
       WRITE(3,2000) STPSIZE,VLIKE
2000   FORMAT(/,'STEPSIZE = ',F5.3,'    LOG-LIKELIHOOD VALUE = ',E13.6)
10   IF(IPARAM.EQ.0) GO TO 20
       WRITE(3,3000)
3000   FORMAT(/,'PARAMETERS :')
       CALL WRTVEC(N,X)
20   RETURN
     END

     SUBROUTINE ISCORE(I,N,RHO,X,SCOREI)
     DIMENSION XH(3000,40),XW(3000,40),X(40),SCOREI(40)
     DIMENSION DF(40),DI(40)
     INTEGER NYH(3000),NYW(3000)
     COMMON NOBS,NH,NW,XH,XW,NYH,NYW
     INYH=NYH(I)
     INYW=NYW(I)
     CALL DELTA(I,X,DELTAH,DELTAW)
     IF (INYH.EQ.1.OR.INYW.EQ.1) GO TO 11
       CALL IPROB00(RHO,X,DELTAH,DELTAW,PR00)
       CALL ISCORE00(I,N,RHO,X,DELTAH,DELTAW,PR00,SCOREI)
       GO TO 14
11    IF (INYH.EQ.0.OR.INYW.EQ.1) GO TO 12
       CALL IPROB10(RHO,X,DELTAH,DELTAW,PR10)
       CALL ISCORE10(I,N,RHO,X,DELTAH,DELTAW,PR10,SCOREI)
       GO TO 14
12    IF (INYH.EQ.1.OR.INYW.EQ.0) GO TO 13
       CALL IPROB01(RHO,X,DELTAH,DELTAW,PR01)
```

```
        CALL ISCORE01(I,N,RHO,X,DELTAH,DELTAW,PR01,SCOREI)
        GO TO 14
13    CALL IPROB11(RHO,X,DELTAH,DELTAW,PR11)
      CALL ISCORE11(I,N,RHO,X,DELTAH,DELTAW,PR11,SCOREI)
14    CONTINUE
      RETURN
      END

      SUBROUTINE IPROB00(RHO,X,DELTAH,DELTAW,PR00)
      DIMENSION X(40)
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(1)
      W2=-DELTAW-X(2)
      CALL MDBNOR(H1,W1,RHO,PR00,IER)
      X1=X(1)
      X2=X(2)
      IF (X1.LT.0..AND.X2.LT.0.) GO TO 13
        CALL INTEGRAL(H1,W1,H2,W2,RHO,PROB,IER)
        IF (X1.LT.0..OR.X2.LT.0.) GO TO 10
          PR00=PR00-.5*PROB
          GO TO 13
10      IF (X1.LT.0..OR.X2.GE.0.) GO TO 11
          PR00=PR00-.25*PROB
          GO TO 13
11      PR00=PR00-.25*PROB
13    RETURN
      END

      SUBROUTINE ISCORE00(I,N,RHO,X,DELTAH,DELTAW,PR00,SCOREI)
      DIMENSION XH(3000,40),XW(3000,40),X(40),SCOREI(40)
      DIMENSION DI(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      SCOREI(1)=0.
      SCOREI(2)=0.
      H1=-DELTAH
      ZH=-DELTAW+RHO*DELTAH
      CALL PARTIAL(RHO,H1,ZH,PARTH)
      PARTH=-PARTH
      DO 10 J=1,NH
        JH=J+2
        SCOREI(JH)=PARTH*XH(I,J)
10    CONTINUE
      W1=-DELTAW
      ZW=-DELTAH+RHO*DELTAW
      CALL PARTIAL(RHO,W1,ZW,PARTW)
      PARTW=-PARTW
      DO 20 J=1,NW
        JW=J+NH+2
```

```
      SCOREI(JW)=PARTW*XW(I,J)
20    CONTINUE
      X1=X(1)
      X2=X(2)
      IF (X1.LT.0..AND.X2.LT.0.) GO TO 30
        CALL PARTIALI(I,N,RHO,DELTAH,DELTAW,DI,X)
        IF (X1.LT.0..OR.X2.LT.0.) GO TO 40
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-.5*DI(J)
31        CONTINUE
          GO TO 30
40      IF (X1.LT.0..OR.X2.GE.0.) GO TO 50
          DO 41 J=1,N
            SCOREI(J)=SCOREI(J)-.25*DI(J)
41        CONTINUE
          GO TO 30
50      CONTINUE
        DO 51 J=1,N
          SCOREI(J)=SCOREI(J)-.25*DI(J)
51      CONTINUE
30    CONTINUE
      CALL TOL(PROO)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PROO
60    CONTINUE
      RETURN
      END

      SUBROUTINE IPROB10(RHO,X,DELTAH,DELTAW,PR10)
      DIMENSION X(40)
      XNEGRHO=-RHO
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(1)
      W2=-DELTAW-X(2)
      CALL MDBNOR(DELTAH,W2,XNEGRHO,PR10,IER)
      X1=X(1)
      X2=X(2)
      IF (X1.GE.0..AND.X2.GE.0.) GO TO 13
        CALL INTEGRAL(H1,W1,H2,W2,RHO,PROB,IER)
        IF (X1.LT.0..OR.X2.GE.0.) GO TO 10
          PR10=PR10-.25*PROB
          GO TO 13
10      IF (X1.GE.0..OR.X2.LT.0.) GO TO 11
          PR10=PR10-.25*PROB
          GO TO 13
11      PR10=PR10-.5*PROB
13    RETURN
      END
```

```
      SUBROUTINE ISCORE10(I,N,RHO,X,DELTAH,DELTAW,PR10,SCOREI)
      DIMENSION XH(3000,40),XW(3000,40),X(40),SCOREI(40)
      DIMENSION DI(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      SCOREI(1)=0.
      W2=-DELTAW-X(2)
      W3=-W2
      ZW=DELTAH-RHO*W3
      CALL PARTIAL(RHO,W2,ZW,PARTW)
      SCOREI(2)=-PARTW
      ZH=W2+RHO*DELTAH
      CALL PARTIAL(RHO,DELTAH,ZH,PARTH)
      DO 10 J=1,NH
        JH=J+2
        SCOREI(JH)=PARTH*XH(I,J)
10    CONTINUE
      PARTW=-PARTW
      DO 20 J=1,NW
        JW=J+NH+2
        SCOREI(JW)=PARTW*XW(I,J)
20    CONTINUE
      X1=X(1)
      X2=X(2)
      IF (X1.GE.0..AND.X2.GE.0.) GO TO 30
        CALL PARTIALI(I,N,RHO,DELTAH,DELTAW,DI,X)
        IF (X1.LT.0..OR.X2.GE.0.) GO TO 40
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-.25*DI(J)
31        CONTINUE
          GO TO 30
40      IF (X1.GE.0..OR.X2.LT.0.) GO TO 50
          DO 41 J=1,N
            SCOREI(J)=SCOREI(J)-.25*DI(J)
41        CONTINUE
          GO TO 30
50      CONTINUE
        DO 51 J=1,N
          SCOREI(J)=SCOREI(J)-.5*DI(J)
51      CONTINUE
30    CONTINUE
      CALL TOL(PR10)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PR10
60    CONTINUE
      RETURN
      END

      SUBROUTINE IPROB01(RHO,X,DELTAH,DELTAW,PR01)
      DIMENSION X(40)
```

```
      XNEGRHO=-RHO
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(1)
      W2=-DELTAW-X(2)
      CALL MDBNOR(H2,DELTAW,XNEGRHO,PR01,IER)
      X1=X(1)
      X2=X(2)
      IF (X1.GE.0..AND.X2.GE.0.) GO TO 13
        CALL INTEGRAL(H1,W1,H2,W2,RHO,PROB,IER)
        IF (X1.LT.0..OR.X2.GE.0.) GO TO 10
          PR01=PR01-.25*PROB
          GO TO 13
10      IF (X1.GE.0..OR.X2.LT.0.) GO TO 11
          PR01=PR01-.25*PROB
          GO TO 13
11      PR01=PR01-.5*PROB
13    RETURN
      END

      SUBROUTINE ISCORE01(I,N,RHO,X,DELTAH,DELTAW,PR01,SCOREI)
      DIMENSION XH(3000,40),XW(3000,40),X(40),SCOREI(40)
      DIMENSION DI(40)
      INTEGER NYH(3000),NYW(3000)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      H2=-DELTAH-X(1)
      H3=-H2
      ZW=DELTAW-RHO*H3
      CALL PARTIAL(RHO,H2,ZW,PARTW)
      SCOREI(1)=-PARTW
      SCOREI(2)=0.
      DO 10 J=1,NH
        JH=J+2
        SCOREI(JH)=-PARTW*XH(I,J)
10    CONTINUE
      ZH=H2+RHO*DELTAW
      CALL PARTIAL(RHO,DELTAW,ZH,PARTH)
      DO 20 J=1,NW
        JW=J+NH+2
        SCOREI(JW)=PARTH*XW(I,J)
20    CONTINUE
      X1=X(1)
      X2=X(2)
      IF (X1.GE.0..AND.X2.GE.0.) GO TO 30
        CALL PARTIALI(I,N,RHO,DELTAH,DELTAW,DI,X)
        IF (X1.LT.0..OR.X2.GE.0.) GO TO 40
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-.25*DI(J)
31        CONTINUE
          GO TO 30
```

```
 40      IF (X1.GE.0..OR.X2.LT.0.) GO TO 50
           DO 41 J=1,N
             SCOREI(J)=SCOREI(J)-.25*DI(J)
 41        CONTINUE
           GO TO 30
 50      CONTINUE
         DO 51 J=1,N
           SCOREI(J)=SCOREI(J)-.5*DI(J)
 51      CONTINUE
30     CONTINUE
       CALL TOL(PR01)
       DO 60 J=1,N
         SCOREI(J)=SCOREI(J)/PR01
 60    CONTINUE
       RETURN
       END

       SUBROUTINE IPROB11(RHO,X,DELTAH,DELTAW,PR11)
       DIMENSION X(40)
       H1=-DELTAH
       W1=-DELTAW
       H2=-DELTAH-X(1)
       W2=-DELTAW-X(2)
       H3=-H2
       W3=-W2
       CALL MDBNOR(H3,W3,RHO,PR11,IER)
       X1=X(1)
       X2=X(2)
       IF (X1.LT.0..AND.X2.LT.0.) GO TO 13
         CALL INTEGRAL(H1,W1,H2,W2,RHO,PROB,IER)
         IF (X1.LT.0..OR.X2.LT.0.) GO TO 10
           PR11=PR11-.5*PROB
           GO TO 13
 10      IF (X1.LT.0..OR.X2.GE.0.) GO TO 11
           PR11=PR11-.25*PROB
           GO TO 13
 11      PR11=PR11-.25*PROB
 13    RETURN
       END

       SUBROUTINE ISCORE11(I,N,RHO,X,DELTAH,DELTAW,PR11,SCOREI)
       DIMENSION XH(3000,40),XW(3000,40),X(40),SCOREI(40)
       DIMENSION DI(40)
       INTEGER NYH(3000),NYW(3000)
       COMMON NOBS,NH,NW,XH,XW,NYH,NYW
       H3=DELTAH+X(1)
       W3=DELTAW+X(2)
       ZH=W3-RHO*H3
       CALL PARTIAL(RHO,H3,ZH,PARTH)
       SCOREI(1)=PARTH
```

```
      ZW=H3-RHO*W3
      CALL PARTIAL(RHO,W3,ZW,PARTW)
      SCOREI(2)=PARTW
      DO 10 J=1,NH
        JH=J+2
        SCOREI(JH)=PARTH*XH(I,J)
10    CONTINUE
      DO 20 J=1,NW
        JW=J+NH+2
        SCOREI(JW)=PARTW*XW(I,J)
20    CONTINUE
      X1=X(1)
      X2=X(2)
      IF (X1.LT.0..AND.X2.LT.0.) GO TO 30
        CALL PARTIALI(I,N,RHO,DELTAH,DELTAW,DI,X)
        IF (X1.LT.0..OR.X2.LT.0.) GO TO 40
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-.5*DI(J)
31        CONTINUE
          GO TO 30
40      IF (X1.LT.0..OR.X2.GE.0.) GO TO 50
          DO 41 J=1,N
            SCOREI(J)=SCOREI(J)-.25*DI(J)
41        CONTINUE
          GO TO 30
50      CONTINUE
        DO 51 J=1,N
          SCOREI(J)=SCOREI(J)-.25*DI(J)
51      CONTINUE
30    CONTINUE
      CALL TOL(PR11)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PR11
60    CONTINUE
      RETURN
      END

      SUBROUTINE TOL(XPROB)
      XMIN=1.E-20
      IF(XPROB.LE.XMIN) XPROB=XMIN
      RETURN
      END

      SUBROUTINE IQ(N,SCOREI,QI)
      DIMENSION SCOREI(40),QI(820)
      IR=0
      DO 10 IX=1,N
        DO 20 JX=1,IX
          IL=IR+JX
          QI(IL)=SCOREI(IX)*SCOREI(JX)
```

```
20        CONTINUE
          IR=IR+IX
10      CONTINUE
        NN=N*(N+1)/2
        RETURN
        END

        SUBROUTINE PARTIAL(RHO,Y,Z,PART)
        ZX=Z/SQRT(1.-RHO**2)
        DENSY=.3989422*EXP(-Y**2/2.)
        CALL MDNOR(ZX,PHIZX)
        PART=DENSY*PHIZX
        RETURN
        END

        SUBROUTINE PARTIALI(I,N,RHO,DELTAH,DELTAW,DI,X)
        DIMENSION DI(40),XH(3000,40),XW(3000,40),X(40)
        DIMENSION NYH(3000),NYW(3000)
        COMMON NOBS,NH,NW,XH,XW,NYH,NYW
        H1=-DELTAH
        W1=-DELTAW
        H2=-DELTAH-X(1)
        W2=-DELTAW-X(2)
        H3=-H2
        W3=-W2
        Z1=W2+RHO*H3
        Z2=W1+RHO*H3
        Z3=H2+RHO*W3
        Z4=H1+RHO*W3
        Z5=W1+RHO*DELTAH
        Z6=W2+RHO*DELTAH
        Z7=H1+RHO*DELTAW
        Z8=H2+RHO*DELTAW
        CALL PARTIAL(RHO,H2,Z1,PART1)
        CALL PARTIAL(RHO,H2,Z2,PART2)
        CALL PARTIAL(RHO,W2,Z3,PART3)
        CALL PARTIAL(RHO,W2,Z4,PART4)
        CALL PARTIAL(RHO,H1,Z5,PART5)
        CALL PARTIAL(RHO,H1,Z6,PART6)
        CALL PARTIAL(RHO,W1,Z7,PART7)
        CALL PARTIAL(RHO,W1,Z8,PART8)
        DI(1)=-PART1+PART2
        DI(2)=-PART3+PART4
        PARTH=-PART5-PART1+PART6+PART2
        DO 10 J=1,NH
          JH=J+2
          DI(JH)=PARTH*XH(I,J)
10      CONTINUE
        PARTW=-PART7-PART3+PART4+PART8
        DO 20 J=1,NW
```

```
        JW=J+2+NH
        DI(JW)=PARTW*XW(I,J)
20    CONTINUE
      RETURN
      END

      SUBROUTINE INTEGRAL(ARG1,ARG2,ARG3,ARG4,ARG5,PROB,IER)
      CALL MDBNOR(ARG1,ARG2,ARG5,PROB1,IER)
      CALL MDBNOR(ARG3,ARG4,ARG5,PROB2,IER)
      CALL MDBNOR(ARG1,ARG4,ARG5,PROB3,IER)
      CALL MDBNOR(ARG2,ARG3,ARG5,PROB4,IER)
      PROB=PROB1+PROB2-PROB3-PROB4
      RETURN
      END

      SUBROUTINE WRTE(N,NH,NW,RHO,VLIKE,STPSIZE,NITER,X,Q,SCORE,IVAR)
      DIMENSION X(40),Q(820),SCORE(40)
      WRITE(3,1000) RHO
1000  FORMAT(//,'RHO = ',F5.3)
      WRITE(3,2000) NITER,STPSIZE
2000  FORMAT(/,'NUMBER OF ITERATIONS = ',I3,'  FINAL STEPSIZE =',F5.3)
      WRITE(3,2001)
2001  FORMAT(/,'FINAL SCORE : ')
      CALL WRTVEC(N,SCORE)
      WRITE(3,3000) VLIKE
3000  FORMAT(/,'LOG-LIKELIHOOD VALUE = ',E13.6)
      TSTAT=X(1)/SQRT(Q(1))
      WRITE(3,4000) X(1),TSTAT
4000  FORMAT(//,'BETA H     = ',E10.3,'  T-STATISTIC = ',E10.3)
      TSTAT=X(2)/SQRT(Q(3))
      WRITE(3,5000) X(2),TSTAT
5000  FORMAT(/,'BETA W     = ',E10.3,'  T-STATISTIC = ',E10.3)
      NHH=NH+2
      DO 100 I=3,NHH
        IH=I-2
        IN=I*(I+1)/2
        TSTAT=X(I)/SQRT(Q(IN))
        WRITE (3,6000) IH,X(I),TSTAT
100   CONTINUE
6000  FORMAT(/,'GAMMA H ',I2,' = ',E10.3,'  T-STATISTIC = ',E10.3)
      K=3+NH
      DO 200 I=K,N
        IW=I-K+1
        IN=I*(I+1)/2
        TSTAT=X(I)/SQRT(Q(IN))
        WRITE(3,7000) IW,X(I),TSTAT
200   CONTINUE
7000  FORMAT(/,'GAMMA W ',I2,' = ',E10.3,'  T-STATISTIC = ',E10.3)
      IF(IVAR.EQ.0) GO TO 50
        CALL WRTVAR(N,Q)
```

```
50    CONTINUE
      RETURN
      END

      SUBROUTINE WRTVAR(N,Q)
      DIMENSION Q(840),ROW(40)
      WRITE(3,8000)
8000  FORMAT(//,'ASYMPTOTIC COVARIANCE MATRIX')
      DO 20 I=1,N
        IL=I*(I-1)/2 + 1
        IU=IL+I-1
        IJ=0
        DO 30 J=IL,IU
          IJ=IJ+1
          ROW(IJ)=Q(J)
30      CONTINUE
        WRITE(3,4000)
4000    FORMAT(/)
        CALL WRTVEC(I,ROW)
20    CONTINUE
99    RETURN
      END

      SUBROUTINE WRTVEC(NDIM,VEC)
      DIMENSION VEC(40)
      MDIM=NDIM/8
      IF(MDIM.EQ.0) GO TO 10
        DO 20 IM=1,MDIM
          M1=(IM-1)*8+1
          M2=IM*8
          WRITE(3,3000) (VEC(J),J=M1,M2)
3000      FORMAT(8E10.3)
20      CONTINUE
10    M3=MDIM*8+1
      IF(M3.GT.NDIM) GO TO 30
        WRITE(3,3000) (VEC(J),J=M3,NDIM)
30    RETURN
      END

      SUBROUTINE FIT(RHO,X,VLIKE,IFIT1,IFIT2)
      DIMENSION XH(3000,40),XW(3000,40),X(40),RTAB(4)
      DIMENSION NYH(3000),NYW(3000),INDEXO(3000),INDEXP(3000)
      INTEGER TABLE(4),ESTTAB(4),NGTAB(4)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      NGH0=0
      NGH1=0
      NGW0=0
      NGW1=0
      DO 10 J=1,4
        NGTAB(J)=0
```

```
   10 CONTINUE
      IF(IFIT1.EQ.1) GO TO 12
   11   CALL TREAT1(RHO,X,INDEXO,INDEXP,NGHO,NGH1,NGWO,NGW1,NGTAB)
        GO TO 13
   12   CALL TREAT2(RHO,X,INDEXO,INDEXP,NGHO,NGH1,NGWO,NGW1,NGTAB)
   13 IF(IFIT2.EQ.0) GO TO 14
      CALL FITSTAT(NOBS,VLIKE,INDEXO,INDEXP,NGHO,NGH1,NGWO,NGW1,NGTAB,
     1             TABLE,ESTTAB,R,RH,RW,RTAB,RHO,RH1,RW0,RW1,RLIKE)
      CALL WRTSTAT(TABLE,ESTTAB,R,RH,RW,RTAB,RHO,RH1,RW0,RW1,RLIKE)
      NX=NH+NW+2
      CALL WRTVEC(NX,X)
   14 RETURN
      END

      SUBROUTINE TREAT1(RHO,X,INDEXO,INDEXP,NGHO,NGH1,NGWO,NGW1,NGTAB)
      DIMENSION XH(3000,40),XW(3000,40),X(40)
      DIMENSION NYH(3000),NYW(3000),INDEXO(3000),INDEXP(3000)
      INTEGER NGTAB(4)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      DO 100 I=1,NOBS
        CALL DELTA(I,X,DELTAH,DELTAW)
        CALL IPROB00(RHO,X,DELTAH,DELTAW,PR00)
        CALL IPROB01(RHO,X,DELTAH,DELTAW,PR01)
        CALL IPROB10(RHO,X,DELTAH,DELTAW,PR10)
        CALL IPROB11(RHO,X,DELTAH,DELTAW,PR11)
        CALL YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
        INDEXP(I)=INDEXI
        IYH=NYH(I)
        IYW=NYW(I)
        CALL IND(IYH,IYW,INDEXOI)
        INDEXO(I)=INDEXOI
        CALL COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,NGHO,NGH1,
     1             NGWO,NGW1,NGTAB)
  100 CONTINUE
      RETURN
      END

      SUBROUTINE TREAT2(RHO,X,INDEXO,INDEXP,NGHO,NGH1,NGWO,NGW1,NGTAB)
      DIMENSION XH(3000,40),XW(3000,40),X(40)
      DIMENSION NYH(3000),NYW(3000),INDEXO(3000),INDEXP(3000)
      INTEGER NGTAB(4)
      COMMON NOBS,NH,NW,XH,XW,NYH,NYW
      WRITE(3,1000)
 1000 FORMAT(//,'  YH  ','  YW  ',' HHAT ',' WHAT ',' PR00 ',
     1       ' PR01 ',' PR10 ',' PR11 ',/)
      DO 100 I=1,NOBS
        CALL DELTA(I,X,DELTAH,DELTAW)
        CALL IPROB00(RHO,X,DELTAH,DELTAW,PR00)
        CALL IPROB01(RHO,X,DELTAH,DELTAW,PR01)
        CALL IPROB10(RHO,X,DELTAH,DELTAW,PR10)
```

```
        CALL IPROB11(RHO,X,DELTAH,DELTAW,PR11)
        CALL YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
        INDEXP(I)=INDEXI
        IYH=NYH(I)
        IYW=NYW(I)
        WRITE(3,1001) IYH,IYW,IYHHAT,IYWHAT,PR00,PR01,PR10,PR11
1001    FORMAT(4(3X,I1,2X),4(1X,F4.3,1X))
        CALL IND(IYH,IYW,INDEXOI)
        INDEXO(I)=INDEXOI
        CALL COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,NGH0,NGH1,
     1             NGW0,NGW1,NGTAB)
  100 CONTINUE
      RETURN
      END

      SUBROUTINE YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
      INDEXI=1
      IF(PR00.GT.PR01) GO TO 200
        PRMAX=PR01
        INDEXI=2
        GO TO 110
  200 PRMAX=PR00
  110 IF(PRMAX.GT.PR10)  GO TO 100
        PRMAX=PR10
        INDEXI=3
  100 IF(PRMAX.GT.PR11) GO TO 500
        INDEXI=4
  500 GO TO (501,502,503,504),INDEXI
  501 IYHHAT=0
      IYWHAT=0
      GO TO 900
  502 IYHHAT=0
      IYWHAT=1
      GO TO 900
  503 IYHHAT=1
      IYWHAT=0
      GO TO 900
  504 IYHHAT=1
      IYWHAT=1
  900 RETURN
      END

      SUBROUTINE IND(IYH,IYW,INDEXOI)
      INDEXOI=1
      IF(IYW.EQ.1) GO TO 200
        IF(IYH.EQ.1) INDEXOI=3
          GO TO 100
  200 INDEXOI=2
      IF(IYH.EQ.1) INDEXOI=4
  100 CONTINUE
```

```
      RETURN
      END

      SUBROUTINE COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,
     1                 NGH0,NGH1,NGW0,NGW1,NGTAB)
      INTEGER NGTAB(4)
      IF(IYH.EQ.0.AND.IYHHAT.EQ.0) NGH0=NGH0+1
      IF(IYH.EQ.1.AND.IYHHAT.EQ.1) NGH1=NGH1+1
      IF(IYW.EQ.0.AND.IYWHAT.EQ.0) NGW0=NGW0+1
      IF(IYW.EQ.1.AND.IYWHAT.EQ.1) NGW1=NGW1+1
      DO 100 J=1,4
        IF(INDEXOI.EQ.J.AND.INDEXI.EQ.J) NGTAB(J)=NGTAB(J)+1
  100 CONTINUE
      RETURN
      END

      SUBROUTINE FITSTAT(NOBS,VLIKE,INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,
     1           NGTAB,TABLE,ESTTAB,R,RH,RW,RTAB,RH0,RH1,RW0,RW1,RLIKE)
      DIMENSION XNGTAB(4),RTAB(4)
      INTEGER NGTAB(4),TABLE(4),ESTTAB(4),INDEXO(3000),INDEXP(3000)
      CALL CONTIN(NOBS,INDEXO,TABLE)
      CALL CONTIN(NOBS,INDEXP,ESTTAB)
      XNGH0=NGH0
      XNGH1=NGH1
      XNGH=XNGH0+XNGH1
      XNGW0=NGW0
      XNGW1=NGW1
      XNGW=XNGW0+XNGW1
      XNOBS=NOBS
      RH=XNGH/XNOBS
      RH0=XNGH0/(TABLE(1)+TABLE(2))
      RH1=XNGH1/(TABLE(3)+TABLE(4))
      RW=XNGW/XNOBS
      RW0=XNGW0/(TABLE(1)+TABLE(3))
      RW1=XNGW1/(TABLE(2)+TABLE(4))
      DO 10 J=1,4
        XNGTAB(J)=NGTAB(J)
        RTAB(J)=XNGTAB(J)/TABLE(J)
   10 CONTINUE
      XNG=XNGTAB(1)+XNGTAB(2)+XNGTAB(3)+XNGTAB(4)
      R=XNG/XNOBS
      RLIKE=1.+ VLIKE/(XNOBS*ALOG(4.))
      RETURN
      END

      SUBROUTINE CONTIN(NOBS,NVEC,NTAB)
      INTEGER NVEC(3000),NTAB(4)
      DO 100 J=1,4
        NTAB(J)=0
  100 CONTINUE
```

```
      DO 200 I=1,NOBS
        NVECI=NVEC(I)
        GO TO (101,102,103,104),NVECI
 101      NTAB(1)=NTAB(1)+1
          GO TO 201
 102      NTAB(2)=NTAB(2)+1
          GO TO 201
 103      NTAB(3)=NTAB(3)+1
          GO TO 201
 104      NTAB(4)=NTAB(4)+1
 201    CONTINUE
 200  CONTINUE
      RETURN
      END

      SUBROUTINE WRTSTAT(TABLE,ESTTAB,R,RH,RW,RTAB,RH0,RH1,RW0,RW1,
     1                   RLIKE)
      DIMENSION RTAB(4)
      INTEGER TABLE(4),ESTTAB(4)
      WRITE(3,2004) RLIKE,R
2004  FORMAT(//,' RLIKE = ',F4.3,3X,' R = ',F4.3)
      WRITE(3,2000)
2000  FORMAT(//,' OBSERVED AND PREDICTED CONTINGENCY TABLES :')
      WRITE(3,2001)
2001  FORMAT(/,'     CELLS :',' 00   ','  01   ','  10   ','  11')
      WRITE(3,2002) (TABLE(J),J=1,4)
2002  FORMAT(/,'  OBSERVED :',2X,I4,4X,I4,4X,I4,4X,I4)
      WRITE(3,2003) (ESTTAB(J),J=1,4)
2003  FORMAT(/,' PREDICTED :',2X,I4,4X,I4,4X,I4,4X,I4)
      WRITE(3,2005) (RTAB(J),J=1,4)
2005  FORMAT(/,' HW-RATIOS :',2X,F4.3,4X,F4.3,4X,F4.3,4X,F4.3)
      WRITE(3,2006)
2006  FORMAT(//,'   RATIOS : WORK & NOTWORK   WORK ONLY   NOT WORK ')
      WRITE(3,2007) RH,RH1,RH0
2007  FORMAT(/,'  HUSBAND :',6X,F5.3,11X,F5.3,9X,F5.3)
      WRITE(3,2008) RW,RW1,RW0
2008  FORMAT(/,'     WIFE :',6X,F4.3,11X,F4.3,9X,F4.3)
      RETURN
      END
```

APPENDIX B: DOCUMENTATION AND COMPUTER PROGRAM FOR THE STACKELBERG MODEL

The computational procedures required for the estimates obtained in Chapter IV for the Stackelberg models involve the formulation of the log-likelihood function in the parameters involved and the maximization of this log-likelihood function given observations on the two jointly dependent dichotomous endogenous variables, and whatever exogenous explanatory variables are thought to affect the occurrence or nonoccurrence of the qualitative dependent variables. The program, consistent with the model described in Chapter III, assumes that the disturbance pair $(\varepsilon_h, \varepsilon_w)$ is bivariate normally distributed with zero means, unit variances and correlation coefficient ρ. For computational ease, the program provides for a grid search over possible values of ρ. In addition, to economize on the number of lines of code, four IMSL routines are used: MDNOR and MDBNOR, the univariate and bivariate normal cumulative distribution functions, respectively; ZXLSF, which provides the minimum of a one-dimensional function; and LINV3P which inverts a positive definite matrix. To control input and output, three files are required by the program. First, an input file called PARAM.DAT consists of a number of lines of control parameters; these parameters will be discussed later. Second, another input file, INPUT.DAT, is the user's data set.

INPUT.DAT (SUBROUTINE RDFILE2); as a check, the first four observations are printed out. The following steps are now performed for each value of the correlation coefficient, ρ, used in the grid search. SUBROUTINE INIT now calculates and prints the initial values of the coefficients and the initial value of the log-likelihood function. Control is now transferred to SUBROUTINE BHHH, the routine which performs the iterations for each value of the correlation coefficient. Now, for each iteration, SUBROUTINE QSCORE performs two tasks. First, the total score is calculated; that is, the individual score vectors are summed over the number of observations. This is performed using subroutines ISCORE00, ISCORE01, ISCORE10 and ISCORE11 since the functional form for the score will obviously depend on the values taken jointly by the dichotomous endogenous variables. Second, the outer product of each individual score vector is computed and these outer product matrices are then summed over the number of observations producing a new matrix Q. This is done with the aid of SUBROUTINE IQ. Next, IMSL routine LINV3P checks to see whether or not matrix Q is nonsingular. If Q is nonsingular, LINV3P calculates the inverse of the Q matrix and postmultiplies it by the score vector; in this case the BHHH algorithm will be used. If Q is singular, SUBROUTINE CHOICE divides each element of the score vector by its norm; in this case, the method of steepest ascent is used. Control is now transferred to SUBROUTINE OPTSTP where the optimal step size is calculated; once the step size if found, the new parameter values are calculated. SUBROUTINE QSCORE is now called again and SUBROUTINE TEST checks for convergence. If convergence is not attained, the preceding sequence will be repeated either until convergence is attained or, failing that, until the maximum number of iterations has been completed. If convergence is attained, routine LINV3P is again called to calculate the inverse of the Q matrix, the asymptotic covariance matrix. Control is then transferred to SUBROUTINE WRIE which prints out the number of iterations, the final score vector, the log-likelihood value, and the estimated coefficients along with the associated t-statistics.

The program described here is subject to a number of limitations, most of which may be relaxed easily. The program has been written for data sets with up to 2100 observations, where the behavior of the leader and follower in the Stackelberg game can each be estimated by up to 13 explanatory variables. Note also, as mentioned earlier, that the algorithm has no built-in constant term; if one wishes to include a constant term as an explanatory variable, a vector of "ones" should be included in the input file INPUT.DAT.

We now describe the control cards required by the program. Each control card should be entered on a separate line in PARAM.DAT.

1. First control card (4I2)

(a) INDS: One or zero, indicating whether or not the score is to be written after each iteration on output file OUTPUT.DAT.

(b) ILOG: One or zero, indicating whether or not the log-likelihood value is to be written out after each iteration; if ILOG = 0, the log-likelihood value is written only after the final iteration, whether or not convergence is achieved.

(c) IOPT: The value of this parameter, either zero or one, determines whether or not a fixed stepsize is automatically taken at each iteration. Although in theory a fixed stepsize (equal to one) is asymptotically efficient, in many numerical problems a variable stepsize is often required. The parameters required to control the stepsize are entered on the third and fourth control cards.

(d) IPARAM: One or zero, indicating whether or not the estimated parameters are to be written out after each iteration. IF IPARAM = 0, the parameter estimates are written out only at the end of the final iteration.

2. Second control card (3I2, E10.3)

(a) IVAR: One or zero, depending on whether or not the asymptotic covariance matrix is to be written after the final iteration.

(b) IFIT1: One or zero, depending on whether or not the predicted probabilities for each observation are to written after the final iteration.

(c) IFIT2: One or zero, depending on whether or not the observed and predicted two by two contingency tables are to be written after the final iteration.

(d) BADSC: It is occasionally the case that for particular values of the parameters, the score vector for certain observations may contain extremely large elements. When this occurs, it is often the case that the sum over the observations of the outer product of the score is a singular matrix. In such a case, the BHHH algorithm will not be used at that particular iteration. Therefore, by setting BADSC to a particular value, any observations for which one or more elements of the score vector exceeds BADSC will not be used in the calculation of the total score and Q matrix for that iteration. Note, however, that the score for each observation is checked against BADSC at each iteration; if an observation is deleted during a particular iteration, it is not necessarily deleted from successive iterations. If BADSC = 0, it will default to 1·E + 20.

3. Third control card (I2, 4F7.5, I3)

The parameters on this control card are used to determine the optimal step size when using the BHHH algorithm. Parameters (b) - (f) are arguments in IMSL subroutine ZXLSF, a routine for finding the maximum of a general one-dimensional function. If IOPT = 0, only STPSIZ1 needs to be determined; all other parameters on this control can be left blank. In this case, it is suggested that STPSIZE be set to one since a fixed full step is asymptotically efficient. When IOPT = 1, all parameters on this control card need to be set.

(a) NINC1: A positive integer, indicating how many times the log-likelihood function is to be evaluated over the maximum step size allowed.

(b) STPSIZ1: A positive real number, indicating the maximum step size allowed.

(c) STEP1: An order of magnitude estimate, either positive or negative, of the required change in STPSIZ1.

(d) BOUND1: A limit, which must be set to a positive number, on the amount by which STPSIZ1 may be changed from its initial value.

(e) XACC1: The required absolute accuracy in the final value of STPSIZ1. Normally there are points on either side of STPSIZ1 within a distance XACC1 at which the value of the log-likelihood function is no greater than the value of the log-likelihood function evaluated at STPSIZ1.

(f) MAXFN1: A limit, which must be set to a positive integer, on the number of attempts to find the optimal step size.

An example should help clarify things. Assume one wishes to allow a maximum possible step size of one at each iteration but would like to evaluate the log-likelihood function at step size increments of .25, .50, .75, and 1.0. Therefore, STPSIZE1 is set to 1.0 and NINC1 = 4. In addition, assume that one would like to search over the entire range of each interval, e.g., (0, .25), but also insure that the minimum step size taken is .0001. Therefore, STEP1 should be set at -.24990 and BOUND1 should be set at .24990. The IMSL documentation on ZXLSF suggests setting XACC1 = .001 and MAXFN1 = 50.

4. Fourth control card (I2, 4F7.5, I3)

The parameters on this control card are used to determine the optimal step size when using the method of steepest ascent algorithm. The description for these control cards is the same as for the third control card.

5. Fifth control card (I6, 5I3, F4.3)

(a) NOBS: Number of observations in data set INPUT.DAT.

(b) NHW: Total number of unique explanatory variables in data set INPUT.DAT.

(c) NH: Number of explanatory variables used to describe the behavior of the leader in the Stackelberg model.

(d) NW: Number of explanatory variables used to describe the behavior of the follower in the Stackelberg model.

(e) NRHO: Number of values of the correlation coefficient, ρ, used in the grid search.

(f) MITER: Limit on the number of iterations for each value of the correlation coefficient.

(g) EPS: The convergence criterion. The mean taken over the number of observations, for each element of the score must be less then or equal to EPS for convergence to be attained.

6. Sixth control card (20F4.2)

VRHO(I), I = 1, NRHO: The values of the correlation coefficient, ρ, which comprise the grid search. Up to 20 values are allowed.

7. Seventh control card (8E10.3)

XO(I), I = 1, NH+NW+3: Initial values of the parameters. Eight values are allowed per line.

8. Eighth control card (40I2)

KH(I), I = 1,NH: Positive integers corresponding to the column locations in INPUT.DAT of the explanatory variables for the Stackelberg leader.

9. Ninth control card (40I2)

KW(I), I=1,NW: Positive integers corresponding to the column locations in INPUT.DAT of the explanatory variables for the Stackelberg follower.

10. Tenth control card

Format in which data file INPUT.DAT is to be read. The format must be placed in parentheses, (•).

```
      DIMENSION X(30),Q(465),SCORE(30),XO(30),VRHO(20)
      DIMENSION XH(2100,13),XW(2100,13),X1(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100),KH(13),KW(13),MAXFN
      REAL STEP,BOUND,XACC
      CHARACTER*4 RDFT(20)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT,/CMN2/ RATIO
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      OPEN(UNIT=1,FILE='PARAM.DAT',STATUS='OLD',ERR=10)
      OPEN(UNIT=2,FILE='INPUT.DAT',STATUS='OLD',ERR=10)
      OPEN(UNIT=3,FILE='OUTPUT.DAT',STATUS='OLD',ERR=10)
      CALL RDFILE1(NOBS,NHW,NH,NW,KH,KW,XO,RDFT,
     1MITER,EPS,INDS,ILOG,IOPT,IPARAM,IVAR,IFIT1,IFIT2,NRHO,VRHO,
     2NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1,
     3NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,BADSC)
      CALL RDFILE2(NOBS,NHW,NH,NW,KH,KW,RDFT,XH,XW,
     1  NYH,NYW)
      DO 11 KRHO=1,NRHO
        RHO=VRHO(KRHO)
        WRITE(3,1000) KRHO,RHO
1000    FORMAT(//,' CASE ',I2,' RHO = ',F5.3)
        CALL INIT(XO,VLIKE,ILOG,IOPT)
        CALL BHHH(MITER,EPS,Q,VLIKE,NITER,INDS,ILOG,IOPT,IPARAM,IER,
     2  NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1,STPSIZE,
     3  NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,BADSC)
        CALL WRTE(NH,NW,VLIKE,STPSIZE,NITER,Q,IVAR,IER)
        IF(IFIT1.EQ.1.OR.IFIT2.EQ.1) CALL FIT(VLIKE,IFIT1,IFIT2)
11    CONTINUE
10    CONTINUE
      STOP
      END

      SUBROUTINE RDFILE1(NOBS,NHW,NH,NW,KH,KW,XO,RDFT,
     1MITER,EPS,INDS,ILOG,IOPT,IPARAM,IVAR,IFIT1,IFIT2,NRHO,VRHO,
     2NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1,
     3NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,BADSC)
      DIMENSION XO(30),VRHO(20),X(30),X1(30),SCORE(30)
      INTEGER KH(13),KW(13),MAXFN
      REAL STEP,BOUND,XACC
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      CHARACTER*4 RDFT(20)
      READ(1,1003) INDS,ILOG,IOPT,IPARAM
 1003 FORMAT(4I2)
      READ(1,1004) IVAR,IFIT1,IFIT2,BADSC
 1004 FORMAT(3I2,E10.3)
      READ(1,1005) NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1
 1005 FORMAT(I2,4F7.5,I3)
      READ(1,1006) NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2
 1006 FORMAT(I2,4F7.5,I3)
      IF(BADSC.EQ.0.) BADSC=1.E+20
      READ(1,1000) NOBS,NHW,NH,NW,NRHO,MITER,EPS
```

```
1000 FORMAT(I6,5I3,F4.3)
     READ(1,2000) (VRHO(I),I=1,NRHO)
2000 FORMAT(20F4.2)
     N=NH+NW+3
     M=N/8
     IF(M.EQ.0) GO TO 11
       DO 15 IM=1,M
         M1=(IM-1)*8+1
         M2=IM*8
         READ(1,5000) (XO(J),J=M1,M2)
5000     FORMAT(8E10.3)
15     CONTINUE
11   M3=M*8+1
     IF(M3.GT.N) GO TO 12
       READ(1,5000) (XO(J),J=M3,N)
12   IF(NH.EQ.0) GO TO 10
       READ(1,3000) (KH(J),J=1,NH)
10   IF(NW.EQ.0) GO TO 20
       READ(1,3000) (KW(J),J=1,NW)
3000   FORMAT(40I2)
20   CONTINUE
     READ(1,4000) RDFT
4000 FORMAT(20A4)
     CALL WRTRD1(NOBS,NHW,NH,NW,MITER,EPS,RDFT)
     RETURN
     END

     SUBROUTINE WRTRD1(NOBS,NHW,NH,NW,MITER,EPS,RDFT)
     CHARACTER*4 RDFT(20)
     WRITE(3,1001)
1001 FORMAT(' NOBS ','NHW',' NH',' NW',' MITER',' EPS')
     WRITE(3,1002) NOBS,NHW,NH,NW,MITER,EPS
1002 FORMAT(I6,3I3,3X,I3,F4.3)
     WRITE(3,4001) RDFT
4001 FORMAT(/,'READING FORMAT :',20A4)
     RETURN
     END

     SUBROUTINE RDFILE2(NOBS,NHW,NH,NW,KH,KW,RDFT,XH,XW,
    1                  NYH,NYW)
     DIMENSION XX(30),XH(2100,13),XW(2100,13)
     INTEGER KH(13),KW(13),NYH(2100),NYW(2100)
     CHARACTER*4 RDFT(20)
     DO 100 I=1,NOBS
       READ(2,RDFT) NYH(I),NYW(I),(XX(J),J=1,NHW)
       IF(NH.EQ.0) GO TO 10
       DO 110 J=1,NH
         KJ=KH(J)
         XH(I,J)=XX(KJ)
110    CONTINUE
```

```
 10     IF(NW.EQ.0) GO TO 100
        DO 120 J=1,NW
          KJ=KW(J)
          XW(I,J)=XX(KJ)
120     CONTINUE
100   CONTINUE
      CALL WRTOBS(NH,NW,NYH,NYW,XH,XW)
      RETURN
      END

      SUBROUTINE WRTOBS(NH,NW,NYH,NYW,XH,XW)
      DIMENSION NYH(2100),NYW(2100),XH(2100,13),XW(2100,13)
      DIMENSION XXH(13),XXW(13)
      WRITE(3,1003)
1003  FORMAT(/,'FIRST 4 OBSERVATIONS :')
      DO 100 I=1,4
        IYH=NYH(I)
        IYW=NYW(I)
        WRITE(3,1000) I,IYH,IYW
1000    FORMAT(/,'OBSERVATION ',I4,' :',' IYH =',I2,' IYW =',I2)
        IF(NH.EQ.0) GO TO 10
        DO 110 J=1,NH
          XXH(J)=XH(I,J)
110     CONTINUE
        WRITE(3,1001)
1001    FORMAT(/,'  XH :')
        CALL WRTVEC(NH,XXH)
 10     IF(NW.EQ.0) GO TO 100
        DO 120 J=1,NW
          XXW(J)=XW(I,J)
120     CONTINUE
        WRITE(3,1002)
1002    FORMAT(/,'  XW :')
        CALL WRTVEC(NW,XXW)
 100  CONTINUE
      RETURN
      END

      SUBROUTINE INIT(XO,VLIKE,ILOG,IOPT)
      DIMENSION XH(2100,13),XW(2100,13),XO(30),X(30),X1(30),SCORE(30)
      DIMENSION PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DO 20 J=1,N
        X(J)=XO(J)
20    CONTINUE
      WRITE(3,2001)
2001  FORMAT(/,'STARTING VALUES :')
      CALL WRTVEC(N,X)
```

```
        CALL VALUE(N,X,VLIKE)
      IF(ILOG.EQ.0) GO TO 14
        WRITE(3,3000) VLIKE
3000    FORMAT(/,'INITIAL LOG-LIKELIHOOD VALUE = ',E13.6)
14    RETURN
      END

      SUBROUTINE VALUE(N,X,VLIKE)
      DIMENSION XH(2100,13),XW(2100,13),X(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      VLIKE=0.
      DO 100 I=1,NOBS
        INYH=NYH(I)
        INYW=NYW(I)
        CALL DELTA(I,DELTAH,DELTAW)
        IF (INYH.EQ.1.OR.INYW.EQ.1) GO TO 11
          CALL IPROB00(DELTAH,DELTAW,PR00)
          PROBHAT(I)=PR00
          CALL TOL(PR00)
          VLIKE=VLIKE+ALOG(PR00)
          GO TO 99
11      IF (INYH.EQ.0.OR.INYW.EQ.1) GO TO 12
          CALL IPROB10(DELTAH,DELTAW,PR10)
          PROBHAT(I)=PR10
          CALL TOL(PR10)
          VLIKE=VLIKE+ALOG(PR10)
          GO TO 99
12      IF (INYH.EQ.1.OR.INYW.EQ.0) GO TO 13
          CALL IPROB01(DELTAH,DELTAW,PR01)
          PROBHAT(I)=PR01
          CALL TOL(PR01)
          VLIKE=VLIKE+ALOG(PR01)
          GO TO 99
13      CALL IPROB11(DELTAH,DELTAW,PR11)
        PROBHAT(I)=PR11
        CALL TOL(PR11)
        VLIKE=VLIKE+ALOG(PR11)
99      CONTINUE
100   CONTINUE
      RETURN
      END

      SUBROUTINE DELTA(I,DELTAH,DELTAW)
      DIMENSION X(30),XH(2100,13),XW(2100,13),X1(30),SCORE(30)
      DIMENSION PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DELTAH=0.0
```

```
      DELTAW=0.0
      DO 100 J=1,NH
        JH=J+3
        DELTAH=DELTAH+X(JH)*XH(I,J)
100   CONTINUE
      DO 200 J=1,NW
        JW=J+NH+3
        DELTAW=DELTAW+X(JW)*XW(I,J)
200   CONTINUE
      RETURN
      END

      SUBROUTINE BHHH(MITER,EPS,Q,VLIKE,NITER,INDS,ILOG,IOPT,IPARAM,IER,
     2NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1,STPSIZE,
     3NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,BADSC)
      DIMENSION XH(2100,13),XW(2100,13),X(30),Q(465),SCORE(30),
     1SCORE1(30),X1(30),PROBHAT(2100),D(30),Z(30,30),WK(495),QEIG(465)
      INTEGER NYH(2100),NYW(2100),MAXFN
      REAL STEP,BOUND,XACC
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      NITER=999
      NN=N*(N+1)/2
      CALL QSCORE(Q,BADSC)
      IF(INDS.EQ.0) GO TO 1002
        CALL WRTBH1
1002  CONTINUE
      DO 10 ITER=1,MITER
        WRITE(3,1000) ITER
1000    FORMAT(/,' ITERATION ',I3,'  STARTS')
        DO 70 I=1,N
          SCORE1(I)=SCORE(I)
70      CONTINUE
        CALL AVEC(NN,A,Q)
        CALL LINV3P(Q,SCORE1,3,N,IER)
        CALL AVEC1(N,A,SCORE1)
        CALL AVEC(NN,A,Q)
        CALL CHOICE(IER,SCORE1,NINC1,STPSIZ1,STEP1,BOUND1,XACC1,MAXFN1,
     1               NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,
     2               NINC,STPSIZ0,STEP,BOUND,XACC,MAXFN)
        CALL OPTSTP(NINC,STPSIZ0,STPSIZE,VLIKE,ILOG,IOPT,STEP,BOUND,
     1               XACC,MAXFN,IER)
        CALL QSCORE(Q,BADSC)
        CALL TEST(NOBS,EPS,IEND)
        IF (IEND.EQ.1) GO TO 20
        IF(ITER.EQ.MITER) GO TO 11
          CALL WRTBHH(ITER,STPSIZE,VLIKE,INDS,ILOG,IPARAM,IOPT,IER)
11      CONTINUE
10    CONTINUE
      GO TO 50
```

```
20     NITER=ITER
50     IF(ILOG.EQ.1.OR.IOPT.EQ.1) GO TO 40
         CALL VALUE(N,X,VLIKE)
40     IF(IOPT.EQ.0) GO TO 60
         WRITE(3,6000) IER
6000     FORMAT(/,'LAST STEP OPTIMIZATION, IER =',I3)
60     CONTINUE
       CALL AVEC(NN,A,Q)
       CALL LINV3P(Q,SCORE,1,N,IER)
       CALL AVEC(NN,A,Q)
       IF(IER.EQ.129.OR.IER.EQ.130) IER=128
       IF(IER.EQ.128) GO TO 999
       GO TO 998
999    WRITE(3,1001) IER
1001   FORMAT(/,1X,'IER = ',I3,
      1/,'THE Q MATRIX IS SINGULAR OR ILL-CONDITIONED',
      2/,'THE Q MATRIX BEFORE INVERSION IS :')
       CALL WRTVAR(Q)
998    RETURN
       END

       SUBROUTINE AVEC(NDIM,A,VEC)
       DIMENSION VEC(465)
       A=1.E+04
       DO 100 I=1,NDIM
       VEC(I)=A*VEC(I)
100    CONTINUE
       RETURN
       END

       SUBROUTINE AVEC1(NDIM,A,VEC)
       DIMENSION VEC(30)
       A=1.E+04
       DO 100 I=1,NDIM
       VEC(I)=A*VEC(I)
100    CONTINUE
       RETURN
       END

       SUBROUTINE CHOICE(IER,SCORE1,NINC1,STPSIZ1,STEP1,BOUND1,XACC1,
      1           MAXFN1,NINC2,STPSIZ2,STEP2,BOUND2,XACC2,MAXFN2,
      2           NINC,STPSIZ0,STEP,BOUND,XACC,MAXFN)
       DIMENSION X(30),X1(30),SCORE(30),SCORE1(30)
       COMMON /CMN3/ N,RHO,X,X1,SCORE
       IF (IER.NE.0) GO TO 10
         WRITE(3,1000)
1000     FORMAT(/,' BHHH ALGORITHM ')
         NINC=NINC1
         STPSIZ0=STPSIZ1
         STEP=STEP1
```

```
          BOUND=BOUND1
          XACC=XACC1
          MAXFN=MAXFN1
          DO 20 I=1,N
            SCORE(I)=SCORE1(I)
 20       CONTINUE
          GO TO 30
 10     WRITE(3,2000)
2000    FORMAT(/,' SCORE ALGORITHM ')
        NINC=NINC2
        STPSIZ0=STPSIZ2
        STEP=STEP2
        BOUND=BOUND2
        XACC=XACC2
        MAXFN=MAXFN2
        CALL NORM(N,SCORE,XNSCOR)
        DO 40 I=1,N
          SCORE(I)=SCORE(I)/XNSCOR
 40     CONTINUE
 30     RETURN
        END

        SUBROUTINE QSCORE(Q,BADSC)
        DIMENSION XH(2100,13),XW(2100,13),X(30),Q(465),SCORE(30)
        DIMENSION QI(465),SCOREI(30),X1(30),PROBHAT(2100)
        INTEGER NYH(2100),NYW(2100)
        COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
        COMMON /CMN3/ N,RHO,X,X1,SCORE
        NN=N*(N+1)/2
        DO 400 J=1,N
          SCORE(J)=0.
400     CONTINUE
        DO 500 K=1,NN
          Q(K)=0.
500     CONTINUE
        NBAD=0
        DO 100 I=1,NOBS
          CALL ISCORE(I,SCOREI)
          CALL IQ(SCOREI,QI)
          DO 101 J=1,N
            IF(SCOREI(J).GE.BADSC) GO TO 102
101       CONTINUE
          DO 200 J=1,N
            SCORE(J)=SCORE(J)+SCOREI(J)
200       CONTINUE
          DO 300 K=1,NN
            Q(K)=Q(K)+QI(K)
300       CONTINUE
          GO TO 100
102       WRITE(3,1000) I
```

```
1000   FORMAT(/,' BAD OBSERVATION = ',I4)
       CALL WRTVEC(N,SCOREI)
       NBAD=NBAD+1
100  CONTINUE
     WRITE(3,2000) NBAD
2000 FORMAT(/,' NUMBER OF BAD OBSERVATIONS :',I6)
     RETURN
     END

     SUBROUTINE WRTBHH(ITER,STPSIZE,VLIKE,INDS,ILOG,IPARAM,IOPT,IER)
     DIMENSION X(30),SCORE(30),X1(30)
     COMMON /CMN3/ N,RHO,X,X1,SCORE
     IF(INDS.EQ.0.AND.ILOG.EQ.0.AND.IPARAM.EQ.0)
    1  GO TO 11
       CALL WRTBH2(ITER,STPSIZE,VLIKE,ILOG,IPARAM,IOPT,IER)
       IF(INDS.EQ.1) CALL WRTBH1
11   RETURN
     END

     SUBROUTINE WRTBH1
     DIMENSION SCORE(30),X(30),X1(30)
     COMMON /CMN3/ N,RHO,X,X1,SCORE
     WRITE(3,4000)
4000 FORMAT(/,'    SCORE    : ')
     CALL WRTVEC(N,SCORE)
     CALL WRTVEC1(N,SCORE)
     RETURN
     END

     SUBROUTINE OPTSTP(NINC,STPSIZ0,STPSIZE,VLIKE,ILOG,IOPT,STEP,BOUND,
    1  XACC,MAXFN,IER)
     DIMENSION XH(2100,13),XW(2100,13),X(30),X1(30),SCORE(30)
     DIMENSION PROBHAT(2100)
     INTEGER NYH(2100),NYW(2100),MAXFN,IER
     REAL STPSIZE,STEP,BOUND,XACC,MLOGLIK
     COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
     COMMON /CMN3/ N,RHO,X,X1,SCORE
     EXTERNAL MLOGLIK
     IF(IOPT.EQ.0) GO TO 10
       STPINC=STPSIZ0/NINC
       INCMAX=0
       VLIKEM=VLIKE
       DO 20 J=1,N
         X1(J)=X(J)
20     CONTINUE
       DO 60 INC=1,NINC
         STPSIZE=INC*STPINC
         DO 70 J=1,N
           X(J)=X1(J)+STPSIZE*SCORE(J)
70       CONTINUE
```

```
          CALL VALUE(N,X,VLIKE)
          WRITE(3,1000) STPSIZE,VLIKE
1000      FORMAT(' STEPSIZE =',F7.5,'  LOGLIK =',E13.6)
          IF(VLIKE.LE.VLIKEM) GO TO 80
            INCMAX=INC
            VLIKEM=VLIKE
80        CONTINUE
60      CONTINUE
        IF(INCMAX.EQ.0) GO TO 90
          STPSIZE=INCMAX*STPINC
          CALL ZXLSF(MLOGLIK,STPSIZE,STEP,BOUND,XACC,MAXFN,IER)
          GO TO 200
90      STPSIZE=STPINC/2
        STPO=STEP/2
        BOUNDO=BOUND/2
        CALL ZXLSF(MLOGLIK,STPSIZE,STPO,BOUNDO,XACC,MAXFN,IER)
200     DO 40 J=1,N
          X(J)=X1(J)
40      CONTINUE
        GO TO 50
10    STPSIZE=STPSIZO
50    CONTINUE
      DO 30 J=1,N
        X(J)=X(J)+STPSIZE*SCORE(J)
30    CONTINUE
      IF(ILOG.EQ.0.OR.IOPT.EQ.0) GO TO 100
        CALL VALUE(N,X,VLIKE)
100   CONTINUE
      RETURN
      END

      SUBROUTINE NORM(NV,VECTOR,XNVEC)
      DIMENSION VECTOR(30)
      XNVEC=0.
      DO 100 I=1,NV
      XNVEC=XNVEC+VECTOR(I)**2
100   CONTINUE
      XNVEC=SQRT(XNVEC)
      RETURN
      END

      SUBROUTINE TEST(NOBS,EPS,IEND)
      DIMENSION SCORE(30),X(30),X1(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      IEND=0
      DO 10 J=1,N
        A=ABS(SCORE(J))/NOBS
        IF (A.GT.EPS) GO TO 20
10    CONTINUE
      IEND=1
```

```
20    RETURN
      END

      SUBROUTINE WRTBH2(ITER,STPSIZE,VLIKE,ILOG,IPARAM,IOPT,IER)
      DIMENSION X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      WRITE(3,1000) ITER
1000  FORMAT(//,'ITERATION ',I3)
      IF(ILOG.EQ.0) GO TO 10
        WRITE(3,2000) STPSIZE,VLIKE
2000    FORMAT(/,'STEPSIZE = ',F7.5,'   LOG-LIKELIHOOD VALUE = ',E13.6)
        IF(IOPT.EQ.0) GO TO 10
          WRITE(3,2001) IER
2001      FORMAT(/,' STEP OPTIMIZATION, IER = ',I3)
10    IF(IPARAM.EQ.0) GO TO 20
        WRITE(3,3000)
3000    FORMAT(/,'PARAMETERS :')
        CALL WRTVEC(N,X)
        CALL WRTVEC1(N,X)
20    RETURN
      END

      SUBROUTINE ISCORE(I,SCOREI)
      DIMENSION XH(2100,13),XW(2100,13),X(30),SCOREI(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100),X1(30),SCORE(30)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      INYH=NYH(I)
      INYW=NYW(I)
      CALL DELTA(I,DELTAH,DELTAW)
      IF (INYH.EQ.1.OR.INYW.EQ.1) GO TO 11
        CALL IPROB00(DELTAH,DELTAW,PR00)
        CALL ISCORE00(I,DELTAH,DELTAW,PR00,SCOREI)
        GO TO 14
11    IF (INYH.EQ.0.OR.INYW.EQ.1) GO TO 12
        CALL IPROB10(DELTAH,DELTAW,PR10)
        CALL ISCORE10(I,DELTAH,DELTAW,PR10,SCOREI)
        GO TO 14
12    IF (INYH.EQ.1.OR.INYW.EQ.0) GO TO 13
        CALL IPROB01(DELTAH,DELTAW,PR01)
        CALL ISCORE01(I,DELTAH,DELTAW,PR01,SCOREI)
        GO TO 14
13    CALL IPROB11(DELTAH,DELTAW,PR11)
      CALL ISCORE11(I,DELTAH,DELTAW,PR11,SCOREI)
14    CONTINUE
      RETURN
      END

      SUBROUTINE IPROB00(DELTAH,DELTAW,PR00)
      DIMENSION X(30),X1(30),SCORE(30)
```

```
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(2)
      W2=-DELTAW-X(3)
      CALL MDBNOR(H1,W1,RHO,PROO,IER)
      X4=X(3)
      IF (X4.LT.0.) GO TO 13
        CALL INTEGRAL(H1,W1,H2,W2,PROB,IER)
          PROO=PROO-PROB
13    RETURN
      END

      SUBROUTINE ISCORE00(I,DELTAH,DELTAW,PROO,SCOREI)
      DIMENSION XH(2100,13),XW(2100,13),X(30),SCOREI(30)
      DIMENSION DI(30),X1(30),SCORE(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      SCOREI(1)=0.
      SCOREI(2)=0.
      SCOREI(3)=0.
      H1=DELTAH
      ZH=-DELTAW+RHO*DELTAH
      CALL PARTIAL(H1,ZH,PARTH)
      PARTH=-PARTH
      DO 10 J=1,NH
        JH=J+3
        SCOREI(JH)=PARTH*XH(I,J)
10    CONTINUE
      W1=DELTAW
      ZW=-DELTAH+RHO*DELTAW
      CALL PARTIAL(W1,ZW,PARTW)
      PARTW=-PARTW
      DO 20 J=1,NW
        JW=J+NH+3
        SCOREI(JW)=PARTW*XW(I,J)
20    CONTINUE
      X4=X(3)
      IF (X4.LT.0.) GO TO 30
        CALL PARTLI00(I,DELTAH,DELTAW,DI)
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-DI(J)
31        CONTINUE
30    CONTINUE
      CALL TOL(PROO)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PROO
        CALL TOL1(SCOREI(J),ITRUN)
        CALL BIGN1(SCOREI(J),ITRUN)
```

```
        IF(ITRUN.GT.0) WRITE(*,1000) I,J,ITRUN
1000    FORMAT(' OBSERVATION ',I6,' PARAMETER ',I3,' ITRUN ',I2)
60      CONTINUE
        RETURN
        END

        SUBROUTINE IPROB10(DELTAH,DELTAW,PR10)
        DIMENSION X(30),X1(30),SCORE(30)
        COMMON /CMN3/ N,RHO,X,X1,SCORE
        XNEGRHO=-RHO
        H1=-DELTAH
        W1=-DELTAW
        H2=-DELTAH+X(1)
        W2=-DELTAW-X(3)
        CALL MDBNOR(DELTAH,W2,XNEGRHO,PR10,IER)
        X4=X(3)
        IF (X4.GE.0.) GO TO 13
          CALL INTEGRAL(H1,W2,H2,W1,PROB,IER)
            PR10=PR10+PROB
13      RETURN
        END

        SUBROUTINE ISCORE10(I,DELTAH,DELTAW,PR10,SCOREI)
        DIMENSION XH(2100,13),XW(2100,13),X(30),SCOREI(30)
        DIMENSION DI(30),X1(30),SCORE(30),PROBHAT(2100)
        INTEGER NYH(2100),NYW(2100)
        COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
        COMMON /CMN3/ N,RHO,X,X1,SCORE
        ARGH=DELTAH
        ARGW=DELTAW+X(3)
        SCOREI(1)=0.
        SCOREI(2)=0.
        W2=DELTAW+X(3)
        ZW=DELTAH-RHO*W2
        CALL PARTIAL(W2,ZW,PARTW)
        SCOREI(3)=-PARTW
        ZH=-W2+RHO*DELTAH
        CALL PARTIAL(DELTAH,ZH,PARTH)
        DO 10 J=1,NH
          JH=J+3
          SCOREI(JH)=PARTH*XH(I,J)
10      CONTINUE
        PARTW=-PARTW
        DO 20 J=1,NW
          JW=J+NH+3
          SCOREI(JW)=PARTW*XW(I,J)
20      CONTINUE
        X4=X(3)
        IF (X4.GE.0.) GO TO 30
          CALL PARTLI10(I,DELTAH,DELTAW,DI)
```

```
        DO 31 J=1,N
          SCOREI(J)=SCOREI(J)+DI(J)
31      CONTINUE
30    CONTINUE
      CALL TOL(PR10)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PR10
        CALL TOL1(SCOREI(J),ITRUN)
        CALL BIGN1(SCOREI(J),ITRUN)
60    CONTINUE
      RETURN
      END

      SUBROUTINE IPROB01 DELTAH,DELTAW,PR01)
      DIMENSION X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      XNEGRHO=-RHO
      H1=-DELTAH+X(1)
      W1=-DELTAW
      H2=-DELTAH-X(2)+X(1)
      W2=-DELTAW-X(3)
      CALL MDBNOR(H2,DELTAW,XNEGRHO,PR01,IER)
      X4=X(3)
      IF (X4.GE.0.) GO TO 13
        CALL INTEGRAL(H1,W2,H2,W1,PROB,IER)
          PR01=PR01+PROB
13    RETURN
      END

      SUBROUTINE ISCORE01(I,DELTAH,DELTAW,PR01,SCOREI)
      DIMENSION XH(2100,13),XW(2100,13),X(30),SCOREI(30)
      DIMENSION DI(30),X1(30),SCORE(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      ARGH=DELTAH+X(2)-X(1)
      ARGW=DELTAW
      H2=DELTAH+X(2)-X(1)
      H3=-H2
      ZW=DELTAW-RHO*H2
      CALL PARTIAL(H2,ZW,PARTW)
      SCOREI(1)=PARTW
      SCOREI(2)=-PARTW
      SCOREI(3)=0.
      DO 10 J=1,NH
        JH=J+3
        SCOREI(JH)=-PARTW*XH(I,J)
10    CONTINUE
      ZH=H3+RHO*DELTAW
      CALL PARTIAL(DELTAW,ZH,PARTH)
```

```
      DO 20 J=1,NW
        JW=J+NH+3
        SCOREI(JW)=PARTH*XW(I,J)
 20   CONTINUE
      X4=X(3)
      IF (X4.GE.0.) GO TO 30
        CALL PARTLI01(I,DELTAH,DELTAW,DI)
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)+DI(J)
 31       CONTINUE
30    CONTINUE
      CALL TOL(PR01)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PR01
        CALL TOL1(SCOREI(J),ITRUN)
        CALL BIGN1(SCOREI(J),ITRUN)
 60   CONTINUE
      RETURN
      END

      SUBROUTINE IPROB11(DELTAH,DELTAW,PR11)
      DIMENSION X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH-X(2)
      W1=-DELTAW
      H2=-DELTAH-X(2)+X(1)
      W2=-DELTAW-X(3)
      H3=-H2
      W3=-W2
      CALL MDBNOR(H3,W3,RHO,PR11,IER)
      X4=X(3)
      IF (X4.LT.0.) GO TO 13
        CALL INTEGRAL(H1,W1,H2,W2,PROB,IER)
          PR11=PR11-PROB
 13   RETURN
      END

      SUBROUTINE ISCORE11(I,DELTAH,DELTAW,PR11,SCOREI)
      DIMENSION XH(2100,13),XW(2100,13),X(30),SCOREI(30)
      DIMENSION DI(30),X1(30),SCORE(30),PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      ARGH=DELTAH+X(2)-X(1)
      ARGW=DELTAW+X(3)
      H3=DELTAH+X(2)-X(1)
      W3=DELTAW+X(3)
      ZH=W3-RHO*H3
      CALL PARTIAL(H3,ZH,PARTH)
      SCOREI(1)=-PARTH
```

```
      SCOREI(2)=PARTH
      ZW=H3-RHO*W3
      CALL PARTIAL(W3,ZW,PARTW)
      SCOREI(3)=PARTW
      DO 10 J=1,NH
        JH=J+3
        SCOREI(JH)=PARTH*XH(I,J)
10    CONTINUE
      DO 20 J=1,NW
        JW=J+NH+3
        SCOREI(JW)=PARTW*XW(I,J)
20    CONTINUE
      X4=X(3)
      IF (X4.LT.0.) GO TO 30
        CALL PARTLI11(I,DELTAH,DELTAW,DI)
          DO 31 J=1,N
            SCOREI(J)=SCOREI(J)-DI(J)
31        CONTINUE
30    CONTINUE
      CALL TOL(PR11)
      DO 60 J=1,N
        SCOREI(J)=SCOREI(J)/PR11
        CALL TOL1(SCOREI(J),ITRUN)
        CALL BIGN1(SCOREI(J),ITRUN)
60    CONTINUE
      RETURN
      END

      SUBROUTINE TOL(XPROB)
      XMIN=1.E-20
      IF(XPROB.LE.XMIN) XPROB=XMIN
      RETURN
      END

      SUBROUTINE TOL1(XNUM,ITRUN)
      ITRUN=0
      ASML=1.0E-20
      ANUM=ABS(XNUM)
      IF(ANUM.GE.ASML) GOTO 100
        ITRUN=1
        SIGN=1.
        IF(XNUM.LT.0.) SIGN=-1.
        XNUM=SIGN*ASML
  100 RETURN
      END

      SUBROUTINE IQ(SCOREI,QI)
      DIMENSION SCOREI(30),QI(465),X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      IR=0
```

```
      DO 10 IX=1,N
        DO 20 JX=1,IX
          IL=IR+JX
          QI(IL)=SCOREI(IX)*SCOREI(JX)
20      CONTINUE
        IR=IR+IX
10    CONTINUE
      NN=N*(N+1)/2
      RETURN
      END

      SUBROUTINE PARTIAL(Y,Z,PART)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DIMENSION X(30),X1(30),SCORE(30)
      SNUM=SQRT(1.-RHO**2)
      CALL TOL(SNUM)
      ZX=Z/SNUM
      XNUM=-Y**2/2.
      CALL BIGN(XNUM)
      DENSY=(EXP(XNUM))/SQRT(2*ACOS(-1.))
      CALL MDNOR(ZX,PHIZX)
      PART=DENSY*PHIZX
      RETURN
      END

      SUBROUTINE PARTLI00(I,DELTAH,DELTAW,DI)
      DIMENSION DI(30),XH(2100,13),XW(2100,13),X(30)
      DIMENSION NYH(2100),NYW(2100),X1(30),SCORE(30),PROBHAT(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(2)
      W2=-DELTAW-X(3)
      H3=-H2
      W3=-W2
      Z1=W1+RHO*H3
      Z2=W2+RHO*H3
      Z3=H1+RHO*W3
      Z4=H2+RHO*W3
      Z5=W1+RHO*DELTAH
      Z6=W2+RHO*DELTAH
      Z7=H1+RHO*DELTAW
      Z8=H2+RHO*DELTAW
      CALL PARTIAL(H2,Z1,PART1)
      CALL PARTIAL(H2,Z2,PART2)
      CALL PARTIAL(W2,Z3,PART3)
      CALL PARTIAL(W2,Z4,PART4)
      CALL PARTIAL(H1,Z5,PART5)
      CALL PARTIAL(H1,Z6,PART6)
```

```
      CALL PARTIAL(W1,Z7,PART7)
      CALL PARTIAL(W1,Z8,PART8)
      DI(1)=0.
      DI(2)=PART1-PART2
      DI(3)=PART3-PART4
      PARTH=-PART5+PART6+PART1-PART2
      DO 10 J=1,NH
        JH=J+3
        DI(JH)=PARTH*XH(I,J)
10    CONTINUE
      PARTW=-PART7+PART3-PART4+PART8
      DO 20 J=1,NW
        JW=J+3+NH
        DI(JW)=PARTW*XW(I,J)
20    CONTINUE
      RETURN
      END

      SUBROUTINE PARTLI10(I,DELTAH,DELTAW,DI)
      DIMENSION DI(30),XH(2100,13),XW(2100,13),X(30)
      DIMENSION NYH(2100),NYW(2100),X1(30),SCORE(30),PROBHAT(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH+X(1)
      W2=-DELTAW-X(3)
      H3=-H2
      W3=-W2
      Z1=W2+RHO*H3
      Z2=W1+RHO*H3
      Z3=H1+RHO*W3
      Z4=H2+RHO*W3
      Z5=W2+RHO*DELTAH
      Z6=W1+RHO*DELTAH
      Z7=H1+RHO*DELTAW
      Z8=H2+RHO*DELTAW
      CALL PARTIAL(H2,Z1,PART1)
      CALL PARTIAL(H2,Z2,PART2)
      CALL PARTIAL(W2,Z3,PART3)
      CALL PARTIAL(W2,Z4,PART4)
      CALL PARTIAL(H1,Z5,PART5)
      CALL PARTIAL(H1,Z6,PART6)
      CALL PARTIAL(W1,Z7,PART7)
      CALL PARTIAL(W1,Z8,PART8)
      DI(1)=-PART1+PART2
      DI(2)=0.
      DI(3)=-PART3+PART4
      PARTH=-PART5+PART1+PART6-PART2
```

```
      DO 10 J=1,NH
        JH=J+3
        DI(JH)=PARTH*XH(I,J)
10    CONTINUE
      PARTW=PART7-PART3+PART4-PART8
      DO 20 J=1,NW
        JW=J+3+NH
        DI(JW)=PARTW*XW(I,J)
20    CONTINUE
      RETURN
      END


      SUBROUTINE PARTLI01(I,DELTAH,DELTAW,DI)
      DIMENSION DI(30),XH(2100,13),XW(2100,13),X(30)
      DIMENSION NYH(2100),NYW(2100),X1(30),SCORE(30),PROBHAT(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH+X(1)
      H3=-DELTAH-X(2)+X(1)
      W2=-DELTAW-X(3)
      H4=-H2
      H5=-H3
      W3=-W2
      Z1=W2+RHO*H4
      Z2=W1+RHO*H4
      Z3=W2+RHO*H5
      Z4=W1+RHO*H5
      Z5=H2+RHO*W3
      Z6=H3+RHO*W3
      Z7=H2+RHO*DELTAW
      Z8=H3+RHO*DELTAW
      CALL PARTIAL(H2,Z1,PART1)
      CALL PARTIAL(H2,Z2,PART2)
      CALL PARTIAL(H3,Z3,PART3)
      CALL PARTIAL(H3,Z4,PART4)
      CALL PARTIAL(W3,Z5,PART5)
      CALL PARTIAL(W3,Z6,PART6)
      CALL PARTIAL(W1,Z7,PART7)
      CALL PARTIAL(W1,Z8,PART8)
      DI(1)=PART1-PART2-PART3+PART4
      DI(2)=PART3-PART4
      DI(3)=-PART5+PART6
      PARTH=-PART4-PART1+PART3+PART2
      DO 10 J=1,NH
        JH=J+3
        DI(JH)=PARTH*XH(I,J)
10    CONTINUE
```

```
      PARTW=-PART5-PART8+PART7+PART6
      DO 20 J=1,NW
        JW=J+3+NH
        DI(JW)=PARTW*XW(I,J)
20    CONTINUE
      RETURN
      END

      SUBROUTINE PARTLI11(I,DELTAH,DELTAW,DI)
      DIMENSION DI(30),XH(2100,13),XW(2100,13),X(30)
      DIMENSION NYH(2100),NYW(2100),X1(30),SCORE(30),PROBHAT(2100)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      H1=-DELTAH
      W1=-DELTAW
      H2=-DELTAH-X(2)
      H3=-DELTAH-X(2)+X(1)
      W2=-DELTAW-X(3)
      H4=-H2
      H5=-H3
      W3=-W2
      Z1=W1+RHO*H5
      Z2=W2+RHO*H5
      Z3=W1+RHO*H4
      Z4=W2+RHO*H4
      Z5=H2+RHO*W3
      Z6=H3+RHO*W3
      Z7=H2+RHO*DELTAW
      Z8=H3+RHO*DELTAW
      CALL PARTIAL(H5,Z1,PART1)
      CALL PARTIAL(H5,Z2,PART2)
      CALL PARTIAL(H4,Z3,PART3)
      CALL PARTIAL(H4,Z4,PART4)
      CALL PARTIAL(W3,Z5,PART5)
      CALL PARTIAL(W3,Z6,PART6)
      CALL PARTIAL(W1,Z7,PART7)
      CALL PARTIAL(W1,Z8,PART8)
      DI(1)=-PART1+PART2
      DI(2)=-PART3+PART4+PART1-PART2
      DI(3)=PART5-PART6
      PARTH=-PART3-PART2+PART4+PART1
      DO 10 J=1,NH
        JH=J+3
        DI(JH)=PARTH*XH(I,J)
10    CONTINUE
      PARTW=-PART7-PART6+PART5+PART8
      DO 20 J=1,NW
        JW=J+3+NH
        DI(JW)=PARTW*XW(I,J)
20    CONTINUE
```

```
      RETURN
      END

      SUBROUTINE INTEGRAL(ARG1,ARG2,ARG3,ARG4,PROB,IER)
      DIMENSION X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      CALL MDBNOR(ARG1,ARG2,RHO,PROB1,IER)
      CALL MDBNOR(ARG3,ARG4,RHO,PROB2,IER)
      CALL MDBNOR(ARG1,ARG4,RHO,PROB3,IER)
      CALL MDBNOR(ARG3,ARG2,RHO,PROB4,IER)
      PROB=PROB1+PROB2-PROB3-PROB4
      RETURN
      END

      SUBROUTINE WRTE(NH,NW,VLIKE,STPSIZE,NITER,Q,IVAR,IER)
      DIMENSION X(30),Q(465),SCORE(30),X1(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      WRITE(3,2000) NITER,STPSIZE
2000  FORMAT(/,'NUMBER OF ITERATIONS = ',I3,'  FINAL STEPSIZE =',F7.5)
      WRITE(3,2001)
2001  FORMAT(/,'FINAL SCORE : ')
      CALL WRTVEC(N,SCORE)
      WRITE(3,3000) VLIKE
3000  FORMAT(/,'LOG-LIKELIHOOD VALUE = ',E13.6)
      WRITE(3,4000) RHO
4000  FORMAT(//,'RHO          = ',E10.3)
      TSTAT=X(1)/SQRT(Q(1))
      IF(IER.EQ.128) TSTAT=0.
      WRITE(3,5000) X(1),TSTAT
5000  FORMAT(/,'ALPHA H-0  = ',E10.3,'  T-STATISTIC = ',E10.3)
      TSTAT=X(2)/SQRT(Q(3))
      IF(IER.EQ.128) TSTAT=0.
      WRITE(3,6000) X(2),TSTAT
6000  FORMAT(/,'ALPHA H-1  = ',E10.3,'  T-STATISTIC = ',E10.3)
      TSTAT=X(3)/SQRT(Q(6))
      IF(IER.EQ.128) TSTAT=0.
      WRITE(3,7000) X(3),TSTAT
7000  FORMAT(/,'DEL ALPHA W= ',E10.3,'  T-STATISTIC = ',E10.3)
      NHH=NH+3
      DO 100 I=4,NHH
        IH=I-3
        IN=I*(I+1)/2
        TSTAT=X(I)/SQRT(Q(IN))
        IF(IER.EQ.128) TSTAT=0.
        WRITE (3,8000) IH,X(I),TSTAT
100   CONTINUE
8000  FORMAT(/,'DELTA UH',I2,' = ',E10.3,'  T-STATISTIC = ',E10.3)
      K=4+NH
      DO 200 I=K,N
```

```
        IW=I-K+1
        IN=I*(I+1)/2
        TSTAT=X(I)/SQRT(Q(IN))
        IF(IER.EQ.128) TSTAT=0.
        WRITE(3,9000) IW,X(I),TSTAT
200   CONTINUE
9000  FORMAT(/,'DELTA UW',I2,' = ',E10.3,'  T-STATISTIC = ',E10.3)
      IF(IVAR.EQ.0.OR.IER.EQ.128) GO TO 50
        WRITE(3,1000)
1000    FORMAT(//,'ASYMPTOTIC COVARIANCE MATRIX')
        CALL WRTVAR(Q)
50    CONTINUE
      RETURN
      END

      SUBROUTINE WRTVAR(Q)
      DIMENSION Q(465),ROW(30),X(30),X1(30),SCORE(30)
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DO 20 I=1,N
        IL=I*(I-1)/2 + 1
        IU=IL+I-1
        IJ=0
        DO 30 J=IL,IU
          IJ=IJ+1
          ROW(IJ)=Q(J)
30      CONTINUE
        WRITE(3,4000)
4000    FORMAT(/)
        CALL WRTVEC(I,ROW)
20    CONTINUE
99    RETURN
      END

      SUBROUTINE WRTVEC(NDIM,VEC)
      DIMENSION VEC(30)
      MDIM=NDIM/8
      IF(MDIM.EQ.0) GO TO 10
        DO 20 IM=1,MDIM
          M1=(IM-1)*8+1
          M2=IM*8
          WRITE(3,3000) (VEC(J),J=M1,M2)
3000      FORMAT(8E10.3)
20      CONTINUE
10    M3=MDIM*8+1
      IF(M3.GT.NDIM) GO TO 30
        WRITE(3,3000) (VEC(J),J=M3,NDIM)
30    RETURN
      END

      SUBROUTINE WRTVEC1(NDIM,VEC)
```

```
      DIMENSION VEC(30)
      MDIM=NDIM/8
      IF(MDIM.EQ.0) GO TO 10
        DO 20 IM=1,MDIM
          M1=(IM-1)*8+1
          M2=IM*8
          WRITE(*,3000) (VEC(J),J=M1,M2)
3000      FORMAT(1X,8E10.3)
20      CONTINUE
10    M3=MDIM*8+1
      IF(M3.GT.NDIM) GO TO 30
        WRITE(*,3000) (VEC(J),J=M3,NDIM)
30    RETURN
      END

      SUBROUTINE FIT(VLIKE,IFIT1,IFIT2)
      DIMENSION XH(2100,13),XW(2100,13),X(30),RTAB(4),X1(30),SCORE(30)
      DIMENSION NYH(2100),NYW(2100),INDEXO(2100),INDEXP(2100)
      DIMENSION PROBHAT(2100)
      INTEGER TABLE(4),ESTTAB(4),NGTAB(4)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      NGH0=0
      NGH1=0
      NGW0=0
      NGW1=0
      DO 10 J=1,4
        NGTAB(J)=0
   10 CONTINUE
      IF(IFIT1.EQ.1) GO TO 12
   11   CALL TREAT1(INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,NGTAB)
        GO TO 13
   12   CALL TREAT2(INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,NGTAB)
   13 IF(IFIT2.EQ.0) GO TO 14
      CALL FITSTAT(NOBS,VLIKE,INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,NGTAB,
     1             TABLE,ESTTAB,R,RH,RW,RTAB,RH0,RH1,RW0,RW1,RLIKE)
      CALL WRTSTAT(TABLE,ESTTAB,R,RH,RW,RTAB,RH0,RH1,RW0,RW1,RLIKE)
      NX=NH+NW+4
      CALL WRTVEC(NX,X)
   14 RETURN
      END

      SUBROUTINE TREAT1(INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,NGTAB)
      DIMENSION XH(2100,13),XW(2100,13),X(30),X1(30),SCORE(30)
      DIMENSION NYH(2100),NYW(2100),INDEXO(2100),INDEXP(2100)
      DIMENSION PROBHAT(2100)
      INTEGER NGTAB(4)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DO 100 I=1,NOBS
```

```
        CALL DELTA(I,DELTAH,DELTAW)
        CALL IPROB00(DELTAH,DELTAW,PR00)
        CALL IPROB01(DELTAH,DELTAW,PR01)
        CALL IPROB10(DELTAH,DELTAW,PR10)
        CALL IPROB11(DELTAH,DELTAW,PR11)
        CALL YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
        INDEXP(I)=INDEXI
        IYH=NYH(I)
        IYW=NYW(I)
        CALL IND(IYH,IYW,INDEXOI)
        INDEXO(I)=INDEXOI
        CALL COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,NGH0,NGH1,
     1             NGW0,NGW1,NGTAB)
  100 CONTINUE
      RETURN
      END

      SUBROUTINE TREAT2(INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,NGTAB)
      DIMENSION XH(2100,13),XW(2100,13),X(30),X1(30),SCORE(30)
      DIMENSION NYH(2100),NYW(2100),INDEXO(2100),INDEXP(2100)
      DIMENSION PROBHAT(2100)
      INTEGER NGTAB(4)
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      WRITE(3,1000)
 1000 FORMAT(//,'  YH  ','  YW  ',' HHAT ',' WHAT ',' PR00 ',
     1       ' PR01 ',' PR10 ',' PR11 ',/)
      DO 100 I=1,NOBS
        CALL DELTA(I,DELTAH,DELTAW)
        CALL IPROB00(DELTAH,DELTAW,PR00)
        CALL IPROB01(DELTAH,DELTAW,PR01)
        CALL IPROB10(DELTAH,DELTAW,PR10)
        CALL IPROB11(DELTAH,DELTAW,PR11)
        CALL YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
        INDEXP(I)=INDEXI
        IYH=NYH(I)
        IYW=NYW(I)
        WRITE(3,1001) IYH,IYW,IYHHAT,IYWHAT,PR00,PR01,PR10,PR11
 1001   FORMAT(4(3X,I1,2X),4(1X,F4.3,1X))
        CALL IND(IYH,IYW,INDEXOI)
        INDEXO(I)=INDEXOI
        CALL COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,NGH0,NGH1,
     1             NGW0,NGW1,NGTAB)
  100 CONTINUE
      RETURN
      END

      SUBROUTINE YHAT(PR00,PR01,PR10,PR11,IYHHAT,IYWHAT,INDEXI)
      INDEXI=1
      IF(PR00.GT.PR01) GO TO 200
```

```
      PRMAX=PR01
      INDEXI=2
      GO TO 110
200 PRMAX=PR00
110 IF(PRMAX.GT.PR10)  GO TO 100
      PRMAX=PR10
      INDEXI=3
100 IF(PRMAX.GT.PR11) GO TO 500
      INDEXI=4
500 GO TO (501,502,503,504),INDEXI
501 IYHHAT=0
    IYWHAT=0
    GO TO 900
502 IYHHAT=0
    IYWHAT=1
    GO TO 900
503 IYHHAT=1
    IYWHAT=0
    GO TO 900
504 IYHHAT=1
    IYWHAT=1
900 RETURN
    END

    SUBROUTINE IND(IYH,IYW,INDEXOI)
    INDEXOI=1
    IF(IYW.EQ.1) GO TO 200
      IF(IYH.EQ.1) INDEXOI=3
        GO TO 100
200 INDEXOI=2
    IF(IYH.EQ.1) INDEXOI=4
100 CONTINUE
    RETURN
    END

    SUBROUTINE COUNT(IYH,IYHHAT,IYW,IYWHAT,INDEXOI,INDEXI,
   1                 NGH0,NGH1,NGW0,NGW1,NGTAB)
    INTEGER NGTAB(4)
    IF(IYH.EQ.0.AND.IYHHAT.EQ.0) NGH0=NGH0+1
    IF(IYH.EQ.1.AND.IYHHAT.EQ.1) NGH1=NGH1+1
    IF(IYW.EQ.0.AND.IYWHAT.EQ.0) NGW0=NGW0+1
    IF(IYW.EQ.1.AND.IYWHAT.EQ.1) NGW1=NGW1+1
    DO 100 J=1,4
      IF(INDEXOI.EQ.J.AND.INDEXI.EQ.J) NGTAB(J)=NGTAB(J)+1
100 CONTINUE
    RETURN
    END

    SUBROUTINE FITSTAT(NOBS,VLIKE,INDEXO,INDEXP,NGH0,NGH1,NGW0,NGW1,
   1         NGTAB,TABLE,ESTTAB,R,RH,RW,RTAB,RH0,RH1,RW0,RW1,RLIKE)
```

```
      DIMENSION XNGTAB(4),RTAB(4)
      INTEGER NGTAB(4),TABLE(4),ESTTAB(4),INDEXO(2100),INDEXP(2100)
      CALL CONTIN(NOBS,INDEXO,TABLE)
      CALL CONTIN(NOBS,INDEXP,ESTTAB)
      XNGHO=NGHO
      XNGH1=NGH1
      XNGH=XNGHO+XNGH1
      XNGWO=NGWO
      XNGW1=NGW1
      XNGW=XNGWO+XNGW1
      XNOBS=NOBS
      RH=XNGH/XNOBS
      RHO=XNGHO/(TABLE(1)+TABLE(2))
      RH1=XNGH1/(TABLE(3)+TABLE(4))
      RW=XNGW/XNOBS
      RWO=XNGWO/(TABLE(1)+TABLE(3))
      RW1=XNGW1/(TABLE(2)+TABLE(4))
      DO 10 J=1,4
        XNGTAB(J)=NGTAB(J)
        RTAB(J)=XNGTAB(J)/TABLE(J)
   10 CONTINUE
      XNG=XNGTAB(1)+XNGTAB(2)+XNGTAB(3)+XNGTAB(4)
      R=XNG/XNOBS
      RLIKE=1.+ VLIKE/(XNOBS*ALOG(4.))
      RETURN
      END

      SUBROUTINE CONTIN(NOBS,NVEC,NTAB)
      INTEGER NVEC(2100),NTAB(4)
      DO 100 J=1,4
        NTAB(J)=0
  100 CONTINUE
      DO 200 I=1,NOBS
        NVECI=NVEC(I)
        GO TO (101,102,103,104),NVECI
  101     NTAB(1)=NTAB(1)+1
          GO TO 201
  102     NTAB(2)=NTAB(2)+1
          GO TO 201
  103     NTAB(3)=NTAB(3)+1
          GO TO 201
  104     NTAB(4)=NTAB(4)+1
  201   CONTINUE
  200 CONTINUE
      RETURN
      END

      SUBROUTINE WRTSTAT(TABLE,ESTTAB,R,RH,RW,RTAB,RHO,RH1,RWO,RW1,
     1                   RLIKE)
      DIMENSION RTAB(4)
```

```
      INTEGER TABLE(4),ESTTAB(4)
      WRITE(3,2004) RLIKE,R
 2004 FORMAT(//,' RLIKE = ',F4.3,3X,' R = ',F4.3)
      WRITE(3,2000)
 2000 FORMAT(//,' OBSERVED AND PREDICTED CONTINGENCY TABLES :')
      WRITE(3,2001)
 2001 FORMAT(/,'     CELLS :','   00    ','   01    ','   10    ','   11')
      WRITE(3,2002) (TABLE(J),J=1,4)
 2002 FORMAT(/,'  OBSERVED :',2X,I4,4X,I4,4X,I4,4X,I4)
      WRITE(3,2003) (ESTTAB(J),J=1,4)
 2003 FORMAT(/,' PREDICTED :',2X,I4,4X,I4,4X,I4,4X,I4)
      WRITE(3,2005) (RTAB(J),J=1,4)
 2005 FORMAT(/,' HW-RATIOS :',2X,F4.3,4X,F4.3,4X,F4.3,4X,F4.3)
      WRITE(3,2006)
 2006 FORMAT(//,'    RATIOS : WORK & NOTWORK    WORK ONLY    NOT WORK ')
      WRITE(3,2007) RH,RH1,RHO
 2007 FORMAT(/,'   HUSBAND :',6X,F5.3,11X,F5.3,9X,F5.3)
      WRITE(3,2008) RW,RW1,RWO
 2008 FORMAT(/,'      WIFE :',6X,F5.3,11X,F5.3,9X,F5.3)
      RETURN
      END

      SUBROUTINE BIGN(XNUM)
      ABIG=85.
      ANUM=ABS(XNUM)
      IF(ANUM.LT.ABIG) GOTO 100
        SIGN=1.
        IF(XNUM.LT.0.) SIGN=-1.
        XNUM=SIGN*ABIG
  100 RETURN
      END

      SUBROUTINE BIGN1(XNUM,ITRUN)
      ITRUN=0
      ABIG=1.0E+20
      ANUM=ABS(XNUM)
      IF(ANUM.LT.ABIG) GOTO 100
        ITRUN=2
        SIGN=1.
        IF(XNUM.LT.0.) SIGN=-1.
        XNUM=SIGN*ABIG
  100 RETURN
      END

      REAL FUNCTION F(N,Z)
      COMMON /CMN2/ RATIO
      REAL Z(1)
      PI=ACOS(-1.)
      XNUM=-.5*Z(1)**2
      CALL BIGN(XNUM)
```

```
      DENS=(1./SQRT(2.*PI))*EXP(XNUM)
      Y=RATIO*Z(1)
      CALL MDNOR(Y,CDF)
      F=DENS*CDF
      RETURN
      END

      REAL FUNCTION MLOGLIK(STPSIZE)
      DIMENSION X(30),X1(30),SCORE(30),XH(2100,13),XW(2100,13)
      DIMENSION PROBHAT(2100)
      INTEGER NYH(2100),NYW(2100)
      REAL STPSIZE
      COMMON /CMN1/ NOBS,NH,NW,XH,XW,NYH,NYW,PROBHAT
      COMMON /CMN3/ N,RHO,X,X1,SCORE
      DO 10 J=1,N
      X(J)=X1(J)+STPSIZE*SCORE(J)
 10   CONTINUE
      CALL VALUE(N,X,VLIKE)
      WRITE(3,102) STPSIZE,VLIKE
102   FORMAT(1X,'STEPSIZE =',F7.5,'  LOGLIK =',E13.6)
      MLOGLIK=-VLIKE
      RETURN
      END
```

REFERENCES

Aguirre-Torres, V. and Gallant, A. R. "The Null and Non-Null Asymptotic Distribution of the Cox Test for Multivariate Nonlinear Regression." Journal of Econometrics 21 (1983):5-33.

Akaike, H. "Information Theory and an Extension of the Maximum Likelihood Principle." In Proceedings of the 2nd International Symposium on Information Theory, edited by N. Petrov and F. Csadki. Budapest: Akademia: Kiako, 1973.

__________. "A New Look at Statistical Model Identification." IEEE Transactions on Automatic Control AC-19 (1974):716-723.

Amemiya, T. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal." Econometrica 42 (1974):999-1012.

__________. "The Estimation of a Simultaneous Equation Generalized Probit Model." Econometrica 46 (1978):1193-1205.

__________. "Selection of Regressors." International Economic Review 21 (1980):331-354.

__________. "Qualitative Response Models: A Survey." Journal of Economic Literature 19 (1981):1483-1536.

Ashenfelter, O. and Heckman, J. "The Estimation of Income and Substitution Effects in a Model of Family Labor Supply." Econometrica 42 (1974):73-85.

Ashford, J. R. and Sowden, R. R. "Multivariate Probit Analysis." Biometrics 26 (1970):535-546.

Atkinson, A. C. "A Method for Discriminating Between Models." Journal of the Royal Statistical Society 32-B (1970):323-345.

Bartlett, M. S. "Approximate Confidence Intervals--II. More then One Unknown Parameter." Biometrica 40 (1953):306-317.

Berndt, E. R.; Hall, B. H.; Hall, R. E. and Hausman, J. A. "Estimation and Inference in Non-Linear Structural Models." Annals of Economic and Social Measurement 3 (1974):653-665.

Bishop, Y. M. M.; Fienberg, S. E. and Holland, P. W. Discrete Multivariate Analysis: Theory and Practice. Cambridge: Massachusetts Institute of Technology Press, 1975.

Bjorn, P. A. and Vuong, Q. H. "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation." Social Science Working Paper No. 537. Pasadena: California Institute of Technology, July, 1984.

Brown, M. and Manser, M. "Neoclassical and Bargaining Approaches to Household Decision Making—With Application to the Household Labor Supply Decision." Discussion Paper No. 401. Buffalo: State University of New York at Buffalo, 1978.

Chow, G. C. "Selection of Econometric Models by the Information Criterion." In Proceedings of the Econometric Society European Meetings, edited by E. G. Charatsis. Amsterdam: North Holland, 1981.

__________. Econometrics. New York: McGraw Hill, 1983.

Cotterman, R. F. "The Role of Assets in Labor Supply Functions." Economic Inquiry 19 (1981):495-505.

Cox, D. R. "Tests of Separate Families of Hypotheses." In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley: University of California Press, 1961.

__________. "Further Results on Tests of Separate Families of Hypotheses." Journal of the Royal Statistical Society 24-B (1962):406-424.

Cragg, S. G. and Uhler, R. S. "The Demand for Automobiles." Canadian Journal of Economics 3 (1970):386-406.

Dastoor, N. K. "A Note on the Interpretation of the Cox Procedure for Non-Nested Hypotheses." Economics Letters 8 (1981):113-119.

Dhrymes, P. J. Econometrics: Statistical Foundations and Applications. New York: Springer-Verlag, 1974.

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

Dubin, J. A. and McFadden, D. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumptions." Econometrica 52 (1984):345-362.

Fisher, G. and McAleer, M. "On the Interpretation of the Cox Test in Econometrics." Economics Letters 4 (1979):145-150.

Goodman, L. A. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." American Journal of Sociology 78 (1972):1135-1191.

Gourieroux, C.; Laffont, J. J. and Monfort, A. "Coherency Conditions in Simultaneous Linear Equations Models with Endogenous Switching Regimes." Econometrica 48 (1980):675-695.

Gourieroux, C.; Monfort, A. and Tragnon, A. "Testing Nested or Non-Nested Hypotheses." Journal of Econometrics 21 (1983):83-115.

__________. "Pseudo Maximum Likelihood Methods: Theory." Econometrica 52 (1984):681-700.

Gronau, R. "The Effect of Children on the Housewife's Value of Time." Journal of Political Economy Supplement, 81 (1973):168-199.

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

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

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

Heckman, J. "Shadow Prices, Market Wages, and Labor Supply." Econometrica 42(1974):679-694.

__________. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models." Annals of Economic and Social Measurement 5(1976):475-492.

__________. "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica 46 (1978):931-959.

Heckman, J. and Macurdy, T. "A Life Cycle Model of Female Labour Supply." Review of Economic Studies 47(1980):47-74.

Kinal, T. and Kajal, L. "On the Power of Various Model Selection Criteria." Mimeographed paper, 1983.

Koopmans, T. C. and Hood, W. C. Studies in Econometric Method. New York: Wiley, 1953.

Kooreman, P. and Kapteyn, A. "Estimation of a Game Theoretic Model of Household Labor Supply." Department of Economics Discussion Paper No. 180. Tilburg: Tilburg University, June, 1985.

Kullback, S. and Leibler, R. A. "On Information and Sufficiency." Annals of Mathematical Statistics. 22 (1951):79-86.

Lazear, E. "Education: Consumption or Production?" Journal of Political Economy 85(1977):569-597.

Leamer, E. E. "Information Criteria for Choice of Regression Models: A Comment." Econometrica 47 (1979):507-510.

Loh, W. "A New Method for Testing Separate Families of Hypotheses." Journal of the American Statistical Association 80 (1985):362-268.

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

Maddala, G. S. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press, 1983.

Mallows, C. L. "Some Comments on Cp." Technometrics 16 (1973):661-675.

Manser, M. and Brown, M. "Marriage and Household Decision-Making: A Bargaining Analysis." International Economics Review 21 (1980):31-44.

Manski, C. and McFadden, D. Structural Analysis of Discrete Data with Econometric Applications. Cambridge: Massachusetts Institute of Technology Press, 1981.

McElroy, M. B. and Horney, M. J. "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand." International Economic Review 22 (1981):333-347.

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

__________. "Econometric Models of Probabilistic Choice." In Structural Analysis of Discrete Data with Econometric Applications, edited by C. Manski and D. McFadden. Cambridge: Massachusetts Institute of Technology Press, 1981.

Mickle, M. H. and Sze, T. W. Optimization in Systems Engineering Scranton: Intext Educational Publishers, 1972.

Nakamura, A. and Nakamura, M. "A Comparison of the Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes." Econometrica 49(19811):451-489.

Nelson, F. D. "Structural Models for Qualitative Variables." Mimeographed paper, 1979.

Nelson, F. D. and Olsen, L. "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables." International Economic Review 19 (1978):695-710.

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 no. 1. Center for Statistics and Probability, Northwestern University, 1976.

Owen, G. Game Theory. New York: Academic Press, 1982.

Parsons, D. A. "The Decline of Male Labor Force Participation." Journal of Political Economy 88 (1980):117-134.

Pearson, K. "Mathematical Contributions to the Theory of Evolution in the Inheritance of Characters Note Capable of Exact Qualitative Measurement, VIII." Philosophical Transactions of the Royal Society. 195-A (1900):79-150.

Pesaran, M. H. "On the General Problem of Model Selection." Review of Economic Studies 41 (1974):153-171.

Rothenberg, T. J. "Identification in Parametric Models." Econometrica 39 (1971):577-591.

Sawa, T. "Information Criteria for Discriminating Among Alternative Regression Models." Econometrica 46 (1978):1273-1291.

Sawyer, K. R. "Testing Separate Families of Hypotheses: An Information Criterion." Journal of the Royal Statistical Society 45-B (1983):89-99.

Schmidt, P. "Constraints on the Parameters in Simultaneous Tobit and Probit Models." In Structural Analysis of Discrete Data with Econometric Applications, edited by C. Manski and D. McFadden. Cambridge: Massachusetts Institute of Technology Press, 1981.

Schmidt, P. and Strauss, R. P. "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach." Econometria 43 (1975a):745-755.

__________. "The Prediction of Occupation Using Multiple Logit Models." International Economic Review 16 (1975b):471-485.

Schwartz, G. "Estimating the Dimension of a Model." Annals of Statistics 6 (1978):461-464.

Vuong, Q. H. "Heckman's Criticism of the Nerlove-Press Model: A Reply." Mimeographed paper, 1980.

__________. "Structural log-linear Probability Models." Working Paper 8114, GREMAQ, Universite' des Sciences Sociales, Toulouse 1981.

__________. "Probability Feedback in a Recursive System of Probability Models." Social Science Working Paper 443. California Institute of Technolody, 1982a.

__________. "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application." Ph.D. dissertation, Northwestern University, 1982b.

__________. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." Mimeographed paper, 1985.

Waldman, D. M. "An Economic Interpretation of Parameter Constraints in a Simultaneous Equations Model with Limited Dependent Variables." International Economic Review 22 (1981):731-739.

Williams, D. A. Discussion of "A Method for Discriminating Between Models," by A. C. Atkinson, Journal of the Royal Statistical Society 32-B (1970):350.

Yule, G. U. "On the Association of Attributes in Statistics: With Illustration from the Material of the Childhood Society." Philosophical Transactions of the Royal Society. 194-A (1900):257-319.