FULLY RECURSIVE PROBABILITY MODELS AND MULTIVARIATE LOG-LINEAR PROBABILITY MODELS FOR THE ANALYSIS OF QUALITATIVE DATA

by

Lung Fei-Lee

Discussion Paper No. 80 - 132, May 1980

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

April 1980

FULLY RECURSIVE PROBABILITY MODELS AND MULTIVARIATE LOG-LINEAR PROBABILITY MODELS FOR THE ANALYSIS OF QUALITATIVE DATA

Lung-Fei Lee*

1. INTRODUCTION

Econometricans have developed multiple equation models to analyze multivariate continuous endogenous variables (see, e.g. Theil [1971]). Seemingly unrelated regressions models, recursive models and simultaneous equations models are valuable tools for analyzing associations, causation and joint dependency among the endogenous variables conditional on exogenous variables. Recently, analogous models have been developed. for the analysis of multivariate qualitative variables.

For the analysis of bivariate dichotomous dependent variables, bivariate normal (or bivariate probit) models have been proposed in Ashford and Sowden [1970] and Amemiya [1974]. This model is useful for analyzing correlations between the dichotomous dependent variables. For analyzing causal structures of discrete responses, recursive models have been formulated in Maddala and Lee [1976]. A bivariate resursive model has been specified as

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = R\alpha \tag{1.1}$$

---

*I appreciate having financial support from the National Science Foundation under Grant SOC-78-07304.

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \tag{1.2}$$

for the analysis of two dichotomous responses where Y_1 is defined as

$$Y_1 = 1 \quad \text{iff} \quad Y = 1 \tag{1.3}$$

$$Y_1 = 0 \quad \text{otherwise.}$$

Similarly, we define Z_1 corresponding to Z. The model forms a causual system of logistic equations as in Goodman [1973a, 1973b]. In this model, Y has a direct effect on the probability of occurence of Z but not vice versa, i.e., the reciprocal effect (of Z on Y) is zero.

Multiple equation models analogous to the classical simultaneous equation models have been proposed in Nerlove and Press [1973, 1976] and Schmidt and Strauss [1975]. A simple representation of their formulations for the analysis of a bivariate dichotomous dependent variable is the two equation model:

$$\ell n \frac{P(Y = 1|Z, X)}{P(Y = 2|Z, X)} = X\alpha + \gamma Z_1 \tag{1.4}$$

$$\ell n \frac{P(Z = 1|Y, X)}{P(Z = 2|Y, X)} = X\beta + \gamma Y_1 \tag{1.5}$$

where $P(Y|Z, X)$ is the conditional probability of Y given Z and X, etc. Nerlove and Press refer to them as a multivariate logit model while Schmidt and Strauss label them as a simultaneous logit model. As opposed to classical simultaneous equation models, the derivation of the two structural equations (1.4) and (1.5) is based on conditional distributions; furthermore, the direct effect of Y on Z and the reciprocal effect are equal.

While the log-linear models are useful for analyzing associations among the discrete variables, one may desire to detect possible causation among the responses. For these cases, recursive models as in (1.1), (1.2) will be useful.

In this article, we will explore the connections between the recursive probability models and the log-linear models. The analysis will be helpful in clarifying the structures of the two models and for interpreting equations (1.4) and (1.5) when analyzing discrete variables. A statistical test procedure is also proposed as a possible way to test the causal chain specification. Our analysis concentrates on two equation models with dichotomous dependent variables. Some of the analysis can be extended to higher dimensional contingency tables as well as polychotomous dependent variables.

2. General Multivariate Log-Linear (or Logistic) Models

In this section, we review the specification of the log linear models and provide some general relations between the fully recursive and the log linear models for subsequent analysis. It is convenient for our analysis to distinguish models with all explanatory variables being qualitative variables and models with continuous explanatory variables. In the statistics literatures, only models with all discrete variables have been analyzed.

Let Y and Z be two dichotomous dependent variables and x be a vector of qualitative explanatory variables. Without loss of generality, for our analysis, it is assumed that $X = (1, C, D)$ where C and D are two distinct qualitative variables with K and L levels, respectively. A bivariate log-linear model (or logistic model) is specified as

$$\begin{aligned} \ell n p_{ijk\ell}^{YZ\overline{CD}} = \mu_{k\ell} &+ \theta_i^Y + \theta_j^Z \\ &+ \theta_{ik}^{YC} + \theta_{i\ell}^{YD} + \theta_{jk}^{ZC} + \theta_{j\ell}^{ZD} + \theta_{ij}^{YZ} \qquad (2.1) \\ &+ \theta_{ik\ell}^{YCD} + \theta_{jk\ell}^{ZCD} + \theta_{ijk}^{YZC} + \theta_{ij\ell}^{YZD} + \theta_{ijk\ell}^{YZCD} \end{aligned}$$

$$\begin{aligned} i&=1,2 \\ j&=1,2 \\ k&=1,\ldots,K \\ \ell&=1,\ldots,L \end{aligned}$$

where $p_{ijk\ell}^{YZ\overline{CD}}$ denotes the conditional joint probability of Y and Z at levels i and j, respectively, conditional on C and D at levels k and ℓ; θ_i^Y denotes an effect due to Y (at level i); θ_j^Z denotes an effect due to Z; θ_{ik}^{YC} denotes a second order interaction effect between Y and C (at levels i and k respectively);....; $\theta_{ijk\ell}^{YZCD}$ denotes a 4th order interaction effect among all of the variables at levels i, j, k, ℓ. This notation originated in Goodman [1972b]. The overall effect $\mu_{k\ell}$ is a function of all of the other effects. As $\Sigma_i \Sigma_j \, p_{ijk\ell}^{\overline{YZ}CD} = 1$,

$$\mu_{k\ell} = -\ell n \Sigma_i \Sigma_j \exp \{\theta_i^Y + \theta_j^Z + \ldots + \theta_{ijk\ell}^{YZCD}\} \quad (2.2).$$

The probabilities $p_{ijk\ell}^{\overline{YZ}CD}$ have the logistic functional form. The parameters of the other effects are subject to the constraints as in ANOVA models, i.e.,

$$\Sigma_i \theta_i^Y = \Sigma_j \theta_j^Z = 0$$

$$\Sigma_i \theta_{ik}^{YC} = \Sigma_k \theta_{ik}^{YC} = \ldots = \Sigma_j \theta_{ij}^{YZ} = 0$$

$$\Sigma_i \theta_{ik\ell}^{YCD} = \Sigma_k \theta_{ik\ell}^{YCD} = \ldots = \Sigma_\ell \theta_{ij\ell}^{YZD} = 0$$

$$\Sigma_i \theta_{ijk\ell}^{YZCD} = \Sigma_j \theta_{ijk\ell}^{YZCD} = \ldots = \Sigma_\ell \theta_{ijk\ell}^{YZCD} = 0$$

The total number of free parameters in this model is 3KL. Since there are only 3KL free cells for this 4 way table, this model is saturated (Goodman [1972a]). A log-linear model is called nonsaturated if some of the effects vanish. An equivalent representation of the log-linear model is the conditional log odds equation (or conditional structural equation) model as in Goodman [1972b] and Nerlove and Press [1976].

$$\ell n \frac{p_{1jk\ell}^{\overline{Y}ZCD}}{p_{2jk\ell}^{\overline{Y}ZCD}} = \delta_1^Y + \delta_{1k}^{YC} + \delta_{1\ell}^{YD} + \delta_{1j}^{YZ} + \delta_{1k\ell}^{YCD} + \delta_{1jk}^{YZC} + \delta_{1j\ell}^{YZD} + \delta_{1jk\ell}^{YZCD} \quad (2.4)$$

$$\ell n \frac{p_{i1k\ell}^{Y\overline{Z}CD}}{p_{i2k\ell}^{Y\overline{Z}CD}} = \delta_1^Z + \delta_{1k}^{ZC} + \delta_{1\ell}^{ZD} + \delta_{i1}^{YZ} + \delta_{1k\ell}^{ZCD} + \delta_{i1\ell}^{YZC} + \delta_{i1\ell}^{YZD} + \delta_{i1k\ell}^{YZCD} \quad (2.5)$$

where $p_{ijk\ell}^{\overline{Y}ZCD}$ denotes the conditional probability of Y at level i conditional on Z, C, D at levels j, k, ℓ respectively,...; $\delta_1^Y = 2\theta_1^Y$, $\delta_{1k}^{YC} = 2\theta_{1k}^{YC}, \ldots, \delta_{i1k\ell}^{YZCD} = 2\theta_{i1k\ell}^{YZCD}$.

When the discrete variables are dichotomous, equations (2.1) or (2.4), (2.5) can be written in a more appealing form. Define

$$\begin{aligned} Y_1^* &= 1 \text{ iff the level 1 of } Y \text{ occurs;} \\ Y_2^* &= -1 \text{ iff the level 2 of } Y \text{ occurs.} \end{aligned} \tag{2.6}$$

Similarly, define Z_j^*, C_k^* and D_ℓ^* for the variables Z, C, and D respectively. Equation (2.1) can be rewritten as

$$\begin{aligned} \ell n\, p_{ijk\ell}^{\overline{YZCD}} = {} & \mu_{k\ell} + \gamma_1 Y_i^* + \gamma_2 Z_j^* \\ & + \gamma_{13} Y_i^* C_k^* + \gamma_{14} Y_i^* D_\ell^* + \gamma_{23} Z_j^* C_k^* + \gamma_{24} Z_j^* D_\ell^* + \gamma_{12} Y_i^* Z_j^* \\ & + \gamma_{134} Y_i^* C_k^* D_\ell^* + \gamma_{234} Z_j^* C_k^* D_\ell^* + \gamma_{123} Y_i^* Z_j^* C_k^* + \gamma_{124} Y_i^* Z_j^* D_\ell^* \\ & + \gamma_{1234} Y_i^* Z_j^* C_k^* D_\ell^* \end{aligned} \tag{2.7}$$

The conditional log odds equations become

$$\begin{aligned} \ell n \frac{p_{1jk\ell}^{\bar{Y}ZCD}}{p_{2jk\ell}^{\bar{Y}ZCD}} = {} & \omega_1 + \omega_{13} C_k^* + \omega_{14} D_\ell^* + \omega_{134} C_k^* D_\ell^* + \omega_{12} Z_j^* + \omega_{123} Z_j^* C_k^* \\ & + \omega_{124} Z_j^* D_\ell^* + \omega_{1234} Z_j^* C_k^* D_\ell^* \end{aligned} \tag{2.8}$$

$$\begin{aligned} \ell n \frac{p_{i1k\ell}^{Y\bar{Z}CD}}{p_{j2k\ell}^{Y\bar{Z}CD}} = {} & \omega_2 + \omega_{23} C_k^* + \omega_{24} D_\ell^* + \omega_{234} C_k^* D_\ell^* + \omega_{12} Y_i^* + \omega_{123} Y_i^* C_k^* \\ & + \omega_{124} Y_i^* D_\ell^* + \omega_{1234} Y_i^* C_k^* D_\ell^* \end{aligned} \tag{2.9}$$

where $\omega_1 = 2\gamma_1$, $\omega_{13} = 2\gamma_{13}, \ldots, \omega_{1234} = 2\gamma_{1234}$. These expressions are similar to the logistic regression models (see, e.g. McFadden [1974]) and are analogous to classical simultaneous equations model as pointed out in Goodman [1972b], Nerlove and Press [1973] and Schmidt and Strauss [1975].

Nerlove and Press [1973] generalized the model to allow continuous as well as qualitative explanatory variables. For the two dichotomous dependent variables with explanatory variable vector X, their models specified in equations (1.4) and (1.5) obviously generalize the specification of the above equations (2.8) and (2.9).

In comparing the various recursive models with the log-linear models, the following lemma will be used.

<u>Lemma</u>: Let Y and Z be two dichotomous dependent variables with values 1 and 2 and R and Q are explanatory variables. Let $P(Y)$ denote the marginal probability of Y conditional on the explanatory variables, $P(Y|Z)$ and $P(Z|Y)$ be the conditional probabilities of Y conditional on Z, R and Q, and Z conditional on Y, R and Q, respectively. Denote $Y^* = 1$ iff $Y = 1$; $Y^* = -1$ iff $Y = 2$; similarly Z^* is defined as corresponding to Z. Then the two equations

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = R\alpha \tag{2.10}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y^* \tag{2.11}$$

are equivalent to

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + \lambda Z^* + G(Q\beta, \lambda) \tag{2.12}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y^* \tag{2.13}$$

where

$$G(Q\beta, \lambda) = \lambda + \ell n\left(\frac{1 + \exp(Q\beta - \lambda)}{1 + \exp(Q\beta + \lambda)}\right) \tag{2.14}$$

<u>Proof</u>: ($\Rightarrow$:) From the Bayes theorem,

$$\frac{P(Y = 1|Z)}{P(Y = 2|Z)} = \frac{P(Z|Y = 1)}{P(Z|Y = 2)} \frac{P(Y = 1)}{P(Y = 2)} \tag{2.15}$$

It follows from (2.11),

$$P(Z|Y) = \frac{\exp[(Q\beta + \lambda Y^*)I]}{1 + \exp(Q\beta + \lambda Y^*)} \tag{2.16}$$

where $I = 1$ iff $Z = 1$; $I = 0$ iff $Z = 2$; and hence

$$\frac{P(Z|Y = 1)}{P(Z|Y = 2)} = \exp(2\lambda I) \frac{1 + \exp(Q\beta - \lambda)}{1 + \exp(Q\beta + \lambda)} \tag{2.17}$$

Substituting the probability odds in (2.10) and (2.17) into (2.16),

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + 2\lambda I + \ell n(\frac{1 + \exp(Q\beta - \lambda)}{1 + \exp(Q\beta + \lambda)})$$

$$= R\alpha + \lambda Z^* + G(Q\beta, \lambda)$$

where G is the function defined in (2.14).

($\Leftarrow$:) From (2.15),

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = \ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} - \ell n \frac{P(Z|Y = 1)}{P(Z|Y = 2)} \tag{2.18}$$

Since (2.17) holds from (2.13), equations (2.17), (2.12) and (2.18) imply

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = R\alpha + \lambda Z^* + G(Q\beta, \lambda) - 2\lambda I - \ell n (\frac{1 + \exp(Q\beta - \lambda)}{1 + \exp(Q\beta + \lambda)})$$

$$= R\alpha \qquad\qquad \text{Q.E.D.}$$

<u>Corollary</u>: Denote $Y_1 = 1$ iff $Y = 1$; $Y_1 = 0$ iff $Y = 2$; similarly Z_1 is defined as corresponding to Z. Then the two equations

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = R\alpha \tag{2.19}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \tag{2.20}$$

are equivalent to

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + \lambda Z_1 + G_1(Q\beta, \lambda) \tag{2.21}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \tag{2.22}$$

where

$$G_1(Q\beta, \lambda) = \ell n \left(\frac{1 + \exp(Q\beta)}{1 + \exp(Q\beta + \lambda)}\right) \tag{2.23}$$

3. Relationships among the Various Probability Models

In comparing the two models, we consider the models with all exogenous variables being qualitative, first. To simplify notation, we consider $X = (1, C, D)$, where C and D are two distinct qualitative variables. The analysis holds for any vector of qualitative explanatory variables. Let $\tilde{Y}$ and $\tilde{Z}$ be two underlying indexes which are related to the observed responses as $Y = 1$ iff $\tilde{Y} > 0$, and $Y = 2$ otherwise; similarly, $Z = 1$ iff $\tilde{Z} > 0$, and $Z = 2$ otherwise. The general recursive model to be considered is,

$$\tilde{Y} = \alpha + \alpha_k^C + \alpha_\ell^D - \varepsilon_1 \tag{3.1}$$

$$\tilde{Z} = \beta + \beta_k^C + \beta_\ell^D + \lambda_{i1}^{YZ} - \varepsilon_2 \tag{3.2}$$

where the disturbances ε_1, ε_2 have zero means; α is the constant term, α_k^C is the effect of C at level k on $\tilde{Y}$ (or the probability of Y),..., β_ℓ^D is the effect of D at level ℓ on $\tilde{Z}$ and λ_{i1}^{YZ} is the direct effect of Y at level i in influencing $\tilde{Z}$. The effects satisfy the ANOVA constraints,

$$\Sigma_{k=1}^K \alpha_k^C = \Sigma_{\ell=1}^L \alpha_\ell^D = 0$$

$$\Sigma_{k=1}^K \beta_k^C = \Sigma_{\ell=1}^L \beta_\ell^D = \Sigma_{i=1}^L \lambda_{i1}^{YZ} = 0$$

If the equations involve higher order interaction terms, these effects will also be subject to the ANOVA constraints, e.g. if C and D have interaction effects $\alpha_{k\ell}^{CD}$ on $\tilde{Y}$, the effects will satisfy the constraints $\Sigma_{k=1}^K \alpha_{k\ell}^{CD} = \Sigma_{\ell=1}^L \alpha_{k\ell}^{CD} = 0$. For some cases, it is desirable to differentiate between cases when the discrete variables are ordered or unordered. If the ordered discrete variables have linear effects, the model should incorporate such information. For example, if D is an ordered polychotomous variable with values $1,\ldots,L$ and it has a linear effect on $\tilde{Y}$, the effects α_ℓ^D can be specified as $\alpha_\ell^D = \alpha^D(\ell - \frac{1}{L}\Sigma_{i=1}^L i)$ where α^D is an unknown coefficient on the explanatory variable D. In the following analysis, the discrete variables are implicitly assumed to be unordered unless specifically stated otherwise.

When $\lambda_{i1}^{YZ} = 0$ and the disturbances ε_1 and ε_2 in (3.1) and (3.2) are independent, the two dependent variables Y and Z are independent (conditional on the explanatory variables). In addition, if ε_1 and ε_2 are logistically distributed i.e., with distributions $\frac{1}{1+e^{-\varepsilon_1}}$ and $\frac{1}{1+e^{-\varepsilon_2}}$ respectively, the two equations are equivalent to

$$\ell n \frac{p_{1k\ell}^{\bar{Y}CD}}{p_{2k\ell}^{\bar{Y}CD}} = \alpha + \alpha_k^C + \alpha_\ell^D \tag{3.3}$$

$$\ell n \frac{p_{1k\ell}^{\bar{Z}CD}}{p_{2k\ell}^{\bar{Z}CD}} = \beta + \beta_k^C + \beta_\ell^D . \tag{3.4}$$

This independence model is equivalent to the log linear model

$$\ell n\ p_{ijk\ell}^{\overline{YZ}CD} = \mu_{k\ell} + \theta_i^Y + \theta_{ik}^{YC} + \theta_{i\ell}^{YD} + \theta_j^Z + \theta_{jk}^{ZC} + \theta_{j\ell}^{ZD} \tag{3.5}$$

with all third order interaction effects removed. The α's and β's are related to the θ's as

$$\begin{aligned} \alpha &= 2\theta_1^Y, & \alpha_k^C &= 2\theta_{1k}^{YC}, & \alpha_\ell^D &= 2\theta_{1\ell}^{YD} \\ \beta &= 2\theta_1^Z, & \beta_k^C &= 2\theta_{1k}^{ZC}, & \beta_\ell^D &= 2\theta_{1\ell}^{ZD} \end{aligned} \tag{3.6}$$

In general, as shown in Nerlove and Press [1976] among others, the two dependent variables, conditional on the explanatory variables, are independent if and only if $\theta^{YZ}_{ij} = \theta^{YZC}_{ijk} = \theta^{YZD}_{ij\ell} = \theta^{YZCD}_{ijk\ell} = 0$ for all i, j, k and ℓ.

When the disturbances ε_1, ε_2 are independent and are logistic distributed, equations (3.1) and (3.2) imply a fully recursive logit equation model,

$$\ell n \frac{P^{\bar{Y}CD}_{1k\ell}}{P^{\bar{Y}CD}_{2k\ell}} = \alpha + \alpha^C_k + \alpha^D_\ell \tag{3.7}$$

$$\ell n \frac{P^{Y\bar{Z}CD}_{i1k\ell}}{P^{Y\bar{Z}CD}_{i2k\ell}} = \ell n \frac{P^{\bar{Y}\bar{Z}CD}_{i1k\ell}}{P^{\bar{Y}\bar{Z}CD}_{i2k\ell}} = \beta + \beta^C_k + \beta^D_\ell + \lambda^{YZ}_{i1} \tag{3.8}$$

$$= \beta + \beta^C_k + \beta^D_\ell + \lambda Y^*_i$$

where $\lambda = \lambda^{YZ}_{11}$.

To investigate the relationships between this model and the log linear model, the lemma can be used. It follows from equations (2.12) and (2.14)

$$\ell n \frac{P^{\bar{Y}ZCD}_{1jk\ell}}{P^{\bar{Y}ZCD}_{2jk\ell}} = \alpha + \alpha^C_k + \alpha^D_\ell + \lambda Z^*_j + G(k, \ell) \tag{3.9}$$

where

$$G(k, \ell) = \lambda + \ell n\left(\frac{1 + \exp(\beta + \beta^C_k + \beta^D_\ell - \lambda)}{1 + \exp(\beta + \beta^C_k + \beta^D_\ell + \lambda)}\right) \tag{3.10}$$

The logarithmic probability odds equation (3.9) can be reparameterized:

$$\ell n \frac{P^{\bar{Y}ZCD}_{1jk\ell}}{P^{\bar{Y}ZCD}_{2jk\ell}} = \zeta + \zeta^C_k + \zeta^D_\ell + \zeta^{CD}_{k\ell} + \lambda Z^*_j \tag{3.11}$$

where the parameters satisfy the ANOVA constraints and are defined by equations (3.9) and (3.10). Let $G(k, \cdot) = \frac{1}{L} \Sigma_{\ell=1}^{L} G(k, \ell)$, $G(\cdot, \ell) = \frac{1}{K} \Sigma_{k=1}^{K} G(k, \ell)$, and $G(\cdot, \cdot) = \frac{1}{KL} \Sigma_{k=1}^{K} \Sigma_{\ell=1}^{L} G(k. \ell)$. Explicitly,

$$\zeta = \alpha + G(\cdot, \cdot)$$

$$\zeta_k^C = \alpha_k^C + G(k, \cdot) - G(\cdot, \cdot)$$

$$\zeta_\ell^D = \alpha_\ell^D + G(\cdot, \ell) - G(\cdot, \cdot)$$

$$\zeta_{k\ell}^{CD} = \alpha_{k\ell}^{CD} + G(k, \ell) - G(\cdot, \ell) - G(k, \cdot) + G(\cdot, \cdot) \tag{3.12}$$

where $\alpha_{k\ell}^{CD} = 0$.

Thus, the fully recursive logit model in (3.7) and (3.8) is contained in the log-linear model given by (3.11) and (3.8). These two models are nested. The recursive model implies nonlinear cross equation constraints on the two conditional structural equations (3.10) and (3.11) and additional interaction among explanatory variables, i.e., $\zeta_{k\ell}^{CD}$, is introduced. The effective constraint in (3.12) is that the interaction effect $\zeta_{k\ell}^{CD}$ is a function of all the parameters in the other conditional structural equation (3.8). Thus the fully recursive model can be discriminated from the unconstrained log-linear model (3.8) and (3.11). Note that if the explanatory variable C (or D) is ordered and has linear effects in (3.7), the implied effects ζ_k^C (or ζ_ℓ^D) will not be linear.

For the three dimensional marginal table of Y, C and D, equation (3.7) is nonsaturated. In the saturated equation model, i.e.,

$$\ln \frac{p_{1k\ell}^{\bar{Y}CD}}{p_{2k\ell}^{\bar{Y}CD}} = \alpha + \alpha_k^C + \alpha_\ell^D + \alpha_{k\ell}^{CD} \tag{3.7$'$}$$

where the effects are nonlinear, there will be no effective constraints imposed on (3.11). In this case, the fully recursive model given by (3.7) and (3.8) is statistically equivalent to the log-linear model given by (3.11) and (3.8). This follows immediately from the lemma, and the relations (3.12).

If the explanatory variables C and D have interaction effects in equation (3.8), i.e.,

$$\ln \frac{P_{i1k\ell}^{Y\bar{Z}CD}}{P_{i2k\ell}^{Y\bar{Z}CD}} = \beta + \beta_k^C + \beta_\ell^D + \beta_{k\ell}^{CD} + \lambda Y_i^* \qquad (3.8)'$$

and if the first equation is saturated as in (3.7)' the fully recursive model in (3.7)' and (3.8)' will be statistically equivalent to an alternative fully recursive model as in (3.13) and (3.14).

$$\ln \frac{P_{1k\ell}^{\bar{Z}CD}}{P_{2k\ell}^{\bar{Z}CD}} = \beta^* + \beta_k^{*C} + \beta_\ell^{*D} + \beta_{k\ell}^{*CD} \qquad (3.13)$$

$$\ln \frac{P_{1jk\ell}^{\bar{Y}ZCD}}{P_{2jk\ell}^{\bar{Y}ZCD}} = \alpha^* + \alpha_k^{*C} + \alpha_\ell^{*D} + \alpha_{k\ell}^{*CD} + \lambda Z_j^* \qquad (3.14)$$

This follows from similar arguments in the last paragraph.

Let us consider now the Nerlove and Press models with continuous explanatory variables. The fully recursive model with continuous variables that will be compared with the Nerlove and Press models can be specified as

$$\ln \frac{P(Y = 1)}{P(Y = 2)} = R\alpha \qquad (3.15)$$

$$\ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \qquad (3.16)$$

The Nerlove-Press model is specified as

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + \lambda Z_1 \tag{3.17}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \tag{3.18}$$

where Y_1 and Z_1 are dummy variables as defined in the corollary. To analyze the differences of these two models, the corollary can be used. It follows from this corollary, the equation (3.15) is equivalent to

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + \lambda Z_1 + G_1(Q\beta, \lambda) \tag{3.19$'$}$$

where $G_1(Q\beta, \lambda)$ is defined in (2.23). Thus the two models are different in that the implied equation (3.15)' is nonlinear in parameters as well as nonlinear in variables. When Q contains continuous explanatory variables, it is obvious that $G_1(Q\beta, \lambda)$ can not be an exact linear function of the explanatory variables in R. Hence, in principle, the specification of the fully recursive logit model in (3.15) and (3.16) can be discriminated from the Nerlove-Press model in (3.17) and (3.18).

4. Estimation and Statistical Test

To discriminate the fully recursive model in (3.19) and (3.20) as compared with the log-linear model in (3.21) and (3.22), we suggest the following procedure. The two models can be nested into the following generalized model,

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + \lambda Z_1 + \delta G_1(Q\beta, \lambda) \tag{4.1}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + \lambda Y_1 \tag{4.2}$$

The fully recursive model corresponds to the above model with $\delta = 1$ and the log-linear model as specified in (3.17) and (3.18) corresponds to the above model with $\delta = 0$. So to test the hypothesis, we can test the corresponding values of δ in the generalized model if the parameter δ is identifiable in the above model. The identification of δ requires that the function $G_1(Q\beta, \lambda)$ is not an exact linear combination of the explanatory variables in R. As we have pointed out in the previous paragraph, if all the explanatory variables in Q are qualitative, depending on the variables in R, G_1 may be an exact linear combination of R and δ will not be identifiable.

Let us assume that (Y_i, Z_i, R_i, Q_i), $i = 1,\ldots,N$, be the observed random samples. There are several different ways that we can estimate the model. One approach is the maximum likelihood procedure. Unfortunately, the log likelihood function for the above model will not be concave even though the likelihood functions for the fully recursive and the Nerlove-Press model specified in (3.17) and (3.18) are concave. An alternative procedure which is relatively simpler but inefficient is to estimate each equation separately as follows. First, we estimate equation (4.2) by the logit maximum likelihood procedure. The log likelihood function is

$$\ell nL_2(\theta_2) = \Sigma_{i=1}^{N}\{Z_{1i}(Q_i\beta + \lambda Y_{1i}) - \ell n(1 + \exp(Q_i\beta + \lambda Y_{1i}))\}$$

where $\theta_2' = (\beta', \lambda)$. Let $X_{2i} = (Q_i, Y_{1i})$. The first and second derivatives of $\ell nL_2(\theta_2)$ are

$$\frac{\partial nL_2(\theta_2)}{\partial\theta_2} = \Sigma_{i=1}^{N}\{Z_{1i} - \frac{\exp(X_{2i}\theta_2)}{1 + \exp(X_{2i}\theta_2)}\} X_{2i}'$$

and

$$\frac{\partial^2 \ell nL_2(\theta_2)}{\partial\theta_2\partial\theta_2'} = -\Sigma_{i=1}^{N} \frac{\exp(X_{2i}\theta_2)}{(1 + \exp(X_{2i}\theta_2))^2} X_{2i}'X_{2i}$$

Assume that the first two moments of X_2 exist, its variance matrix is nonsingular and the parameter space of (β, λ) is compact. It can be easily shown that the maximum likelihood estimate $\hat{\theta}_2$ is strongly consistent and is asymptotic normal. The asymptotic covariance matrix can be estimated by

$$\left[-\frac{\partial^2 \ell nL_2(\hat{\theta}_2)}{\partial\theta_2\partial\theta_2'}\right]^{-1} .$$

To estimate the equation (4.1), we can use a two stage procedure similar to the one suggested in Lee [1978, 1979]. The two stage estimate $\hat{\theta}_1$ of $\theta_1 = (\alpha', \delta, \lambda)'$ is derived by maximing the following function

$$\ell nL_1(\theta_1,\hat{\theta}_2) = \Sigma_{i=1}^{N}\{Y_{1i}(R_i\alpha+\lambda Z_{1i}+\delta G_1(Q_i\hat{\beta},\hat{\lambda}))$$
$$- \ell n(1+\exp(R_i\alpha+\lambda Z_{1i}+\delta G_1(Q_i\hat{\beta},\hat{\lambda})))\}$$

Denote $X_{1i} = (R_i, G_1(Q_i\beta, \lambda), Z_{1i})$ and $\hat{X}_{1i} = (R_i, G_1(Q_i\hat{\beta}, \hat{\lambda}), Z_{1i})$. The first and second derivatives of $\ell nL_1(\theta_1, \hat{\theta}_2)$ are

$$\frac{\partial \ell nL_1(\theta_1, \hat{\theta}_2)}{\partial \theta_1} = \Sigma_{i=1}^{N} \{Y_{1i} - \frac{\exp(\hat{X}_{1i}\theta_1)}{1+\exp(\hat{X}_{1i}\theta_1)}\} \hat{X}'_{1i}$$

and

$$\frac{\partial^2 \ell nL_i(\theta_1, \hat{\theta}_2)}{\partial \theta_1 \partial \theta'_1} = -\Sigma_{i=1}^{N} \frac{\exp(\hat{X}_{1i}\theta_1)}{(1+\exp(\hat{X}_{1i}\theta_1))^2} \hat{X}'_{1i}\hat{X}_{1i}$$

In addition to the assumptions we have made, we assume also that the parameter space of θ_1 is compact, the first two moments of X_1 exist and its variance matrix is non-singular. Similar to the arguments in Lee [1979], the estimate $\hat{\theta}_1$ is strongly consistent. Asymptotic normality can also be proved. However the asymptotic covariance matrix is not the matrix

$$\left[- \frac{\partial^2 \ell nL_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta'_1}\right]^{-1} .$$

The correct asymptotic covariance matrix, which is derived as in Amemiya [1979], is

$$V(\hat{\theta}_1) = \frac{1}{N}[- \frac{1}{N} \frac{\partial^2 \ell nL_1(\theta_1,\theta_2)}{\partial \theta_1 \partial \theta'_1}]^{-1} \cdot \{[- \frac{1}{N} \frac{\partial^2 \ell nL_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta'_1}]$$

$$+ [\frac{1}{N} \frac{\partial^2 \ell nL_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta'_2}][- \frac{1}{N} \frac{\partial^2 \ell nL_2(\theta_2)}{\partial \theta_2 \partial \theta'_2}]^{-1}[\frac{1}{N} \frac{\partial^2 \ell nL_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta'_2}]'$$

$$+ [\frac{1}{N} \frac{\partial^2 \ell nL_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta'_2}][- \frac{1}{N} \frac{\partial^2 \ell nL_2(\theta_2)}{\partial \theta_2 \partial \theta'_2}]^{-1}(\frac{1}{N} \frac{\partial \ell nL_2(\theta_2)}{\partial \theta_2} \frac{\partial \ell nL_1(\theta_1, \theta_2)}{\partial \theta'_1})$$

$$+ \left(\frac{1}{N} \frac{\partial \ell n L_1(\theta_1, \theta_2)}{\partial \theta_1} \frac{\partial \ell n L_2(\theta_2)}{\partial \theta_2'}\right) \left[- \frac{1}{N} \frac{\partial^2 \ell n L_2(\theta_2)}{\partial \theta_2 \partial \theta_2'}\right]^{-1}$$

$$\cdot \left[\frac{1}{N} \frac{\partial^2 \ell n L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'}\right]' \} \cdot \left[- \frac{1}{N} \frac{\partial^2 \ell n L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_1'}\right]^{-1}$$

The expression can be simplified with the relations

$$\frac{1}{N} \frac{\partial \ell n L_1(\theta_1, \theta_2)}{\partial \theta_1} \cdot \frac{\partial \ell n L_2(\theta_2)}{\partial \theta_2'} = \frac{1}{N} \Sigma_{i=1}^{N} \left(Y_{i1} - \frac{\exp(X_{1i}\theta_1)}{1+\exp(X_{1i}\theta_2)}\right)$$

$$\cdot \left(Z_{1i} - \frac{\exp(X_{2i}\theta_2)}{1+\exp(X_{2i}\theta_2)}\right)$$

and

$$\frac{1}{N} \frac{\partial^2 \ell n L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'} = - \frac{1}{N} \Sigma_{i=1}^{N} \delta \frac{\exp(X_{1i}\theta_1 + Q_i\beta)}{(1+\exp(X_{1i}\theta_1))^2 (1+\exp(Q_i\beta+\lambda))(1+\exp(Q_i\beta))} \cdot$$

$$X_{1i}'[(1 - \exp(\lambda))Q_i, - \exp(\lambda)(1+\exp(Q_i\beta))]$$

which are valid asymptotically. Statistical test for the hypothesis $\delta = 1$ or $\delta = 0$ can be derived based on the asymptotic normality distribution of the estimate $\hat{\delta}$.

5. Fully Recursive Logit Model with Interaction Terms

The fully recursive model in equations (3.1) and (3.2) and the model in equations (3.15) and (3.16) analyzed in previous sections do not include interaction terms of the dependent variable Y (or Z) with the exogenous variables. For the log linear models, such interaction terms are usually included in the models for the analysis of goodness-of-fit. In the Nerlove and Press models, they also consider such possible interaction effects,

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + (S\theta)Z_1 \tag{5.1}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + (S\theta)Y_1 \tag{5.2}$$

where the vector S consisting of exogenous variables which may or may not be subset of (R, Q). This model generalizes the log-linear model in equations (3.17) and (3.18). When all the exogenous variables are discrete, these correspond to the equations (2.4) and (2.5).

The full recursive model can similarly be generalized to include such interactions,

$$\ell n \frac{P(Y = 1)}{P(Y = 2)} = R\alpha \tag{5.3}$$

$$\ell n \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q\beta + (S\theta)Y_1 \tag{5.4}$$

Similar to the arguments in the lemma, the two equations imply

$$\ell n \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R\alpha + (S\theta)Z_1 + \ell n\left(\frac{1+\exp(Q\beta)}{1+\exp(Q\beta+S\theta)}\right) \tag{5.5}$$

The estimation and testing procedure discussed in the previous paragraph can be generalized in a straightforward manner to these equations.

When all the exogenous variables are discrete and there are no excluded variables restrictions in the equation (5.4), one should be careful in interpreting the causal relations since they may be equivalent to some other bivariate probability models where the dependence of Y and Z are purely spurious. To illustrate this possibility, we will consider some bivariate probit models and compare its structure with the fully recursive logit models.

Conditional on the explanatory variables, if Z is not dependent on Y, i.e., $\lambda_{i1}^{YZ} = 0$ in (3.2), the model (3.1) and (3.2) is a bivariate probability model. If ε_1 and ε_2 are bivariate normal variates with zero means and correlation coefficient ρ, the model becomes a bivariate normal (or probit) model. Let $\Phi(\cdot)$ be the standard normal distribution and $F(\cdot, \cdot, \rho)$ be the bivariate standard normal distribution with correlation coefficient ρ. The bivariate normal model is

$$P_{1k\ell}^{\bar{Y}CD} = \Phi(\alpha + \alpha_k^C + \alpha_\ell^D) \tag{5.6}$$

$$P_{1k\ell}^{\bar{Z}CD} = \Phi(\beta + \beta_k^C + \beta_\ell^D) \tag{5.7}$$

$$P_{11k\ell}^{\bar{Y}\bar{Z}CD} = F(\alpha + \alpha_k^C + \alpha_\ell^D,\ \beta + \beta_k^C + \beta_\ell^D,\ \rho) \tag{5.8}$$

These probabilities can be rewritten as

$$P_{ijk\ell}^{\bar{Y}\bar{Z}CD} = G(Y, Z, C, D)$$

$$\equiv \int_{-\infty}^{(-1)^{1+Y}(\alpha+\alpha_k^C+\alpha_\ell^D)} \int_{-\infty}^{(-1)^{1+Z}(\beta+\beta_k^C+\beta_\ell^D)} f((-1)^{1+Y}\varepsilon_1,$$

$$(-1)^{1+Z}\varepsilon_2,\ \rho)d\varepsilon_1 d\varepsilon_2$$

where $f(\cdot,\cdot,\rho)$ is the bivariate standard normal density with correlation coefficient ρ.

To compare the bivariate normal model with the log linear model, we consider

$$\ell n \frac{p_{1jk\ell}^{\bar{Y}ZCD}}{p_{2jk\ell}^{\bar{Y}ZCD}} = \ell n \frac{G(1,\ Z,\ C,\ D)}{G(2,\ Z,\ C,\ D)} \tag{5.9}$$

$$= \zeta + \zeta_k^C + \zeta_\ell^D + \zeta_{k\ell}^{CD} + \zeta_j^Z + \zeta_{jk}^{ZC} + \zeta_{j\ell}^{ZD} + \zeta_{jk\ell}^{ZCD}$$

$$\ell n \frac{p_{i1k\ell}^{Y\bar{Z}CD}}{p_{i2k\ell}^{Y\bar{Z}CD}} = \ell n \frac{G(Y,\ 1,\ C,\ D)}{G(Y,\ 2,\ C,\ D)} \tag{5.10}$$

$$= \eta + \eta_k^C + \eta_\ell^D + \eta_{k\ell}^{CD} + \eta_i^Y + \eta_{ik}^{YC} + \eta_{i\ell}^{YD} + \eta_{ik\ell}^{YCD}$$

where the parameters ζ's and η's are defined from the equation and satisfy the ANOVA constraints. The parameters ζ and η are nonlinear functions of the parameters in the bivariate normal model. When $\rho = 0$, the model becomes the independent probit model and all the interaction terms containing both Y and Z vanish. Therefore, correlation between the disturbances implies the inclusion of second order terms θ_{ij}^{YZ} between the dependent variables and all of their higher order interaction terms with the explanatory variables, θ_{ijk}^{YZC}, $\theta_{ij\ell}^{YZD}$ and $\theta_{ijk\ell}^{YZCD}$, in the general log-linear model. Since the effects in the log linear models are highly nonlinear functions of the effects in the bivariate normal model, the effects in the log linear models will be nonlinear even if C (or D) are ordered variables and have linear effects in (5.6) - (5.8).

In the above bivariate normal model, the correlation coefficient ρ is constant; more generally, the model can be extended to allow the correlation coefficient ρ to depend on the levels of C and D as $\rho_1 + \rho_k^C + \rho_\ell^D + \rho_{k\ell}^{CD}$ with the parameters ρ_k^C, ρ_ℓ^D, and $\rho_{k\ell}^{CD}$ satisfying the ANOVA constraint. Using the same arguments, this model is also a nested model of the general saturated log-linear model. The most general bivariate normal model with heteroscedastic correlation coefficients,

$$p_{1k\ell}^{\bar{Y}CD} = \Phi(\alpha + \alpha_k^C + \alpha_\ell^D + \alpha_{k\ell}^{CD}) \tag{5.11}$$

$$p_{1k\ell}^{\bar{Z}CD} = \Phi(\beta + \beta_k^C + \beta_\ell^D + \beta_{k\ell}^{CD}) \tag{5.12}$$

$$p_{11k\ell}^{\bar{Y}\bar{Z}CD} = F(\alpha+\alpha_k^C+\alpha_\ell^D+\alpha_{k\ell}^{CD},\ \beta+\beta_k^C+\beta_\ell^D+\beta_{k\ell}^{CD},\ \rho_1+\rho_k^C+\rho_\ell^D+\rho_{k\ell}^{CD}), \tag{5.13}$$

is statistically equivalent to the saturated log linear model given in (2.4) and (2.5). This equivalence relation has been proved in Lee [1979b].

The bivariate normal model can also be written in a fully recursive form as in equations (5.6) and (5.10). We have pointed out that the general bivariate normal model is equivalent to the saturated log-linear model since the general bivariate normal model is itself saturated. Equation (5.11) is saturated for the three way marginal table Y, C and D with fixed two way margins C and D. Consider now other cases where equation (5.11) need not be saturated,

$$p_{1k\ell}^{\bar{Y}CD} = \Phi(a_{k\ell}) \tag{5.14}$$

$$p_{1k\ell}^{\bar{Z}CD} = \Phi(\beta + \beta_k^C + \beta_\ell^D + \beta_{k\ell}^{CD}) \tag{5.15}$$

$$p_{11k\ell}^{\bar{Y}\bar{Z}CD} = F(a_{k\ell},\ \beta + \beta_k^C + \beta_\ell^D + \beta_{k\ell}^{CD},\ \rho + \rho_k^C + \rho_\ell^D + \rho_{k\ell}^{CD}) \tag{5.16}$$

where $a_{k\ell}$ denote some unspecified combination of effects. This model is saturated for the three way marginal table Z, C, D with fixed margins C, D; however, it need not be saturated for the full four way table, nor the three way table Y, C, D with fixed margins, C, D. Consider another recursive model,

$$p_{1k\ell}^{\bar{Y}CD} = \Phi(a_{k\ell}) \tag{5.17}$$

$$\ell n \frac{p_{i1k\ell}^{Y\bar{Z}CD}}{p_{i2k\ell}^{Y\bar{Z}CD}} = \eta + \eta_k^C + \eta_\ell^D + \eta_{k\ell}^{CD} + \eta_i^Y + \eta_{ik}^{YC} + \eta_{i\ell}^{YD} + \eta_{ik\ell}^{YCD} \tag{5.18}$$

where the second equation is saturated for the four way table with fixed margins C, D, and Y. We would like to show that, conditional on Y, C and D, equations (5.15) and (5.16) are equivalent to equation (5.18). This can be done as follows.

Since $p_{11k\ell}^{\bar{Y}\bar{Z}CD} = p_{11k\ell}^{Y\bar{Z}CD} p_{1k\ell}^{\bar{Y}CD}$ and

$$p_{1k\ell}^{\bar{Z}CD} = p_{11k\ell}^{\bar{Y}\bar{Z}CD} + p_{21k\ell}^{\bar{Y}\bar{Z}CD}$$

$$= p_{11k\ell}^{Y\bar{Z}CD}\, p_{1k\ell}^{\bar{Y}CD} + p_{21k\ell}^{Y\bar{Z}CD}\,(1 - p_{1k\ell}^{\bar{Y}CD})$$

it follows

$$\begin{pmatrix} p_{1k\ell}^{\bar{Y}CD} & 0 \\ p_{1k\ell}^{\bar{Y}CD} & 1-p_{1k\ell}^{\bar{Y}CD} \end{pmatrix} \begin{pmatrix} p_{11k\ell}^{Y\bar{Z}CD} \\ p_{21k\ell}^{Y\bar{Z}CD} \end{pmatrix} = \begin{pmatrix} p_{11k\ell}^{\bar{Y}\bar{Z}CD} \\ p_{1k\ell}^{\bar{Z}CD} \end{pmatrix} \tag{5.19}$$

As long as $p_{1k\ell}^{\bar{Y}CD}$ is fixed and is neither zero nor one, there is a one-to-one correspondence between the set of conditional probabilities $p_{i1k\ell}^{Y\bar{Z}CD}$ and $(p_{11k\ell}^{\bar{Y}\bar{Z}CD}, p_{1k\ell}^{\bar{Z}CD})$ as in (5.19). Consider equation (5.18) which implies

$$p^{Y\bar{Z}CD}_{i1k\ell} = \frac{\exp(\eta + \eta^{C}_{k} + \eta^{D}_{\ell} + \eta^{CD}_{k\ell} + \eta^{Y}_{i} + \eta^{YC}_{ik} + \eta^{YD}_{i\ell} + \eta^{YCD}_{ik\ell})}{1 + \exp(\eta + \eta^{C}_{k} + \eta^{D}_{\ell} + \eta^{CD}_{k\ell} + \eta^{Y}_{i} + \eta^{YC}_{ik} + \eta^{YD}_{i\ell} + \eta^{YCD}_{ik\ell})}$$

Thus the effects η define a unique set of conditional probabilities $p^{Y\bar{Z}CD}_{i1k\ell}$. Conversely, the effects η can be solved for uniquely from the conditional probabilities as equation (5.18) is saturated for the 4 way table with fixed three way margins Y, C and D. Let

$$a_{ik\ell} = \ell n \frac{p^{Y\bar{Z}CD}_{i1k\ell}}{p^{Y\bar{Z}CD}_{i2k\ell}}, \quad a_{\cdot k\ell} = \frac{1}{2} \Sigma^{2}_{i=1} a_{ik\ell}, \quad a_{i\cdot\ell} = \frac{1}{K} \Sigma^{K}_{k=1} a_{ik\ell},$$

$$a_{ik\cdot} = \frac{1}{L} \Sigma^{L}_{\ell=1} a_{ik\ell}, \quad a_{i\cdot\cdot} = \frac{1}{KL} \Sigma^{K}_{k=1} \Sigma^{L}_{\ell=1} a_{ik\ell}, \quad a_{\cdot k\cdot} = \frac{1}{2L} \Sigma^{2}_{i=1} \Sigma^{L}_{\ell=1} a_{ik\ell}$$

$$a_{\cdot\cdot\ell} = \frac{1}{2K} \Sigma^{2}_{i=1} \Sigma^{K}_{k=1} a_{ik\ell} \quad \text{and} \quad a_{\cdot\cdot\cdot} = \frac{1}{2KL} \Sigma^{2}_{i=1} \Sigma^{K}_{k=1} \Sigma^{L}_{\ell=1} a_{ik\ell}.$$

The effects η can be solved uniquely as

$$\eta = a_{\cdot\cdot\cdot}$$

$$\eta^{C}_{k} = a_{\cdot k\cdot} - a_{\cdot\cdot\cdot}$$

$$\eta^{D}_{\ell} = a_{\cdot\cdot\ell} - a_{\cdot\cdot}$$

$$\eta^{Y}_{i} = a_{i\cdot\cdot} - a_{\cdot\cdot}$$

$$\eta^{YC}_{ik} = a_{ik\cdot} - a_{i\cdot\cdot} - a_{\cdot k\cdot} + a_{\cdot\cdot\cdot}$$

$$\eta^{YD}_{i\ell} = a_{i\cdot\ell} - a_{i\cdot\cdot} - a_{\cdot\cdot\ell} + a_{\cdot\cdot\cdot}$$

$$\eta^{CD}_{k\ell} = a_{\cdot k\ell} - a_{\cdot k\cdot} - a_{\cdot\cdot\ell} + a_{\cdot\cdot\cdot}$$

$$\eta^{YCD}_{ik\ell} = a_{ik\ell} - a_{ik\cdot} - a_{i\cdot\ell} - a_{\cdot k\ell} + a_{i\cdot\cdot} + a_{\cdot k\cdot} + a_{\cdot\cdot\ell} - a_{\cdot\cdot\cdot}$$

Therefore, there is a one-to-one correspondence between the η's and the $p_{i1k\ell}^{Y\bar{Z}CD}$. Now consider equations (5.15) and (5.16) conditional on given $p_{1k\ell}^{\bar{Y}CD}$ or equivalently on fixed $a_{k\ell}$. The bivariate normal distribution can be rewritten in an alternative form. Since $\partial F/\partial\rho = f$, the standard bivariate normal density function, it follows by the Fundamental Theorem of Calculus,

$$F(h,\ k,\ \rho) = \int_0^{\rho} f(h,\ k,\ \varepsilon)d\varepsilon + \Phi(h)\Phi(k).$$

Since f is a positive density function, the integral $I_{h,k}(\rho) = \int_0^{\rho} f(h,\ k,\ \varepsilon)d\varepsilon$ is a strictly positive function of ρ on $(-1,\ 1)$. Hence equations (5.15) and (5.16) can be solved as

$$\beta + \beta_k^C + \beta_\ell^D + \beta_{k\ell}^{CD} = \Phi^{-1}(p_{1k\ell}^{\bar{Z}CD}) \qquad (5.20)$$

$$\rho + \rho_k^C + \rho_\ell^D + \rho_{k\ell}^{CD} = I^{-1}_{a_{k\ell},\Phi^{-1}(p_{1k\ell}^{\bar{Z}CD})}(p_{11k\ell}^{\bar{Y}\bar{Z}CD} - p_{1k\ell}^{\bar{Y}CD}\ p_{1k\ell}^{\bar{Z}CD}) \qquad (5.21)$$

The effects, β's and ρ's, can be solved for uniquely from the ANOVA formulae (5.20) and (5.21). Since these effects also uniquely define the probabilities $p_{1k\ell}^{\bar{Z}CD}$ and $p_{11k\ell}^{\bar{Y}\bar{Z}CD}$ given $a_{k\ell}$ as in (5.15) and (5.16), there is one-to-one correspondence between the effects and the probabilities. Therefore we conclude that the model in (5.14) - (5.16) is equivalent to models in (5.17) - (5.18).

The equation (5.17) is in the probit form but it can be reparametered into logit form. For example, if $a_{k\ell} = \alpha + \alpha_k^c$ we have

$$\ell n \frac{P^{\bar{Y}CD}_{1k\ell}}{P^{\bar{Y}CD}_{2k\ell}} = \ell n \frac{\Phi_N(\alpha + \alpha^c_k)}{1-\Phi_N(\alpha+\alpha^c_k)}$$

$$= \zeta + \zeta^c_k \tag{5.22}$$

where the parameters ζ and ζ^c_k are defined from the above equation and satisfy the ANOVA constraints. Thus the dependence of Z on Y in the fully recursive model (5.22) and (5.18) may be spurious and not causal.

6. An empirical illustration

To illustrate the analysis and the statistical test for comparing the model, we provide an empirical example. In this example, we investigate the association of workers' job satisfaction and subsequent job quitting behavior in the labor market. An analysis of this sort is in Freeman [1978] for several different data sets. Here we use only the 1972-1973 A Panel Study of Income Dynamics [1972] data for our analysis. Our samples are restricted to the head of family in 1972-1973, employed in 1972 and employed or unemployed in 1973. The list of variables used and the detail descriptions are in the appendix. Observations with incomplete information on the list of variables are deleted. The total samples used are 2606. All information on the list of variables except the QUIT variable are derived from the 1972 survey. The explanatory variables included in the models are similar to the variables used in Freeman [1978] but are not identically constructed.

The Nerlove-Press log-linear model and the logit causal chain model are both estimated. This data set has the advantage of linking past job satisfaction to job mobility which provides a fix on the lines of causality (Freeman [1978]). The job satisfaction variable may be related to unobservable object variables such as working conditions, etc. On the other hand, as pointed out in Freeman [1978], in industrial psychology literatures, it is defined as a "positive emotional state resulting from the appraisal of one's job." Based solely on the latter interpretation, this psychological state might present causal effect on job mobility. This argument might provide the justification for causal modelling.

The estimated results are presented in TABLE 1. The fully recursive model consists of the Satisfaction Eq. I and the Quit Equation. The Nerlove-Press log linear model consists of the Satisfaction Eq. II and

Quit Equation. These equations are estimated by the logit maximum likelihood procedure applied separately to each equation. The QUIT and SAT variables have a significant association and is not independent. The job satisfaction, unions, job tenure, education and age all have significant negative effects on job quitting behavior. The other variables are of interest even though they are not significant at the coventional five percent level of significance. The good labor market conditions provide attractive job alternatives and increase the probability of job mobility. The two estimated satisfaction equations I and II have similar results. The only significant exogenous variable is the UNION variable. All the estimated coefficients have the same signs in the two equations and are not greatly different. The negative effect of unions on job satisfaction has been argued by Freeman [1978] that it reflects the role of unions as a "voice" institution, encouraging workers to express discontent related to the job. It is of interest to note that even though for our estimation technique, the equality constraint on the coefficients of SAT and QUIT in the Nerlove-Press model has not been imposed, the estimated coefficients are quite close.

To discriminate these two models, we estimate the generalized Satisfaction eq. III. This equation is estimated by the two stage method described in the text. The estimated coefficient of $G = \ell n[\,(1+\exp(Q\hat{\beta}))/(1+\exp(Q\hat{\beta}+\hat{\lambda}))]$ is -6.8438 with estimated standard error 5.0797. As the standard error is large, neither the null hypothesis of composite hypothesis H_0: $\delta = 0$ vs H_1: $\delta \neq 0$, i.e., the Nerlove-Press model, nor the null hypothesis of the composite hypothesis H_0: $\delta = 1$, vs $H_1 \neq 1$, i.e., the full recursive logit model, can be rejected. Similarly, the null hypothesis H_o of the simple hypothesis H_o: $\delta = 0$ vs. H_1: $\delta = 1$ cannot be rejected with one-side-test at conventional level of significance.

TABLE 1: SINGLE EQUATION PROBIT ESTIMATES

Variables	Quit Eq.	Satisfaction Eqs. I	II	III	IV
Constant	.0290 (.5141)	1.8015 (.3316)*	2.0443 (.3415)*	3.6042 (1.0174)*	3.1076 (.1633)*
UNION	-.4592 (.1818)*	-.6984 (.1595)*	-.7335 (.1603)*	-.9317 (.2124)*	-.8213 (.1669)*
LNWAGE	-.1846 (.1291)	.2465 (.1366)	.2338 (.1380)	.1374 (.1662)	
TENURE	-1.0435 (.1721)*	.1668 (.1675)	.0935 (.1709)	-.3495 (.3394)	
EDUC1	-.1037 (.1597)	-.0947 (.1669)	-.1025 (.1667)	-.1472 (.1906)	
EDUC2	-.5668 (.2835)*	-.0128 (.2905)	-.0442 (.2907)	-.2956 (.3381)	
AGE	-.0358 (.0071)*	.0135 (.0069)	.0112 (.0069)	-.0037 (.0116)	
RACE	.0956 (.1564)	.2046 (.1626)	.2148 (.1628)	.2613 (.1870)	
SEX	.1299 (.2599)	-.0336 (.3064)	-.0206 (.3066)	.0809 (.3651)	
MARITAL	-.2039 (.2301)	-.1981 (.2692)	-.2177 (.2692)	-.3360 (.3298)	
CHILDREN	.0044 (.0455)	.0369 (.0444)	.0356 (.0445)	.0279 (.0482)	
SHORT	.2350 (.1625)				
AWAGE	.1014 (.1502)				
UNRATE	.0181 (.0348)				
SAT	-.6396 (.2045)*				
QUIT			-.6803 (.2028)*	-.6409 (.2046)*	-.6373 (.2038)*
G				-6.8438 (5.0797)	-4.2044 (2.3649)

Notes to Table 1

(1) The figures in the brackets are standard errors; * indicates significance at five percent level of significance.

(2) The variable G is constructed as $G = G_1(Q\hat{\beta}, \hat{\lambda})$ described in the text.

(3) -2 times log-likelihood = 1500.9 for the Quit equation.

= 1448.03 for the Satisfaction Eq. I

= 1437.77 for the Satisfaction Eq. II.

= 1434.78 for the Satisfaction Eq. III.

= 1441.71 for the Satisfaction Eq. IV.

It can be shown analytically that the function $G = \ln[(1+\exp(X))/(1+\exp(X+\lambda))]$ is a monotone function of x bounded by the constant functions 0 and $-\lambda$. It is a strictly increasing function if λ is negative, and is strictly decreasing if λ is positive. Thus the range of the values of G is determined by the value of λ. For our samples, the mean of the nonlinear function G is 0.0486, its standard deviation is 0.0413 and its range is bounded by 0 and 0.6396. The simple correlations of the variable G with each exogenous variables in the equations are not high [1/] but the multiple correlation coefficient R is .90. Apparently, this high multiple correlation induces high standard error for the coefficient estimate of G. To reduce multicollinearity, we estimate an additional equation VI which deletes the insignificant exogenous variables. The estimated coefficient of G for this equation is -4.2044 with estimated standard error 2.3649. With this estimate, we can reject the full recursive model in the composite hypothesis H_0: $\delta = 1$ vs H_1: $\delta \neq 1$ at the five percent level of significance. The Nerlove-Press model specification can not be rejected.

From this analysis, we can conclude that SAT is more likely related to unobserved objective factors such as working conditions, etc. and should not be regarded as a causal factor in job quitting behavior.

Finally, we estimate the Quit equation with all the interaction terms of SAT with all the exogenous variables included. The "-2 log likelihood value" is 1481.099. Its comparison with the "-2 log likelihood value" 1500.878 for the Quit equation without such interaction terms results in a likelihood ratio chi-square test of 19.779 with 13 degrees of freedom. The interaction terms are not significant at the ten percent level of significance.

---

[1] The highest correlation coefficients of G are -.76 with TENURE, -.67 with AGE and -.36 with LNWAGE.

6. Conclusions

In this article, we have investigated the relationship between the fully recursive logit models and the Nerlove-Press multivariate log-linear probability models. We emphasize the important role of the explanatory variables in discriminating those models. An asymptotic test procedure is introduced. An illustrative empirical example is provided.

When the explanatory variables are qualitative, we have demonstrated that there are cases that those models can be statistically equivalent and can not be discriminated. To choose between those models, one needs *a priori* theoretical judgement and reasons.

Our analysis is concentrated on two equation models with dichotomous responses. However, some of the analysis and the testing procedure can be easily extended to general multiple equations models with polychotomous responses.

Appendix: Description of the data from A Panel Study of Income Dynamics 1972-1973; sample size = 2606.

Variables	Description	mean	S.D.
1. QUIT	Quiting previous job, a dichotomous variable; 1 if quit previous job; 0 otherwise	.0986	.2982
2. SAT	Job satisfaction; a dichotomous variable; 1 if job is enjoyable, 0 not enjoyable	.9175	.2752
3. UNION	Labor union status; dichotomous variable; 1 if belongs to a labor union, 0 otherwise	.2698	.4439
4. LNWAGE	Average hourly earnings (dollars) in natural logarithmic scale.	1.1654	.6256
5. TENURE	Years on current (1972) job; a dichotomous variable; 1 if over 3.5 years, 0 if less than 3.5 years	.5491	.4977
6. EDUC1	Schooling completed; a dichotomous variable; 1 if has completed high school but no college degree, 0 otherwise	.4674	.4990
7. EDUC 2	Schooling completed; a dichotomous variable; 1 if college or advanced degree, 0 otherwise	.1186	.3233
8. AGE	Years of age	38.48	12.73
9. RACE	Race, a dichotomous variable, 1 for white, 0 otherwise	.6477	.4778
10. SEX	Sex; 1 for male, 0 for female	.8143	.3890
11. MARITAL	Marital Status; a dichotomous variable; 1 if married and spouse present, 0 otherwise	.7337	.4421
12. CHILDREN	Number of children in family unit - age 0-17	1.6404	1.8254
13. SHORT	Shortage or surplus of unskilled male labor in respondent's country, Aug. 1972; a dichotomous variable; 1 if shortage, 0 if surplus.	.2748	.4465
14. AWAGE	Area wage-typical wage unskilled male worker might receive, Aug. 1972, in respondent's country a coded continuous variable with five discrete values: 1.25, 1.75, 2.25, 2.75 and 3.25.	2.1911	.4796
15. UNRATE	Unemployment rate in respondent's country, Aug. 1972; a roundoff continuous variable with five discrete values: 1.5, 3, 5, 8, 11	5.5027	2.1303

References

Amemiya, T. (1974), "Bivariate Probit Analysis: Minimum Chi-Square Methods", Journal of the American Statistical Association, 69, 940-944.

_______, (1979), "The Estimation of a Simultaneous-Equation Tobit Model", International Economic Review, 20, no. 1, 169-181.

Ashford, J. R. and R. P. Sowden (1970), "Multivariate Probit Analysis", Biometrics 26, 535-546.

Freeman, R. B. (1978), "Job Satisfaction as an Economic Variable", The American Economic Review, Papers and Proceedings, 68, no.2, 135-141.

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

_______ (1972b), "A General Model for tha Analysis of Surveys", American Journal of Sociology 71, 1035-1086.

_______ (1973a), "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys", American Journal of Sociology 78, 1135-1191.

_______ (1973b), "The Analysis of Multidimensional Contingency Tables When Some Variables are Posterior to Others: A Modified Path Analysis Approach", Biometrika 60, 179-192.

Lee, L. F. (1978), "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables", International Economic Review 19, 415-433.

_______ (1979a), "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables", Econometrica 47, no. 4, 977-996.

_______ (1979b), "On Comparisons of Normal and Logistic Models in the Bivariate Dichotomous Analysis", forthcoming in Economics Letters.

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

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

Nerlove, M. and S. J. Press (1973), "Univariate and Multivariate Log-Linear and Logistic Models", R-1306-eda/nih, Rand Corporation, Santa Monica, California.

Nerlove, M. and S. J. Press (1976), "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data", Discussion Paper No. 1, Center for Statistics and Probability, Northwestern University.

Schmidt, P. and R. P. Strauss (1975), "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach", Econometrica 43, 745-755.

Survey Research Center (1972), A Penel Study of Income Dynamics: Study Design, Procedures, Available Data, Ann Arbor, Michigan.

Theil, H. (1971), Principles of Econometrics, New York: Wiley and Sons, Inc.