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

Lung-Fei Lee

1. INTRODUCTION

Econometricians 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 recursive model has been specified as

\[ \ln \frac{P(Y = 1)}{P(Y = 2)} = \alpha \]

* I appreciate having financial support from the National Science Foundation under Grant SOC-78-07304.
\[ \ln \frac{P(Z = 1 | Y)}{P(Z = 2 | Y)} = \beta + \gamma Y_1 \]  

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

\[ Y_1 = 1 \text{ iff } Y = 1 \]  
\[ Y_1 = 0 \text{ otherwise.} \]  

Similarly, we define \( Z_1 \) corresponding to \( Z \). The model forms a causal system of logistic equations as in Goodman [1973a, 1973b]. In this model, \( Y \) has a direct effect on the probability of occurrence 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:

\[ \ln \frac{P(Y = 1 | Z, X)}{P(Y = 2 | Z, X)} = \alpha + \gamma Z_1 \]  
\[ \ln \frac{P(Z = 1 | Y, X)}{P(Z = 2 | Y, X)} = \beta + \gamma Y_1 \]  

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

\[
\log p_{ijkl} = \mu + \theta_i + \theta_j + \theta_{ik} + \theta_{ij} + \theta_{ik} + \theta_{ijk} + \theta_{ijl} + \theta_{ijkl} \quad (2.1)
\]

where \( p_{ijkl} \) denotes the conditional joint probability of \( Y \) and \( Z \) at levels \( i \) and \( j \), respectively, conditional on \( C \) and \( D \) at levels \( k \) and \( \ell \); \( \theta_i \) denotes an effect due to \( Y \) (at level \( i \)); \( \theta_j \) denotes an effect due to \( Z \); \( \theta_{ik} \) denotes a second order interaction effect between \( Y \) and \( C \) (at levels \( i \) and \( k \) respectively); \( \theta_{ijl} \) denotes a 4th order interaction effect among all of the variables at levels \( i, j, k, \ell \). This notation originated in Goodman [1972b].
overall effect $\mu_{jk}$ is a function of all of the other effects. As
$\Sigma_j \Sigma_j P_{ijk} = 1$,

$$\mu_{jk} = -\ln \Sigma_j \exp \{ \theta_1 + \theta_j + ... + \theta_{ijk} \} \quad (2.2).$$

The probabilities $P_{ijk}$ have the logistic functional form. The parameters of the other effects are subject to the constraints as in ANOVA models, i.e.,

$$\Sigma_i \theta_{1i} = \Sigma_j \theta_{1j} = 0$$
$$\Sigma_i \theta_{ik} = \Sigma_k \theta_{ik} = ... = \Sigma_j \theta_{ij} = 0$$
$$\Sigma_i \theta_{ikl} = \Sigma_k \theta_{ikl} = ... = \Sigma_j \theta_{ijkl} = 0$$
$$\Sigma_i \theta_{ijk} = \Sigma_j \theta_{ijk} = ... = \Sigma_k \theta_{ijk} = 0$$

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].

$$\ln P_{ijk} = \delta_1 + \delta_{1k} + \delta_{1l} + \delta_{1lk} + \delta_{2j} + \delta_{2lj} + \delta_{2lk} + \delta_{2ljk}$$
$$\ln P_{ij} = \delta_1 + \delta_{1k} + \delta_{1l} + \delta_{1lk} + \delta_{2j} + \delta_{2lj} + \delta_{2lk} + \delta_{2ljk}$$

where $P_{ijk}$ denotes the conditional probability of $Y$ at level $i$ conditional on $Z, C, D$ at levels $j, k, l$ respectively,...; $\delta_1 = 2\theta_1$, $\delta_{1k} = 2\theta_{1k}, ...$, $\delta_{1lk} = 2\theta_{1lk}$. 
When the discrete variables are dichotomous, equations (2.1) or (2.4), (2.5) can be written in a more appealing form. Define

\[ Y^*_1 = 1 \text{ iff the level 1 of } Y \text{ occurs; } \]
\[ Y^*_2 = -1 \text{ iff the level 2 of } Y \text{ occurs.} \]  

Similarly, define \( Z^*_j, C^*_k \) and \( D^*_l \) for the variables \( Z, C, \) and \( D \) respectively. Equation (2.1) can be rewritten as

\[
\ln \frac{P_{1jkl}}{P_{2jkl}} = u_{kl} + \gamma_1 Y^*_1 + \gamma_2 Y^*_2 + \gamma_{13} Y^*_1 C^*_k + \gamma_{14} Y^*_1 D^*_l + \gamma_{23} Y^*_2 C^*_k + \gamma_{24} Y^*_2 D^*_l \]

\[
+ \gamma_{134} Y^*_1 C^*_k D^*_l + \gamma_{123} Y^*_2 C^*_k D^*_l + \gamma_{124} Y^*_1 D^*_l C^*_k \]
\[
+ \gamma_{1234} Y^*_1 C^*_k D^*_l C^*_k \]  

The conditional log odds equations become

\[
\ln \frac{P_{1jkl}}{P_{2jkl}} = \omega_1 + \omega_{13} C^*_k + \omega_{14} D^*_l + \omega_{134} C^*_k D^*_l \]
\[
+ \omega_{123} Y^*_1 C^*_k D^*_l + \omega_{124} Y^*_1 D^*_l C^*_k \]
\[
+ \omega_{1234} Y^*_1 C^*_k D^*_l C^*_k \]  

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.

**Lemma**: 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

\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = Ra
\tag{2.10}
\]

\[
\ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Qb + \lambda Y^*
\tag{2.11}
\]

are equivalent to

\[
\ln \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = Ra + \lambda Z^* + G(Qb, \lambda)
\tag{2.12}
\]

\[
\ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Qb + \lambda Y^*
\tag{2.13}
\]

where

\[
G(Qb, \lambda) = \lambda + \ln \left( \frac{1 + \exp(Qb - \lambda)}{1 + \exp(Qb + \lambda)} \right)
\tag{2.14}
\]
Proof: \((\Leftrightarrow):\) 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),
\[
P(Z | Y) = \frac{\exp \left[ (Q \beta + \lambda Y^*) I \right]}{1 + \exp (Q \beta + \lambda Y^*)} \tag{2.16}
\]
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),
\[
\ln \frac{P(Y = 1 | Z)}{P(Y = 2 | Z)} = \frac{\ln \left( \frac{1 + \exp (Q \beta - \lambda)}{1 + \exp (Q \beta + \lambda)} \right)}{1 + \exp (Q \beta + \lambda)} = \frac{\ln \left( \frac{1 + \exp (Q \beta - \lambda)}{1 + \exp (Q \beta + \lambda)} \right)}{1 + \exp (Q \beta + \lambda)} + G(Q \beta, \lambda)
\]
where \(G\) is the function defined in (2.14).
\((\Leftarrow):\) From (2.15),
\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = \ln \frac{P(Y = 1 | Z)}{P(Y = 2 | Z)} - \ln \frac{P(Z | Y = 1)}{P(Z | Y = 2)} \tag{2.18}
\]
Since (2.17) holds from (2.13), equations (2.17), (2.12) and (2.18) imply
\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = \frac{\ln \frac{1 + \exp (Q \beta - \lambda)}{1 + \exp (Q \beta + \lambda)}}{1 + \exp (Q \beta + \lambda)} = \frac{\ln \frac{1 + \exp (Q \beta - \lambda)}{1 + \exp (Q \beta + \lambda)}}{1 + \exp (Q \beta + \lambda)} + G(Q \beta, \lambda) = \frac{\ln \left( \frac{1 + \exp (Q \beta - \lambda)}{1 + \exp (Q \beta + \lambda)} \right)}{1 + \exp (Q \beta + \lambda)} + G(Q \beta, \lambda)
\]
\[
= R \alpha \quad Q.E.D.
\]
Corollary: 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

\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = R_0 \tag{2.19}
\]

\[
\ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q_\beta + \lambda Y_1 \tag{2.20}
\]

are equivalent to

\[
\ln \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = R_0 + \lambda Z_1 + G_1(Q_\beta, \lambda) \tag{2.21}
\]

\[
\ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = Q_\beta + \lambda Y_1 \tag{2.22}
\]

where

\[
G_1(Q_\beta, \lambda) = \ln \left[ \frac{1 + \exp(Q_\beta)}{1 + \exp(Q_\beta + \lambda)} \right] \tag{2.23}
\]
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 $Y$ and $Z$ be two underlying indexes which are related to the observed responses as $Y = 1$ iff $Y > 0$, and $Y = 2$ otherwise; similarly, $Z = 1$ iff $Z > 0$, and $Z = 2$ otherwise. The general recursive model to be considered is,

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

$$
\tilde{Z} = \beta + \beta_k^C + \beta_k^D + \lambda_{k1}^{YZ} - \varepsilon_2 \tag{3.2}
$$

where the disturbances $\varepsilon_1$, $\varepsilon_2$ have zero means; $\alpha$ is the constant term, $\alpha_k^C$ is the effect of $C$ at level $k$ on $\tilde{Y}$ (or the probability of $Y$), ..., $\beta_k^D$ is the effect of $D$ at level $k$ on $\tilde{Z}$ and $\lambda_{k1}^{YZ}$ is the direct effect of $Y$ at level $i$ in influencing $\tilde{Z}$. The effects satisfy the ANOVA constraints,

$$
\sum_{k=1}^{K} \alpha_k^C = \sum_{k=1}^{L} \alpha_k^D = 0
$$

$$
\sum_{k=1}^{K} \beta_k^C = \sum_{k=1}^{L} \beta_k^D = \sum_{i=1}^{L} \lambda_{k1}^{YZ} = 0
$$

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

$$
\sum_{k=1}^{K} \alpha_{k\ell}^{CD} = \sum_{\ell=1}^{L} \alpha_{k\ell}^{CD} = 0. \text{ For some cases, it is desirable to differentiate}
$$
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 \( Y \), the effects \( \alpha^D_l \) can be specified as \( \alpha^D_l = \alpha^D ( l - \frac{1}{L} \sum_{i=1}^{L} i ) \) where \( \alpha^D \) 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 \( \lambda_{11}^{YZ} = 0 \) and the disturbances \( \varepsilon_1 \) and \( \varepsilon_2 \) 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 \( \varepsilon_1 \) and \( \varepsilon_2 \) are logistically distributed i.e., with distributions

\[
\frac{1}{1 + e^{-\varepsilon_1}} \quad \text{and} \quad \frac{1}{1 + e^{-\varepsilon_2}}
\]

respectively, the two equations are equivalent to

\[
\ln \frac{\bar{Y}_{CD}}{P_{1kL}} = \alpha + \alpha_k + \alpha_{kL} \quad \text{(3.3)}
\]

\[
\ln \frac{\bar{Z}_{CD}}{P_{1kL}} = \beta + \beta_k + \beta_{kL} \quad \text{(3.4)}
\]

This independence model is equivalent to the log linear model

\[
\ln p_{ijkl} = \mu_{kl} + \theta^Y_{1l} + \theta^Y_{ik} + \theta^Y_{ikl} + \theta^Z_{j} + \theta^Z_{jk} + \theta^Z_{jkl} \quad \text{(3.5)}
\]

with all third order interaction effects removed. The \( \alpha \)'s and \( \beta \)'s are related to the \( \theta \)'s as

\[
\begin{align*}
\alpha &= 2\theta^Y_{1l}, \quad \alpha_k = 2\theta^Y_{1k}, \quad \alpha_{kL} = 2\theta^Y_{1kL} \\
\beta &= 2\theta^Z_{1l}, \quad \beta_k = 2\theta^Z_{1k}, \quad \beta_{kL} = 2\theta^Z_{1kL}
\end{align*}
\]
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 \( \phi_{1j} = \phi_{1jk} = \phi_{1jkl} = 0 \) for all
i, j, k and \( \lambda \).

When the disturbances \( \epsilon_1, \epsilon_2 \) are independent and are logistic
distributed, equations (3.1) and (3.2) imply a fully recursive logit
equation model,

\[
\ln \frac{p_{1kl}}{p_{2kl}} = \alpha + \alpha_k + \alpha_\lambda
\]

(3.7)

\[
\ln \frac{p_{1jk}}{p_{2jk}} = \ln \frac{p_{1jk}}{p_{2jk}} = \beta + \beta_k + \beta_\lambda + \lambda
\]

(3.8)

where \( \lambda = \frac{\lambda}{11} \).

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)

\[
\ln \frac{p_{1jk}}{p_{2jk}} = \alpha + \alpha_k + \alpha_\lambda + \lambda + G(k, \lambda)
\]

(3.9)

where

\[
G(k, \lambda) = \lambda + \ln \left( \frac{1 + \exp(\beta + \beta_k + \beta_\lambda - \lambda)}{1 + \exp(\beta + \beta_k + \beta_\lambda + \lambda)} \right)
\]

(3.10)

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

\[
\ln \frac{p_{1jk}}{p_{2jk}} = \zeta + \zeta_k + \zeta_\lambda + \zeta_{k\lambda} + \lambda_\zeta
\]

(3.11)
where the parameters satisfy the ANOVA constraints and are defined by equations (3.9) and (3.10). Let \( G(k, \cdot) = \frac{1}{L} \sum_{\ell=1}^{L} G(k, \ell) \), \( G(\cdot, \ell) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{\ell=1}^{L} G(k, \ell) \). Explicitly,

\[
\begin{align*}
\zeta &= \alpha + G(\cdot, \cdot) \\
\zeta_{k}^{C} &= \alpha_{k}^{C} + G(k, \cdot) - G(\cdot, \cdot) \\
\zeta_{\ell}^{D} &= \alpha_{\ell}^{D} + G(\cdot, \ell) - G(\cdot, \cdot) \\
\zeta_{k\ell}^{CD} &= \alpha_{k\ell}^{CD} + G(k, \ell) - G(\cdot, \ell) - G(k, \cdot) + G(\cdot, \cdot) \quad (3.12)
\end{align*}
\]

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 \( \zeta_{k}^{C} \) (or \( \zeta_{\ell}^{D} \)) 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.,

\[
\ln \frac{P_{1k\ell}^{C}}{P_{2k\ell}^{CDF}} = \alpha + \alpha_{k}^{C} + \alpha_{\ell}^{D} + \alpha_{k\ell}^{CD} \quad (3.7)'
\]
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.,

\[
\ln \frac{P_{1kk\ell}}{P_{2k\ell}} = \beta + \beta C + \beta D + \beta CD + \lambda Y_1
\]  

(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).

\[
\ln \frac{P_{1kk\ell}}{P_{2k\ell}} = \beta + \beta C + \beta D + \beta CD
\]  

(3.13)

\[
\ln \frac{P_{1jk\ell}}{P_{2jkl}} = \alpha + \alpha C + \alpha D + \alpha CD + \lambda Z_1
\]  

(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

\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = R\alpha
\]  

(3.15)

\[
\ln \frac{P(Z = 1 | Y)}{P(Z = 2 | Y)} = Q\beta + \lambda Y_1
\]  

(3.16)
The Nerlove-Press model is specified as

\[ \ln \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = \alpha \gamma + \lambda Z_1 \]  

(3.17)

\[ \ln \frac{P(Z = 1|Y)}{P(Z = 2|Y)} = \beta \gamma + \lambda Y_1 \]  

(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

\[ \ln \frac{P(Y = 1|Z)}{P(Y = 2|Z)} = \alpha \gamma + \lambda Z_1 + G_1(Q\beta, \lambda) \]  

(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,

\[
\ln \frac{P(Y = 1 | Z)}{P(Y = 2 | Z)} = Ra + \lambda Z_1 + \delta G_1(Q \beta, \lambda) \tag{4.1}
\]

\[
\ln \frac{P(Z = 1 | Y)}{P(Z = 2 | Y)} = Q \beta + \lambda Y_1 \tag{4.2}
\]

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 \( \delta \) in the generalized model if the parameter \( \delta \) is identifiable in the above model. The identification of \( \delta \) 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 \( \delta \) 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

$$\ln L_2(\theta_2) = \sum_{i=1}^{N} \left( Z_{1i}(Q_1 \beta + \lambda Y_{1i}) - \ln(1 + \exp(Q_1 \beta + \lambda Y_{1i})) \right)$$

where $\theta_2' = (\beta', \lambda)$. Let $X_{2i} = (Q_1, Y_{1i})$. The first and second derivatives of $\ln L_2(\theta_2)$ are

$$\frac{\partial \ln L_2(\theta_2)}{\partial \theta_2} = \sum_{i=1}^{N} \left( Z_{1i} - \frac{\exp(X_{2i} \theta_2)}{1 + \exp(X_{2i} \theta_2)} \right) X_{2i}'$$

and

$$\frac{\partial^2 \ln L_2(\theta_2)}{\partial \theta_2 \partial \theta_2'} = -\sum_{i=1}^{N} \frac{\exp(X_{2i} \theta_2)}{(1 + \exp(X_{2i} \theta_2))^2} X_{2i}' X_{2i}$$

Assume that the first two moments of $X_2$ exist, its variance matrix is nonsingular and the parameter space of $(\beta, \lambda)$ 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

$$\left[ -\frac{\partial^2 \ln L_2(\hat{\theta}_2)}{\partial \theta_2 \partial \theta_2'} \right]^{-1}.$$

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 maximizing the following function

$$\ln L_1(\theta_1, \theta_2) = \sum_{i=1}^{N} \left( Y_{1i}(R_1 \alpha + \lambda Z_{1i}) + \delta G_1(Q_1 \beta, \lambda) \right)$$

$$- \ln(1 + \exp(R_1 \alpha + \lambda Z_{1i} + \delta G_1(Q_1 \beta, \lambda)))$$
Denote $X_{1i} = (R_i, G_i(Q_i, \lambda), Z_{1i})$ and $\hat{X}_{1i} = (R_i, G_i(Q_i, \hat{\lambda}), Z_{1i})$.

The first and second derivatives of $\ln L_1(\theta_1, \theta_2)$ are

$$\frac{\partial \ln L_1(\theta_1, \theta_2)}{\partial \theta_1} = \sum_{i=1}^{N} \{y_{1i} - \frac{\exp(X_{1i}theta_1)}{1+\exp(X_{1i}theta_1)}\} \hat{x}_{1i}^t$$

and

$$\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_1'} = -\sum_{i=1}^{N} \frac{\exp(X_{1i}theta_1)}{(1+\exp(X_{1i}theta_1))^2} \hat{x}_{1i}^t \hat{x}_{1i}$$

In addition to the assumptions we have made, we assume also that the parameter space of $\theta_1$ 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

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

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

$$V(\hat{\theta}_1) = \frac{1}{N} \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_1'}\right]^{-1} \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_2 \partial \theta_2'}\right]^{-1} \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'}\right]^{-1}$$

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

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

$$+ \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'}\right] \left[\frac{\partial^2 \ln L_2(\theta_2)}{\partial \theta_2 \partial \theta_2'}\right]^{-1} \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'}\right]^{-1} \left[\frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_2 \partial \theta_2'}\right]^{-1}$$
The expression can be simplified with the relations
\[
+ \left( \frac{1}{N} \frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_1'} \right) \cdot \left( \frac{1}{N} \frac{\partial^2 \ln L_2(\theta_2)}{\partial \theta_2 \partial \theta_2'} \right)^{-1} \left[ \frac{1}{N} \frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1' \partial \theta_2} \right] \cdot \left[ \frac{1}{N} \frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_2' \partial \theta_1} \right]^{-1}
\]

The expression can be simplified with the relations
\[
\frac{1}{N} \frac{\partial \ln L_1(\theta_1, \theta_2)}{\partial \theta_1} \cdot \frac{\partial \ln L_2(\theta_2)}{\partial \theta_2'} = \frac{1}{N} \sum_{i=1}^{N} \left( y_{1i} - \frac{\exp(x_{1i} \theta_1)}{1 + \exp(x_{1i} \theta_2)} \right)
\]
\[
\cdot \left( z_{1i} - \frac{\exp(x_{2i} \theta_2)}{1 + \exp(x_{2i} \theta_2)} \right)
\]

and
\[
\frac{1}{N} \frac{\partial^2 \ln L_1(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2'} = -\frac{1}{N} \sum_{i=1}^{N} \frac{\exp(x_{1i} \theta_1 + Q_i \beta)}{(1 + \exp(x_{1i} \theta_1))^2 (1 + \exp(Q_i \beta + \lambda))(1 + \exp(Q_i \beta))}.
\]
\[
(1 - \exp(\lambda))Q_i, - \exp(\lambda)(1 + \exp(Q_i \beta))
\]

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,

\[
\ln \frac{P(Y = 1 | Z)}{P(Y = 2 | Z)} = \alpha + (S\theta)Z_1 \tag{5.1}
\]

\[
\ln \frac{P(Z = 1 | Y)}{P(Z = 2 | Y)} = \beta + (S\theta)Y_1 \tag{5.2}
\]

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,

\[
\ln \frac{P(Y = 1)}{P(Y = 2)} = \alpha \tag{5.3}
\]

\[
\ln \frac{P(Z = 1 | Y)}{P(Z = 2 | Y)} = \beta + (S\theta)Y_1 \tag{5.4}
\]

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

\[
\ln \frac{P(Y = 1 | Z)}{P(Y = 2 | Z)} = \alpha + (S\theta)Z_1 + \ln \left( \frac{1 + \exp(Q\beta)}{1 + \exp(Q\beta + S\theta)} \right) \tag{5.5}
\]

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., $\lambda_{YI} = 0$ in (3.2), the model (3.1) and (3.2) is a bivariate
probability model. If $\varepsilon_1$ and $\varepsilon_2$ are bivariate normal variates with
zero means and correlation coefficient $\rho$, 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 $\rho$. The bivariate normal model is

\[
\begin{align*}
\tilde{Y}_{ik} & = \Phi(a + \alpha_k + \alpha_l) \\
\tilde{Z}_{ik} & = \Phi(b + \beta_k + \beta_l) \\
\tilde{Y}_{ik} \tilde{Z}_{ik} & = F(a + \alpha_k + \beta_k + \alpha_l + \beta_l, \rho)
\end{align*}
\]

These probabilities can be rewritten as

\[
\begin{align*}
\tilde{Y}_{ij} \tilde{Z}_{ij} & = G(Y, Z, C, D) \\
& = f(-1)^{1+Y}(a + \alpha_k + \alpha_l) \quad f(-1)^{1+Z}(b + \beta_k + \beta_l) \\
& = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(-1)^{1+Y} e_1 \quad f(-1)^{1+Z} e_2 \rho d e_1 d e_2
\end{align*}
\]

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

To compare the bivariate normal model with the log linear model,
we consider
where the parameters $\zeta$'s and $\eta$'s are defined from the equation and satisfy the ANOVA constraints. The parameters $\zeta$ and $\eta$ 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 $\theta_{ij}$ between the dependent variables and all of their higher order interaction terms with the explanatory variables, $\theta_{Y}^{YZC}$, $\theta_{ijl}^{YZD}$ and $\theta_{ijkl}^{YZCD}$, 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 $\rho$ is constant; more generally, the model can be extended to allow the correlation coefficient $\rho$ to depend on the levels of $C$ and $D$ as $\rho_{1} + \rho_{k}^{C} + \rho_{l}^{D} + \rho_{kl}^{CD}$ with the parameters $\rho_{k}^{C}$, $\rho_{l}^{D}$ and $\rho_{kl}^{CD}$ 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,

\[ \frac{\bar{Y}_{CD}}{p_{1k\ell}} = \phi(a + a_k + a_{\ell} + a_{k\ell}) \]  

(5.11)

\[ \frac{\bar{Z}_{CD}}{p_{1k\ell}} = \phi(\beta + \beta_k - \beta_{\ell} + \beta_{k\ell}) \]  

(5.12)

\[ \frac{\bar{Y}_{ZCD}}{p_{11k\ell}} = F(a + a_k + a_{\ell} + a_{k\ell} + a_{CD} + \beta_k + \beta_{\ell} + \beta_{k\ell}, \rho + \rho_k + \rho_{\ell} + \rho_{k\ell}) \]  

(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,

\[ \frac{\bar{Y}_{CD}}{p_{1k\ell}} = \phi(a_{k\ell}) \]  

(5.14)

\[ \frac{\bar{Z}_{CD}}{p_{1k\ell}} = \phi(\beta + \beta_k - \beta_{\ell} + \beta_{k\ell}) \]  

(5.15)

\[ \frac{\bar{Y}_{ZCD}}{p_{11k\ell}} = F(a_{k\ell}, \beta + \beta_k - \beta_{\ell} + \beta_{k\ell}, \rho + \rho_k + \rho_{\ell} + \rho_{k\ell}) \]  

(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,

\[
\bar{Y}_{CD} = \phi(a_{k\ell}) \tag{5.17}
\]

\[
\ln \frac{\bar{Y}_{ZCD}}{\bar{Y}_{ZCD}} = \eta + \eta_C + \eta_D + \eta_{CD} + \eta_Y + \eta_{YC} + \eta_{YD} + \eta_{YCD} \tag{5.18}
\]

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 \( \bar{Y}_{ZCD} = \bar{Y}_{ZCD} \bar{Y}_{CD} \) and

\[
\bar{Z}_{CD} = \frac{\bar{Y}_{ZCD} \bar{Y}_{CD}}{\bar{Y}_{ZCD}} = \frac{\bar{Y}_{ZCD}}{\bar{Z}_{CD}} + \bar{Y}_{ZCD}
\]

\[
= \frac{\bar{Y}_{ZCD}}{\bar{Z}_{CD}} \bar{P}_{11k\ell} + \bar{Y}_{ZCD} (1 - \bar{P}_{1k\ell})
\]

it follows

\[
\begin{pmatrix}
\bar{Y}_{CD} \\
\bar{P}_{1k\ell}
\end{pmatrix}
= \begin{pmatrix}
\bar{Y}_{ZCD} \\
\bar{P}_{11k\ell}
\end{pmatrix}
\begin{pmatrix}
\bar{Y}_{ZCD} \\
\bar{P}_{21k\ell}
\end{pmatrix}
\]

\[
\begin{pmatrix}
\bar{Y}_{CD} \\
\bar{P}_{1k\ell}
\end{pmatrix}
= \begin{pmatrix}
\bar{Y}_{ZCD} \\
\bar{P}_{11k\ell} (1 - \bar{P}_{1k\ell})
\end{pmatrix}
\begin{pmatrix}
\bar{Z}_{CD} \\
\bar{P}_{1k\ell}
\end{pmatrix}
\]

As long as \( \bar{P}_{1k\ell} \) is fixed and is neither zero nor one, there is a one-to-one correspondence between the set of conditional probabilities \( \bar{Y}_{ZCD} \) and \( (\bar{Y}_{ZCD}, \bar{Z}_{CD}, \bar{P}_{1k\ell}) \) as in (5.19). Consider equation (5.18) which implies
Thus the effects $\eta$ define a unique set of conditional probabilities $P_{ikl}$. Conversely, the effects $\eta$ 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

$$a_{ikl} = \ln \frac{P_{ikl}}{P_{ikl}}$$

and

$$a_{ik.} = \frac{1}{L} \sum_{l=1}^{L} a_{ikl}, \quad a_{i..} = \frac{1}{KL} \sum_{k=1}^{K} a_{ikl}, \quad a_{..l} = \frac{1}{KL} \sum_{i=1}^{L} a_{ikl}, \quad a_{..l} = \frac{1}{KL} \sum_{i=1}^{L} a_{ikl}$$

The effects $\eta$ can be solved uniquely as

$$\eta = a_{..l}$$

$$\eta_{k} = a_{..l} - a_{..l}$$

$$\eta_{l} = a_{..l} - a_{..l}$$

$$\eta_{i} = a_{i..} - a_{i..}$$

$$\eta_{ik} = a_{ik..} - a_{ik..} - a_{..l} + a_{ik..}$$

$$\eta_{ikl} = a_{ik..} - a_{ik..} - a_{..l} - a_{ik..} + a_{ik..} + a_{..l} - a_{..l} - a_{ik..}$$
Therefore, there is a one-to-one correspondence between the $\eta$'s and the $p_{\text{YCD}_{\text{Pilk},Q}}$'s. Now consider equations (5.15) and (5.16) conditional on given $p_{\text{YCD}_{\text{Pilk},Q}}$ 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 $\rho$ on $(-1, 1)$. Hence equations (5.15) and (5.16) can be solved as

$$\beta + \beta_k^C + \beta_{\lambda}^D + \beta_{\text{CD}}^k = \Phi^{-1}(\text{p}_{\text{YCD}_{\text{Pilk},Q}})$$

$$\rho + \rho_k^C + \rho_{\lambda}^D + \rho_{\text{CD}}^k = I_{a_{k\ell}}^{-1}(\Phi^{-1}(\text{p}_{\text{YCD}_{\text{Pilk},Q}})(\text{p}_{\text{YCD}_{\text{Pilk},Q}}) - \text{p}_{\text{YCD}_{\text{Pilk},Q}} \text{p}_{\text{YCD}_{\text{Pilk},Q}})$$

The effects, $\beta$'s and $\rho$'s, can be solved for uniquely from the ANOVA formulae (5.20) and (5.21). Since these effects also uniquely define the probabilities $\text{p}_{\text{YCD}_{\text{Pilk},Q}}$ and $\text{p}_{\text{YCD}_{\text{Pilk},Q}}$ 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 reparameterized into logit form. For example, if $a_{k\ell} = \alpha + \alpha^c_k$ we have
\[
\ln \frac{-Y_{CD}}{p_{1k}} = \ln \frac{\phi_N (\alpha + \alpha^c_k)}{1 - \phi_N (\alpha + \alpha^c_k)}
\]

\[
= \zeta + \zeta^c_k
\]

(5.22)

where the parameters \( \zeta \) and \( \zeta^c_k \) 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 conventional 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

\[ G = \ln\left(\frac{1+\exp(Q_{\hat{\delta}})}{1+\exp(Q_{\hat{\delta}+\lambda})}\right) \]

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 \( H_1: \delta \neq 1 \), i.e., the full recursive logit model, can be rejected. Similarly, the null hypothesis \( H_0 \) of the simple hypothesis \( H_0: \delta = 0 \) vs. \( H_1: \delta = 1 \) cannot be rejected with one-side-test at conventional level of significance.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Quit Eq.</th>
<th>Satisfaction Eqs.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Constant</td>
<td>.0290</td>
<td>1.8015</td>
</tr>
<tr>
<td></td>
<td>(.5141)</td>
<td>(.3316)*</td>
</tr>
<tr>
<td>UNION</td>
<td>-.4592</td>
<td>-.6984</td>
</tr>
<tr>
<td></td>
<td>(.1818)*</td>
<td>(.1595)*</td>
</tr>
<tr>
<td>LNWAGE</td>
<td>-.1846</td>
<td>.2465</td>
</tr>
<tr>
<td></td>
<td>(.1291)</td>
<td>(.1366)</td>
</tr>
<tr>
<td>TENURE</td>
<td>-1.0435</td>
<td>.1668</td>
</tr>
<tr>
<td></td>
<td>(.1721)*</td>
<td>(.1675)</td>
</tr>
<tr>
<td>EDUC1</td>
<td>-.1037</td>
<td>-.0947</td>
</tr>
<tr>
<td></td>
<td>(.1597)</td>
<td>(.1669)</td>
</tr>
<tr>
<td>EDUC2</td>
<td>-.5668</td>
<td>-.0128</td>
</tr>
<tr>
<td></td>
<td>(.2835)*</td>
<td>(.2905)</td>
</tr>
<tr>
<td>AGE</td>
<td>-.0358</td>
<td>.0135</td>
</tr>
<tr>
<td></td>
<td>(.0071)*</td>
<td>(.0069)</td>
</tr>
<tr>
<td>RACE</td>
<td>.0956</td>
<td>.2046</td>
</tr>
<tr>
<td></td>
<td>(.1564)</td>
<td>(.1626)</td>
</tr>
<tr>
<td>SEX</td>
<td>.1299</td>
<td>-.0336</td>
</tr>
<tr>
<td></td>
<td>(.2599)</td>
<td>(.3066)</td>
</tr>
<tr>
<td>MARITAL</td>
<td>-.2039</td>
<td>-.1981</td>
</tr>
<tr>
<td></td>
<td>(.2301)</td>
<td>(.2692)</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>.0044</td>
<td>.0369</td>
</tr>
<tr>
<td></td>
<td>(.0455)</td>
<td>(.0444)</td>
</tr>
<tr>
<td>SHORT</td>
<td>.2350</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.1625)</td>
<td></td>
</tr>
<tr>
<td>AWAGE</td>
<td>.1014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.1502)</td>
<td></td>
</tr>
<tr>
<td>UNRATE</td>
<td>.0181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.0348)</td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>-.6396</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.2045)*</td>
<td></td>
</tr>
<tr>
<td>QUIT</td>
<td>-.6803</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.2028)*</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>-6.8438</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.0797)</td>
</tr>
</tbody>
</table>
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 \( G = \ln\left(\frac{1+\exp(X)}{1+\exp(X+\lambda)}\right) \) is a monotone function of \( x \) bounded by the constant functions 0 and \( -\lambda \). It is a strictly increasing function if \( \lambda \) is negative, and is strictly decreasing if \( \lambda \) is positive. Thus the range of the values of \( G \) is determined by the value of \( \lambda \). 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 "\(-2 \text{ log likelihood value}\)" is 1481.099. Its comparison with the "\(-2 \text{ log likelihood value}\)" 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.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. QUIT</td>
<td>Quitting previous job, a dichotomous variable; 1 if quit previous job; 0 otherwise</td>
<td>.0986</td>
<td>.2982</td>
</tr>
<tr>
<td>2. SAT</td>
<td>Job satisfaction; a dichotomous variable; 1 if job is enjoyable, 0 not enjoyable</td>
<td>.9175</td>
<td>.2752</td>
</tr>
<tr>
<td>3. UNION</td>
<td>Labor union status; dichotomous variable; 1 if belongs to a labor union, 0 otherwise</td>
<td>.2698</td>
<td>.4439</td>
</tr>
<tr>
<td>4. LNWAGE</td>
<td>Average hourly earnings (dollars) in natural logarithmic scale.</td>
<td>1.1654</td>
<td>.6256</td>
</tr>
<tr>
<td>5. TENURE</td>
<td>Years on current (1972) job; a dichotomous variable; 1 if over 3.5 years, 0 if less than 3.5 years</td>
<td>.5491</td>
<td>.4977</td>
</tr>
<tr>
<td>6. EDUCL</td>
<td>Schooling completed; a dichotomous variable; 1 if has completed high school but no college degree, 0 otherwise</td>
<td>.4674</td>
<td>.4990</td>
</tr>
<tr>
<td>7. EDUC2</td>
<td>Schooling completed; a dichotomous variable; 1 if college or advanced degree, 0 otherwise</td>
<td>.1186</td>
<td>.3233</td>
</tr>
<tr>
<td>8. AGE</td>
<td>Years of age</td>
<td>38.48</td>
<td>12.73</td>
</tr>
<tr>
<td>9. RACE</td>
<td>Race, a dichotomous variable, 1 for white, 0 otherwise</td>
<td>.6477</td>
<td>.4778</td>
</tr>
<tr>
<td>10. SEX</td>
<td>Sex; 1 for male, 0 for female</td>
<td>.8143</td>
<td>.3890</td>
</tr>
<tr>
<td>11. MARITAL</td>
<td>Marital Status; a dichotomous variable; 1 if married and spouse present, 0 otherwise</td>
<td>.7337</td>
<td>.4421</td>
</tr>
<tr>
<td>12. CHILDREN</td>
<td>Number of children in family unit - age 0-17</td>
<td>1.6404</td>
<td>1.8254</td>
</tr>
<tr>
<td>13. SHORT</td>
<td>Shortage or surplus of unskilled male labor in respondent's country, Aug. 1972; a dichotomous variable; 1 if shortage, 0 if surplus.</td>
<td>.2748</td>
<td>.4465</td>
</tr>
<tr>
<td>14. AWAGE</td>
<td>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.</td>
<td>2.1911</td>
<td>.4796</td>
</tr>
<tr>
<td>15. UNRATE</td>
<td>Unemployment rate in respondent's country, Aug. 1972; a roundoff continuous variable with five discrete values: 1.5, 3, 5, 8, 11</td>
<td>5.5027</td>
<td>2.1303</td>
</tr>
</tbody>
</table>
References


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


