A Structural Probit Model With Latent Variables

BENGT MUTHEN*

A model with dichotomous indicators of latent variables is developed. The latent variables are related to each other and to a set of exogenous variables in a system of structural relations. Identification and maximum likelihood estimation of the model are treated. A sociological application is presented in which a theoretical construct (an attitude) is related to a set of background variables. The construct is not measured directly, but is indicated by the answers to a pair of questionnaire statements.

KEY WORDS: Dichotomous response; Multiple indicators; Linear structural relations; Maximum likelihood estimation.

1. INTRODUCTION

Consider the situation of regression analysis with several dichotomous dependent variables. For this case, Ashford and Sowden (1970) proposed the multivariate probit model, and, using the multivariate logistic function, Nerlove and Press (1973) and Schmidt and Strauss (1975) treated polytomous variables and models similar to simultaneous-equation models for quantitative variables, that is, interval or ratio scaled variables (see Amemiya 1975 for a review).

In this article, a structural equation model involving latent variables underlying the dichotomous responses is developed. This allows for a general representation of the causal relations between the response and the exogenous variables.

2. THE MODEL

Let v be a dichotomous variable with the two possible values 0 and 1. The traditional model of probit analysis assumes

$$\Pr(v = 1 | x) = \Phi(\alpha + \boldsymbol{\theta}'\mathbf{x}) \ , \qquad (2.1)$$

where Φ denotes the standardized normal distribution function and $\mathbf{x}$ is a q-dimensional random vector (e.g., see Finney 1971).

A different model is

$$\Pr(v = 1 | \eta) = \Phi(\kappa + \lambda\eta) \ , \qquad (2.2)$$

where η is a latent quantitative variable such that

$$\eta = \boldsymbol{\gamma}'\mathbf{x} + \zeta \ , \qquad (2.3)$$

and ζ is a random disturbance. Here, v acts as an indicator of the value of η, in the sense that the conditional probability of a v response is a monotonically increasing or decreasing function of η. Attention is now focused on η and the structural relation (2.3). The specification may be more reasonable than (2.1) in problems such as attitude studies, in which the subject's response to a certain question or statement is designed to measure a hypothetical construct, the subject's attitude toward something. The primary concern is how the attitude, rather than the observed response, is related to other variables of interest.

In this situation, there are usually many response variables that can be used to measure the same construct, that is, we can have multiple indicators of the same latent variable. A similar situation occurs in medical studies, in which questionnaire information on symptoms is used to indicate the development of some disease. The same specification may also be relevant in econometric studies with characteristics such as the purchase or nonpurchase of some commodity, the work or leisure of a person, and so on. We may then think of the latent variable as a desire to consume, a desire to work, and so forth. In such applications, however, the existence of multiple indicators is less frequent, because often a single response variable may be the only relevant indicator. As will be shown in Section 2, the availability of multiple indicators for each latent variable is, in general, necessary for the identification of the model.

It is assumed that ζ in (2.3) is independent of $\mathbf{x}$ and normally distributed with mean zero and variance ψ. Because the scale for η is arbitrary, the model has an indeterminacy such that the parameter λ may be multiplied by a constant and η divided by the same constant without affecting $\Pr(v = 1 | \eta)$. We eliminate this indeterminacy by setting $\lambda = 1$ (also see the general formulation of the model given later).

As shown in the Appendix, an interesting property of the model is that $\Pr(v = 1 | \mathbf{x})$ can be expressed in terms of the normal distribution function as

$$\Pr(v = 1 | \mathbf{x}) = \Phi(\delta\kappa + \delta\boldsymbol{\lambda}'\mathbf{x}) \ , \qquad (2.4)$$

where

$$\delta = (\psi + 1)^{-\frac{1}{2}} \ .$$

Thus the present model has the same functional form as the traditional probit model of (2.1), and when $\psi = 0$ they are equivalent. With $\psi = 0$, the disturbance ζ, representing the effect of omitted exogenous variables, is excluded from the model. In most social science applications, however, this disturbance is needed; this motivated Amemiya and Nold (1975) to attempt a modification of the logit model. With one response

* Bengt Muthén is Research Associate, Department of Statistics, University of Uppsala, P.O. Box 513, S-751 20 Uppsala, Sweden. Research was supported by the Bank of Sweden Tercentenary Foundation under project "Structural Equation Models in the Social Sciences," project director Karl G. Jöreskog.

© Journal of the American Statistical Association
December 1979, Volume 74, Number 368
Theory and Methods Section

variable, a disturbance of this type is not identified because it is only possible to identify $\alpha = \delta\kappa$ and $\boldsymbol{\theta} = \delta\boldsymbol{\gamma}$. As shown in Section 3, however, two or more indicators of the latent variable make it possible to identify the disturbance.

The simple model shown may be generalized to the case of p indicators and m structural relations. Let $\mathbf{v}$ be a p-dimensional vector of dichotomous variables, and let $\boldsymbol{\eta}$ be an m-dimensional vector of latent variables. The following measurement specification is the same as in the factor analysis model for dichotomous variables (e.g., see Bock and Lieberman 1970 and Muthén 1978), and, in the case of $m = 1$, it is the same as the classical latent trait model (e.g., see Lord and Novick 1968).

The relation between $\mathbf{v}$ and $\boldsymbol{\eta}$ is described by the parameters of a vector $\boldsymbol{\kappa}(p \times 1)$ and a matrix $\boldsymbol{\Lambda}(p \times m)$. Let $\boldsymbol{\lambda}'_i$ be the ith row of $\boldsymbol{\Lambda}$. Then, (2.2) is generalized to

$$\Pr(v_i = 1 \mid \boldsymbol{\eta}) = \Phi(\kappa_i + \boldsymbol{\lambda}'_i\boldsymbol{\eta}) \;, \quad i = 1, 2, \ldots, p \;. \tag{2.5}$$

Usually different (possibly overlapping) groups of v's are chosen to measure different η's, so that certain elements of $\boldsymbol{\Lambda}$ are zero a priori. As before, the scale of the η's may be determined by setting one element equal to 1 in each column of $\boldsymbol{\Lambda}$. The latent variables are assumed to account for all the interdependencies among the indicators, so that conditional on $\boldsymbol{\eta}$, the v's are independent (e.g., see Anderson 1959). Thus,

$$f(\mathbf{v} \mid \boldsymbol{\eta}) = f_1(v_1 \mid \boldsymbol{\eta}) f_2(v_2 \mid \boldsymbol{\eta}) \ldots f_p(v_p \mid \boldsymbol{\eta}) \;, \tag{2.6}$$

where $f, f_1, f_2, \ldots, f_p$ are conditional probability distributions.

We wish to study a linear structural equation system

$$\mathbf{B}\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \;, \tag{2.7}$$

where $\mathbf{B}(m \times m)$ is nonsingular, $\mathbf{B}$ and $\boldsymbol{\Gamma}(m \times q)$ are parameter matrices of structural coefficients, $\boldsymbol{\eta}(m \times 1)$ is the vector of latent variables, $\mathbf{x}(q \times 1)$ is a vector of observed fixed or random variables, and $\boldsymbol{\zeta}(m \times 1)$ is a vector of disturbances that is independent of $\mathbf{x}$ and has a multivariate normal distribution, with mean vector zero and covariance matrix $\boldsymbol{\Psi}$. Both $\mathbf{B}$ and $\boldsymbol{\Gamma}$ may contain some fixed elements (zeroes and ones) specified a priori. The structural equation system is of the same form as studied in econometrics, but the difference is that $\boldsymbol{\eta}$ is not directly observed. Structural relations involving such latent variables are often useful in the social sciences. Currently, such models are used in the case in which $\boldsymbol{\eta}$ has quantitative indicators, as in the so called LISREL model (see Jöreskog 1973, 1977).

The model is now completely specified, and we can deduce the probability distribution of $\mathbf{v}$ for given $\mathbf{x}$. To describe this, we let $F(\mathbf{a}, \mathbf{C})$ be the p-variate normal distribution function with argument $\mathbf{a}$, mean vector zero, and covariance matrix $\mathbf{C}$. We note that for a diagonal matrix $\mathbf{D}(p \times p)$, with positive diagonal elements,

$F(\mathbf{Da}, \mathbf{DCD}) = F(\mathbf{a}, \mathbf{C})$. When $p = 1$, $F(a, 1) = \Phi(a)$.

The probabilities of different $\mathbf{v}$ observations conditional on $\mathbf{x}$ may all be expressed as functions of the different marginal probabilities for observing one or several v's equal to 1, conditional on $\mathbf{x}$. If we use the result in the Appendix, these marginal probabilities are obtained as the marginal counterparts of the following p-variate normal distribution function

$$\Pr(v_1 = 1, \ldots, v_p = 1 \mid \mathbf{x}) = F(\boldsymbol{\kappa} + \boldsymbol{\Lambda}\mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{x}, \boldsymbol{\Lambda}\mathbf{B}^{-1}\boldsymbol{\Psi}\mathbf{B}^{-1\prime}\boldsymbol{\Lambda}' + \mathbf{I}) \;. \tag{2.8}$$

It is interesting to relate this model to the multivariate probit model of Ashford and Sowden (1970). In the latter, we have

$$\Pr(v_1 = 1, \ldots, v_p = 1 \mid \mathbf{x}) = F(\boldsymbol{\alpha} + \boldsymbol{\Theta}\mathbf{x}, \boldsymbol{\Xi}) \;, \tag{2.9}$$

where $\boldsymbol{\alpha}(p \times 1)$, $\boldsymbol{\Theta}(p \times q)$, and $\boldsymbol{\Xi}(p \times p)$ are parameter matrices, with $\operatorname{diag}(\boldsymbol{\Xi}) = \mathbf{I}$. The probabilities for other $\mathbf{v}$ observations are obtained in the same way as previously described.

Put

$$\boldsymbol{\Pi} = \mathbf{B}^{-1}\boldsymbol{\Gamma} \;, \tag{2.10}$$

$$\boldsymbol{\Omega} = \mathbf{B}^{-1}\boldsymbol{\Psi}\mathbf{B}^{-1\prime} \;, \tag{2.11}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Omega}\boldsymbol{\Lambda}' + \mathbf{I} \;, \tag{2.12}$$

$$\boldsymbol{\Delta} = \operatorname{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}} \;. \tag{2.13}$$

A comparison between (2.8) and (2.9) yields

$$\boldsymbol{\alpha} = \boldsymbol{\Delta}\boldsymbol{\kappa} \;, \tag{2.14}$$

$$\boldsymbol{\Theta} = \boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Pi} \;, \tag{2.15}$$

$$\boldsymbol{\Xi} = \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta} \;. \tag{2.16}$$

Thus, the two models have the same functional form. Our model is, however, stronger and more restricted because it imposes the structure of (2.15) and (2.16) on $\boldsymbol{\Theta}$ and $\boldsymbol{\Xi}$, where the number of parameters in $\boldsymbol{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ is less than or equal to the number of parameters in $\boldsymbol{\Theta}$ and $\boldsymbol{\Xi}$.

3. IDENTIFICATION AND MAXIMUM LIKELIHOOD ESTIMATION

To study the identification of parameters in the general formulation of our model, we may consider

$$F(\boldsymbol{\alpha} + \boldsymbol{\Theta}\mathbf{x}, \boldsymbol{\Xi}) = F(\boldsymbol{\Delta}\boldsymbol{\kappa} + \boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Pi}\mathbf{x}, \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}) \;. \tag{3.1}$$

In the Ashford and Sowden (1970) model we can identify the $p(p-1)/2$ correlations of $\boldsymbol{\Xi}$, the p elements of $\boldsymbol{\alpha}$, and the $p \times q$ elements of $\boldsymbol{\Theta}$ (assuming that there are at least $q + 1$ different $\mathbf{x}$ values). We note that this is the maximum number of identifiable parameters under the present model and that $\boldsymbol{\kappa}$, $\boldsymbol{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ necessarily have to be identified in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Xi}$ to yield an identified model.

First, consider the case in which a certain latent variable η_t has a single indicator v_s, in the sense that the element λ_{st} of $\boldsymbol{\Lambda}$ is the only nonzero element in the sth row and the tth column of $\boldsymbol{\Lambda}$. Then, it is seen that multiplication of the sth row of $\boldsymbol{\kappa}$ and the tth row of $\boldsymbol{\Pi}$ by the same scalar can be compensated by a change in $\boldsymbol{\Sigma}$ through the elements of the tth row (and column) of $\boldsymbol{\Omega}$, leaving (3.1) unaltered. Because the imposed changes in $\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$ generally can be absorbed in $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$, such a model is not identified.

Identification rules are difficult to establish in the general case. By considering parts of the model separately, however, using results pertaining to factor-analysis models and simultaneous-equation models, we can obtain a useful set of sufficient conditions for the identification of the model.

Consider the correlation matrix of (3.1):

$$\boldsymbol{\Xi} = \boldsymbol{\Delta\Sigma\Delta} = \boldsymbol{\Delta}(\boldsymbol{\Lambda\Omega\Lambda}' + \mathbf{I})\boldsymbol{\Delta} \ . \tag{3.2}$$

We note that the scale factors of $\boldsymbol{\Delta}$ are functions of the elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$, but of no other parameters. Are $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ identified in terms of the correlations of $\boldsymbol{\Xi}$? This is the case if $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are identified in terms of the covariances of $\boldsymbol{\Sigma}$, that is, the off-diagonal elements of $\boldsymbol{\Lambda\Omega\Lambda}'$. These covariances have the well-known structure of a restricted factor-analysis model (e.g., see Lawley and Maxwell 1971), for which the identification status is known for common applications. Identification can be obtained through a suitable arrangement of at least $m - 1$ zeroes and one element set equal to 1 in each column of $\boldsymbol{\Lambda}$ (also see Jöreskog 1969 and Dunn 1973). The identification of the factor model thus implies that $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are identified.

Now, consider the argument of (3.1),

$$\boldsymbol{\alpha} + \boldsymbol{\Theta}\mathbf{x} = \boldsymbol{\Delta\kappa} + \boldsymbol{\Delta\Lambda\Pi}\mathbf{x} \ . \tag{3.3}$$

Given the identification of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$, $\boldsymbol{\Delta}$ is given, and we can identify $\boldsymbol{\kappa}$ and $\boldsymbol{\Pi}$ (where $\boldsymbol{\Pi} = (\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Delta}^{-1}\boldsymbol{\Theta}$, assuming that $\boldsymbol{\Lambda}$ has full-column rank). The remaining parameter matrices $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ are identified in terms of $\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$ under conditions that are well known in simultaneous-equation modeling (e.g., see Goldberger 1964, Ch. 7.)

The fact that this set of conditions is sufficient and not necessary for identification is illustrated by the case of $p = 2$. Here, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ obviously are not identified in terms of $\boldsymbol{\Xi}$, but for $m = 1$ the model is still identified due to the restrictions on $\boldsymbol{\Theta}$. Let ρ be the single correlation coefficient of $\boldsymbol{\Xi}$. The parameters are κ_1, κ_2, $\lambda_1 = 1$, λ_2, γ_1, γ_2, $\ldots$, γ_q, ψ. After some algebra, we find that ψ is identified as

$$\psi = \theta_{1j}\theta_{2j}^{-1}\rho/(1 - \theta_{1j}\theta_{2j}^{-1}\rho) \ , \tag{3.4}$$

the γ's as

$$\gamma_i = \theta_{1i}(1 - \theta_{1j}\theta_{2j}^{-1}\rho)^{-\frac{1}{2}} \ , \tag{3.5}$$

and λ_2 as

$$\lambda_2{}^2 = \theta_{1j}{}^{-2}\theta_{2j}{}^2(1 - \theta_{1j}\theta_{2j}{}^{-1}\rho)/(1 - \theta_{1j}{}^{-1}\theta_{2j}\rho) \ , \tag{3.6}$$

where i, $j = 1, 2, \ldots, q$, and the sign of λ_2 is determined by the notation of the two v_2 alternatives (to obtain real solutions, $1 - \theta_{1j}\theta_{2j}{}^{-2}\rho > 0$ and $1 - \theta_{1j}{}^{-1}\theta_{2j}\rho > 0$). This determines $\boldsymbol{\Delta}$ and thus κ_1, κ_2.

Now consider the maximum likelihood estimation (MLE) of the model. The data are a sample of N independent observations on the random vector $(\mathbf{v}', \mathbf{x}')$. Let D be the number of distinct $\mathbf{x}$ values observed in a given sample, and let R be the number of possible v responses ($R = 2^p$). Denote by Pr_{dr} and n_{dr} the probability and the observed frequency, respectively, of the rth $\mathbf{v}$ response, given the dth distinct $\mathbf{x}$ value. The Pr_{dr}'s are functions of the parameter in the way described in Section 2. Because the parameters of the $\mathbf{x}$ distribution are not restricted, maximizing the likelihood L of the sample is equivalent to minimizing

$$F = -\sum_{d=1}^{D}\sum_{r=1}^{R} n_{dr} \log \text{Pr}_{dr} \ , \tag{3.7}$$

with respect to $\boldsymbol{\kappa}$, $\boldsymbol{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$. It is clear that the numerical minimization of F involves heavy computations for large values of D or p. In the application of Section 3, $D = 76$ and $p = 2$ yield moderate computational work.

For the MLE's of the vector of parameters $\boldsymbol{\theta}$, say, large-sample standard errors in principle may be obtained from the inverse of the Fisher information matrix

$$E(\partial \log L/\partial\boldsymbol{\theta}\ \partial \log L/\partial\boldsymbol{\theta}') \ , \tag{3.8}$$

where $\partial \log L/\partial\boldsymbol{\theta} = -\partial F/\partial\boldsymbol{\theta}$. Let the subscript s denote the $\mathbf{v}$-response pattern for the ith $\mathbf{x}$ observation and put

$$\mathbf{A} = \sum_{i=1}^{N} \partial \log \text{Pr}_{is}/\partial\boldsymbol{\theta}\partial \log \text{Pr}_{is}/\partial\boldsymbol{\theta}' \tag{3.9}$$

$$= \sum_{d=1}^{D}\sum_{r=1}^{R} n_{dr}/\text{Pr}_{dr}{}^2\partial \text{ Pr}_{dr}/\partial\boldsymbol{\theta}\partial \text{ Pr}_{dr}/\partial\boldsymbol{\theta}' \ . \tag{3.10}$$

The matrix $\mathbf{A}$ is an asymptotically valid approximation of (3.8). Thus, $\mathbf{A}^{-1}$ evaluated at the minimum of F will be used as the estimate of the covariance matrix of the MLE's.

4. A SOCIOLOGICAL APPLICATION

As an example, a model with one structural relation and two dichotomous indicators of a single latent variable will now be estimated. The data and the basic idea of the model formulation have been kindly supplied by Otis Dudley Duncan, Department of Sociology, University of Arizona. The data are from a 1971 Detroit area study, and the responses of 659 married women with complete data regarding the following variables will be used.

The dichotomous responses come from the interview question:

> Here are some things that might be done by a boy or a girl. Suppose the person were about 13 years old. As I read each of these to you, I would like you to tell me if it should be done as a regular task by a boy, by a girl or both.
> a. Shoveling walks
> b. Washing the car
> c. Dusting furniture
> d. Making beds.

Maximum Likelihood Estimates

$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\psi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$
.238 (.078)[b]	.181 (.067)	−.094 (.046)	.757 (.270)	1.000[a] —	1.695 (.751)	−.289 (.271)	−.880 (.258)

[a] Fixed value.
[b] Standard errors in parentheses.

For this illustration, we will use statements b. (CAR) and d. (BEDS), corresponding to each of the two domains of the traditional sexual division of labor. The response was practically limited to the alternative "both" (coded as 1) and the alternative "boys" for CAR, "girls" for BEDS (each coded as 0). The observed proportions of "both" answers are .689, .653, and .543 for CAR, BEDS, CAR and BEDS, respectively.

The latent variable of interest is the attitude regarding "sex typing," or rather "the propensity to reject sex typing" (Duncan 1975). This propensity is assumed to be a quantitative variable, large positive values being connected with a "liberal" view. Let v_1 correspond to the CAR response and v_2 correspond to the BEDS response. Here, v_1 and v_2 are two different indicators of the sex-typing propensity, where, for each indicator, the probability of a "both" answer is supposed to increase for increasing value of the propensity, as in (2.5). The sex-typing propensity is linearly related to three quantitative x variables, x_1 and x_2 being the number of years of schooling completed by the respondent and the spouse, respectively (as reported by the respondent), and x_3 as the number of years married (which may well act as a proxy for age). The education variables are scored from 1 to 4, with the categories: elementary (0–8 years); high school, nongraduate (9–11); high school, graduate (12); and college (13 and more). Years married is scored 1 to 6, corresponding to the intervals 1–9, 10–19, 20–29, ..., 50–59.

It is interesting to note that the model in this example is analogous—and generalizes—to the situation of the "multiple indicators—multiple causes" model studied, for example, by Jöreskog and Goldberger (1975). The essential difference is that in our case quantitative indicators are not available, but only dichotomous ones.

The estimates and their standard errors are shown in the table. The "explained" variance in the sex-typing propensity, $V(\boldsymbol{\gamma}'\mathbf{x})/V(\eta)$, is estimated as 18 percent, emphasizing the role of the disturbance parameter, as discussed in Section 2. Muthén (1976) considers a more complex application of the general model to this set of data. Here, the response of the spouse is also included, with reciprocal interaction between the spouses' sex-typing propensities.

5. CONCLUSION

In this article, we have proposed a latent-variable model to deal with situations involving multivariate dichotomous response. The latent variables were used in a set of structural equations to model the causal relations underlying the response. Identification and MLE of the model were considered. The ML approach was found to involve too heavy computations in the general case.

APPENDIX

Denote by $F(\mathbf{a}, \mathbf{C})$ the distribution function with argument $\mathbf{a}$ for a multivariate normal vector with mean zero and covariance matrix $\mathbf{C}$, and denote by $\phi(\mathbf{z}; \mathbf{a}, \mathbf{C})$ the density of a multivariate normal distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{C}$. Consider the k-dimensional dichotomous vector $\mathbf{u}$ and the n-dimensional multivariate normal vector $\mathbf{y}$.

Theorem: If

$$\Pr(u_1 = 1, \ldots, u_k = 1 | \mathbf{y}) = F(\mathbf{a} + \mathbf{B}\mathbf{y}, \mathbf{C}) \ , \quad \text{(A.1)}$$

where $\mathbf{y}$ has mean $\mathbf{d}$ and covariance matrix $\mathbf{E}$, then

$$\Pr(u_1 = 1, \ldots, u_k = 1) = F(\mathbf{a} + \mathbf{B}\mathbf{d}, \mathbf{B}\mathbf{E}\mathbf{B}' + \mathbf{C}) \ . \quad \text{(A.2)}$$

Proof: Integrating (A.1) over $\mathbf{y}$, we find after a variable transformation and a change of the order of integration,

$$\Pr(u_1 = 1, \ldots, u_k = 1) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_k} f(\mathbf{z}) dz_1 \ldots dz_k \ , \quad \text{(A.3)}$$

where

$$f(\mathbf{z}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(\mathbf{y}; \mathbf{d}, \mathbf{E}) \phi(\mathbf{z}; -\mathbf{B}\mathbf{y}, \mathbf{C}) dy_1 \ldots dy_n . \quad \text{(A.4)}$$

Using standard results on normal distributions, we find that

$$f(\mathbf{z}) = \phi(\mathbf{z}; -\mathbf{B}\mathbf{d}, \mathbf{B}\mathbf{E}\mathbf{B}' + \mathbf{C}) \ . \quad \text{(A.5)}$$

Inserting (A.5) in (A.3) and (A.2) immediately follows.

In the text, the theorem is applied to the distribution of $\boldsymbol{\eta}$ conditional on $\mathbf{x}$, which is multivariate normal with mean $\mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{x}$, corresponding to $\mathbf{d}$, and covariance matrix $\mathbf{B}^{-1}\boldsymbol{\Psi}\mathbf{B}^{-1\prime}$, corresponding to $\mathbf{E}$. Furthermore, it is seen that $\boldsymbol{\kappa}$, $\boldsymbol{\Lambda}$, and $\mathbf{I}$ correspond to $\mathbf{a}$, $\mathbf{B}$, and $\mathbf{C}$, respectively.

[*Received November 1976. Revised August 1978.*]

REFERENCES

Amemiya, T. (1975), "Qualitative Response Models," *Annals of Economic and Social Measurement*, 4, 363–372.
———, and Nold, F. (1975), "A Modified Logit Model," *Review of Economics and Statistics*, 57, 255–257.

Anderson, T.W. (1959), "Some Scaling Models and Estimation Procedures in the Latent Class Model," in *Probability and Statistics*, The Harald Cramér Volume, ed. U. Grenander, New York: John Wiley & Sons, 9–38.

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

Bock, R.D., and Lieberman, M. (1970), "Fitting a Response Model for *n* Dichotomously Scored Items," *Psychometrika*, 35, 179–197.

Duncan, O. D. (1975), Personal communication.

Dunn, J.E. (1973), "A Note on a Sufficiency Condition for Uniqueness of a Restricted Factor Matrix," *Psychometrika*, 38, 141–143.

Finney, D.J. (1971), *Probit Analysis* (3rd ed.), Cambridge, England: Cambridge University Press.

Goldberger, A.S. (1964), *Econometric Theory*, New York: John Wiley & Sons.

Jöreskog, K.G. (1969), "A General Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psychometrika*, 34, 183–202.

——— (1973), "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, eds. A.S. Goldberger and O.D. Duncan, New York: Seminar Press, 85–112.

——— (1977), "Structural Equation Models in the Social Sciences: Specification, Estimation and Testing," in *Proceedings of the Symposium on Applications of Statistics*, ed. P.R. Krishnaiah, Amsterdam: North-Holland Publishing Co., 265–286.

———, and Goldberger, A.S. (1975), "Estimation of a Model With Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 79, 631–639.

Lawley, D.N., and Maxwell, A.E. (1971), *Factor Analysis As a Statistical Method*, London: Butterworths.

Lord, F., and Novick, H. (1968), *Statistical Theories of Mental Test Scores*, Reading, Mass.: Addison-Wesley.

Muthén, B. (1976), "Structural Equation Models With Dichotomous Dependent Variables: A Sociological Analysis Problem Formulated by O.D. Duncan," Research Report 76-19, Dept. of Statistics, University of Uppsala, Uppsala, Sweden.

——— (1978), "Contributions to Factor Analysis of Dichotomous Variables," *Psychometrika*, 43, 551–560.

Nerlove, M., and Press, S.J. (1973), "Univariate and Multivariate Log-linear and Logistic Models," Santa Monica: The Rand Corp., R: 1306-EDA/NIH.

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