A New Estimator for Multivariate Binary Data

Shengfei Fu
PhD Student
Department of Agricultural and Applied Economics
University of Georgia
306 Conner Hall, Athens, GA, 30605
shengfei@uga.edu

J. Scott Shonkwiler
Professor
Department of Agricultural and Applied Economics
University of Georgia
314C Conner Hall, Athens, GA, 30605
jss1@uga.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.

A New Estimator for Multivariate Binary Response Data

Abstract

There is growing interest concerning the analysis of correlated binary data in the study of consumer behavior. The Multivariate Probit model is widely regarded as the preferred estimator of correlated binary response variables. Unfortunately, exact maximum likelihood estimation of the Multivariate Probit requires the evaluation of an M^{th} order integral when there are M correlated binary responses. Simulation estimators are computationally demanding and results may be sensitive to the number of random draws. This study proposes a new estimator for multivariate binary response data. This study considers binary responses as being generated from a truncated multivariate discrete distribution. Specifically, the discrete normal probability mass function, which has support on all integers, is extended to a multivariate form. Truncating this point probability mass function below zero and above one results the multivariate binary discrete normal distribution. This distribution has a number of attractive properties. Monte Carlo simulation and empirical applications are performed to show the properties of this new estimator; comparisons are made to the traditional Multivariate Probit model. Because multivariate binary response modeling is frequently required in areas such as marketing, household behavior, crop selection, and conservation practices, among others, we believe that our findings are of interest to both econometricians and practitioners.

Key words: Multivariate binary response, discrete normal distribution, Multivariate Probit

JEL classification: B23, Q13, D1

1. Introduction

There is growing interest concerning the analysis of correlated binary data in the study of consumer behavior. The purchase of specific products and/or brands, the choice of shopping venues, and the selection of certain activities are amenable to binary response modeling.

The Multivariate Probit (MVP) model is widely regarded as the preferred estimator of correlated binary response variables. Unfortunately, exact maximum likelihood estimation of the Multivariate Probit requires the evaluation of an M^{th} order integral when there are M correlated binary responses. Thus considerable research has been undertaken to evaluate simulation estimators for this model. It appears that the Geweke-Hajivassiliou-Keane (GHK) smooth recursive simulator (Geweke 1989; Hajivassiliou and McFaddem 1998; and Keane 1994) dominates all the simulation methods proposed to date. Yet this estimator is computationally demanding in large systems and results may be sensitive to the number of random draws.

This study considers binary responses as being generated from a truncated multivariate discrete normal distribution. The discrete normal distribution – as defined by Kemp (1997) – has support on all integers. The discrete normal probability mass function is extended to a multivariate form. Doubly truncating this joint probability mass function below zero and above one results the multivariate binary discrete normal distribution for a system of binary response variables. Maximum likelihood estimation is straightforward because for M response variables, only 2^M support points need to be evaluated to obtain the normalizing factor for the multivariate binary normal probability mass function. We term this new estimator for multivariate binary response data as Multivariate Binary Discrete Normal (MVBDN) Estimator.

The multivariate binary discrete normal distribution has a number of attractive properties: it is a member of the quadratic exponential family; it nests a system of independent binary logits;

it does not require conditioning to eliminate nuisance parameters; and exact maximum likelihood estimation is feasible since the normalizing factor only requires the evaluation of 2^M support points for a system of M response variables. Monte Carlo simulation and empirical application are performed to show the properties of this new estimator and comparisons are made to the traditional Multivariate Probit model.

The remainder of this study is organized as follows: section 2 derives the new estimator from the Multivariate Binary Normal distribution and discusses its maximum likelihood estimation. Section 3 explains the simulation setup and compares MVBDN estimates to the traditional Multivariate Probit model. Section 4 reports empirical applications. Finally, section 5 concludes with a discussion.

2. Multivariate Binary Discrete Normal Estimator

2.1 The Discrete Normal Distribution

Kemp (1997, p.224) characterizes the probability mass function (pmf) of a discrete normal random variable Y with parameters (λ, q) as

(1) $$P(Y = y) = \frac{\lambda^y q^{y(y-1)/2}}{\sum_{x=-\infty}^{\infty} \lambda^y q^{y(y-1)/2}}$$, y =…, −2, −1, 0, 1, 2, …; $\lambda>0$ and $0< q <1$.

The discrete normal distribution has a number of attractive properties: 1) the discrete normal distribution is analogous to the normal distribution in that it is the only two-parameter discrete distribution on (−∞, ∞) for which the first two moment equations are the maximum-likelihood equations; 2) the distribution is unimodal like the normal distribution; and 3) the distribution is log-concave like the normal distribution (Kemp 1997, p.225).

Let $\lambda = \exp((\mu - 0.5)/\sigma^2)$ and $q = \exp(-1/\sigma^2)$ such that $-\infty< \mu <\infty$ and $\sigma^2 >0$. The discrete normal may now be represented as

(2) $$P(Y=y)=\frac{\exp(-0.5(y-\mu)^2/\sigma^2)}{\sum_Y \exp(-0.5(Y-\mu)^2/\sigma^2)}$$

where y = … , −2, −1, 0, 1, 2, …; $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

The discrete normal can handle binary data by doubly truncating outcomes below zero and above one. Since location and not scale is of interest we set $\sigma^2=1$. Then P(Y=0)=1−P(Y=1) and

(3) $$P(Y=1)=\frac{\exp(-0.5(1-\mu)^2\ )}{\exp(-0.5(0-\mu)^2)+\exp(-0.5(1-\mu)^2)}=\frac{\exp(\mu-0.5)}{1+\exp(\mu-0.5)}.$$

It is obvious that this model is indistinguishable from the conventional binary logit model under the parameterizations $\mu=X\beta$ when X contains a constant. This feature indicates that the binary discrete normal distribution may be more applicable to data with thicker tails than the normal distribution.

The univariate discrete normal distribution can be generalized to the multivariate case of M integer responses using the following representation for a single observation:

(4) $$(Y_1=y_1, Y_2=y_2, \dots, Y_M=y_M)=\frac{\exp(-0.5(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\prime)}{\sum_{Y_1,\dots,Y_M} \exp(-0.5(\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\prime)}$$

y_m = …, −2, −1, 0, 1, 2, … ∀ m=1, 2, … M.

where **Σ** is assumed to be a positive definite M x M symmetric matrix and **y** and **μ** are 1 x M vectors. The summation term in the denominator represents all points of support of the distribution.

2.2 Multivariate Binary Discrete Normal Distribution

For multivariate binary responses, the random variables are doubly truncated below zero and above one so that the support becomes y_m = 0, 1 ∀ m=1, 2, … , M. The diagonal elements of **Σ** are constrained to unity for identification, and under independence ($\sigma_{ij}=0$; $i \neq j$), it clearly nests independent binary logit models. We term this model the multivariate binary discrete normal (MVBDN) estimator and its probability mass function is

(5) $$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_M = y_M) = \frac{\exp(-0.5(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\prime)}{\sum_{Y_1,\dots,Y_M} \exp(-0.5(\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\prime)}$$

$y_m = 0, 1 \ \forall \ m=1, 2, \dots M$.

Note that because of the truncation, the number of points of support needed to calculate the normalizing factor in the denominator amounts to just 2^M. Thus exact maximum likelihood estimation of multivariate binary response models is quite feasible on systems with M as large as 20.

To show that the discrete normal distribution is a member of the exponential family recall the pmf in terms of parameters μ and σ^2 as expressed in Equation (2). It can be rewritten as

(6) $$P(Y = y) = \frac{\exp(-\mu^2/2\sigma^2)\exp(y\mu/\sigma^2 - y^2/2\sigma^2)}{\exp(-\mu^2/2\sigma^2)\sum_Y \exp(Y\mu/\sigma^2 - Y^2/2\sigma^2)}.$$

Then the canonical representation of the pmf is

(7) $$P(Y = y) = \frac{\exp(\theta_1 y + \theta_2 y^2)}{\sum_Y \exp(\theta_1 Y + \theta_2 Y^2)}, \text{ where } \theta_1 = \mu/\sigma^2 \text{ and } \theta_2 = -1/2\sigma^2.$$

This representation has the form $a(\boldsymbol{\theta})\exp(\boldsymbol{\theta}'T(\mathbf{y}))$ where $a(\boldsymbol{\theta})$ is the normalizing factor and $T(\mathbf{y}) = [y|y^2]'$, then the discrete normal distribution is a member of the exponential family. Consequently if we define $\kappa(\mathbf{t}) = \ln(\exp(tY)a(\boldsymbol{\theta})^{-1})$, then the derivatives of $\kappa(\mathbf{t})$ with respect to t evaluated at t=0 are the cumulants of y. This is particularly important because a truncated distribution from the exponential family merely has the domain of Y restricted to a subspace and remains a member of this family (Lindsey 1996, p.37).

By analogy to the continuous multivariate normal distribution, the joint discrete normal pmf has, upon canceling the common term $\exp(-1/2\mu\Sigma^{-1}\mu')$, the canonical form as below:

(8) $$P(Y = y) = \exp\left[\sum_{i=1}^{k} c_i(\Theta) T_i(y) - \ln a(\Theta)\right]$$

where Θ is a k=m+m(m+1)/2 parameter vector ; $c(\Theta)'=\{\mu'\Sigma^{-1}, -1/2\,\Sigma^{11}, -\Sigma^{1m}, \ldots, -1/2\Sigma^{22}, -\Sigma^{23}, \ldots, -1/2\Sigma^{mm}\}$ where $\Sigma^{ij}=[\Sigma^{-1}]_{ij}$; $T(\mathbf{y})'==\{\mathbf{y}', y_1^2, y_1y_2, \ldots, y_1y_m, y_2^2, y_{23}, \ldots, y_m^2\}$; and $a(\Theta) = \sum_{Y_m} \ldots \sum_{y_1} \exp(c(\Theta)'T(Y))$ is a finite, real-valued function which does not depend on y. Therefore this joint pmf is a member of the exponential family.

The MVBDN model is a special case of the quadratic exponential models developed by Prentice and Zhao (1990) and Fitzmaurice and Laird (1993) for the analysis of multivariate binary observations. As such, the maximum likelihood estimator can possess attractive properties even under distributional misspecification (Gourieroux, Monfort, and Trognon, 1984). Worthy of note is that as a result of double truncation, the variance of the MVBDN model is not identified because y_m^2 in $T(\mathbf{y})$ reduces to be y_m, given the only possible values for y_m are 0 and 1. This is why we constrain the diagonal elements of $\boldsymbol{\Sigma}$ in Equation (5) to unity for identification.

2.3 Maximum Likelihood Estimation of MVBDN Estimator

It can be shown that for the multivariate binary discrete normal distribution the first two moment equations are the maximum likelihood equations when we have identical regressors across equations.

The multivariate binary discrete normal joint probability mass function can be rewritten as

(9) $$P(\boldsymbol{Y}=\boldsymbol{y}) = \frac{\exp\left(0.5\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}' + 0.5\boldsymbol{\mu}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}' - 0.5\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}'\right)}{\sum_{\boldsymbol{Y}}\exp\left(0.5\boldsymbol{Y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}' + 0.5\boldsymbol{\mu}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\right)}.$$

The log likelihood for the i^{th} observation (suppressing subscripts) is as follows:

(10) $$l_{(i)} = 0.5\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}' + 0.5\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}' - 0.5\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}' - \ln\left(\sum_{\boldsymbol{Y}}\exp(0.5\boldsymbol{Y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}' + 0.5\boldsymbol{\mu}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')\right).$$

Let's first illustrate through a simple case where $\boldsymbol{\mu}$ is a vector of constants. Solving for the maximum likelihood estimators:

(11) $$\frac{\partial l_{(i)}}{\partial \boldsymbol{\mu}} = \boldsymbol{y}\boldsymbol{\Sigma}^{-1} - \frac{\sum_Y \boldsymbol{Y}\boldsymbol{\Sigma}^{-1} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}$$

(12) $$\sum_i \frac{\partial l_{(i)}}{\partial \boldsymbol{\mu}} = \sum_i \boldsymbol{y_i}\boldsymbol{\Sigma}^{-1} - \frac{n\sum_Y \boldsymbol{Y}\boldsymbol{\Sigma}^{-1} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')} = 0$$

The first-order condition for the estimator of $\boldsymbol{\mu}$ is:

(13) $$\overline{\boldsymbol{y}} = \frac{\sum_Y \boldsymbol{Y} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}$$

which is the definition of the mean of the vector $\mathbf{y}$ over the sample. Note that is $\boldsymbol{\mu}$ not the estimator of $\overline{\boldsymbol{y}}$ even when $\boldsymbol{\Sigma}$ is an identity matrix. This result stems from the representation in Equation (3).

Solving for the ML estimator of $\boldsymbol{\Sigma}$, we get:

(14) $$\frac{\partial l_{(i)}}{\partial \boldsymbol{\Sigma}} = -0.5\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1} + 0.5\boldsymbol{\Sigma}^{-1}\boldsymbol{y}'\boldsymbol{y}\boldsymbol{\Sigma}^{-1}$$

$$-\frac{\sum_Y(-0.5\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}'\boldsymbol{Y}+\boldsymbol{Y}'\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}+0.5\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}) \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}$$, and

(15) $$\sum_i \frac{\partial l_{(i)}}{\partial \boldsymbol{\Sigma}} = 0.5\boldsymbol{\Sigma}^{-1}\left[\left(\sum_i \boldsymbol{y}_i'\boldsymbol{y_i} - n\frac{\sum_Y \boldsymbol{Y}'\boldsymbol{Y} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}\right)\right.$$

$$\left.-\boldsymbol{\mu}'\left(\sum_i \boldsymbol{y_i} - n\frac{\sum_Y \boldsymbol{Y} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}\right) - \left(\sum_i \boldsymbol{y_i} - n\frac{\sum_Y \boldsymbol{Y} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}\right)'\boldsymbol{\mu}\right]\boldsymbol{\Sigma}^{-1} = 0.$$

Upon simplification and using the previous first order condition in Eq. (9) and defining $\boldsymbol{S_{yy}} = \sum_i \boldsymbol{y}_i'\boldsymbol{y_i}/n$, we have that the ML estimator of $\boldsymbol{\Sigma}$ satisfies

(16) $$\boldsymbol{S} = \frac{\sum_Y \boldsymbol{Y}'\boldsymbol{Y} \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_Y \exp(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')} \equiv \boldsymbol{S_{yy}}$$

which is the definition of the un-centered sample variance-covariance matrix $\mathbf{S_{yy}}$.

With Equations (13) and (16), it is obvious that the MLE estimate of the correlation matrix from MVBDN model equals sample correlation matrix, since we have $\boldsymbol{S} - \bar{y}'\bar{y} = E(y'y) - (Ey)'Ey$.

Let's generalize by defining $\boldsymbol{\mu}_i=(\mathbf{X}_i\boldsymbol{\beta})'$, $\mathbf{X}_i$=diagonal($\mathbf{x}_{1i}$, $\mathbf{x}_{2i}$, …, $\mathbf{x}_{Mi}$), and $\boldsymbol{\beta}$=[$\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, …, $\boldsymbol{\beta}_M$]´, where $\mathbf{x}_{mi}$ ∀m=1, 2, …, M is a 1 x k_m vector, containing a constant term and (k_m −1) explanatory variables for the m^{th} response. $\mathbf{X}_i$ is a (M x $\sum_{m=1}^{M} k_m$) matrix and $\boldsymbol{\beta}$ is a $\sum_{m=1}^{M} k_m$ elements column vector. Then the first-order conditions with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are as follows, respectively:

(17) $\frac{1}{n}\sum_i(\boldsymbol{y}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i) = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_i E(\boldsymbol{y}_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i$

(18) $\boldsymbol{S} = \frac{1}{n}\sum_i \frac{\sum_{\boldsymbol{Y}} \boldsymbol{Y}'\boldsymbol{Y} \exp(\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_{\boldsymbol{Y}} \exp(\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')} + \boldsymbol{A} + \boldsymbol{A}' = \boldsymbol{S}_{yy} + \boldsymbol{A} + \boldsymbol{A}'$

Where $\boldsymbol{A} = \frac{1}{n}\sum_i \boldsymbol{\mu}_i'(\boldsymbol{y}_i - \boldsymbol{E}(\boldsymbol{y}_i))$ and $E(\boldsymbol{y}_i) = \frac{\sum_{\boldsymbol{Y}} \boldsymbol{Y} \exp(\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}{\sum_{\boldsymbol{Y}} \exp(\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}' - 0.5\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}')}$.

Equation (17) is essentially a collection of $\sum_{m=1}^{M} k_m$ equations. Let $\boldsymbol{x}_{mi}^*$ be a 1 x **(**k_m-1) vector containing the (k_m −1) explanatory variables for the m^{th} response. Then we can re-organize $\mathbf{X}_i$ such that $\widetilde{\boldsymbol{X}_i} = \begin{bmatrix} \boldsymbol{I}_M & \\ & \boldsymbol{X}_i^* \end{bmatrix}$ where $\boldsymbol{X}_i^* = diag(\boldsymbol{x}_{1i}^*, \dots, \boldsymbol{x}_{Mi}^*)$. The new-ordered first M equations then reduce to the first moment condition as in Equation (13). The remaining ($\sum_{m=1}^{M} k_m - M$) equations provide additional information that we will explore later:

(19) $\frac{1}{n}\sum_i(\boldsymbol{y}_i - \boldsymbol{E}(\boldsymbol{y}_i))\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i^*) = \mathbf{0}$

Note in Equation (18) the first part of the right-hand side (RHS) is the un-centered sample variance-covariance. The two additional components, **A** and **A'**, form a symmetric M x M matrix, which becomes zero when X is identical across equations. This is easily seen when $\boldsymbol{\mu}$ is a vector of constants. A proof is given in Appendix that this is also the case when X is identical across the M response variables.

2.4 Marginal Effects

In many economic studies, we are interested in the marginal effect of an explanatory variable. Using the definition of the mean as given in Equation (10), we get the derivatives of E(**y**) with respect to the **μ** is an M x M matrix:

$$(20)\qquad \frac{\partial \boldsymbol{E}(\boldsymbol{y})}{\partial \boldsymbol{\mu}} = \begin{bmatrix} \frac{\partial E(y_1)}{\partial \mu_1} & \frac{\partial E(y_1)}{\partial \mu_2} & \cdots & \frac{\partial E(y_1)}{\partial \mu_M} \\ \frac{\partial E(y_2)}{\partial \mu_1} & \frac{\partial E(y_2)}{\partial \mu_2} & \cdots & \frac{\partial E(y_2)}{\partial \mu_M} \\ \vdots & \vdots & & \vdots \\ \frac{\partial E(y_M)}{\partial \mu_1} & \frac{\partial E(y_M)}{\partial \mu_2} & \cdots & \frac{\partial E(y_M)}{\partial \mu_M} \end{bmatrix} = \frac{\sum_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\bar{\boldsymbol{y}})'\boldsymbol{Y}\exp\left((\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})'\right)}{\sum_{\boldsymbol{Y}} \exp\left((\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})'\right)}.$$

Let **μ**=(**Xβ**)´ as previously defined and $\boldsymbol{\beta}_{\mathbf{mk}}$ be the parameter of the kth regressor in **X** for the mth response. The derivative of the **μ** with respect to x_k is an M-element column vector:

$$(21)\qquad \frac{\partial \boldsymbol{\mu}}{\partial x_k} = \begin{bmatrix} \frac{\partial E(\mu_1)}{\partial x_k} & \frac{\partial E(\mu_2)}{\partial x_k} & \cdots & \frac{\partial E(\mu_M)}{\partial x_k} \end{bmatrix}' = [\mathbf{1}_{\mathbf{1}}(x_k)\beta_{1k} \quad \mathbf{1}_{\mathbf{2}}(x_k)\beta_{2k} \quad \ldots \quad \mathbf{1}_{\boldsymbol{M}}(x_k)\beta_{Mk}]'$$

where $\mathbf{1}_{\boldsymbol{m}}(x_k) = \begin{cases} 1 \; if \; x_k \in x_m \\ 0 \; otherwise \end{cases} \qquad \forall\, m = 1, 2, \ldots, M$.

The marginal effects of x_k on the M response variables are then obtained by multiplying Equations (20) and (21) according to the chain rule of derivatives.

Note the marginal effects involve all M sets of regressors if there are common variables. The marginal effects are the addition of direct effects on the response variable and indirect effects through the other (M-1) response variables. When event 1 indicates purchase, the marginal effect of x_k is interpreted as change in the probability of purchasing product m, corresponding to a one-unit change in x_k. We report average marginal effects in this study.

A straightforward way to derive marginal effects under the MVP model is to use the unconditional distributions (the univariate normal distributions). The unconditional mean functions are:

(22) $$E(y_m|x_m) = Prob\big((y_m = 1|x_m)\big) = \Phi(\boldsymbol{x_m\beta_m})\,; m = 1, 2, \dots, M.$$

And thus the marginal effects are straightforward:

(23) $$\frac{\partial y E(y_m|x_m)}{\partial x_k} = \phi(\boldsymbol{x_m\beta_m})\beta_{mk}\,; m = 1, 2, \dots, M.$$

As Mullahy (2011) points out, while such aggregation approaches may be informative for some purposes, it should be emphasized that they fail fundamentally to represent the properties of the underlying probability structure of the multivariate model. More appropriate marginal effects are based on joint conditional probabilities or probabilities conditional on subvectors of **y**. This complicates the computation of marginal effects because there is an ambiguity in the conditional distributions. Given the dimension of M response variables, there are 2^M probability outcomes based on the joint conditional probabilities. And the outcomes based on probabilities conditional on subvectors of **y** mount to $\sum_{k=1}^{M-1}\binom{M}{k}2^k$ (which is 4, 18, 64 for M=2, 3, 4, respectively). See Mullahy (2011) for a general analytical formula for such marginal effects. As a comparison, the marginal effects under MVBDN are much easier to computation, since an easy to implement formula of the expected mean functions are given by the definition. This adds another attractive property to the MVBDN model.

For simplicity, we only compute the marginal effects for the MVP model using the unconditional distributions in this paper. We point out that the marginal effect under MVBDN accounts for the underlying property of the multivariate model while unconditional MVP marginal effect does not. Therefore, these two sets of marginal effects are not directly comparable.

3. Monte Carlo Simulations

Monte Carlo simulations are performed to show the properties of the MVBDN estimator and provide comparisons to the traditional Multivariate Probit model. In the first set of simulations, we generate multivariate binary responses assuming that they are drawn from an underlying normal distribution. We will compare the average marginal and conditional probabilities from the multivariate binary discrete normal to those estimated by the Multivariate Probit. Thus we should be able to provide some guidelines as to the applicability of the multivariate binary discrete normal when the data are known to come from a multivariate normal distribution. Next we will carry out simulations where the binary responses come from an underlying multivariate t-distribution. The idea here is that the multivariate binary discrete normal may actually outperform the Multivariate Probit model in such circumstances.

3.1 Data Generating Process

Consider the M-equation multivariate binary model, where, for convenience, the individual observation index is omitted:

(24) $y_m^* = \boldsymbol{\beta'_m X_m} + \varepsilon_m, m = 1, \ldots, M;\;\; y_m = 1\; if\; y_m^*\; and\; 0\; otherwise.$

In the first set of simulations, ε_m, m=1, …, M are error terms distributed as multivariate normal (MVN), each with a mean of zero, and variance-covariance matrix **Σ**, where **Σ** has values of 1 on the leading diagonal and correlations $\sigma_{jk} = \sigma_{kj}$ as off-diagonal elements. A random draw from an MVN distribution can be obtained using the Cholesky decomposition of **Σ**, the lower triangular factor **A**, for which **AA**´= **Σ**, and a vector of univariate normal draws, z. Specifically, Z=**A**z is a random draw from the MVN distribution with mean vector zero and covariance matrix **Σ**.

In the second set of simulations, $\varepsilon_{m,}$, m=1, …, M are error terms distributed as multivariate t-distribution (MVT), each with mean of zero, and variance-covariance matrix **Σ** as defined previously. The degree of freedom is set to four such that the variance of the multivariate t-distribution is twice of the standard normal distribution.

Two scenarios – identical and different regressors for the M-equations – are examined. Therefore, there are four sets of simulations in total. We generate 400 observations for each simulation and repeat the process up to 300 times. Since the values of X are fixed through repetition, we use the empirical distributions of the parameter to provide standard errors for inference.

Identical Regressors

Regressors are identical across all four equations, including a constant term and an explanatory variable x, generated as two times a random uniform variable. The variance-covariance matrix to generate the correlated error terms is $\begin{bmatrix} 1 & 0.3 & -0.2 & 0.1 \\ 0.3 & 1 & 0.25 & 0.5 \\ -0.2 & 0.25 & 1 & 0.75 \\ 0.1 & 0.5 & 0.75 & 1 \end{bmatrix}$. And the values of beta are **β**=[[3.45 -1],[5 -1.3],[0.8 -0.5],[-2.8 1.4]]´.

Different Regressors

Following Cappellari and Jenkins (2003), we use $\mathbf{\Sigma}=\begin{bmatrix} 1 & 0.25 & 0.5 & 0.75 \\ 0.25 & 1 & 0.75 & 0.5 \\ 0.5 & 0.75 & 1 & 0.75 \\ 0.75 & 0.5 & 0.75 & 1 \end{bmatrix}$, x_1=uniform ()-0.5; x_2=uniform ()+1/3; x_3=2*uniform ()+0.5; x_4=0.5*uniform ()-1/3;

y1s=.5+4*x_1+u(:,1);

y2s=3+0.5*x_1-3*x_2+u(:,2);

y3s = 1 - 2*x_1 + .4*x_2 - .75*x_3 + u(:,3);

y4s = -3.5 + 1*x_1 - .3*x_2 + 3*x_3 - .4*x_4 + u(:,4);

y1 = (y1s>0);

y2 = (y2s>0);

y3 = (y3s>0);

y4 = (y4s>0);

Note for y4, we use -3.5 as the intercept instead of -6 in Cappellari and Jenkins (2003). Table 1 reports the mean and cross correlation among the generated response variables under above four simulation scenarios.

[Table 1 here]

3.2 Simulation Results

Parameter estimates and marginal effects, as well as model fit measured by log likelihood values are compared across Multivariate Binary Discrete Normal model and Multivariate Probit model (GHK simulation with 500 draws).

Tables 2 to 5 report parameter estimates for both models under each of the four simulation scenarios, respectively. A comparison across these two sets of parameter estimates shows that the pattern of signs and significance are similar between the two models, but by no means identical. The fits of the two models are almost identical – the 95% empirical confidence intervals of the difference in log-likelihood value between MVBDN and MVP are not significantly different from zero (statistics not reported). In addition, Maximum Likelihood Estimation of the MVP model sometimes runs into problem because the combination of data set and initial values leads to a non-positive-definite covariance matrix for the GHK simulation. The MVP model experienced a failure rate of 28 percent, while the MVBDN successfully ran for all 300 repetitions. The average time required to run MVP model was about five times of that to run the MVBDN model for our simulated data sets.

[Tables 2-5 here]

We now focus on average marginal effects and some conditional expectations. Table 6 presents average marginal effects for the case of identical regressors. Regardless whether the error terms are generated from MVN or MVT, the average marginal effects derived from unconditional distributions are very close in values under MVBDN and MVP. Although some of them are statistically significant at 5% level, it's arguable that economically they are not substantially different.

[Table 6 here]

Table 7 reports average marginal effects for the case of different regressors. Note because of the multivariate nature of the model, the marginal effects of x_2, x_3, and x_4 on y_1 are nonzero, even though they are not regressors for y_1. This is because there are indirect effects channeled through the correlation between y_1 and the other three response variables. In contrast, the unconditional marginal effects from MVP fail to represent the underlying properties of multivariate model. Marginal effects of x_4, the regressor that enters the fourth equation only and thus have direct effect only, are estimated to have the same sign under MVBDN and MVP. However, marginal effects of x_1, which enters all four equations, are very different under MVBDN (direct and indirect effects) versus MVP (direct effects only).

[Table 7 here]

3.3 Correlations among Response Variables

As shown in Table 8, when the regressors are identical across equations, the MVBDN estimates exactly match the observed correlation. This is because the first two moment equations are the maximum likelihood equations as proved in Section 2.3. When there are different regressors across equations, the first moment equation is still the maximum likelihood equation as long as there is a constant term. The second moment from the Multivariate Binary Discrete Normal distribution is slightly different from the sample moment (Equation 18). Therefore in simulation cases 3 and 4, MVBDN estimates of the correlation among response variables are slightly different from observed correlation. However, they are still much closer than the correlation among MVP residuals (correlation associated with the latent variables).

[Table 8 Here]

4. Applications to the Ketchup Brands Data

Using data provided by the James M. Kilts Center, University of Chicago Booth School of Business that was originally collected by the now-defunct ERIM division of A. C. Nielsen, we examine the ketchup purchasing behavior of 1651 households in Sioux Falls, S.D. Data are from full calendar year 1986. The five brands studied represent more than 98% of all reported ketchup purchases. If the household is observed to purchase a given brand of ketchup at any time(s) during the year, then the response variable for that brand and household is coded one; otherwise it is coded zero. Table 9 summarizes the data and Table 10 suggests some merit to considering a multivariate approach.

[Tables 9 and 10 here]

Using the GHK simulator, a Multivariate Probit model was estimated. The simple model posits that the decision to purchase a given brand of ketchup is related to a polynomial in household size, whether the household lives in a house, the income of the household in $1000's, and the highest grade achieved by the head of the household. Results are reported in Table 11. Of particular interest is whether the response variables are correlated, and the test reported in the table suggests this is the case.

[Table 11 here]

Table 12 reports the exact maximum likelihood results using the multivariate binary discrete normal distribution. Note that the pattern of signs and significance are similar between the two models, but by no means identical. The fits of the two models are almost identical and since both models estimate the same number of parameters, application of an information criterion is unproductive.

[Table 12 here]

If we focus on estimation of the marginal effects, we see almost an exact correspondence (at least to the first three decimal places) between the two models as shown in Table 13. Given that these marginal effects are simpler to derive from the multivariate binary discrete normal model and that estimation was achieved more than 20 times faster, this estimator appears to have merit in the analysis of large data sets.

[Table 13 here]

Finally, the implied correlations are presented in Table 14. The multivariate binary discrete normal reproduces the raw observed correlations. The Multivariate Probit's correlations are associated with the latent responses.

[Table 14 here]

5. Conclusions and Discussions

A multivariate binary response model, what we termed the Multivariate Binary Discrete Normal model, is obtained from the multivariate discrete normal distribution. The statistical model is a member of the quadratic exponential family and as such, under proper specification of conditional means, the maximum likelihood estimator possesses desirable properties even under distributional misspecification. Maximum likelihood estimation of the multivariate binary discrete normal model is straightforward because the normalizing factor for the joint probability mass function is obtained via the evaluation of 2^M support points for M binary response variables. The MVBDN model nests the independent logit model. The MVBDN estimates of the correlations among response variables coincide with observed ones when the regressors (including a constant term) are identical across equations. Lastly, marginal effects that count for the underlying property of multivariate model are much easier to derive and compute under the MVBDN model.

Application of the statistical model to simulation data and to a well-known empirical data set suggests that the estimator can produce results with fits comparable to other estimators; and in our empirical data computation time is reduced by a factor of 20 relative to a Multivariate Probit model.

References

Cappellari, L. and S.P. Jenkins. 2003. "Multivariate Probit Regression Using Simulated Maximum Likelihood." The Stata Journal 3(3): 278-294.

Fitzmaurice, G.M., N. M. Laird, and A.G. Rotnitzky. 1993. "Regression Models for Discrete Longitudinal Responses." *Statistical Science* 8(3): 284-299.

Geweke, J. 1989. "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica* 57(6):1317-1339.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52(3):681-700.

Hajivassiliou, V. A. and D. L. McFadden. 1998. "The Method of Simulated Scores for the Estimation of LDV Models." *Econometrica* 66(4): 863-896.

Keane, M. P. 1994. "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62 (1):95-116.

Kemp, A. W. 1997. "Characterizations of a Discrete Normal Distribution." *Journal of Statistical Planning and Inference* 63 (2):223-229.

Lindsey, J.K. 1996. *Parametric Statistical Inference*. New York: Oxford Science Publications.

Mullahy, J. 2011. "Marginal Effects in Multivariate Probit and Kindred Discrete and Count Outcome Models, with Applications in Health Economics." National Bureau of Economic Research (NBER) working paper 17588.

Pearson K. 1900. "Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable." *Philosophical Transactions of the Royal Society of* London, Series A, Containing Papers of a Mathematical or Physical Character: 195:1-47.

Prentice, R.L. and L.P., Zhao. 1991. "Estimating Equation for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses." *Biometrics* 47 (3): 825-839.

Table 1. Summary of Descriptive Statistics of the Generated Response Variables

		y_1	y_2	y_3	y_4
Mean (average over 300 repetitions)					
Case 1: Identical regressors + MVN		0.9250	0.9775	0.5400	0.4750
Case 2: Identical regressors + MVT		0.9150	0.9600	0.5475	0.4825
Case 3: Different regressors + MVN		0.6194	0.6394	0.7783	0.1537
Case 4: Different regressors + MVT		0.6750	0.6900	0.8150	0.2075
Correlation (the last data set for example)					
Case 1: Identical regressors + MVN	y_1		0.3177	0.0683	-0.2651
	y_2			0.1221	-0.1965
	y_3				-0.0912
Case 2: Identical regressors + MVT	y_1		0.2100	0.0452	-0.2434
	y_2			0.1335	-0.0978
	y_3				-0.0374
Case 3: Different regressors + MVN	y_1		0.0646	-0.1206	0.1533
	y_2			0.1271	0.1168
	y_3				0.0382
Case 4: Different regressors + MVT	y_1		0.0964	-0.0415	0.2072
	y_2			0.1480	0.1282
	y_3				0.0458

Table 2. Parameter Estimates: Identical Regressor and MVN

		Btrue	MVP [a]	95% Empirical CI		MVBDN [b]	95% Empirical CI	
y_1	Intercept	3.45	**3.4872***	2.8074	4.5319	**9.4064***	4.1471	13.5649
	X	-1	**-1.0083***	-1.3815	-0.7545	**-2.4984***	-3.7249	-1.1026
y_2	Intercept	5	**5.3019***	3.8622	8.0935	**10.4651***	6.6867	15.4808
	X	-1.3	**-1.3913***	-2.3174	-0.8547	**-2.5537***	-4.2306	-1.2714
y_3	Intercept	0.8	**0.7975***	0.5636	1.0205	**4.7663**	-1.0632	10.8475
	X	-0.5	**-0.4971***	-0.6141	-0.3629	**-1.0765**	-2.9187	0.7417
y_4	Intercept	-2.8	**-2.86***	-3.3343	-2.4073	**4.8362**	-1.4653	10.7174
	X	1.4	**1.4277***	1.1933	1.6788	**-0.7035**	-2.7888	1.482
	σ_{12}	0.3	**0.2918**	-0.0309	0.6248	**0.4205**	-0.3168	0.7886
	σ_{13}	-0.2	**-0.188**	-0.4045	0.0258	**0.0853**	-0.541	0.6073
	σ_{14}	0.1	**0.114**	-0.1602	0.3792	**0.2713**	-0.4693	0.7126
	σ_{23}	0.25	**0.2407**	-0.0723	0.7234	**0.6082***	0.0881	0.8879
	σ_{24}	0.5	**0.4626***	0.0591	0.7674	**0.6992***	0.232	0.8605
	σ_{34}	0.75	**0.748***	0.5878	0.8889	**0.7925***	0.6731	0.8685
Log likelihood			**-512.75**	-551.03	-481.12	**-517.37**	-557.8	-487.65
Repetitions			240			240		

[a] Average estimates over all repetitions under the Multivariate Probit model Average
[b] Average estimates over all repetitions under the Multivariate Binary Discrete Normal model
* Estimates are significantly different from zero at 5% level

Table 3. Parameter Estimates: Identical Regressor and MVT

		Btrue	MVP [a]	95% Empirical CI		MVBDN [b]	95% Empirical CI	
y_1	Intercept	3.45	**2.571***	2.0707	3.3645	**6.7215***	4.3643	8.7547
	X	-1	**-0.7062***	-0.9978	-0.5091	**-1.6343***	-2.3137	-1.0343
y_2	Intercept	5	**3.2118***	2.4947	4.4541	**6.6968***	4.2842	9.2862
	X	-1.3	**-0.7449***	-1.1824	-0.4509	**-1.443***	-2.316	-0.6201
y_3	Intercept	0.8	**0.738***	0.4996	0.9566	**2.2552**	-1.4681	5.3254
	X	-0.5	**-0.4588***	-0.5926	-0.33	**-0.4275**	-1.4218	0.6292
y_4	Intercept	-2.8	**-2.3154***	-2.7663	-1.9508	**2.6129**	-0.8039	5.3878
	X	1.4	**1.1614***	0.9587	1.3864	**-0.1403**	-1.0875	1.0451
	σ_{12}	0.3	**0.3616***	0.0663	0.6306	**0.5368**	-0.1302	0.7824
	σ_{13}	-0.2	**-0.2013**	-0.4455	0.0076	**0.043**	-0.5111	0.4898
	σ_{14}	0.1	**0.0874**	-0.1649	0.3658	**0.2933**	-0.3416	0.6791
	σ_{23}	0.25	**0.1768**	-0.1265	0.4455	**0.5035***	0.0386	0.7662
	σ_{24}	0.5	**0.4861***	0.1631	0.7553	**0.7030***	0.3906	0.8324
	σ_{34}	0.75	**0.7066***	0.5292	0.8582	**0.7646***	0.6136	0.8431
Log likelihood			**-590.28**	-630.3	-549.73	**-592.28**	-633.56	-554.36
Repetitions			200			200		

* Estimates are significantly different from zero at 5% level

Table 4. Parameter Estimates: Different Regressors and MVN

		Btrue	MVP	95% Empirical CI		MVBDN	95% Empirical CI	
y_1	Intercept	0.5	**0.5137***	0.3661	0.6886	**1.1455***	0.6944	1.7673
	X_1	4	**4.0676***	3.4671	4.8008	**5.0473***	2.9378	7.5436
y_2	Intercept	3	**3.0192***	2.4951	3.5958	**2.8278***	1.8504	5.5827
	X_1	0.5	**0.4941***	0.0655	0.9527	**2.2065**	-0.7852	4.2439
	X_2	-3	**-3.0194***	-3.6123	-2.4614	**-1.9964***	-5.1332	-1.1018
y_3	Intercept	1	**1.012***	0.5963	1.4785	**1.5551***	0.3053	2.6503
	X_1	-2	**-2.051***	-2.6518	-1.4927	**1.5753**	-0.9511	3.5665
	X_2	0.4	**0.3988**	-0.071	0.9236	**-0.4046**	-3.012	0.2812
	X_3	-0.75	**-0.7532***	-1.0262	-0.5097	**-0.0485**	-0.4729	1.5728
y_4	Intercept	-3.5	**-3.6124***	-4.7337	-2.6899	**-0.7686**	-6.2506	0.7897
	X_1	1	**1.0218***	0.3132	1.7595	**2.8142***	0.3301	4.8789
	X_2	-0.3	**-0.3209**	-0.9396	0.3787	**-0.363**	-1.7262	0.1221
	X_3	3	**3.1054***	2.5166	3.7969	**1.5117***	0.5885	5.8239
	X_4	-0.4	**-0.4174**	-1.7784	0.9407	**-0.1112**	-0.9933	0.7121
	σ_{12}	0.25	**0.2346***	0.0033	0.4331	**0.35**	-0.0172	0.4316
	σ_{13}	0.5	**0.5099***	0.3031	0.743	**0.3855***	0.0530	0.4428
	σ_{14}	0.75	**0.7287***	0.4930	0.8937	**0.3634***	0.0003	0.4479
	σ_{23}	0.75	**0.7462***	0.6001	0.8622	**0.4373***	0.3658	0.4575
	σ_{24}	0.5	**0.5046***	0.2380	0.7251	**0.3689***	0.0715	0.4471
	σ_{34}	0.75	**0.7597***	0.5691	0.9231	**0.4118***	0.2281	0.453
Log likelihood			**-559.01**	-595.98	-523.34	**-585.86**	-625.73	-547.91
Repetitions			**187**			187		

* Estimates are significantly different from zero at 5% level

Table 5. Comparison of MVBDN and MVP estimates: Different Regressors and MVT

		Btrue	MVP	95% Empirical CI		MVBDN	95% Empirical CI	
y_1	Intercept	0.5	**0.4100***	0.2694	0.5736	**0.9077**	0.4835	1.4443
	X_1	4	**3.3900***	2.8533	4.0514	**4.1058**	2.5171	5.9076
y_2	Intercept	3	**2.6088***	2.1324	3.1098	**2.2563**	1.6002	4.289
	X_1	0.5	**0.415**	-0.0209	0.8798	**2.0941**	0.3818	3.7578
	X_2	-3	**-2.6142***	-3.1427	-2.0951	**-1.5642**	-3.6269	-1.0036
y_3	Intercept	1	**0.8635***	0.4451	1.3802	**1.3561**	0.7897	2.4262
	X_1	-2	**-1.6433***	-2.2285	-1.0481	**1.4665**	-0.2331	3.2123
	X_2	0.4	**0.3303**	-0.1877	0.8552	**-0.1847**	-1.626	0.277
	X_3	-0.75	**-0.6389***	-0.888	-0.3788	**-0.2147**	-0.4326	0.6999
y_4	Intercept	-3.5	**-2.6432***	-3.5946	-1.847	**-0.2316**	-3.6073	0.5882
	X_1	1	**0.8471***	0.2273	1.582	**2.5821**	0.7679	4.2904
	X_2	-0.3	**-0.2264**	-0.8476	0.3595	**-0.2537**	-1.2418	0.1273
	X_3	3	**2.241***	1.691	3.1243	**0.8858**	0.5043	4.2019
	X_4	-0.4	**-0.3069**	-1.4722	0.8466	**-0.1073**	-0.8153	0.3915
	σ_{12}	0.25	**0.2396***	0.0358	0.428	**0.6918**	0.3062	0.7952
	σ_{13}	0.5	**0.4825***	0.2780	0.6846	**0.7565**	0.5736	0.8347
	σ_{14}	0.75	**0.7159***	0.4967	0.904	**0.759**	0.0877	0.8511
	σ_{23}	0.75	**0.7353***	0.6006	0.8674	**0.8373**	0.7481	0.8774
	σ_{24}	0.5	**0.4697***	0.2319	0.7204	**0.7431**	0.1735	0.8398
	σ_{34}	0.75	**0.7171***	0.5020	0.8795	**0.7883**	0.4023	0.8576
Log likelihood			**-626.1806**	-662.68	-587.71	**-647.51**	-686.94	-603.9
Repetitions			240			240		

* Estimates are significantly different from zero at 5% level

Table 6. Average Marginal Effects: Identical Regressors

Error terms ~ MVN

		MVBDN	95% CI		MVP $E(y_{\mathbf{m}})$ [a]	95% CI		MVBDN -MVP	95% CI	
me	x on y_1	**-0.1611***	-0.2025	-0.1197	**-0.1448***	-0.1814	-0.1056	**-0.0164***	-0.0342	-0.0034
	x on y_2	**-0.1057***	-0.1557	-0.0563	**-0.0922***	-0.1396	-0.046	**-0.0135**	-0.0293	0.0019
	x on y_3	**-0.2072***	-0.2325	-0.1724	**-0.1749***	-0.2053	-0.1355	**-0.0323***	-0.0478	-0.0179
	x on y_4	**0.2759***	0.2617	0.284	**0.282***	0.2733	0.2866	**-0.0061***	-0.0166	-0.0011

Error terms ~ MVT

		MVBDN	95% CI		MVP $E(y_{\mathbf{m}})$ [a]	95% CI		MVBDN -MVP	95% CI	
me	x on y_1	**-0.1409***	-0.1800	-0.1034	**-0.1282***	-0.1639	-0.0938	**-0.0127***	-0.0233	-0.003
	x on y_2	**-0.0899***	-0.1418	-0.0517	**-0.0767***	-0.1157	-0.0459	**-0.0132***	-0.0268	-0.0024
	x on y_3	**-0.1915***	-0.2206	-0.1573	**-0.1641***	-0.199	-0.1248	**-0.0274***	-0.044	-0.0151
	x on y_4	**0.2668***	0.2497	0.2791	**0.2718***	0.2568	0.2822	**-0.0051**	-0.0134	0.0004

[a] Marginal effects derived from unconditional distributions.
*****The estimate is significantly different from zero at 5% level.

Table 7. Average Marginal Effects: Different Regressors

Error terms ~ MVN							
		MVBDN	95%	CI	MVP $E(y_{\mathbf{m}})$ [a]	95%	CI
me of x_1	y_1	**1.0319***	0.6630	2.3125	**0.9513***	0.8109	1.1228
	y_2	**0.4751**	-0.6017	0.999	**0.1301***	0.0172	0.2508
	y_3	**-0.5566***	-0.9945	-0.2524	**-0.4379***	-0.5662	-0.3187
	y_4	**-0.2501***	-0.7971	-0.022	**0.1477***	0.0453	0.2543
me of x_2	y_1	**0.1694***	0.0311	0.6647			
	y_2	**-0.8708***	-1.9168	-0.5499	**-0.7947***	-0.9508	-0.6479
	y_3	**0.2642**	-0.0905	0.3806	**0.0851**	-0.0152	0.1972
	y_4	**0.2532***	0.0846	0.7471	**-0.0464**	-0.1358	0.0547
me of x_3	y_1	**-0.1503***	-0.7131	-0.0261			
	y_2	**-0.127***	-0.6732	-0.0046			
	y_3	**-0.1728**	-0.2803	0.1013	**-0.1608**	-0.2191	-0.1088
	y_4	**0.5069***	0.2314	1.7978	**0.4489***	0.3638	0.5488
me of x_4	y_1	**0.0114**	-0.0728	0.1016			
	y_2	**0.0098**	-0.0626	0.0873			
	y_3	**0.0114**	-0.0728	0.1015			
	y_4	**-0.0368**	-0.3291	0.2359	**-0.0603**	-0.2571	0.136
Error terms ~ MVT							
		MVBDN	95%	CI	MVP $E(y_{\mathbf{m}})$	95%	CI
meof x_1	y_1	**1.1933***	0.8677	1.9184	**0.9019***	0.7591	1.0779
	y_2	**0.4412**	-0.2133	0.8905	**0.1225**	-0.0062	0.2596
	y_3	**-0.5869***	-0.7913	-0.4198	**-0.3866***	-0.5243	-0.2466
	y_4	**-0.345***	-0.7927	-0.1254	**0.1416***	0.038	0.2645
me of x_2	y_1	**0.193***	0.0559	0.5959			
	y_2	**-0.7791***	-1.6079	-0.5457	**-0.7715***	-0.9274	-0.6183
	y_3	**0.2162***	0.0103	0.3237	**0.0777**	-0.0442	0.2012
	y_4	**0.2149***	0.0956	0.4984	**-0.0378**	-0.1417	0.0601
me of x_3	y_1	**-0.086***	-0.581	-0.0276			
	y_2	**-0.0622***	-0.455	-0.0156			
	y_3	**-0.1762***	-0.2741	-0.0877	**-0.1503***	-0.2089	-0.0891
	y_4	**0.357***	0.2253	1.4939	**0.3747***	0.2827	0.5224
me of x_4	y_1	**0.0133**	-0.0486	0.1011			
	y_2	**0.0102**	-0.0371	0.0773			
	y_3	**0.0111**	-0.0403	0.084			
	y_4	**-0.0401**	-0.3046	0.1463	**-0.0513**	-0.2461	0.1416

Table 8. Correlation Estimates among Response Variables

Case 1. Identical Regressors + MVN

	MVBDN				MVP Residuals		
	y_2	y_3	y_4		y_2	y_3	y_4
y_1	0.3177	0.0683	-0.2651	y_1	0.9084	0.0862	-0.6396
y_2		0.1221	-0.1965	y_2		0.1340	-0.6825
y_3			-0.0912	y_3			0.0945

Case 2. Identical Regressors + MVT

	MVBDN				MVP Residuals		
	y_2	y_3	y_4		y_2	y_3	y_4
y_1	0.2100	0.0452	-0.2434	y_1	0.6530	0.0260	-0.4079
y2		0.1335	-0.0978	y2		0.1367	-0.4491
y_3			-0.0374	y_3			0.1536

Case 3. Different Regressors + MVN

	MVBDN				MVP Residuals		
		y_2	y_3	y_4	y_2	y_3	y_4
y_1	0.0514	-0.1060	0.1319	y_1	0.0836	-0.1493	0.1517
y_2		0.1266	0.1077	y_2		0.1123	0.1342
y_3			0.0414	y_3			-0.2231

Case 4. Different Regressors + MVT

	MVBDN				MVP Residuals		
		y_2	y_3	y_4	y_2	y_3	y_4
y_1	0.1001	-0.0221	0.1790	y_1	0.1680	0.0032	0.1597
y_2		0.1456	0.1416	y_2		0.1658	0.1497
y3			0.0519	y3			-0.1123

Note: Estimates from the last data set for illustration.

Table 9. Counts of Households Which Purchase Ketchup (216 Households Make No Purchases)

Combination of Brand(s)	With **Brand1**	**Brand2**	**Brand3**	**Brand4**	**Brand5**
1	**1258**				
2	343	**426**			
3	465	223	**555**		
4	96	56	76	**123**	
5	134	81	102	19	**173**
2&3	198				
2&4	51				
2&5	69				
3&4	68	43			
3&5	83	56			
4&5	16	12	16		
2&3&4	41				
2&3&5	50				
3&4&5	14	12			
2&3&4&5	11				

Table 10. Correlations of Observed Binary Responses: Ketchup Brands

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
Brand 1	1	0.0598	0.1268	0.0123	0.0101
Brand 2		1	0.2338	0.1279	0.1644
Brand 3			1	0.1692	0.1836
Brand 4				1	0.046
Brand 5					1

Table 11. Ketchup Brands Maximum Likelihood Results—Multivariate Probit (GHK Simulator)

Log Likelihood = -3607.05

Brand1	**Coefficient**	**Std. Error**	**z-Value**	**Brand2**	Coefficient	Std. Error	z-Value
Intercept	-0.6103	0.1821	-3.3511	Intercept	-1.448	0.192	-7.5417
Size	0.7284	0.1137	6.4051	Size	0.3729	0.1059	3.5204
Size^2	-0.0649	0.0172	-3.7623	Size^2	-0.0304	0.0151	-2.0168
House	0.1357	0.0992	1.368	House	0.3055	0.111	2.7519
Income	0.0002	0.0025	0.0753	Income	-0.0040	0.0023	-1.7478
Educ	-0.0192	0.0188	-1.0181	Educ	-0.0174	0.018	-0.9648

Brand3	Coefficient	Std. Error	z-value	**Brand4**	Coefficient	Std. Error	z-value
Intercept	-1.3703	0.1803	-7.6001	Intercept	-1.2189	0.2273	-5.3622
Size	0.524	0.1013	5.1709	Size	0.0738	0.1440	0.5124
Size^2	-0.0424	0.0145	-2.9304	Size^2	0.0045	0.0197	0.2262
House	-0.0246	0.1005	-0.2448	House	-0.1044	0.1353	-0.7715
Income	-0.0058	0.0024	-2.4558	Income	-0.0095	0.0051	-1.8599
Educ	0.0064	0.0171	0.3761	Educ	-0.0200	0.0260	-0.7675

Brand5	Coefficient	Std. Error	z-Value
Intercept	-2.0977	0.2529	-8.2932
Size	0.3852	0.1335	2.8858
Size^2	-0.0307	0.0183	-1.6753
House	0.1651	0.1435	1.150
Income	-0.0044	0.0030	-1.4577
Educ	-0.0004	0.0231	-0.0182

r_{12}	-0.0052	0.0473	-0.1109	r_{24}	0.2840	0.0580	4.8963
r_{13}	0.1017	0.0456	2.2324	r_{25}	0.2967	0.0527	5.6274
r_{14}	-0.0157	0.0637	-0.2472	r_{34}	0.3720	0.0543	6.8564
r_{15}	-0.1004	0.0592	-1.6963	r_{35}	0.3178	0.0500	6.3553
r_{23}	0.3403	0.0408	8.3504	r_{45}	0.1155	0.0729	1.5850

LR Test of H_0: $\rho_{ij}=0$ all i<j

X^2 = 172.3 with 10 df.
p = 0.0000

Table 12. Ketchup Brands Maximum Likelihood Results—Multivariate Binary Discrete Normal

Log Likelihood = -3606.03

Brand1	Coefficient	Std. Error	z-Value	**Brand2**	Coefficient	Std. Error	z-Value
Intercept	0.0566	1.384	0.0409	Intercept	-4.7266	0.5202	-9.0869
Size	1.0734	0.3984	2.694	Size	1.2382	0.3090	4.0065
Size^2	-0.0911	0.0425	-2.1451	Size^2	-0.0928	0.0382	-2.4323
House	0.1465	0.1901	0.7706	House	0.4604	0.3059	1.5051
Income	0.002	0.0086	0.2353	Income	-0.0259	0.0091	-2.8383
Educ	-0.0344	0.0344	-1.000	Educ	-0.0346	0.0509	-0.6790

Brand3	Coefficient	Std. Error	z-Value	**Brand4**	Coefficient	Std. Error	z-Value
Intercept	-4.7816	0.5160	-9.2664	Intercept	-4.3343	0.5511	-7.8642
Size	1.4267	0.3110	4.5883	Size	1.0388	0.3588	2.895
Size^2	-0.1075	0.0384	-2.7967	Size^2	-0.0700	0.0434	-1.6129
House	0.2670	0.3027	0.8822	House	0.1505	0.3214	0.4683
Income	-0.0279	0.0096	-2.8955	Income	-0.0308	0.0122	-2.5302
Educ	-0.0250	0.0518	-0.4826	Educ	-0.0402	0.0593	-0.6776

Brand5	Coefficient	Std. Error	z-Value
Intercept	-4.951	0.5848	-8.4662
Size	1.085	0.3269	3.3192
Size^2	-0.0811	0.0410	-1.9799
House	0.3868	0.3379	1.1444
Income	-0.0236	0.0090	-2.6087
Educ	-0.0144	0.0544	-0.2641

s_{15}	-0.1864	0.2088	-0.8926
s_{23}	0.7088	0.0232	30.5638
s_{24}	0.6627	0.0448	14.7817
s_{25}	0.6461	0.0482	13.3995
s_{34}	0.7124	0.0367	19.4051
s_{35}	0.6472	0.0529	12.2345
s_{45}	0.5278	0.1044	5.0575

LR Test of H_0: $\sigma_{ij}=0$ all i<j

$X^2 = 179.8$ with 10 df.
p = 0.0000

Table 13. Comparison of Average Marginal Effects for Ketchup Brands Choice

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
MVP—Size +1	0.08802	0.053	0.08492	0.015345	0.03096
MVBDN—Size +1	0.08874	0.05292	0.08493	0.016987	0.03114
MVP—Income+1[a]	0.00006	-0.00124	-0.00199	-0.001299	-0.00075
MVBDN—Income+1	0.00012	-0.00133	-0.00219	-0.001563	-0.00079

[a]Income in $1,000's

Table 14. Ketchup Brands Choice Correlations Implied by Models—MVDBN Coincides with Observed Correlations

Brand	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
1-MVP	1	-0.0052	0.1017	-0.0157	-0.1004
1-MVBDN		0.0598	0.1268	0.0123	0.0101
2-MVP		1	0.3403	0.2840	0.2967
2-MVBDN			0.2338	0.1279	0.1644
3-MVP			1	0.3720	0.3178
3-MVBDN				0.1692	0.1836
4-MVP				1	0.1155
4-MVBDN					0.0460
5-MVP					1
5-MVBDN					

Appendix

Proposition: $\mathbf{A}$ in Equation (18) is a zero matrix when the regressors are identical across equations.

Proof:

When regressors are identical across equations, we have that $\boldsymbol{X}_{\boldsymbol{i}}^{*} = \boldsymbol{I}_{\boldsymbol{M}} \otimes \boldsymbol{x}_{\boldsymbol{i}}^{*}$. Equation (19) then becomes

(A1) $\frac{1}{n}\sum_{\boldsymbol{i}}(\boldsymbol{y}_{\boldsymbol{i}}\boldsymbol{\Sigma}^{-1}\boldsymbol{I}_{\boldsymbol{M}} \otimes \boldsymbol{x}_{\boldsymbol{i}}^{*}) = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}} E(\boldsymbol{y}_{\boldsymbol{i}})\boldsymbol{\Sigma}^{-1}\boldsymbol{I}_{\boldsymbol{M}} \otimes \boldsymbol{x}_{\boldsymbol{i}}^{*}$

which can be further re-organized as follows:

(A2) $\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{y}} = \boldsymbol{0}$

where $\widetilde{\boldsymbol{y}} = \left[\frac{1}{n}\sum_{i}(y_{1i} - E(y_{1i}))\boldsymbol{x}_{\boldsymbol{i}}^{*};\quad \frac{1}{n}\sum_{i}(y_{2i} - E(y_{2i}))\boldsymbol{x}_{\boldsymbol{i}}^{*};\quad \ldots\quad \frac{1}{n}\sum_{i}(y_{Mi} - E(y_{Mi}))\boldsymbol{x}_{\boldsymbol{i}}^{*}\right]$ is a M x k matrix, where k is the number of explanatory variables. We can solve for the elements of $\widetilde{\boldsymbol{y}}$ by Cramer's Rule. The system in Equation (A2) can be broke down into k systems, where the to-be-solved vector is a column of $\widetilde{\boldsymbol{y}}$ and the answer vector is the corresponding zero column vector of the M x k answer matrix $\boldsymbol{0}$. For each system, let D=det($\boldsymbol{\Sigma}^{-1}$), and D_m be coefficient determinant with answer-column values in m-column of the coefficient matrix $\boldsymbol{\Sigma}^{-1}$. Since the answer-column is a zero vector, $D_k = 0\ \forall$ m=1, 2, … M. This solution holds for all k systems and thus the elements of $\widetilde{\boldsymbol{y}}$ are zeros. In general notation, we have that

(A3) $\frac{1}{n}\sum_{i=1}^{n}\big(y_{mi} - E(y_{mi})\big)\boldsymbol{x}_{\boldsymbol{i}}^{*} = 0;\ \forall \mathrm{m} = 1, 2, \ldots \mathrm{M}.$

Now let's examine the extra component in Equation (18):

$$\boldsymbol{A} = \frac{1}{n}\sum\nolimits_{\boldsymbol{i}} \boldsymbol{X_i\beta}(\,\boldsymbol{y_i} - \boldsymbol{E}(\boldsymbol{y_i}))$$

$$= \frac{1}{n}\sum\nolimits_{\boldsymbol{i}} \begin{bmatrix} 1 & \boldsymbol{x_{1i}} & 0 & \boldsymbol{0} & \cdots & 0 & \boldsymbol{0} \\ 0 & \boldsymbol{0} & 1 & \boldsymbol{x_{2i}} & \cdots & 0 & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \boldsymbol{0} & 0 & \boldsymbol{0} & \cdots & 1 & \boldsymbol{x_{2M}} \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \boldsymbol{\beta_{11}} \\ \beta_{20} \\ \boldsymbol{\beta_{21}} \\ \vdots \\ \beta_{M0} \\ \boldsymbol{\beta_{M1}} \end{bmatrix} [y_{1i} - E(y_{1i}) \quad y_{2i} - E(y_{2i}) \quad \cdots \quad y_{Mi} - E(y_{Mi})]$$

where β_{m0} is the parameter associated with the constant term for the m^{th} response variable, and $\boldsymbol{\beta_{m1}}$ is a vector of parameters of the explanatory variables for equation m (m=1, 2, … M). The (m, k)th element of matrix **A**, a_{mk}, is as follows:

(A4) $\quad a_{mk} = \beta_{m0}\frac{1}{n}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\big(y_{ki} - E(y_{ki})\big) + \frac{1}{n}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\big(y_{ki} - E(y_{ki})\big)\boldsymbol{x}^*_{\boldsymbol{mi}}\,\boldsymbol{\beta_{m1}},\ m, k = 1,2, \dots M.$

The first term of the RHS is zero under the first moment condition, regardless whether the regressors are identical across equations. The second term reduces to zero when there are identical regressors (Equation (A3) as previously proven). Therefore the extra component in Equation (18) is a zero matrix and thus the second moment equation is the maximum likelihood equation under identical regressors. The combination of the first two moment equations leads to the result that the estimated correlation among response variables under the MVBDN model matches the sample correlation.