

## NBER WORKING PAPER SERIES

MARGINAL EFFECTS IN MULTIVARIATE PROBIT AND KINDRED DISCRETE  
AND COUNT OUTCOME MODELS, WITH APPLICATIONS IN HEALTH ECONOMICS

John Mullahy

Working Paper 17588

<http://www.nber.org/papers/w17588>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

November 2011

Many thanks are owed to Bill Greene, Alberto Holly, Mari Palta, and Ron Thisted, and participants in the UW-Madison Health Econometrics Workgroup for their extraordinarily helpful comments, suggestions and insights. They are, of course, blameless for any shortcomings. This work has been supported in part by the RWJ Health & Society Scholars Program at UW-Madison. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by John Mullahy. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Marginal Effects in Multivariate Probit and Kindred Discrete and Count Outcome Models,  
with Applications in Health Economics

John Mullahy

NBER Working Paper No. 17588

November 2011

JEL No. C35,I1

### **ABSTRACT**

Estimation of marginal or partial effects of covariates  $x$  on various conditional parameters or functionals is often the main target of applied microeconomic analysis. In the specific context of probit models such estimation is straightforward in univariate models, and Greene, 1996, 1998, has extended these results to cover the case of quadrant probability marginal effects in bivariate probit models.

The purpose of this paper is to extend these results to the general multivariate probit context for arbitrary orthant probabilities and to demonstrate the applicability of such extensions in contexts of interest in health economics applications. The baseline results are extended to models that condition on subvectors of  $y$ , to count data structures that derive from the probability structure of  $y$ , to multivariate ordered probit data structures, and to multinomial probit models whose marginal effects turn out to be a special case of those of the multivariate probit model. Simulations reveal that analytical formulae versus fully numerical derivatives result in a reduction in computational time as well as an increase in accuracy.

John Mullahy

University of Wisconsin-Madison

Dept. of Population Health Sciences

787 WARF, 610 N. Walnut Street

Madison, WI 53726

and NBER

[jmullahy@facstaff.wisc.edu](mailto:jmullahy@facstaff.wisc.edu)

## 1. Introduction

Given a conditional distribution function  $F(\mathbf{y}|\mathbf{x})$  defined on possibly multivariate outcomes  $\mathbf{y}$  and exogenous covariates  $\mathbf{x}$ , estimation of marginal or partial effects of covariates  $\mathbf{x}$  on various conditional parameters or functionals  $\xi(\mathbf{x})$  of  $F(\mathbf{y}|\mathbf{x})$  is often the main target of applied microeconomic analysis. In general  $\partial\xi(\mathbf{x})/\partial\mathbf{x}$  will describe  $\mathbf{x}$ 's effects on conditional means, quantiles, probabilities, and other conditional functionals.

In the specific context of probit models, estimation of partial effects like  $\partial\text{Prob}(\mathbf{y}\in S|\mathbf{x})/\partial\mathbf{x}$  is typically of central interest. Such estimation is straightforward in univariate models for  $\partial\text{Prob}(y=1|\mathbf{x})/\partial\mathbf{x}$ , and Greene, 1996, 1998,<sup>1</sup> has extended these calculations to handle quadrant probability marginal effects  $\partial\text{Prob}(y_1=k_1, y_2=k_2|\mathbf{x})/\partial\mathbf{x}$ ,  $k_j=0,1$ , in bivariate probit models.

The main purpose of this paper is to extend these results to encompass the general  $m\geq 2$  multivariate probit (MVP) case for arbitrary orthant probabilities. Specifically the paper derives and then demonstrates in several contexts the usefulness of analytical representations of

$$\frac{\partial\text{Prob}(y_1=k_1, \dots, y_m=k_m|\mathbf{x})}{\partial\mathbf{x}} \tag{1}$$

or, in shorthand,  $\partial\text{Prob}(\mathbf{y}=\mathbf{k}|\mathbf{x})/\partial\mathbf{x}$ , where  $\mathbf{y}=[y_j]$  is the  $m$ -variate binary outcome vector,  $\mathbf{k}=[k_j]$  is an  $m$ -vector of zeros or ones indicating any of the  $2^m$  possible outcomes,  $\mathbf{x}$  are conditioning covariates,<sup>2</sup> and  $\text{Prob}(\dots)$  is a joint or orthant probability from a multivariate normal

---

<sup>1</sup> See also Christofides et al., 1997, 1998.

<sup>2</sup> To streamline the analysis and notation the  $\mathbf{x}$ 's will be treated as continuous so that " $\partial\mathbf{x}$ " calculus  
(continued)

distribution.<sup>3</sup>

Greene's formulae for the marginal effects in the bivariate quadrant probability ( $m=2$ ) case are well established, but the analytical formulae describing the general orthant probability result are not evident in the literature. This paper derives the general result which contains Greene's bivariate formula as a special case. While numerical simulation methods like GHK (see Hajivassiliou et al., 1996) are available for obtaining these marginal effects as discussed below, there may arise computational advantages from calculating analytical marginal effects, with the bottom line result being that the dimensionality of the cumulative normal that must be evaluated to obtain the partial effect is reduced by one to  $m-1$  if the analytical formulae are applied. The paper then extends these results in several directions that are described below.

### **Data and Estimation**

The outcomes  $\mathbf{y} = [y_j]$  can be thought of as arising in the standard probit context as binary

---

(continued)

can be used. Discrete  $\mathbf{x}$ 's (e.g. dummy variables or count measures) can be introduced straightforwardly with the understanding that discrete differences in  $\text{Prob}(y_1 = k_1, \dots, y_m = k_m | \mathbf{x})$  due to  $\Delta x_j = 1$  will be of interest; these can be computed by evaluating  $\text{Prob}(y_1 = k_1, \dots, y_m = k_m | \mathbf{x})$  at two different values of  $x_j$  and then differencing. See Stock, 1989, for discussion of partial effects of interest in policy analysis.

<sup>3</sup> Somewhat informally, the paper uses the term "orthant probability" in reference to the vector of binary outcomes  $\mathbf{y}$  to refer to the probabilities that the underlying latent random variables that map into the observed binary  $\mathbf{y}$  (see (2) below) occupy any of the  $2^m$  orthants in  $\mathbb{R}^m$  defined implicitly by  $\mathbf{k}$ .

Some additional notation will also prove useful. Let  $\mathbf{K}$  be the  $m \times 2^m$  matrix whose columns (arranged arbitrarily) are the  $2^m$  possible outcome configurations  $\mathbf{k}$ . Let  $\mathbb{P}$  be a  $2^m$ -element set indexing columns of  $\mathbf{K}$  having typical indexing element  $p$ , so that  $\mathbf{k}_p = \mathbf{K}_{\cdot p}$  will denote a particular ( $p$ -th) outcome configuration. Subject-specific "i" subscripts will be suppressed unless useful for clarity.

indicators of threshold crossings of latent marginal normal variates:<sup>4</sup>

$$y_j^* = \mathbf{x}\boldsymbol{\beta}_j + \varepsilon_j, \quad j=1, \dots, m$$

$$\mathbf{y} = 1(\mathbf{y}^* \geq \mathbf{0})$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_m] \sim \text{MVN}(\mathbf{0}, \mathbf{R}) \quad (2)$$

and

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho_{1m} & \cdots & & 1 \end{bmatrix}.$$

The parameters  $\mathbf{B} = [\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_m^T]$  and  $\mathbf{R}$  can be estimated using algorithms like Stata's *mvprobit* that uses a "full-information" approach (i.e. estimating all elements of  $\mathbf{B}$  and  $\mathbf{R}$  simultaneously) with simulated maximum likelihood. Alternatively  $\mathbf{B}$  and  $\mathbf{R}$  can be estimated consistently using a "limited-information" approach suggested by Avery et al., 1980, in which, e.g., the  $\rho_{jk}$  elements of  $\mathbf{R}$  are estimated one-by-one using bivariate probit estimators (e.g. Stata's *biprobit* algorithm).<sup>5</sup> Which pairwise estimates of  $\mathbf{B}$  to use in the latter instance is not obvious, but note that even univariate probit estimates of  $\mathbf{B}$  would suffice to obtain consistent estimates of the  $\boldsymbol{\beta}_j$  parameters. For present purposes the method of estimation is not of particular concern so long as consistent estimates of  $\mathbf{B}$  and  $\mathbf{R}$  are available. In practice, if considerations of inference about marginal effects are relevant then the method of estimation may become relevant.

---

<sup>4</sup> This paper does not appeal to a common factor error structure for  $\boldsymbol{\varepsilon}$  in (2) although it may be that such an assumption would simplify estimation and, ultimately, computation of the marginal effects.

<sup>5</sup> Avery et al. actually discuss use of a GMM approach, but the limited-information idea can also be implemented straightforwardly using bivariate probit MLE.

## Applications

Why might such marginal effects be of interest in economic applications? In some contexts the sample or population averages of the marginals will be of interest in their own right for all or for some particular patterns of the  $k_j$ 's, i.e.

$$APE_p = \text{Avg}_{\mathbf{x}} \left( \frac{\partial \text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x})}{\partial \mathbf{x}} \right), \quad (3)$$

for some  $p$  or set of  $p$ 's in  $\mathbb{P}$ . In practice, a variety of situations arise where understanding how a  $\Delta \mathbf{x}$  intervention affects the entire pattern of multivariate outcomes (or, possibly, particular patterns of interest) is of central importance.<sup>6</sup>

Beyond this, consider an evaluation context involving a policy question where the focus is on how a change in some  $x_j$  (intervention, incentive, policy, etc.) affects expected utility through its impacts on the distribution of a set of outcomes  $\mathbf{y}$  over which welfare is defined. Let utility be given by  $V(y_1, \dots, y_m) = V(\mathbf{y})$ . Expected utility conditional on  $\mathbf{x}$  is then given by

$$E[V(\mathbf{y}) | \mathbf{x}] = \sum_{k_m=0}^1 \cdots \sum_{k_1=0}^1 \left\{ V(y_1 = k_1, \dots, y_m = k_m) \times \text{Prob}(y_1 = k_1, \dots, y_m = k_m | \mathbf{x}) \right\}. \quad (4)$$

Thus the change in expected utility arising from a change in  $\mathbf{x}$  is

---

<sup>6</sup> Two examples of work in progress are illustrative. The first is an examination of the role of socioeconomic status measures (e.g. parental schooling attainment) as correlates or determinants of patterns of five categories of disabilities in children ages 5-14 using data from the American Community Survey. The second is an analysis of determinants of the temporal persistence of high-quantile Medicare expenditures using data from the Medical Expenditure panel survey, in which persistence is characterized by an elderly individual's expenditure being always or largely in the upper quantiles (.75, .90) of the expenditure distribution in that age group over the survey waves. In each case consideration of how some or all of the orthant probabilities respond to one or more of the  $\mathbf{x}$ 's is of central concern.

$$\frac{\partial E[V(\mathbf{y})|\mathbf{x}]}{\partial \mathbf{x}} = \sum_{k_m=0}^1 \cdots \sum_{k_1=0}^1 \left\{ V(y_1 = k_1, \dots, y_m = k_m) \times \frac{\partial \text{Prob}(y_1 = k_1, \dots, y_m = k_m | \mathbf{x})}{\partial \mathbf{x}} \right\}. \quad (5)$$

As such one must know the full conditional joint probability structure and how it varies with  $\mathbf{x}$  to undertake welfare analysis of interventions in this context.

Quite generally, if one has available consistent estimates of the full conditional probability structure  $\text{Prob}(\mathbf{y} = \mathbf{k} | \mathbf{x})$  for all  $\mathbf{k}$  and of how that structure varies with  $\mathbf{x}$ , then it is possible using the approach described below to investigate a broad class of applied questions involving the role of varying  $\mathbf{x}$ 's on outcomes defined by  $\text{Prob}(\mathbf{y} = \mathbf{k} | \mathbf{x})$  as well as aggregates over or differences between such joint probabilities for different  $\mathbf{k}$ 's of interest as well as particular moment structures based thereon. Sections 4-6 of the paper pursue this idea in greater detail.

### ***Plan for the Paper***

The remainder of the paper is organized in nine short sections. Section 2 derives the results for arbitrary joint distributions. Section 3 presents the specific formulae for the multivariate probit model. Building on the results from section 3, section 4 derives the marginal effects of probabilities that are conditioned on subvectors of  $\mathbf{y}$ . Section 5 constructs a familiar count data model on the foundation of an MVP probability structure and derives several marginal effects relating to the count data structure, including those for that model's conditional mean, section 6 considers issues arising when using univariate models' marginals to represent those of underlying MVP structures, and section 7 extends the results of the previous sections to multivariate ordered probit models. Section 8 demonstrates that the marginal effects in multinomial probit models are essentially a special case of those in the MVP model. Section 9 reports on a simulation exercise comparing the

computational performance of the analytical results obtained here with results obtained using numerical differentiation based on simulated probabilities. Section 10 summarizes.

## 2. Results for Arbitrary Joint Distributions

The paper first establishes the main results on marginal effects for an arbitrary joint distribution and then proceeds in the next section to obtain the particular results for the MVP model. Assume  $(u_1, \dots, u_m)$  are continuously measured random variables with population joint distribution function  $F(u_1, \dots, u_m)$ ; normality is not assumed at this point. A standard result or definition is

$$\frac{\partial^m F(v_1, \dots, v_m)}{\partial v_1 \dots \partial v_m} = f(v_1, \dots, v_m) \quad (6)$$

where  $f(\dots)$  is the joint density and  $\mathbf{v} = [v_j]$  are specific points of evaluation of  $\mathbf{u} = [u_j]$ . Note that

(6) can also be written as:

$$f_j(v_j) \times \frac{\partial^{m-1} F(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m | v_j)}{\partial v_1 \dots \partial v_{j-1} \partial v_{j+1} \dots \partial v_m} = f(v_1, \dots, v_m), \text{ for any } j=1, \dots, m. \quad (7)$$

### Proposition 1

The partial derivative of  $F(v_1, \dots, v_m)$  with respect to  $v_j$  satisfies:

$$\frac{\partial F(\mathbf{v})}{\partial v_j} = f_j(v_j) \times F_{-j}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m | v_j), \quad j=1, \dots, m \quad (8)$$



An explicit general expression for  $\partial F(\mathbf{v})/\partial v_j$  is elusive in the mathematical statistics texts and econometrics literature the author has surveyed, although (8) is obviously trivially correct in the case of mutual independence. For an intuition, note that in the  $m=2$  case the partial derivative w.r.t.  $v_1$  of the function  $g(v_1, v_2) \equiv \partial F(v_1, v_2)/\partial v_2$  must give the joint density  $f(v_1, v_2)$ . One function  $g(v_1, v_2)$  satisfying this is  $g(v_1, v_2) = f_2(v_2) \times F(v_1|v_2)$ , which is of the form (8); this follows since

$$\frac{\partial f_2(v_2) \times F(v_1|v_2)}{\partial v_1} = f_2(v_2) \times \frac{\partial F(v_1|v_2)}{\partial v_1} = f_2(v_2) \times f(v_1|v_2) = f(v_1, v_2). \quad (9)$$

By recursion, this  $m=2$  result generalizes to cases with  $m>2$  by working backwards from the  $m$ -th cross partial derivative to the first. The form of the corresponding forward sequence of cross partials is suggested in the following result.

### Proposition 2

The general (forward) cumulative sequence of partial derivatives of  $F(\dots)$  is (differentiating w.o.l.o.g. in the order  $j=1,2,\dots,m$ ):

$$\frac{\partial F(\mathbf{v})}{\partial v_1} = f_1(v_1) \times F_{-1}(v_2, \dots, v_m | v_1),$$

$$\frac{\partial^r F(\mathbf{v})}{\partial v_1 \dots \partial v_r} = f_1(v_1) \times \left\{ \prod_{k=2}^r f(v_k | v_1, \dots, v_{k-1}) \right\} \times F_{-\{1, \dots, r\}}(v_{r+1}, \dots, v_m | v_1, \dots, v_r), \quad r=2, \dots, m-1, \quad (10)$$

and

$$\frac{\partial^m F(\mathbf{v})}{\partial v_1 \dots \partial v_m} = f(\mathbf{v})$$

The result in (10) is trivial in the case where the  $v_j$  are mutually independent (in which case all the conditionings vanish), but less so when the  $v_j$  are not mutually independent.

Alternatively (8) can be obtained directly using Leibniz's rule for differentiation of integrals whose limits depend on the variable of differentiation, as follows.

**Proposition 3**

Since

$$F(\mathbf{v}) = \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_1} f(\mathbf{u}) \, du_1 \cdots du_m, \quad (11)$$

then one can obtain  $\partial F(\mathbf{v})/\partial v_j$  by noting that  $v_j$  appears in this expression only once, as the upper limit of one integration, so that passing Leibniz's rule into the integral yields:

$$\begin{aligned} \frac{\partial}{\partial v_j} \left( \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_1} f(\mathbf{u}) \, du_1 \cdots du_m \right) &= \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_{j+1}} \int_{-\infty}^{v_{j-1}} \cdots \int_{-\infty}^{v_1} \left( \frac{\partial}{\partial v_j} \int_{-\infty}^{v_j} f(u_1, \dots, u_m) \, du_j \right) du_1 \cdots du_{j-1} du_{j+1} \cdots du_m \\ &= \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_{j+1}} \int_{-\infty}^{v_{j-1}} \cdots \int_{-\infty}^{v_1} \left( f(u_1, \dots, u_{j-1}, v_j, u_{j+1}, \dots, u_m) \right) du_1 \cdots du_{j-1} du_{j+1} \cdots du_m \\ &= \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_{j+1}} \int_{-\infty}^{v_{j-1}} \cdots \int_{-\infty}^{v_1} \left( f(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_m | v_j) \times f_j(v_j) \right) du_1 \cdots du_{j-1} du_{j+1} \cdots du_m \\ &= f_j(v_j) \times \int_{-\infty}^{v_m} \cdots \int_{-\infty}^{v_{j+1}} \int_{-\infty}^{v_{j-1}} \cdots \int_{-\infty}^{v_1} f(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_m | v_j) \, du_1 \cdots du_{j-1} du_{j+1} \cdots du_m \\ &= f_j(v_j) \times F_{-j}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m | v_j) \end{aligned} \quad (12)$$

Note that (12) is a restatement of (8).

Suppose  $F(\mathbf{u})$  is evaluated at  $\mathbf{u} = \mathbf{c}(\boldsymbol{\theta}) = [c_1(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta})]$ , where  $\boldsymbol{\theta}$  is a common parameter (scalar or vector) across the  $j=1, \dots, m$  margins of  $F(\dots)$ , and where all  $c_j(\boldsymbol{\theta})$  are differentiable in  $\boldsymbol{\theta}$ .

This gives

$$F(\mathbf{c}(\boldsymbol{\theta})) = F(c_1(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta})). \quad (13)$$

Using a standard chain rule for differentiation in conjunction with (8) or (12) gives:

$$\begin{aligned} \frac{\partial F(c_1(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \sum_{j=1}^m \left\{ \frac{\partial F(c_1(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta}))}{\partial c_j(\boldsymbol{\theta})} \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\} \\ &= \sum_{j=1}^m \left\{ f_j(c_j(\boldsymbol{\theta})) \times F_{-j}(c_1(\boldsymbol{\theta}), \dots, c_{j-1}(\boldsymbol{\theta}), c_{j+1}(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta})) | c_j(\boldsymbol{\theta}) \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\} \end{aligned} \quad (14)$$

### 3. Results for the Multivariate Probit Model

Recall that there are  $2^m$  possible outcome configurations. For each configuration  $\mathbf{k}_p$ ,  $p \in \mathbb{P}$ , one has a corresponding conditional outcome probability  $\text{Prob}(y_1 = k_{1p}, \dots, y_m = k_{mp} | \mathbf{x})$ . Let  $s_{jp} = 2k_{jp} - 1$  so that  $s_{jp} \in \{-1, +1\}$ , and define correspondingly the  $m \times m$  diagonal transformation matrixes  $\mathbf{T}_p = \text{diag}[s_{jp}]$ ,  $p=1, \dots, 2^m$ ,  $j=1, \dots, m$ . Also define for each  $p$  a transformation  $\mathbf{Q}_p = \mathbf{T}_p \mathbf{R} \mathbf{T}_p$  of the original covariance (i.e correlation) matrix  $\mathbf{R}$  so that  $\mathbf{Q}_p$  is of the form

$$\mathbf{Q}_p = \begin{bmatrix} 1 & s_{1p}s_{2p}\rho_{12} & \cdots & s_{1p}s_{mp}\rho_{1m} \\ s_{1p}s_{2p}\rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ s_{1p}s_{mp}\rho_{1m} & \cdots & & 1 \end{bmatrix} = \begin{bmatrix} 1 & \tau_{12p} & \cdots & \tau_{1mp} \\ \tau_{12p} & 1 & & \vdots \\ \vdots & & \ddots & \\ \tau_{1mp} & \cdots & & 1 \end{bmatrix}. \quad (15)$$

The conditional-on- $\mathbf{x}$  probability of any particular outcome configuration  $\mathbf{k}_p$  is thus given by

$$\text{Prob}(y_1 = k_{1p}, \dots, y_m = k_{mp} | \mathbf{x}) = \Phi_{\mathbf{Q}_p}(s_{1p} \mathbf{x} \boldsymbol{\beta}_1, \dots, s_{mp} \mathbf{x} \boldsymbol{\beta}_m) = \Phi_{\mathbf{Q}_p}(\alpha_{1p}, \dots, \alpha_{mp}), \quad (16)$$

where  $\Phi_{\mathbf{Q}}(\dots)$  denotes the cumulative of an  $\text{MVN}(\mathbf{0}, \mathbf{Q})$  distribution having corresponding density  $\phi_{\mathbf{Q}}(\dots)$  and  $\alpha_{jp}$  is shorthand for  $s_j \mathbf{x} \boldsymbol{\beta}_j$ .<sup>7</sup>

To obtain the specific results for the MVP's marginal effects it thus suffices to obtain the particular expressions corresponding to the second line in (14). Specifically  $f_j(c_j(\boldsymbol{\theta}))$  is simply a univariate  $N(0,1)$  density and  $F_{-j}(c_1(\boldsymbol{\theta}), \dots, c_{j-1}(\boldsymbol{\theta}), c_{j+1}(\boldsymbol{\theta}), \dots, c_m(\boldsymbol{\theta}) | c_j(\boldsymbol{\theta}))$  is the cumulative of a conditional (m-1)-variate multivariate normal distribution. The  $c_j(\boldsymbol{\theta})$  in (14) are equal to  $s_j \mathbf{x} \boldsymbol{\beta}_j$  in the MVP context, with  $\mathbf{x}$  playing the role of the "parameter" that is common across outcomes, so that  $dc_j(\boldsymbol{\theta})/d\boldsymbol{\theta}$  is  $d(s_j \mathbf{x} \boldsymbol{\beta}_j)/d\mathbf{x} = s_j \boldsymbol{\beta}_j$ .

Substituting into (14)  $\phi(\dots)$  for  $f(\dots)$ ,  $\Phi(\dots)$  for  $F(\dots)$ , and  $\alpha_{jp}$  for  $c_j(\boldsymbol{\theta})$  gives:

$$\begin{aligned} \frac{\partial \Phi_{\mathbf{Q}_p}(\alpha_{1p}, \dots, \alpha_{mp})}{\partial \mathbf{x}} &= \sum_{j=1}^m \left\{ \left( \frac{\partial \Phi_{\mathbf{Q}_p}(\alpha_{1p}, \dots, \alpha_{mp})}{\partial \alpha_{jp}} \right) \times \left( \frac{\partial \alpha_{jp}}{\partial \mathbf{x}} \right) \right\} \\ &= \sum_{j=1}^m \left\{ \left( \phi(\alpha_{jp}) \times \Phi_{\mathbf{Q}_p\{-j\}}(\alpha_{1p}, \dots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \dots, \alpha_{mp} | \alpha_{jp}) \right) \times (s_{jp} \boldsymbol{\beta}_j)^T \right\}. \end{aligned} \quad (17)$$

Given consistent estimates  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{Q}}$ , estimation of (17) is complicated only by evaluation of the expression  $\Phi_{\mathbf{Q}_p\{-j\}}(\alpha_{1p}, \dots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \dots, \alpha_{mp} | \alpha_{jp})$ . This result provides the basis of this calculation:

---

<sup>7</sup> Using the transformed matrixes  $\mathbf{Q}$  in place of the original correlation matrixes  $\mathbf{R}$  provides for streamlined notation and computation since for each configuration  $p$  the outcome orthant probability can be described by a joint cumulative rather than by a computationally messy mix of marginal cumulatives and survivor functions. In effect this amounts to a linear change-of-variables operation on  $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_m]$  of the form  $\mathbf{T}_p \boldsymbol{\varepsilon}$  which becomes the effective error structure of model at each  $p$ ; this transformation works due to the symmetry of  $\boldsymbol{\varepsilon}$  around the origin.

**Result: Joint Conditional Distribution of an MVN-Variate, Adapted from Rao, 1973 (8a.2.11)**

Suppose  $\mathbf{z} = [z_1, \dots, z_m] \sim \text{MVN}(\mathbf{0}, \mathbf{\Omega})$  and partition  $\mathbf{\Omega}$  as  $\begin{bmatrix} \omega_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix}$  where  $\omega_{11}$  is scalar.

Then  $\mathbf{z}_{-1} = [z_2, \dots, z_m]$  conditional on  $z_1$  has an  $(m-1)$ -variate  $\text{MVN}(\mathbf{\Omega}_{21}\omega_{11}^{-1}z_1, (\mathbf{\Omega}_{22} - \omega_{11}^{-1}\mathbf{\Omega}_{21}\mathbf{\Omega}_{12}))$  distribution. (This result generalizes in an obvious way to  $\mathbf{z}_{-j} = [z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_m]$ ,  $j=2, \dots, m$ , by defining different partitions of  $\mathbf{\Omega}$ .)

In the particular case of interest here,  $\mathbf{\Omega} = \mathbf{Q}_p$  so that  $\omega_{11} = 1$ . It then follows that the particular form of the joint conditional distribution is

$$\mathbf{z}_{-1} | z_1 \sim \text{MVN} \left( \begin{bmatrix} z_1 \tau_{12p} \\ \vdots \\ z_1 \tau_{1mp} \end{bmatrix}, \begin{bmatrix} 1 - \tau_{12p}^2 & \tau_{23p} - \tau_{12p} \tau_{13p} & \dots & \tau_{2mp} - \tau_{12p} \tau_{1mp} \\ \tau_{23p} - \tau_{12p} \tau_{13p} & 1 - \tau_{13p}^2 & & \vdots \\ \vdots & & \ddots & \\ \tau_{2mp} - \tau_{12p} \tau_{1mp} & \dots & & 1 - \tau_{1mp}^2 \end{bmatrix} \right), \quad (18)$$

again with obvious generalization to the distributions of  $\mathbf{z}_{-j} | z_j$ ,  $j=2, \dots, m$ .

To compute  $\Phi_{\mathbf{Q}_p\{-j\}}(\alpha_{1p}, \dots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \dots, \alpha_{mp} | \alpha_{jp})$ , it is useful to define the  $(m-1)$ -vector of

differences from (conditional) means

$$\Delta_{-j,p} = \left[ (\alpha_{1p} - \alpha_{jp} \tau_{1jp}), \dots, (\alpha_{(j-1)p} - \alpha_{jp} \tau_{(j-1)jp}), (\alpha_{(j+1)p} - \alpha_{jp} \tau_{(j+1)jp}), \dots, (\alpha_{mp} - \alpha_{jp} \tau_{mjp}) \right]^T, \quad (19)$$

and an  $(m-1) \times (m-1)$  diagonal transformation matrix  $\mathbf{H}_{jp} = \text{diag}_{k \neq j} \left[ \left( \sqrt{1 - \tau_{jkp}^2} \right)^{-1} \right]$ . Let  $\mathbf{L}_{jp} = \mathbf{H}_{jp} \Delta_{-j,p}$  be

the corresponding  $(m-1)$ -vector of normalized differences. Then

$\Phi_{\mathbf{Q}_p\{-j\}}(\alpha_{1p}, \dots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \dots, \alpha_{mp} | \alpha_{jp})$  can be computed by referring  $\mathbf{L}_{jp}$  to  $\Phi_{\mathbf{z}, \mathbf{\Sigma}}(\cdot)$ , which is the

cumulative of an  $(m-1)$ -variate semi-standardized MVN( $\mathbf{0}, \mathbf{\Sigma}$ ) distribution in which the off-diagonals of  $\mathbf{\Sigma}$  may be nonzero. In this instance  $\mathbf{\Sigma}$  is the variance-covariance matrix of  $\mathbf{L}_{jp}$  which is in correlation matrix form having typical off-diagonal  $(r,c)$  element  $(\tau_{rcp} - \tau_{jrp} \tau_{jcp}) / \sqrt{(1 - \tau_{jrp}^2)(1 - \tau_{jcp}^2)}$ .

Let this matrix be denoted  $\mathbf{V}_{jp}$ .

Assembling all the components, the MVP model marginal effects are given by

$$\frac{\partial \text{Prob}(y_1 = k_{1p}, \dots, y_m = k_{mp} | \mathbf{x})}{\partial \mathbf{x}} = \sum_{j=1}^m \left\{ \phi(\alpha_{jp}) \times \Phi_{z, \mathbf{V}_{jp}}(\mathbf{L}_{jp}) \times (s_{jp} \boldsymbol{\beta}_j)^\top \right\}. \quad (20)$$

Note that for  $m=2$  this is the result presented by Greene, 1998 (unnumbered equation displayed in the middle of p. 298). Greene's result in his notation,

$$\begin{aligned} \frac{\partial \text{BVN}\Phi(\boldsymbol{\beta}'\mathbf{x}_1 + \gamma, \boldsymbol{\alpha}'\mathbf{x}_2, \rho)}{\partial z_k} &= \left\{ \phi(\boldsymbol{\beta}'\mathbf{x}_1 + \gamma) \Phi\left[\frac{\boldsymbol{\alpha}'\mathbf{x}_2 - \rho(\boldsymbol{\beta}'\mathbf{x}_1 + \gamma)}{\sqrt{1 - \rho^2}}\right] \right\} \beta_z \\ &\quad + \left\{ \phi(\boldsymbol{\alpha}'\mathbf{x}_2) \Phi\left[\frac{(\boldsymbol{\beta}'\mathbf{x}_1 + \gamma) - \rho(\boldsymbol{\alpha}'\mathbf{x}_2)}{\sqrt{1 - \rho^2}}\right] \right\} \alpha_z \end{aligned} \quad (21)$$

translates in the present notation (and for the  $k_1=k_2=1$  case of interest to Greene) into

$$\frac{\partial \Phi_{\mathbf{Q}_{p^*}}(\alpha_{1p^*}, \alpha_{2p^*})}{\partial \mathbf{x}} = \sum_{j=1}^2 \left\{ \phi(\alpha_{jp^*}) \times \Phi_{z, \mathbf{V}_{jp^*}}\left(\frac{\alpha_{(3-j)p^*} - \alpha_{jp^*} \tau_{(3-j)jp^*}}{\sqrt{1 - \tau_{12p^*}^2}}\right) \times (s_{jp^*} \boldsymbol{\beta}_j)^\top \right\}, \quad (22)$$

where  $p^*$  is the element of  $\mathbb{P}$  corresponding to the orthant defined by  $k_1=k_2=1$ .

Finally, from (17) it is noteworthy for computational purposes that it is only an  $(m-1)$ -dimension cumulative normal that must be evaluated to obtain the marginal effects. This may reduce computational time and effort as compared with fully numerical methods that operate on the  $m$ -dimension cumulative normal. Even though numerical methods will typically be required in conjunction with the analytical formulae presented here, the reduction in dimensionality should facilitate computation; this is particularly obvious in the  $m=3$  case in which canned functions like

Stata's *binormal(...)* bivariate normal cumulative can be used in lieu of simulation procedures.

#### 4. Marginal Effects of Probabilities Conditional on Subvectors of $\mathbf{y}$

In the context of bivariate probit models, Greene, 1996, suggests that consideration of the marginal effects of  $\mathbf{x}$  on conditional-on- $\mathbf{y}$  probabilities, e.g.  $\partial \text{Prob}(\mathbf{y}_1 | \mathbf{y}_2, \mathbf{x}) / \partial \mathbf{x}$ , may be of interest in some instances.<sup>8</sup> Using the approach developed above, this idea can be extended straightforwardly to the general multivariate probit context as follows.

Partition the outcome vector  $\mathbf{y}$  as  $[\mathbf{y}_a, \mathbf{y}_b]$  and correspondingly partition  $\mathbf{k}_p$  as  $[\mathbf{k}_{p,a}, \mathbf{k}_{p,b}]$ , where  $\mathbf{y}_a$  and  $\mathbf{k}_{p,a}$  are  $c$ -vectors and  $\mathbf{y}_b$  and  $\mathbf{k}_{p,b}$  are  $(m-c)$ -vectors. Suppose interest is in the quantities  $\text{Prob}(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x})$  and  $\partial \text{Prob}(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}) / \partial \mathbf{x}$ . Note that

$$\text{Prob}(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}) = \frac{\text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x})}{\text{Prob}(\mathbf{y}_b = \mathbf{k}_{p,b} | \mathbf{x})} = \frac{\Phi_{\mathbf{Q}_p}(\alpha_{1p}, \dots, \alpha_{mp})}{\Phi_{\mathbf{Q}_{p,b}}(\alpha_{(c+1)p}, \dots, \alpha_{mp})} \quad (23)$$

where the matrix  $\mathbf{Q}_{p,b}$  is defined in an obvious way as a submatrix of  $\mathbf{Q}_p$ . Using the quotient rule for differentiation,

$$\frac{\partial \text{Prob}(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x})}{\partial \mathbf{x}} = \frac{\text{Prob}(\mathbf{y}_b = \mathbf{k}_{p,b} | \mathbf{x}) \times \left( \frac{\partial \text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x})}{\partial \mathbf{x}} \right) - \text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x}) \times \left( \frac{\partial \text{Prob}(\mathbf{y}_b = \mathbf{k}_{p,b} | \mathbf{x})}{\partial \mathbf{x}} \right)}{\text{Prob}(\mathbf{y}_b = \mathbf{k}_{p,b} | \mathbf{x})^2} \quad (24)$$

---

<sup>8</sup> In applied studies an explicit formulation of the model of interest as  $\text{Prob}(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x})$  is often absent, and this conditional probability may or may not be the parameter whose marginal effects are of particular concern to the analyst. See Greene, 1996, for conceptual discussion of fundamentals. Bhattacharya et al., 2006, provide a broad assessment of such models in the context of treatment effect estimation, illustrating their approach in a model of mortality outcomes; Fichera and Sutton, 2011, present an interesting related application to smoking cessation outcomes.

The component partial derivatives in the numerator of the rhs expression are simply the marginal effects described above for the multivariate outcomes  $\mathbf{y}$  and  $\mathbf{y}_b$ , respectively.

## 5. Count Data Models Based on MVP Probability Structures

Empirical analysis sometimes encounters applications where the scalar sum  $s$  of the vector  $\mathbf{y}$ ,  $s = \sum_{j=1}^m y_j = \mathbf{1}^T \mathbf{y}$ , is the outcome measure of interest.<sup>9</sup> The substantive economic, psychometric, or biometric<sup>10</sup> interpretation of  $s$  notwithstanding,<sup>11</sup> such outcomes are numerically well defined and obviously inherit their statistical properties from those of  $\mathbf{y}$ .

Define  $\mathbb{P}_n = \{p \in \mathbb{P} \mid \mathbf{1}^T \mathbf{k}_p = n\}$ ,  $n=0, \dots, m$ , and consider the count data probability model defined

by

$$\text{Prob}(s = n \mid \mathbf{x}) = \sum_{p \in \mathbb{P}_n} \text{Prob}(\mathbf{y} = \mathbf{k}_p \mid \mathbf{x}), \quad n=0, \dots, m. \quad (25)$$

In some instances, analysis proceeds by regression of the outcome measures  $s$  defined thusly on  $\mathbf{x}$

---

<sup>9</sup> See Dor et al., 2006, for discussion in the context of health outcome measures. More generally one might consider weighted sums  $s = \sum_{j=1}^m w_j y_j = \mathbf{w}^T \mathbf{y}$ ; an example of one such index is the Social Readjustment Rating Scale (Holmes and Rahe, 1967). The discussion that follows should apply equally appropriately for such weighted sums.

<sup>10</sup> A prominent biometric example is that of allostatic load measures (e.g. Seeman et al., 2001) in which binary threshold (quantile) exceedances for each of  $m$  continuous biomarkers are summed to arrive at the comprehensive allostatic load measure. In the measure devised by Seeman et al.,  $m=10$  and the threshold quantiles are specified to be .75 (for biomarkers for which higher levels are harmful, e.g. systolic blood pressure) or .25 (for biomarkers for which lower levels are harmful, e.g. HDL cholesterol). Such additive measures arise outside health contexts as well. For instance, the highly publicized ratings of U.S. Members of Congress by organizations like the League of Conservation Voter and the American Conservative Union are essentially of this nature; asset holdings and composition are also sometimes measured in such a manner (see McKenzie, 2005, for discussion).

<sup>11</sup> Note that even the likert-scale or ordinal properties of such measures may be questionable.



using linear or nonlinear regression,<sup>12</sup> presumably with the goal of recovering an estimate of the conditional mean  $E[s|\mathbf{x}]$  and the marginal effects therein implied. However, when the  $\mathbf{y}$  that beget  $s$  arise according to (2), such approaches fail to respect the underlying structures in (2). As such, it is not clear how to relate an estimate  $\hat{E}[s|\mathbf{x}]$  so obtained to the underlying model structure. Whether or not such considerations are empirically important depends on circumstances, but in any event a linear regression model for  $E[s|\mathbf{x}]$  is unlikely to be an appropriate specification.

If estimation of  $E[s|\mathbf{x}]$  and its marginal effects are of central interest one can, however, specify a proper conditional mean model that respects the underlying probability structure of  $\mathbf{y}$  in (2) and whose marginal effects  $\partial E[s|\mathbf{x}]/\partial \mathbf{x}$  consequently also respect that structure. Moreover, such marginal effects can be computed using exactly the same apparatus as described in section 3, as follows.

Since  $s = \sum_{j=1}^m y_j = \mathbf{1}^T \mathbf{y}$ , (25) implies that

$$E[s|\mathbf{x}] = \sum_{n=0}^m \left( n \times \sum_{\mathbf{p} \in \mathbb{P}_n} \text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x}) \right). \quad (26)$$

Of course since  $s = \sum_{j=1}^m y_j$  then it also holds that

$$E[s|\mathbf{x}] = \sum_{j=1}^m E[y_j|\mathbf{x}] = \sum_{j=1}^m \text{Prob}(y_j = 1|\mathbf{x}), \quad (27)$$

so an equivalent and in some instances more direct representation of the conditional mean is simply as the sum of the  $m$  univariate probit marginals. If (2) holds, then any functional form representation of  $E[s|\mathbf{x}]$  other than (26) or (27) is a misspecification. It follows from (26) that

---

<sup>12</sup> See Evans et al., 2007, for an example in which allostatic load measures of the sort defined in the footnote 10 are the outcomes of interest.

$$\frac{\partial E[s|\mathbf{x}]}{\partial \mathbf{x}} = \sum_{n=0}^m \left( n \times \sum_{\mathbf{p} \in \mathbb{P}_n} \frac{\partial \text{Prob}(\mathbf{y} = \mathbf{k}_p | \mathbf{x})}{\partial \mathbf{x}} \right). \quad (28)$$

The partial derivatives on the rhs of (28) are simply the marginal effects obtained in section 3.

## 6. Univariate Representations of Multivariate Probit Outcomes and Counts

In applied work, multivariate discrete outcomes like those under consideration here may be summarized by univariate discrete outcomes that might be defined quite generally according to

$$v = \begin{cases} 1 & \text{if } \mathbf{y} = \mathbf{k}_p \text{ for } p \in \mathbb{Q} \\ 0 & \text{else} \end{cases}, \quad (29)$$

where  $\mathbb{Q} \subset \mathbb{P}$  is an index set defining a subset of outcome patterns whose entirety is considered for purposes of such analysis to be a "success". Taking such an approach one step further, the analyst may specify that the "univariate" process determining  $v$  is given by a probit model, so that

$$\text{Prob}(v = 1 | \mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\theta}), \quad (30)$$

where  $\Phi(\cdot)$  here denotes a univariate cumulative standard normal distribution.

While such dimension-reduction or aggregation approaches may be informative for some purposes, it should be emphasized that they fail fundamentally to respect the properties of the underlying probability structure of the multivariate model in (2). That is, the parameters  $\boldsymbol{\theta}$  of a standard univariate probit mapping like  $v = 1(\mathbf{x}\boldsymbol{\theta} + \nu > 0)$  implied in (30) cannot readily be interpreted in terms of the parameters  $(\mathbf{B}, \mathbf{R})$  in (2); neither is there any straightforward relationship between the marginal effect of  $\mathbf{x}$  in (30) and the marginal effect defined by summing (20) over  $p \in \mathbb{Q}$ . For instance, one obvious version of such a univariate mapping rule assigns  $v=1$  if  $s \geq n^*$ ,

where  $n^* \leq m$ , i.e. "large" counts are "successes".<sup>13,14</sup> Given (2) and the count data model defined in (25), the proper marginal effect corresponding to (29) is given by defining  $\mathbb{Q} = \{p \in \mathbb{P} \mid \mathbf{1}^T \mathbf{y}_p \geq n^*\}$  and then summing (20) across all  $p \in \mathbb{Q}$ . How such marginal effect estimates compare empirically to those derived from estimates of a model like (30) is an interesting and open question.

## 7. Multivariate Ordered Probit Models

Marginal effects for multivariate ordered probit model (see Greene and Hensher, 2010, chapter 10) are straightforward to compute using essentially the same algebra as derived in section 3 for the multivariate binary probit model.<sup>15</sup> To begin, assume that (2) holds except that for

---

<sup>13</sup> A variety of medical diagnoses are rendered in such a manner. Given  $m$  binary indicators of patient conditions or attributes, a disorder is deemed to be present if  $s \geq n^*$ . One illustrative example is metabolic syndrome:

The metabolic syndrome is a constellation of interrelated risk factors of metabolic origin -- *metabolic risk factors* -- that appear to directly promote the development of atherosclerotic cardiovascular disease (ASCVD). Patients with the metabolic syndrome also are at increased risk for developing type 2 diabetes mellitus. (Grundy et al., 2005).

Metabolic syndrome is diagnosed when at least three of the following five measures satisfy the indicated threshold criteria: waist circumference ( $\geq 35$ " (females),  $\geq 40$ " (males)); triglycerides ( $\geq 150$  mg/dL, or drug treatment for elevated triglycerides); HDL-C ( $< 50$  mg/dL (females),  $< 40$  mg/dL (males), or drug treatment for reduced HDL-C); blood pressure ( $\geq 130$  mm Hg systolic or  $\geq 85$  mm Hg diastolic blood pressure, or antihypertensive drug treatment in patients with hypertension history); and fasting glucose ( $\geq 100$  mg/dL, or drug treatment for elevated glucose). In this paper's notation, therefore,  $m=5$  and  $n^*=3$ . See Behncke, 2011, O'Brien et al., 2006, and Orchard et al., 2005 for examples of economic and clinical studies in which such metabolic syndrome outcomes are analyzed. It should be noted that a variety of psychiatric and substance abuse disorders (e.g. ADHD, alcohol abuse, etc., based on DSM-IV diagnoses), as well as other medical disorders, are diagnosed in analogous fashion.

<sup>14</sup> Geronimus et al., 2006, use such an approach with allostatic load outcome measures as described earlier dichotomized (0-3 vs. 4-10) and analyzed accordingly in regression contexts.

<sup>15</sup> Estimation of the  $m$ -variate multivariate ordered probit model may be challenging. However, consistent estimates of the parameters  $\mathbf{B}$ ,  $\mathbf{R}$ , and  $\mathbf{M}$  can be obtained by estimating bivariate ordered

(continued)

$j=1,\dots,m$  each observed  $y_j$  assumes one of  $g$  possible values,  $y_j \in \{0,\dots,(g-1)\}$ <sup>16</sup> with the mapping given by:

$$y_j = \sum_{c=1}^g (c-1) \times 1 \left[ \left( \mu_{(c-1)j} - \mathbf{x}\boldsymbol{\beta}_j \right) < u_j < \left( \mu_{cj} - \mathbf{x}\boldsymbol{\beta}_j \right) \right], \quad (31)$$

with  $-\infty = \mu_{0j} < \mu_{1j} < \dots < \mu_{gj} = +\infty$ . As such there are  $(g-1)$  free threshold or cutoff parameters

$\{\mu_{1j}, \dots, \mu_{(g-1)j}\}$  for each  $j$ . For shorthand, let  $\boldsymbol{\mu}_j = [\mu_{1j}, \dots, \mu_{(g-1)j}]^T$ ,  $\mathbf{M} = [\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m]$ , and  $\omega_{cj} = \mu_{cj} - \mathbf{x}\boldsymbol{\beta}_j$

for all  $j$ . It follows that

$$\text{Prob}(y_j = (c-1) | \mathbf{x}) = \int_{\omega_{(c-1)j}}^{\omega_{cj}} \phi(u_j) du_j = \int_{-\infty}^{\omega_{cj}} \phi(u_j) du_j - \int_{-\infty}^{\omega_{(c-1)j}} \phi(u_j) du_j, \quad c=1,\dots,g, \quad (32)$$

where  $\phi(\dots)$  is a univariate  $N(0,1)$  density.

Analogous to the definition of  $\mathbf{K}$ , define the  $m \times g^m$  matrix  $\mathbf{C}$  whose columns (arranged arbitrarily) are the  $g^m$  possible outcome configurations  $\mathbf{c}$ , and let  $\mathbb{C}$  be a  $g^m$ -element set indexing columns of  $\mathbf{C}$  having typical indexing element  $r$ , so that  $\mathbf{c}_r = \mathbf{C}_r$  will denote a particular ( $r$ -th) outcome configuration. Thus

$$\text{Prob}(\mathbf{y} = \mathbf{c}_r | \mathbf{x}) = \int_{\omega_{(c_{r1})}}^{\omega_{c_{r1}}} \dots \int_{\omega_{(c_{rm})}}^{\omega_{c_{rm}}} \phi_{\mathbf{R}}(u_1, \dots, u_m) du_1 \dots du_m, \quad r \in \mathbb{C}. \quad (33)$$

Note that (33) will be a sum of positively and negatively signed multivariate normal cdfs (including zeros and ones at lower and upper integration limits), so that marginal effects at any  $\mathbf{c}_r$  will simply

(continued)

probit models for all  $j \neq k$  outcome pairs  $(y_j, y_k)$ , analogous to the discussion in footnote 5. Stata's *bioprobit* is one such estimation procedure.

<sup>16</sup> Allowing the  $y_j$  to have assume different numbers of outcomes is straightforward; the assumption of equal numbers of categories across  $j$  is made solely to keep notation from becoming unwieldy.

be the corresponding sum of the components' marginals, taking care to sign the  $\partial\omega_{c_j}/\partial\mathbf{x}$  correctly.

For example, for the trivariate ( $m=3$ ) ordered probit model the integral in (33) is

$$\int_{\omega_{(c_3-1)}}^{\omega_{c_3}} \int_{\omega_{(c_2-1)}}^{\omega_{c_2}} \int_{\omega_{(c_1-1)}}^{\omega_{c_1}} \phi_{\mathbf{R}}(u_1, \dots, u_m) du_1 du_2 du_3 = \Phi_{\mathbf{R}}(\omega_{c_1}, \omega_{c_2}, \omega_{c_3}) - \Phi_{\mathbf{R}}(\omega_{c_1}, \omega_{c_2}, \omega_{(c_3-1)}) - \Phi_{\mathbf{R}}(\omega_{c_1}, \omega_{(c_2-1)}, \omega_{c_3}) - \Phi_{\mathbf{R}}(\omega_{(c_1-1)}, \omega_{c_2}, \omega_{c_3}) + \Phi_{\mathbf{R}}(\omega_{c_1}, \omega_{(c_2-1)}, \omega_{(c_3-1)}) + \Phi_{\mathbf{R}}(\omega_{(c_1-1)}, \omega_{c_2}, \omega_{(c_3-1)}) + \Phi_{\mathbf{R}}(\omega_{(c_1-1)}, \omega_{(c_2-1)}, \omega_{c_3}) - \Phi_{\mathbf{R}}(\omega_{(c_1-1)}, \omega_{(c_2-1)}, \omega_{(c_3-1)}) \quad (34)$$

Analogous to the count data model (25) that arises in the multivariate binary probit model, one can define a count model corresponding to the multivariate ordered probit model.<sup>17</sup> As before

let  $s = \sum_{j=1}^m y_j = \mathbf{1}^T \mathbf{y}$ , define  $\mathbb{C}_n = \{r \in \mathbb{C} \mid \mathbf{1}^T \mathbf{c}_r = n\}$ ,  $n=0, \dots, (g-1) \times m$ , and consider the count data probability model:

$$\text{Prob}(s = n \mid \mathbf{x}) = \sum_{r \in \mathbb{C}_n} \text{Prob}(\mathbf{y} = \mathbf{c}_r \mid \mathbf{x}), \quad n=0, \dots, (g-1) \times m \quad (35)$$

that has corresponding conditional mean

$$E[s \mid \mathbf{x}] = \sum_{n=0}^{(g-1) \times m} \left( n \times \sum_{r \in \mathbb{C}_n} \text{Prob}(\mathbf{y} = \mathbf{c}_r \mid \mathbf{x}) \right). \quad (36)$$

<sup>17</sup> For example, the CES-D Depression Scale (Radloff, 1977) is based on a data structure that might plausibly be described by a multivariate ordered probit model. For a one-week reference period ("During the past week..."), there are 20 questionnaire items ("I was bothered by things that usually don't bother me", "I did not feel like eating; my appetite was poor", etc.) and for each item a plausibly ordered intensity or frequency dimension ("Rarely or none of the time (less than 1 day)", "Some or a little of the time (1-2 days)", "Occasionally or a moderate amount of time (3-4 days)", and "Most or all of the time (5-7 days)"). Scoring of the CES-D is based on assigning values of 0, 1, 2, or 3 to correspond to higher frequencies for each item, with the overall score or count  $s \in \{0, \dots, 60\}$  being derived as the simple sum of the item scores. In this case there are  $4^{20} = 1,099,511,627,776$  possible outcome patterns, so proper computation of the partial effects is not feasible. Shorter versions of the CES-D based on as few as four items have been proposed (e.g. Melchior et al., 1993); computation of marginal effects is more reasonable in such instances.

Partial effects for the count model outcome probabilities or the conditional mean follow as before, mutatis mutandis. However, it should be noted that as  $g$  and/or  $m$  increase, the computational burden -- in terms of the number of multivariate normal cdf evaluations required -- increases rapidly for any given  $s$  or combinations of  $s$ .<sup>18</sup>

## 8. Marginal Effects in Multinomial Probit Models

It turns out that the marginal effects with respect to the conditioning covariates in a multinomial probit model are special cases of those described above in (20) for the multivariate probit model.<sup>19</sup> This can be shown as follows.

A standard setup for a multinomial probit model is to assume that the value that individual  $i$  attaches to choice  $q$  is given by a random utility model

$$U_{iq} = \mathbf{w}_{iq} \boldsymbol{\pi} + \mathbf{z}_i \boldsymbol{\gamma}_q + \eta_{iq}, \quad q=1, \dots, m, \quad (37)$$

wherein the  $\mathbf{w}_{iq}$  are attributes of choice  $q$  faced by individual  $i$  (e.g. the unit price of acquiring choice  $q$ ), the  $\mathbf{z}_i$  are attribute-invariant characteristics of  $i$  (e.g. income),  $\boldsymbol{\pi}$  and  $\boldsymbol{\gamma}_i$  are corresponding parameters, and  $\eta_{iq}$  is an unobservable contributor to utility distributed jointly normally with the

---

<sup>18</sup> Given the number of ordered categories,  $g$ , for each outcome and the number of outcomes,  $m$ , with  $y_j \in \{0, \dots, (g-1)\}$  for all  $j=1, \dots, m$ , the number of combinations yielding the sum  $s \in \{0, \dots, m \times (g-1)\}$  is the  $s$ -th multinomial coefficient from the expansion of  $\left(\sum_{t=0}^{g-1} x^t\right)^m$ . For example, for  $g=3, 4, 5$  and  $m=2, 3, 4, 5$  the numbers of combinations yielding the indicated values of  $s$  are displayed in table 1. See <http://oeis.org>, entries A027907, A008287, and A035343. Recursions are straightforward; see <http://dlmf.nist.gov/26.4>.

<sup>19</sup> Wooldridge, 2010, p. 649, notes: "Theoretically, the multinomial probit model is attractive, but it has some practical limitations. The response probabilities are very complicated, involving a  $(J+1)$ -dimensional integral. This complexity ... makes it difficult to obtain the partial effects on the response probabilities..." See Deb et al., 1996, for an application of MNP to insurance coverage choice.

other (m-1)  $\eta_{ij}$ . Only pairwise differences in utility are relevant, so define

$$\delta_{ijq} \equiv U_{ij} - U_{iq} = (\mathbf{w}_{ij} - \mathbf{w}_{iq})\boldsymbol{\pi} + \mathbf{z}_i(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_q) + (\eta_{ij} - \eta_{iq}) \quad (38)$$

or, using obvious notational shorthand,

$$\begin{aligned} \delta_{ijq} &= \mathbf{W}_{ijq}\boldsymbol{\pi} + \mathbf{z}_i\boldsymbol{\Gamma}_{jq} + v_{ijq} \\ &= -\boldsymbol{\lambda}_{ijq} + v_{ijq}, \end{aligned} \quad (39)$$

so that  $\boldsymbol{\lambda}_{ijq} = (\mathbf{w}_{iq} - \mathbf{w}_{ij})\boldsymbol{\pi} + \mathbf{z}_i(\boldsymbol{\gamma}_q - \boldsymbol{\gamma}_j)$ . The residuals  $v_{ijq}$  in (39) will have an (m-1)-variate normal

distribution whose particular parameterization is discussed below. The behavioral model typically

assumed is that subject i selects the outcome having the largest  $U_{ij}$  so that

$$\text{Prob}(i \text{ selects } q | \mathbf{W}_i, \mathbf{z}_i) = \text{Prob}\left(\left(v_{i1q} < \boldsymbol{\lambda}_{i1q}\right), \dots, \left(v_{i(q-1)q} < \boldsymbol{\lambda}_{i(q-1)q}\right), \left(v_{i(q+1)q} < \boldsymbol{\lambda}_{i(q+1)q}\right), \dots, \left(v_{imq} < \boldsymbol{\lambda}_{imq}\right) | \mathbf{W}_i, \mathbf{z}_i\right), \quad (40)$$

where the conditioning  $\mathbf{W}_i$  denotes the entire collection of the  $\mathbf{W}_{ijq}$ .

Assuming for the moment that the  $v_{ijq}$  are iid across i and follow an (m-1)-dimension  $N(\mathbf{0}, \boldsymbol{\Psi})$  distribution wherein  $\boldsymbol{\Psi}$  is in correlation matrix form, then the probability expression in (40)

corresponds to an (m-1)-dimension version of the expression (16) in which all  $s_{jp}=1$  and in which the

$\mathbf{x}\boldsymbol{\beta}_j$  are replaced by  $\boldsymbol{\lambda}_{ijq}$ . As such the basic form of the relevant marginal effects

$\partial \text{Prob}(i \text{ selects } q | \mathbf{W}_i, \mathbf{z}_i) / \partial \{\mathbf{w}_{iq}, \mathbf{w}_{ij}, \mathbf{z}_i\}$  corresponds to (20), mutatis mutandis.<sup>20</sup>

---

<sup>20</sup> One computational issue should be noted at this juncture. Normalization of the parameters in a multinomial probit model is in general a complicated matter (StataCorp, 2007, under *asmprobit*; Cameron and Trivedi, 2005, Section 15.8; Monfardini and Santos Silva, 2008). Normalizing  $\boldsymbol{\Psi}$  to have a correlation matrix structure (ones on the diagonal; free correlation parameters elsewhere) is one admissible possibility. However, if the normalization used results in a structure for  $\boldsymbol{\Psi}$  other than one having a correlation matrix structure, then some additional computations may be required to be able to use formulae based on (20); the relevant considerations arise in the discussion  
(continued)

In the case of this general multinomial probit model, there are three distinct partial effects to be considered: own-attribute covariates ( $\mathbf{w}_{iq}$ , e.g. own-price); cross-attribute covariates ( $\mathbf{w}_{ij}$ , e.g. cross-price); and attribute-invariant covariates ( $\mathbf{z}_i$ , e.g. income). These covariates feature differently across the various upper integration limits in (40). Thus, the terms  $(s_{jp}\boldsymbol{\beta}_j)^\top$  that appear in (20) are replaced as follows:

(a) Own-attribute marginal effects,  $\partial\text{Prob}(i \text{ selects } q | \mathbf{w}_i, \mathbf{z}_i) / \partial \mathbf{w}_{iq}$ : Replace  $(s_{jp}\boldsymbol{\beta}_j)^\top$  with  $\boldsymbol{\pi}^\top$  in each summand,  $j=1, \dots, (j-1), (j+1), \dots, m$ .

(b) Cross-attribute marginal effects,  $\partial\text{Prob}(i \text{ selects } q | \mathbf{w}_i, \mathbf{z}_i) / \partial \mathbf{w}_{ik}$ ,  $k=1, \dots, (j-1), (j+1), \dots, m$ : Replace  $(s_{jp}\boldsymbol{\beta}_j)^\top$  with  $-\boldsymbol{\pi}^\top$  only in the  $k$ -th summand; set equal to zero in the other summands.

(c) Attribute-invariant marginal effects,  $\partial\text{Prob}(i \text{ selects } q | \mathbf{w}_i, \mathbf{z}_i) / \partial \mathbf{z}_i$ : Replace  $(s_{jp}\boldsymbol{\beta}_j)^\top$  with  $(\boldsymbol{\gamma}_q - \boldsymbol{\gamma}_j)^\top$  in each summand,  $j=1, \dots, (j-1), (j+1), \dots, m$ .

## 9. Numerical Results

This section reports on a small simulation exercise designed to compare computational performance of marginal effects using the analytical formulae derived here against numerical probability derivatives that can be obtained using the GHK simulator in Stata's Mata programming language (see Hajivassiliou et al., 1996, Gates, 2006). (This is the *ghk(...)* procedure in Mata, Stata

---

(continued)  
appearing between equations (19) and (20) above.



v.10.)  $ghk(\dots)$  is used here both to compute the  $(m-1)$ -dimension cumulative in Result 6 ( $ghk(S,x,V)$ ) as well as to simulate the marginal effects based on fully numerical methods appealing to the  $m$ -dimension cumulative ( $ghk(S,x,V,d1,d2)$ ).

Use of  $ghk(\dots)$  requires specification of the number of simulation points ( $S$ ). To assess how this choice affects computational performance, three sets of results are obtained corresponding to  $S=100$ ,  $S=1,000$ , and  $S=10,000$ . Computational time and accuracy should both increase in  $S$ . Results are obtained here for  $m \in \{3,4,8\}$  using a simple covariate specification where  $x$  is a scalar equal to one. For each  $m$ , one set of illustrative values is chosen for  $\mathbf{B}$  and  $\mathbf{R}$ ; these are detailed in Table 2.

The results of the simulation exercise are presented in Tables 3-5.<sup>21</sup> For each  $S$ , the tables show the simulated orthant probabilities as well as the analytical and the fully numerical marginal effects. While obvious, it is useful to note at this point that

$$\sum_{p \in \mathbb{P}} \frac{\partial \text{Prob}(y_1 = k_{1p}, \dots, y_m = k_{mp} | \mathbf{x})}{\partial \mathbf{x}} \tag{41}$$

i.e. the net effect of a change in  $\mathbf{x}$  on all the orthant probabilities is zero. As such, it should be found empirically that the sum across all orthants of the probabilities should equal one while the sum of the marginal effects across all orthants should equal zero. How the simulated results compare with these benchmarks is one indication of the accuracy of the different computational approaches. Since for  $m=3$  only reference to a bivariate normal cumulative is required to compute the analytical marginal effects, table 3 also shows the results where Stata's canned bivariate normal cumulative function  $binormal(\dots)$  is used instead of the  $ghk(\dots)$  probability simulator.

It is not surprising to observe in the tables that as  $S$  increases the sum across  $p$  of the

---

<sup>21</sup> In the interest of space, only the summaries are shown for  $m=8$ . The full results for the 256 marginal effects are available on request.

probabilities approaches one and that the sum of the marginal effects across  $p$  approaches zero for each  $m \in \{3, 4, 8\}$ . Most noteworthy is that the sum across  $p$  of the analytical marginal effects approaches zero much more rapidly than does the sum across  $p$  of the fully simulated derivatives. In most cases, however, the individual marginal effects for  $j=1, \dots, m$  are for practical purposes relatively similar between methods and across  $S$ .

Computation times are reported in Table 6.<sup>22</sup> The "xb" are generated as  $U(0,1)$  variates. For concreteness, this exercise uses sample sizes in the neighborhood of those commonly encountered in applied microeconomic analysis ( $N=5,000$  and  $N=10,000$ ). At each  $m$ ,  $S$ , and  $N$  the computation time differences between the analytical and fully numerical approaches are striking, ranging from approximately four-fold for  $m=3$  and  $m=4$  to roughly twenty-fold at  $m=8$ .

## 10. Summary

Multivariate probit models are used in a wide variety of cross-sectional and panel data contexts in applied microeconomics. Correspondingly, in practice partial or marginal effects of covariates on various outcome probability configurations are likely to be important quantities to compute based on estimates of such MVP models.

This paper has derived the marginal effects for multivariate probit models of arbitrary dimension  $m \geq 2$ , thus generalizing a result obtained by Greene, 1996, 1998, for the  $m=2$  (bivariate probit) case, and has extended these results to related contexts of interest in sections 4-8. The formulae for these marginal effects are straightforward to program using Stata's Mata language and its *ghk(...)* procedure. Beyond elucidating the mechanics of these marginal effects, one obvious

---

<sup>22</sup> The exercises are conducted using Mata v.10 on a MacBook Pro notebook computer running Mac OS X v.10.6.4 with a 2.4 GHz Intel Core 2 Duo processor and 8 GB of 1067 MHz DDR3 memory.

advantage of the analytical results obtained here is that they reduce the dimension of the multinormal numerical simulation relative to what is required to obtain fully numerical derivatives.<sup>23</sup> The numerical results presented in section 9 show that use of the analytical formulae versus fully numerical derivatives results in both a reduction in computational time as well as an increase in some dimensions of accuracy.

Some of the results obtained above will generalize straightforwardly to related data structures not specifically considered here. For instance, the binary outcomes described in (2) might in some applications be generalized to ordered outcomes, with each marginal distribution having a familiar ordered probit structure. With suitable adaptations, the general approach to computation of marginal effects presented here can be extended to multivariate ordered probit outcome structures. The bivariate model and some of its marginal effects have been analyzed in detail by Greene and Hensher, 2010, Chapter 10. Greene and Hensher also discuss the extension to the general multivariate context, but caution that computational complexity is likely to inhibit applications of such models.

Finally, the paper has not addressed issues regarding sampling variation in the estimates of the marginal effects and corresponding inference considerations. It may be that the results derived here point the way to the derivation of a  $\delta$ -method estimator of the variance of the estimated marginal effects, but the algebra would be quite messy. If computational power is adequate, bootstrapping would provide a far more straightforward approach even though the results in Table 5 suggest that even under best-case circumstances and for a modest number of bootstrap replications such exercises will likely be time-intensive.

---

<sup>23</sup> See Huguenin et al., 2009, for a discussion of other considerations that arise in estimation of MVP models wherein dimension reduction is a primary consideration.

## References

- Avery, R. L.P. Hansen, and V.J. Hotz. 1983. "Multiperiod Probit Models and Orthogonality Condition Estimation." *International Economic Review* 24: 21–35.
- Behncke, S. 2011. "Does Retirement Trigger Ill Health?" *Health Economics* (published online in advance of print).
- Bhattacharya, J. et al. 2006. "Estimating Probit Models with Self-Selected Treatments." *Statistics in Medicine* 25: 389-413.
- Bolduc, D. 1999. "A Practical Technique to Estimate Multinomial Probit Models in Transportation." *Transportation Research Part B* 33: 63-79.
- Cameron, A.C. and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- Christofides, L.N., T. Stengos, and R. Swidinsky. 1997. "On the Calculation of Marginal Effects in the Bivariate Probit Model." *Economics Letters* 54: 203-208.
- Christofides, L.N., T. Stengos, and R. Swidinsky. 2000. "Corrigendum." *Economics Letters* 68: 339.
- Deb, P. et al. 1996. "Choice of Health Insurance by Families of the Mentally Ill." *Health Economics* 5: 61-76.
- Dor, A. et al. 2006. "The Effect of Private Insurance on the Health of Older, Working Age Adults: Evidence from the Health and Retirement Study." *Health Services Research* 41: 759-787.
- Evans, G.W. et al. 2007. "Cumulative Risk, Maternal Responsiveness, and Allostatic Load among Young Adolescents." *Developmental Psychology* 43: 341-351.
- Fichera, E. and M. Sutton. 2011. "State and Self Investments in Health." *Journal of Health Economics* (published online in advance of print).
- Gates, R. 2006. "A Mata Geweke-Hajivassiliou-Keane Multivariate Normal Simulator." *The Stata*

- Journal* 6: 190-213.
- Geronimus, A.T. et al. 2006. "'Weathering' and Age Patterns of Allostatic Load Scores among Blacks and Whites in the United States." *American Journal of Public Health* 96: 826-833.
- Greene, W.H. 1996. "Marginal Effects in the Bivariate Probit Model." Working Paper EC-96-11, NYU, Stern School of Business.
- Greene, W.H. 1998. "Gender Economics Courses in Liberal Arts Colleges: Further Results." *Journal of Economic Education* 29: 291-300.
- Greene, W.H. and D. A. Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.
- Grundy, S.M. et al. 2005. "Diagnosis and Management of the Metabolic Syndrome: An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement." *Circulation* 112: 2735-2752.
- Hajivassiliou, V., D. McFadden, and P. Ruud. 1996. "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results." *Journal of Econometrics* 72: 85-134.
- Holmes, T.H. and R.H. Rahe. 1967. "The Social Readjustment Rating Scale." *Journal of Psychosomatic research*, 11: 213-221.
- Huguenin, J., F. Pelgrin, and A. Holly. 2009. "Estimation of Multivariate Probit Models by Exact Maximum Likelihood." Working Paper 09-02, IEMS, University of Lausanne.
- Kroenke, K. et al. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine* 16: 606-613.
- Machado, J.A.F. and J.M.C. Santos Silva. "Quantiles for Counts." *JASA* 100: 1226-1237.
- McKenzie, D.J. 2005. "Measuring Inequality with Asset Indicators." *Journal of Population Economics*

18: 229-260.

Melchior, L.A. et al. 1993. "A Short Depression Index for Women." *Educational and Psychological Measurement* 53: 1117-1125.

Monfardini, C. and J.M.C. Santos Silva. 2008. "What Can We Learn about Correlations from Multinomial Probit Estimates?" *Economics Bulletin* 3: 1-9.

O'Brien, P.E. et al. 2006. "Treatment of Mild to Moderate Obesity with Laparoscopic Adjustable Gastric Banding or an Intensive Medical Program." *Annals of Internal Medicine* 144: 625-633.

Orchard, T.J. et al. 2005. "The Effect of Metformin and Intensive Lifestyle Intervention on the Metabolic Syndrome: The Diabetes Prevention Program." *Annals of Internal Medicine* 142: 611-619.

Radloff, L.S. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1: 385-401.

Rao, C.R. 1973. *Linear Statistical Inference and Its Applications*, 2nd Edition. New York: Wiley.

Seeman, T.E. et al. 2001. "Allostatic Load as a Marker of Cumulative Biological Risk: MacArthur Studies of Successful Aging." *PNAS* 98: 4770-4775.

StataCorp. 2007. *Stata Statistical Software, Release 10: Reference A-H*. College Station, TX: StataCorp LP.

Stock, J.H. 1989. "Nonparametric Policy Analysis." *JASA* 84: 567-575.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd Edition. Cambridge, MA: MIT Press.

Table 1  
 Number of Combinations Yielding the Sum  $s \in \{0, \dots, m \times (g-1)\}$ ,  $m=2,3,4,5$ ,  $g=3,4,5$

g	m	s																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	2	1	2	3	2	1																
	3	1	3	6	7	6	3	1														
	4	1	4	10	16	19	16	10	4	1												
	5	1	5	15	30	45	51	45	30	15	5	1										
4	2	1	2	3	4	3	2	1														
	3	1	3	6	10	12	12	10	6	3	1											
	4	1	4	10	20	31	40	44	40	31	20	10	4	1								
	5	1	5	15	35	65	101	135	155	155	135	101	65	35	15	5	1					
5	2	1	2	3	4	5	4	3	2	1												
	3	1	3	6	10	15	18	19	18	15	10	6	3	1								
	4	1	4	10	20	35	52	68	80	85	80	68	52	35	20	10	4	1				
	5	1	5	15	35	70	121	185	255	320	365	381	365	320	255	185	121	70	35	15	5	1

Table 2  
Parameter Values for Simulations

<b>m</b>	<b>B</b>	<b>R</b>
3	$[-1,1,1]^T$	$\begin{bmatrix} 1 & -.5 & .25 \\ -.5 & 1 & .5 \\ .25 & .5 & 1 \end{bmatrix}$
4	$[-3,-1,1,2]^T$	$\begin{bmatrix} 1 & -.5 & .25 & -.1 \\ -.5 & 1 & .5 & -.25 \\ .25 & .5 & 1 & .1 \\ -.1 & -.25 & .1 & 1 \end{bmatrix}$
8	$[-2,-1,1,1,1,1,2,2]^T$	$\begin{bmatrix} 1 & -.5 & .25 & -.1 & .05 & .05 & -.05 & -.05 \\ -.5 & 1 & .5 & -.25 & .05 & .05 & -.05 & -.05 \\ .25 & .5 & 1 & .1 & .05 & .05 & .05 & .05 \\ -.1 & -.25 & .1 & 1 & .05 & .05 & .05 & .05 \\ .05 & .05 & .05 & .05 & 1 & -.5 & .25 & -.1 \\ .05 & .05 & .05 & .05 & -.5 & 1 & .5 & -.25 \\ -.05 & -.05 & .05 & .05 & .25 & .5 & 1 & .1 \\ -.05 & -.05 & .05 & .05 & -.1 & -.25 & .1 & 1 \end{bmatrix}$



Table 3  
Simulation Results: m=3

y	Probabilities: GHK S=			Marginal Effects						
	100	1,000	10,000	Analytical: <i>binormal</i>	Analytical: GHK S=			Numerical: GHK S=		
					100	1,000	10,000	100	1,000	10,000
(0,0,0)	0.05130	0.05115	0.05115	-0.09326	-0.09290	-0.09322	-0.09326	-0.09319	-0.09320	-0.09326
(0,0,1)	0.04343	0.04483	0.04496	-0.01231	-0.01186	-0.01226	-0.01230	-0.01069	-0.01215	-0.01229
(0,1,0)	0.09479	0.09507	0.09509	-0.10111	-0.10129	-0.10113	-0.10111	-0.10061	-0.10104	-0.10109
(0,1,1)	0.65370	0.65025	0.65013	0.44865	0.44785	0.44856	0.44863	0.44939	0.44885	0.44868
(1,0,0)	0.01151	0.01136	0.01137	-0.04313	-0.04319	-0.04314	-0.04314	-0.04354	-0.04310	-0.04313
(1,0,1)	0.05142	0.05120	0.05116	-0.09326	-0.09290	-0.09322	-0.09326	-0.09375	-0.09340	-0.09329
(1,1,0)	0.00098	0.00103	0.00104	-0.00446	-0.00440	-0.00446	-0.00446	-0.00421	-0.00444	-0.00446
(1,1,1)	0.09479	0.09507	0.09509	-0.10111	-0.10129	-0.10113	-0.10111	-0.10061	-0.10104	-0.10109
<b>Sum</b>	1.001915	.999957	.999988	-5.55E-17	-1.39E-17	4.16E-17	-1.39E-17	.002784	.000482	.000072

Table 4  
Simulation Results: m=4

y	Probabilities: GHK S=			Marginal Effects					
				Analytical: GHK S=			Numerical: GHK S=		
	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
(0,0,0,0)	0.00457	0.00455	0.00455	-0.02571	-0.02567	-0.02568	-0.02577	-0.02568	-0.02569
(0,0,0,1)	0.15049	0.15013	0.15019	-0.19577	-0.19559	-0.19558	-0.19668	-0.19535	-0.19548
(0,0,1,0)	0.01036	0.01044	0.01045	-0.03996	-0.04016	-0.04020	-0.03975	-0.04013	-0.04019
(0,0,1,1)	0.67317	0.67388	0.67439	0.51686	0.51666	0.51668	0.51414	0.51566	0.51602
(0,1,0,0)	0.00048	0.00049	0.00049	-0.00379	-0.00380	-0.00380	-0.00377	-0.00381	-0.00380
(0,1,0,1)	0.00324	0.00329	0.00329	-0.01636	-0.01635	-0.01634	-0.01604	-0.01631	-0.01634
(0,1,1,0)	0.00722	0.00720	0.00720	-0.03746	-0.03740	-0.03741	-0.03757	-0.03744	-0.03744
(0,1,1,1)	0.14782	0.14754	0.14761	-0.18450	-0.18440	-0.18438	-0.18489	-0.18400	-0.18410
(1,0,0,0)	3.31E-06	3.31E-06	3.31E-06	-5.51E-05	-5.51E-05	-5.51E-05	-5.51E-05	-5.52E-05	-5.51E-05
(1,0,0,1)	0.00004	3.76E-05	3.76E-05	-0.00051	-0.00051	-0.00051	-0.00051	-0.00051	-0.00051
(1,0,1,0)	0.00006	5.90E-05	5.90E-05	-0.00078	-0.00078	-0.00078	-0.00078	-0.00078	-0.00078
(1,0,1,1)	0.00125	0.00125	0.00125	-0.01192	-0.01191	-0.01191	-0.01198	-0.01192	-0.01192
(1,1,0,0)	7.69E-18	5.76E-11	1.47E-11	-1.37E-14	-3.09E-10	-1.00E-10	-5.13E-16	-2.16E-09	-5.54E-10
(1,1,0,1)	7.10E-21	8.22E-21	8.35E-21	-5.60E-19	-5.86E-19	-5.89E-19	-5.00E-19	-5.79E-19	-5.88E-19
(1,1,1,0)	4.81E-07	4.98E-07	4.88E-07	-9.57E-06	-9.56E-06	-9.56E-06	-9.66E-06	-9.92E-06	-9.74E-06
(1,1,1,1)	1.22E-06	1.22E-06	1.22E-06	-2.28E-05	-2.27E-05	-2.27E-05	-2.26E-05	-2.27E-05	-2.27E-05
<b>Sum</b>	.998691	.998875	.999516	3.20E-17	-1.09E-17	1.58E-17	-.003702	-.000359	-.000317

Table 5  
Simulation Results, Summary: m=8

	Probabilities: GHK S=			Marginal Effects					
				Analytical: GHK S=			Numerical: GHK S=		
	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
<b>Sum</b>	.995138	.999038	.999388	2.48E-16	-2.86E-16	-5.17E-17	.079133	.002651	-.000259

Table 6

Computational Time for Partial Derivatives, Samples of Size N (in Seconds):

The Analytical Marginal Effects are computed using  $ghk(S,xb,r)$  with  $xb$   $1 \times (m-1)$  and  $R$   $(m-1) \times (m-1)$ ; the Numerical Marginal Effects are computed using  $ghk(S,xb,r,d1,d2)$  with  $xb$  and  $d1$   $1 \times m$ ,  $R$   $m \times m$ , and  $d2$   $1 \times (.5 * m * (m-1))$ .  
 (In all cases the  $xb$  are generated as the successive rows from a  $uniform(N,m)$  matrix defined in Mata)

m	N	Marginal Effects					
		Analytical: GHK S=			Numerical: GHK S=		
		100	1,000	10,000	100	1,000	10,000
3	5,000	.434	2.28	21.2	1.59	8.53	76.5
	10,000	.729	4.63	43.0	3.15	16.8	154
4	5,000	.538	3.95	37.3	2.61	15.6	144
	10,000	1.09	7.73	77.8	5.07	31.4	293
8	5,000	1.19	10.5	112	23.2	194	2,145
	10,000	2.34	19.6	217	46.1	373	3,930