UCD GEARY INSTITUTE
DISCUSSION PAPER SERIES

Marginal Effects in Multivariate Probit and Kindred Discrete and Count Outcome Models

John Mullahy
University of Wisconsin-Madison,
University College Dublin,
&
The National Bureau of Economic Research

Geary WP2011/35
November 2011

MARGINAL EFFECTS IN MULTIVARIATE PROBIT AND KINDRED DISCRETE AND COUNT OUTCOME MODELS

John Mullahy
University of Wisconsin-Madison, University College Dublin, and NBER

November 13, 2011

Abstract

Estimation of marginal or partial effects of covariates **x** on various conditional parameters or functionals is often the main target of applied microeconometric analysis. In the specific context of probit models, estimation of partial effects involving outcome probabilities will often be of interest. 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 first purpose of this paper is to extend these results to encompass the general $m \geq 2$ multivariate probit (MVP) context for arbitrary orthant probabilities. It is suggested that such partial effects are broadly useful in situations wherein multivariate outcomes are of concern. The paper derives the general result on orthant probability partial effects, which contains Greene's bivariate result as a special case. These results are then 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 the multinomial probit model whose marginal effects turn out to be a special case of those of the multivariate probit model. Numerical simulations suggest that use of the analytical formulae versus fully numerical derivatives results in a reduction in computational time as well as an increase in accuracy.

Acknowledgments

Many thanks are owed to Bill Greene, Alberto Holly, Mari Palta, Ron Thisted, and participants in the UW-Madison Health Econometrics Workgroup for their extraordinarily helpful comments 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.

Contact Information

Prof. John Mullahy
Univ. of Wisconsin-Madison
610 Walnut St., 787 WARF
Madison, WI 53726 USA
jmullahy@wisc.edu
+1-608-265-5410

1. Introduction

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

In the specific context of probit models, estimation of partial effects like $\partial \text{Prob}(\mathbf{y}\in A|\mathbf{x})/\partial\mathbf{x}$, where A is a set of outcomes of interest, is often a central focus. Such estimation is straightforward in univariate models for $\partial \text{Prob}(y=1|\mathbf{x})/\partial\mathbf{x}$, and Greene, 1996, 1998,[1] has extended these calculations to 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 first purpose of this paper is to extend these results to the general $m\geq 2$ multivariate probit (MVP) case for arbitrary orthant probabilities. The paper derives and then demonstrates in several contexts the usefulness of analytical representations of

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

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, **x** are conditioning covariates,[2] and $\text{Prob}(\ldots)$ is a joint or orthant probability from a multivariate normal distribution.[3]

---

[1] See also Christofides et al., 1997, 1998.

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

[3] Somewhat informally, the paper uses the term "orthant probability" in reference to the vector of binary outcomes **y** to refer to the probabilities that the underlying latent random variables that map into (continued)

Greene's results 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 result as a special case. While numerical methods like GHK (see Hajivassiliou et al., 1996) are available for obtaining these marginal effects, there may arise computational advantages from calculating analytical marginal effects, the bottom line 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's second main objective is to extend these results in several directions to be described below.

Data and Estimation

The outcomes $\mathbf{y}=\left[y_j\right]$ can be thought of as arising in the standard probit context as binary indicators of threshold crossings of latent marginal normal variates:[4]

$$
\begin{aligned}
&y_j^* = \mathbf{x}\boldsymbol{\beta}_j + \varepsilon_j, \quad j=1,\ldots,m \\
&\mathbf{y} = 1\left(\mathbf{y}^* \geq \mathbf{0}\right) \\
&\boldsymbol{\varepsilon} = \left[\varepsilon_1,\ldots,\varepsilon_m\right] \sim MVN\left(\mathbf{0},\mathbf{R}\right) \qquad (2) \\
&\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho_{1m} & \cdots & & 1 \end{bmatrix}.
\end{aligned}
$$

The parameters $\mathbf{B}=\left[\boldsymbol{\beta}_1^T,\ldots,\boldsymbol{\beta}_m^T\right]$ and $\mathbf{R}$ can be estimated using algorithms like Stata's *mvprobit* that uses a "full-information" approach (i.e. estimating all elements of **B** and **R** simultaneously) with

---

(continued)

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

[4] 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.

simulated maximum likelihood. Alternatively **B** and **R** can be estimated consistently using a "limited-information" approach suggested by Avery et al., 1980, in which, e.g., the ρ_{jk} elements of **R** are estimated one-by-one using bivariate probit estimators (e.g. Stata's *biprobit* algorithm).[5] Which pairwise estimates of **B** to use in the latter instance is not obvious, but note that even univariate probit estimates of **B** suffice to obtain consistent estimates of the β_j. For present purposes the method of estimation is not of particular concern so long as consistent estimates of **B** and **R** are available.

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 per se for all or some particular k_j's, i.e.

$$APE_p = Avg_{\mathbf{x}}\left(\frac{\partial Prob\left(\mathbf{y}=\mathbf{k}_p \middle| \mathbf{x}\right)}{\partial \mathbf{x}}\right), \tag{3}$$

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. Beyond this, consider an evaluation context where focus is on how a change in some x_j (intervention, policy, etc.) affects expected utility through impacting the distribution outcomes **y** over which welfare is defined. Let utility be $V(y_1,\ldots,y_m)=V(\mathbf{y})$. Expected utility given **x** is then

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

Thus the change in expected utility arising from a change in **x** is

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

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

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

Generally, given consistent estimates of the conditional probability structure $\text{Prob}\left(\mathbf{y}=\mathbf{k}\middle|\mathbf{x}\right)$ for all **k**, then one can use the approach described below to address questions involving the role of varying **x**'s on outcomes defined by $\text{Prob}\left(\mathbf{y}=\mathbf{k}\middle|\mathbf{x}\right)$ as well as aggregates over or differences between such probabilities for different **k**'s of interest. Sections 4-7 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 section 3, section 4 derives the marginal effects of probabilities that are conditioned on subvectors of **y**. Section 5 constructs a count data model on the foundation of an MVP probability structure and derives marginal effects relating to that 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. Section 7 extends the results of the previous sections to multivariate ordered probit models. Section 8 shows the marginal effects in multinomial probit models to be a special case of those in the MVP model. Section 9 reports 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.

Let $\mathbf{u}=\left(u_1,\ldots,u_m\right)$ be continuously measured random variables with population joint distribution function $F\left(u_1,\ldots,u_m\right)$; normality is not assumed at this point. A standard result (or definition) is

$$\frac{\partial^m F\left(v_1,\ldots,v_m\right)}{\partial v_1 \cdots \partial v_m}=f\left(v_1,\ldots,v_m\right), \tag{6}$$

where $f\left(\ldots\right)$ is the joint density and v_j are specific points. Note that (6) can be expressed as:

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

The partial derivative of $F\left(v_1,\ldots,v_m\right)=F\left(\mathbf{v}\right)$ with respect to v_j satisfies:

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

For an intuition[6] for (8), note that in the m=2 case the partial derivative w.r.t. v_1 of the function $g\left(v_1,v_2\right)\equiv\partial F\left(v_1,v_2\right)/\partial v_2$ must in light of (6) yield the joint density $f\left(v_1,v_2\right)$. One function $g\left(v_1,v_2\right)$ satisfying this is $g\left(v_1,v_2\right)=f_2\left(v_2\right)\times F\left(v_1\middle|v_2\right)$, which is of the form (8); this follows since

$$\frac{\partial f_2\left(v_2\right)\times F\left(v_1\middle|v_2\right)}{\partial v_1}=f_2\left(v_2\right)\times\frac{\partial F\left(v_1\middle|v_2\right)}{\partial v_1}=f_2\left(v_2\right)\times f\left(v_1\middle|v_2\right)=f\left(v_1,v_2\right). \tag{9}$$

By recursion, this result generalizes to m>2 by working backwards from the m-th cross partial derivative.

[6] Alternatively (8) can be obtained directly using Leibniz's rule for differentiation of integrals whose limits depend on the variable of differentiation. Since $F\left(\mathbf{v}\right)=\int_{-\infty}^{v_m}\cdots\int_{-\infty}^{v_1}f\left(\mathbf{u}\right)du_1\cdots du_m$, then one can obtain $\partial F\left(\mathbf{v}\right)/\partial v_j$ 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

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

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

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

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

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

Note that the result in (10) is trivial when the v_j are mutually independent, in which case all the conditioning arguments are irrelevant.

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

$$F(\mathbf{c}(\boldsymbol{\theta})) = F(c_1(\boldsymbol{\theta}),...,c_m(\boldsymbol{\theta})). \tag{11}$$

Using a standard chain rule for differentiation in conjunction with (8) yields:

$$\frac{\partial F(c_1(\boldsymbol{\theta}),...,c_m(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{m} \left\{ \frac{\partial F(c_1(\boldsymbol{\theta}),...,c_m(\boldsymbol{\theta}))}{\partial c_j(\boldsymbol{\theta})} \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\}$$

$$= \sum_{j=1}^{m} \left\{ f_j(c_j(\boldsymbol{\theta})) \times F_{-j}(c_1(\boldsymbol{\theta}),...,c_{j-1}(\boldsymbol{\theta}),c_{j+1}(\boldsymbol{\theta}),...,c_m(\boldsymbol{\theta}) | c_j(\boldsymbol{\theta})) \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\} \tag{12}$$

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 $\text{Prob}(y_1 = k_{1p},...,y_m = k_{mp} | \mathbf{x})$. 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,...,$2^m$, j=1,...,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 **R** so that $\mathbf{Q}_p$ is of the form

$$\mathbf{Q}_p = \begin{bmatrix} 1 & s_{1p}s_{2p}\rho_{12} & \cdots & s_{1p}s_{mp}\rho_{1m} \\ s_{1p}s_{2p}\rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ s_{1p}s_{mp}\rho_{1m} & \cdots & & 1 \end{bmatrix} = \begin{bmatrix} 1 & \tau_{12p} & \cdots & \tau_{1mp} \\ \tau_{12p} & 1 & & \vdots \\ \vdots & & \ddots & \\ \tau_{1mp} & \cdots & & 1 \end{bmatrix}. \tag{13}$$

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

$$\text{Prob}\left(y_1 = k_{1p},\ldots,y_m = k_{mp} \middle| \mathbf{x}\right) = \Phi_{\mathbf{Q}_p}\left(s_{1p}\mathbf{x}\boldsymbol{\beta}_1,\ldots,s_{mp}\mathbf{x}\boldsymbol{\beta}_m\right) = \Phi_{\mathbf{Q}_p}\left(\alpha_{1p},\ldots,\alpha_{mp}\right), \tag{14}$$

where $\Phi_{\mathbf{Q}}(\ldots)$ is the cumulative of an $\text{MVN}(\mathbf{0},\mathbf{Q})$ distribution with density $\phi_{\mathbf{Q}}(\ldots)$ and $\alpha_{jp} = s_{jp}\mathbf{x}\boldsymbol{\beta}_j$.[7]

To obtain the MVP's marginal effects it thus suffices to obtain the particular expressions corresponding to the second line in (12). $f_j\left(c_j(\boldsymbol{\theta})\right)$ is a univariate $N(0,1)$ density and $F_{-j}\left(c_1(\boldsymbol{\theta}),\ldots,c_{j-1}(\boldsymbol{\theta}),c_{j+1}(\boldsymbol{\theta}),\ldots,c_m(\boldsymbol{\theta}) \middle| c_j(\boldsymbol{\theta})\right)$ is the cumulative of a conditional (m-1)-variate multivariate normal distribution. The $c_j(\boldsymbol{\theta})$ in (12) are equal to $s_j\mathbf{x}\boldsymbol{\beta}_j$ in the MVP context, with **x** playing the role of the "parameter" that is common across outcomes, so that $dc_j(\boldsymbol{\theta})/d\boldsymbol{\theta}$ is $d\left(s_j\mathbf{x}\boldsymbol{\beta}_j\right)/d\mathbf{x} = s_j\boldsymbol{\beta}_j$. Substituting into (14) $\phi(\ldots)$ for $f(\ldots)$, $\Phi(\ldots)$ for $F(\ldots)$, and α_{jp} for $c_j(\boldsymbol{\theta})$ gives:

$$\begin{aligned} \frac{\partial\Phi_{\mathbf{Q}_p}\left(\alpha_{1p},\ldots,\alpha_{mp}\right)}{\partial\mathbf{x}} &= \sum_{j=1}^{m}\left\{\left(\frac{\partial\Phi_{\mathbf{Q}_p}\left(\alpha_{1p},\ldots,\alpha_{mp}\right)}{\partial\alpha_{jp}}\right)\times\left(\frac{\partial\alpha_{jp}}{\partial\mathbf{x}}\right)\right\} \\ &= \sum_{j=1}^{m}\left\{\left(\phi\left(\alpha_{jp}\right)\times\Phi_{\mathbf{Q}_p\{-j\}}\left(\alpha_{1p},\ldots,\alpha_{(j-1)p},\alpha_{(j+1)p},\ldots,\alpha_{mp}\middle|\alpha_{jp}\right)\right)\times\left(s_{jp}\boldsymbol{\beta}_j\right)^T\right\}. \end{aligned} \tag{15}$$

---

[7] Using the transformed matrixes **Q** in place of the original correlation matrixes **R** streamlines the exposition since for each configuration p the outcome orthant probability can be described by a joint cumulative rather than by a messy mix of cumulatives and survivor functions. This amounts to a linear change-of-variables operation on $\boldsymbol{\varepsilon} = \left[\varepsilon_1,\ldots,\varepsilon_m\right]$ 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.

Given consistent estimates $\hat{\mathbf{B}}$ and $\hat{\mathbf{Q}}$, estimation of (15) is complicated only by evaluation of the term $\Phi_{\mathbf{Q}_p\{-j\}}\left(\alpha_{1p},\ldots,\alpha_{(j-1)p},\alpha_{(j+1)p},\ldots,\alpha_{mp}\middle|\alpha_{jp}\right)$. The following result provides the basis of this calculation:

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

Suppose $\mathbf{z}=\left[z_1,\ldots,z_m\right]\sim MVN\left(\mathbf{0},\boldsymbol{\Omega}\right)$. Partition $\boldsymbol{\Omega}$ as $\begin{bmatrix}\omega_{11} & \boldsymbol{\Omega}_{12}\\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix}$ where ω_{11} is scalar. Then $\mathbf{z}_{-1}=\left[z_2,\ldots,z_m\right]$ conditional on z_1 is (m-1)-variate $MVN\left(\boldsymbol{\Omega}_{21}\omega_{11}^{-1}z_1,\left(\boldsymbol{\Omega}_{22}-\omega_{11}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{12}\right)\right)$. This generalizes straightforwardly to $\mathbf{z}_{-j}=\left[z_1,\ldots,z_{j-1},z_{j+1},\ldots,z_m\right]$, j=2,...,m, by defining different partitions of $\boldsymbol{\Omega}$.

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

$$\mathbf{z}_{-1}\middle|z_1\sim MVN\left(\begin{bmatrix}z_1\tau_{12p}\\ \vdots\\ z_1\tau_{1mp}\end{bmatrix},\begin{bmatrix}1-\tau_{12p}^2 & \tau_{23p}-\tau_{12p}\tau_{13p} & \cdots & \tau_{2mp}-\tau_{12p}\tau_{1mp}\\ \tau_{23p}-\tau_{12p}\tau_{13p} & 1-\tau_{13p}^2 & & \vdots\\ \vdots & & \ddots & \\ \tau_{2mp}-\tau_{12p}\tau_{1mp} & \cdots & & 1-\tau_{1mp}^2\end{bmatrix}\right), \tag{16}$$

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

To compute $\Phi_{\mathbf{Q}_p\{-j\}}\left(\alpha_{1p},\ldots,\alpha_{(j-1)p},\alpha_{(j+1)p},\ldots,\alpha_{mp}\middle|\alpha_{jp}\right)$, define the (m-1)-vector of differences

$$\Delta_{-j,p}=\left[\left(\alpha_{1p}-\alpha_{jp}\tau_{1jp}\right),\ldots,\left(\alpha_{(j-1)p}-\alpha_{jp}\tau_{(j-1)jp}\right),\left(\alpha_{(j+1)p}-\alpha_{jp}\tau_{(j+1)jp}\right),\ldots,\left(\alpha_{mp}-\alpha_{jp}\tau_{mjp}\right)\right]^T, \tag{17}$$

and an (m-1)x(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 $\Phi_{\mathbf{Q}_{p\{-j\}}}\left(\alpha_{1p},\ldots,\alpha_{(j-1)p},\alpha_{(j+1)p},\ldots,\alpha_{mp}\middle|\alpha_{jp}\right)$ can be computed by referring $\mathbf{L}_{jp}$ to $\Phi_{\mathbf{z},\boldsymbol{\Sigma}}\left(\ldots\right)$, which is the cumulative of an (m-1)-variate semi-standardized

$MVN(\mathbf{0},\boldsymbol{\Sigma})$ distribution in which the off-diagonals of $\boldsymbol{\Sigma}$ may be nonzero. In this instance $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $\mathbf{L}_{jp}$ which is in correlation matrix form having typical off-diagonal (r,c) element $\left(\tau_{rcp}-\tau_{jrp}\tau_{jcp}\right)\Big/\sqrt{\left(1-\tau_{jrp}^2\right)\left(1-\tau_{jcp}^2\right)}$. Let this matrix be denoted $\mathbf{V}_{jp}$.

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

$$\frac{\partial \text{Prob}\left(y_1=k_{1p},\ldots,y_m=k_{mp}\middle|\mathbf{x}\right)}{\partial \mathbf{x}}=\sum_{j=1}^{m}\left\{\phi\left(\alpha_{jp}\right)\times\Phi_{z,\mathbf{V}_{jp}}\left(\mathbf{L}_{jp}\right)\times\left(s_{jp}\boldsymbol{\beta}_j\right)^{T}\right\}. \tag{18}$$

For m=2 this is the result obtained by Greene, 1998, p. 298. Greene's result in his notation,

$$\begin{aligned}\frac{\partial \text{BVN}\Phi\left(\boldsymbol{\beta}'\mathbf{x}_1+\gamma,\boldsymbol{\alpha}'\mathbf{x}_2,\rho\right)}{\partial z_k}&=\left\{\phi\left(\boldsymbol{\beta}'\mathbf{x}_1+\gamma\right)\Phi\left[\left(\boldsymbol{\alpha}'\mathbf{x}_2-\rho\left(\boldsymbol{\beta}'\mathbf{x}_1+\gamma\right)\right)\Big/\sqrt{1-\rho^2}\right]\right\}\beta_z\\&+\left\{\phi\left(\boldsymbol{\alpha}'\mathbf{x}_2\right)\Phi\left[\left(\left(\boldsymbol{\beta}'\mathbf{x}_1+\gamma\right)-\rho\left(\boldsymbol{\alpha}'\mathbf{x}_2\right)\right)\Big/\sqrt{1-\rho^2}\right]\right\}\alpha_z\end{aligned} \tag{19}$$

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

$$\frac{\partial \Phi_{\mathbf{Q}_{p^*}}\left(\alpha_{1p^*},\alpha_{2p^*}\right)}{\partial \mathbf{x}}=\sum_{j=1}^{2}\left\{\phi\left(\alpha_{jp^*}\right)\times\Phi_{z,\mathbf{V}_{jp^*}}\left(\frac{\alpha_{(3-j)p^*}-\alpha_{jp^*}\tau_{(3-j)jp^*}}{\sqrt{1-\tau_{12p^*}^2}}\right)\times\left(s_{jp^*}\boldsymbol{\beta}_j\right)^{T}\right\}, \tag{20}$$

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

Finally, from (15) it is noteworthy for computational purposes that only an (m-1)-dimension cumulative normal must be evaluated to obtain the marginal effects. This may reduce computational time as compared with fully numerical methods that are based on an m-dimension cumulative. Even though numerical methods will typically be required in conjunction with the analytical formulae presented here, the dimension reduction should facilitate computation; this is particularly obvious with m=3 where functions like Stata's *binormal(...)* cumulative can be used in lieu of simulation procedures.

4. Marginal Effects of Probabilities Conditional on Subvectors of y

In the context of bivariate probit models, Greene, 1996, suggests that consideration of the marginal effects of **x** on conditional-on-y probabilities, e.g. $\partial \text{Prob}\left(y_1 \middle| y_2, \mathbf{x}\right) \big/ \partial \mathbf{x}$, may be of interest in some instances.[8] Using the approach developed above, this idea can be extended straightforwardly to the general multivariate probit context as follows.

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

$$\text{Prob}\left(\mathbf{y}_a = \mathbf{k}_{p,a} \middle| \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}\right) = \frac{\text{Prob}\left(\mathbf{y} = \mathbf{k}_p \middle| \mathbf{x}\right)}{\text{Prob}\left(\mathbf{y}_b = \mathbf{k}_{p,b} \middle| \mathbf{x}\right)} = \frac{\Phi_{\mathbf{Q}_p}\left(\alpha_{1p}, \ldots, \alpha_{mp}\right)}{\Phi_{\mathbf{Q}_{p,b}}\left(\alpha_{(c+1)p}, \ldots, \alpha_{mp}\right)} \quad (21)$$

where $\mathbf{Q}_{p,b}$ is defined in an obvious way as a submatrix of $\mathbf{Q}_p$. Applying the quotient rule gives:

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

(22)

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

5. Count Data Models Based on MVP Probability Structures

Empirical analysis sometimes involves use of the scalar sum S of the vector **y**, $S = \sum_{j=1}^{m} y_j = \mathbf{1}^T \mathbf{y}$, as the outcome measure of interest.[9] The substantive economic, psychometric, or biometric[10]

---

[8] In applied studies an explicit formulation of the model of interest as $\text{Prob}\left(\mathbf{y}_a = \mathbf{k}_{p,a} \middle| \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}\right)$ is often absent, and this conditional probability may or may not be the parameter whose marginal effects are of interest. See Greene, 1996, for conceptual discussion. Bhattacharya et al., 2006, assess 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.

[9] See Dor et al., 2006, for discussion in the context of health outcome measures. More generally one (continued)

interpretation of S notwithstanding,[11] such outcomes are numerically well defined and obviously inherit their statistical properties from those of **y**.

Define $\mathbb{P}_n = \left\{ p \in \mathbb{P} \middle| \mathbf{1}^T \mathbf{k}_p = n \right\}$, n=0,...,m, and consider the count data probability model defined by

$$\text{Prob}\left(S = n \middle| \mathbf{x}\right) = \sum\nolimits_{p \in \mathbb{P}_n} \text{Prob}\left(\mathbf{y} = \mathbf{k}_p \middle| \mathbf{x}\right), \qquad n=0,...,m. \tag{23}$$

In some instances, analysis proceeds by regression of the outcome measures S defined thusly on **x** using linear or nonlinear regression,[12] presumably with the goal of recovering an estimate of the conditional mean $E\left[S \middle| \mathbf{x}\right]$ and the marginal effects therein implied. However, when the **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 $\hat{E}\left[S \middle| \mathbf{x}\right]$ 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 $E\left[S \middle| \mathbf{x}\right]$ is unlikely to be an appropriate specification.

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

(continued)

might consider weighted sums $S = \sum\nolimits_{j=1}^{m} w_j y_j$; an example of one such index is the Social Readjustment Rating Scale (Holmes and Rahe, 1967).

[10] A prominent biometric example is allostatic load (e.g. Seeman et al., 2001) in which binary threshold (quantile) exceedances for each of m continuously measured biomarkers are summed to arrive at the comprehensive allostatic load measure. In that measure, m=10 and the threshold quantiles are .75 or .25 depending on the particular biomarker. Such additive measures arise in other contexts as well. For instance, the ratings of U.S. Members of Congress by organizations like the League of Conservation Voters 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).

[11] Even the ordinal properties of such measures may be questionable, depending on the nature of what (if any) underlying latent constructs are imagined.

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

The count model (23) implies that

$$E\left[S \middle| \mathbf{x}\right] = \sum_{n=0}^{m} \left(n \times \sum_{p \in \mathbb{P}_n} \text{Prob}\left(\mathbf{y} = \mathbf{k}_p \middle| \mathbf{x}\right)\right). \tag{24}$$

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

$$E\left[S \middle| \mathbf{x}\right] = \sum_{j=1}^{m} E\left[y_j \middle| \mathbf{x}\right] = \sum_{j=1}^{m} \text{Prob}\left(y_j = 1 \middle| \mathbf{x}\right), \tag{25}$$

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 $E\left[S \middle| \mathbf{x}\right]$ other than (24) or (25) is a misspecification. It follows from (24) that

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

The partial derivatives appearing on the rhs of (26) 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

$$v = \begin{cases} 1 \text{ if } \mathbf{y} = \mathbf{k}_p \text{ for } p \in \mathbb{Q} \\ 0 \text{ else} \end{cases}, \tag{27}$$

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

$$\text{Prob}\left(\text{v}=1\middle|\mathbf{x}\right)=\Phi\left(\mathbf{x\theta}\right), \tag{28}$$

where $\Phi(.)$ 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 $\text{v}=1\left(\mathbf{x\theta}+\upsilon>0\right)$ implied in (28) cannot readily be interpreted in terms of the parameters (**B**,**R**) in (2); neither is there any straightforward relationship between the marginal effect of **x** in (28) and the marginal effect defined by summing (18) over $\text{p}\in\mathbb{Q}$. For instance, one obvious version of such a univariate mapping rule assigns v=1 if $\text{S}\geq\text{n}^*$, where $\text{n}^*\leq\text{m}$, i.e. "large" counts are "successes".[13,14] Given (2) and the count data model defined in (23), the proper marginal effect corresponding to (27) is given by defining $\mathbb{Q}=\left\{\text{p}\in\mathbb{P}\middle|\mathbf{1}^{\text{T}}\mathbf{k}_{\text{p}}\geq\text{n}^*\right\}$ and then summing (18) across all $\text{p}\in\mathbb{Q}$. How such marginal effect estimates compare empirically to those derived from estimates of a model like (28) is an interesting and open question.

---

[13] Some medical diagnoses are rendered thusly. Given m binary indicators of patient conditions or attributes, a disorder is deemed to be present if $\text{S}\geq\text{n}^*$. A prominent 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 biomarkers satisfy particular threshold criteria: waist circumference; triglycerides; HDL-C; blood pressure; and fasting glucose. In this paper's notation m=5 and n*=3. Behncke, 2011, O'Brien et al., 2006, and Orchard et al., 2005 are studies involving such metabolic syndrome outcomes. A variety of psychiatric and substance abuse disorders (e.g. ADHD, alcohol abuse, etc., based on DSM-IV criteria) are diagnosed in analogous fashion.

Another common situation is n*=m, so that the binary criterion for a "success" is $\prod_{\text{j}=1}^{\text{m}}\text{y}_{\text{j}}=1$. In some contexts designation of a particular health or disability status is made only if an individual satisfies all of a set of binary criteria. For a child to be considered disabled and eligible for Supplemental Security Income under U.S. Social Security Administration rules, for example, s/he must meet all these criteria: must not be working and earning more than $1,000 a month; must have a physical or mental condition, or a combination of conditions, that results in "marked and severe functional limitations"; and the child's condition(s) must have lasted, or be expected to last, at least 12 months, or must be expected to result in death. (http://www.ssa.gov/pubs/10026.html, accessed Nov. 11, 2011)

[14] 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.

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. To begin, assume that (2) holds except that for j=1,...,m each observed y_j assumes one of g possible values, $y_j \in \{0,...,(g-1)\}$ [15] with the mapping given by:

$$y_j = \sum_{c=1}^{g} (c-1) \times 1\left[\left(\mu_{(c-1)j} - \mathbf{x}\boldsymbol{\beta}_j\right) < u_j < \left(\mu_{cj} - \mathbf{x}\boldsymbol{\beta}_j\right)\right], \tag{29}$$

with $-\infty = \mu_{0j} < \mu_{1j} < ... < \mu_{gj} = +\infty$. As such there are (g-1) free threshold parameters $\left\{\mu_{1j},...,\mu_{(g-1)j}\right\}$ for each j. For shorthand, let $\boldsymbol{\mu}_j = \left[\mu_{1j},...,\mu_{(g-1)j}\right]^T$, $\mathbf{M} = \left[\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_m\right]$, and $\omega_{cj} = \mu_{cj} - \mathbf{x}\boldsymbol{\beta}_j$ for all j. It follows that

$$\text{Prob}\left(y_j = (c-1) \middle| \mathbf{x}\right) = \int_{\omega_{(c-1)j}}^{\omega_{cj}} \phi(u_j) du_j = \int_{-\infty}^{\omega_{cj}} \phi(u_j) du_j - \int_{-\infty}^{\omega_{(c-1)j}} \phi(u_j) du_j, \quad c=1,...,g, \tag{30}$$

where $\phi(...)$ is a univariate N(0,1) density.[16]

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

$$\text{Prob}\left(\mathbf{y} = \mathbf{c}_r \middle| \mathbf{x}\right) = \int_{\omega_{(c_{rm}-1)}}^{\omega_{c_{rm}}} ... \int_{\omega_{(c_{r1}-1)}}^{\omega_{c_{r1}}} \phi_{\mathbf{R}}(u_1,...,u_m) du_1...du_m, \qquad r \in \mathbb{C}. \tag{31}$$

Note that (31) will be a sum of signed multivariate normal cdfs (including zeros and ones at lower and

---

[15] 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.

[16] Estimation of the m-variate multivariate ordered probit model may be computationally challenging. However, consistent estimates of the parameters **B**, **R**, and **M** can be obtained by estimating bivariate ordered 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.

upper integration limits), so that marginal effects at any $\mathbf{c}_r$ will simply be the corresponding sum of the components' marginals. For example, for the trivariate ordered probit model the integral in (31) is

$$\begin{aligned}\int_{\omega_{(c_{r3}-1)}}^{\omega_{c_{r3}}}\int_{\omega_{(c_{r2}-1)}}^{\omega_{c_{r2}}}\int_{\omega_{(c_{r1}-1)}}^{\omega_{c_{r1}}}\phi_{\mathbf{R}}\left(u_1,\ldots,u_m\right)du_1\,du_2\,du_3 = {} & \Phi_{\mathbf{R}}\left(\omega_{c_{r1}},\omega_{c_{r2}},\omega_{c_{r3}}\right)-\Phi_{\mathbf{R}}\left(\omega_{c_{r1}},\omega_{c_{r2}},\omega_{(c_{r3}-1)}\right)- \\ & \Phi_{\mathbf{R}}\left(\omega_{c_{r1}},\omega_{(c_{r2}-1)},\omega_{c_{r3}}\right)-\Phi_{\mathbf{R}}\left(\omega_{(c_{r1}-1)},\omega_{c_{r2}},\omega_{c_{r3}}\right)+ \\ & \Phi_{\mathbf{R}}\left(\omega_{c_{r1}},\omega_{(c_{r2}-1)},\omega_{(c_{r3}-1)}\right)+\Phi_{\mathbf{R}}\left(\omega_{(c_{r1}-1)},\omega_{c_{r2}},\omega_{(c_{r3}-1)}\right)+ \\ & \Phi_{\mathbf{R}}\left(\omega_{(c_{r1}-1)},\omega_{(c_{r2}-1)},\omega_{c_{r3}}\right)-\Phi_{\mathbf{R}}\left(\omega_{(c_{r1}-1)},\omega_{(c_{r2}-1)},\omega_{(c_{r3}-1)}\right)\end{aligned} \quad . \quad (32)$$

Like the count data model (23) that arises in the multivariate binary probit model, one can define a count model corresponding to the multivariate ordered probit model.[17] As before let $S=\sum_{j=1}^{m} y_j$, define $\mathbb{C}_n=\left\{r\in\mathbb{C}\middle|\mathbf{1}^T\mathbf{c}_r=n\right\}$, n=0,..., $(g-1)\times m$, and consider the count data probability model:

$$\mathrm{Prob}\left(S=n\middle|\mathbf{x}\right)=\sum_{r\in\mathbb{C}_n}\mathrm{Prob}\left(\mathbf{y}=\mathbf{c}_r\middle|\mathbf{x}\right),\qquad n=0,\ldots,(g-1)\times m \quad (33)$$

that has corresponding conditional mean

$$E\left[S\middle|\mathbf{x}\right]=\sum_{n=0}^{(g-1)\times m}\left(n\times\sum_{r\in\mathbb{C}_n}\mathrm{Prob}\left(\mathbf{y}=\mathbf{c}_r\middle|\mathbf{x}\right)\right). \quad (34)$$

Partial effects for the count outcome probabilities or conditional mean follow as before, mutatis mutandis. It should be noted that as g and/or m increase, the computational burden -- in terms of the

---

[17] 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 frequency dimension ("Rarely or none of the time (less than 1 day)",..., "Most or all of the time (5-7 days)"). Scoring of the CES-D assigns values $\{0,1,2,3\}$ to correspond to higher frequencies for each item, with the overall score being derived as the simple sum of the item scores. In this case there are 4^{20} 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.

number of multivariate normal cdf evaluations required -- increases rapidly.[18,19]

8. Marginal Effects in Multinomial Probit Models

It turns out that the marginal effects with respect to the conditioning covariates in a multinomial probit (MNP) model are special cases of those described above in (18) for the multivariate probit model.[20] This can be shown as follows. A standard setup for a MNP model is to assume that the value that individual i attaches to choice q is given by a random utility model

$$U_{iq} = \mathbf{w}_{iq}\boldsymbol{\pi} + \mathbf{z}_i\boldsymbol{\gamma}_q + \eta_{iq}, \quad q=1,\ldots,m, \tag{35}$$

where $\mathbf{w}_{iq}$ are attributes of choice q faced by individual i (e.g. the unit price of choice q), $\mathbf{z}_i$ are attribute-invariant characteristics of i (e.g. income), $\boldsymbol{\pi}$ and $\boldsymbol{\gamma}_i$ are parameters, and η_{iq} is an unobservable distributed jointly normally with the other (m-1) η_{ij}. Only pairwise differences in utility matter, so define

$$\delta_{ijq} \equiv U_{ij} - U_{iq} = \left(\mathbf{w}_{ij} - \mathbf{w}_{iq}\right)\boldsymbol{\pi} + \mathbf{z}_i\left(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_q\right) + \left(\eta_{ij} - \eta_{iq}\right) \tag{36}$$

or, using obvious notational shorthand,

$$\begin{aligned}\delta_{ijq} &= \mathbf{W}_{ijq}\boldsymbol{\pi} + \mathbf{z}_i\boldsymbol{\Gamma}_{jq} + \nu_{ijq} \\ &= -\boldsymbol{\lambda}_{ijq} + \nu_{ijq},\end{aligned} \tag{37}$$

---

[18] Given g and m, the number of combinations yielding the sum $S \in \{0,\ldots,m\times(g-1)\}$ is the S-th multinomial coefficient from the expansion of $\left(\sum_{t=0}^{g-1} x^t\right)^m$. For 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.

[19] Using standard count models like Poisson or negative binomial to model either marginal (univariate) or multivariate ordered outcomes is obviously problematic. Extensions to multivariate contexts of alternative approaches such as those proposed by Machado and Santos Silva, 2005, in a univariate context may be interesting to consider in future work.

[20] 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.

with $\lambda_{ijq} = \left(\mathbf{w}_{iq} - \mathbf{w}_{ij}\right)\boldsymbol{\pi} + \mathbf{z}_i\left(\boldsymbol{\gamma}_q - \boldsymbol{\gamma}_j\right)$. The residuals ν_{ijq} in (37) have an (m-1)-variate normal distribution whose parameterization is discussed below. The behavioral model is that subject i selects the outcome with the largest U_{ij} so that

$$\text{Prob}\left(\text{i selects q}\middle|\mathbf{W}_i, \mathbf{z}_i\right) = \text{Prob}\left(\left(\nu_{i1q} < \lambda_{i1q}\right), \ldots, \left(\nu_{i(q-1)q} < \lambda_{i(q-1)q}\right), \left(\nu_{i(q+1)q} < \lambda_{i(q+1)q}\right), \ldots, \left(\nu_{imq} < \lambda_{imq}\right)\middle|\mathbf{W}_i, \mathbf{z}_i\right), \tag{38}$$

where the conditioning $\mathbf{W}_i$ denotes the entire collection of the $\mathbf{W}_{ijq}$. Assuming for the moment that the ν_{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 (38) corresponds to an (m-1)-dimension version of expression (14) in which all $s_{jp}=1$ and in which the $\mathbf{x}\boldsymbol{\beta}_j$ are replaced by λ_{ijq}. As such the basic form of the marginal effects $\partial\text{Prob}\left(\text{i selects q}\middle|\mathbf{W}_i, \mathbf{z}_i\right) / \partial\left\{\mathbf{w}_{iq}, \mathbf{w}_{ij}, \mathbf{z}_i\right\}$ corresponds to (18), mutatis mutandis.[21]

In this general MNP 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 (38). Thus, the terms $\left(s_{jp}\boldsymbol{\beta}_j\right)^T$ that appear in (18) are replaced as follows:

(a) Own-attribute marginal effects, $\partial\text{Prob}(\ldots)/\partial\mathbf{w}_{iq}$: Replace $\left(s_{jp}\boldsymbol{\beta}_j\right)^T$ with $\boldsymbol{\pi}^T$ in each summand, j=1,...,(j-1),(j+1),...,m.

(b) Cross-attribute marginal effects, $\partial\text{Prob}(\ldots)/\partial\mathbf{w}_{ik}$, k=1,...,(j-1),(j+1),...,m: Replace $\left(s_{jp}\boldsymbol{\beta}_j\right)^T$ with $-\boldsymbol{\pi}^T$ only in the k-th summand; set equal to zero otherwise.

(c) Attribute-invariant marginal effects, $\partial\text{Prob}(\ldots)/\partial\mathbf{z}_i$: Replace $\left(s_{jp}\boldsymbol{\beta}_j\right)^T$ with $\left(\boldsymbol{\gamma}_q - \boldsymbol{\gamma}_j\right)^T$ in each summand, j=1,...,(j-1),(j+1),...,m.

---

[21] 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 is one possibility. If the normalization used results in a structure for $\boldsymbol{\Psi}$ other than one having a correlation matrix structure, then additional computations may be required to use formulae based on (18); see the discussion appearing between equations (17) and (18) above.

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 v.10.) *ghk(...)* is used here both to compute the (m-1)-dimension cumulative arising in the discussion that follows equation (15) (*ghk(t,x,V)*) as well as to simulate the marginal effects based on fully numerical methods appealing to the m-dimension cumulative (*ghk(t,x,V,d1,d2)*).

Use of *ghk(...)* requires specification of the number of simulation points (t). To assess how this choice affects computational performance, three sets of results are obtained corresponding to t=100, t=1,000, and t=10,000. Computational time and accuracy should both increase in t. 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 **B** and **R**; these are detailed in Table 2.

The results of the simulation exercise are presented in Tables 3-5.[22] For each t, 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

$$\sum_{p \in \mathbb{P}} \frac{\partial \text{Prob}\left(y_1 = k_{1p}, \ldots, y_m = k_{mp} \middle| \mathbf{x}\right)}{\partial \mathbf{x}} \qquad (39)$$

i.e. the net effect of a change in **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(...)* is used instead of the *ghk(...)* probability simulator.

It is not surprising to observe in the tables that as t increases the sum across p of the probabilities approaches one and that the sum of the marginal effects across p approaches zero for each

[22] 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.

$m \in \{3,4,8\}$. Noteworthy is that the sum across p of the analytical marginal effects approaches zero more rapidly than does the sum of the fully simulated derivatives. In most cases, however, the marginal effects for j=1,...,m are for practical purposes quite similar between methods and across t.

Computation times are reported in Table 6.[23] 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 microeconometric analysis (N=5,000 and N=10,000). At each m, t, and N the computation time differences between the analytical and fully numerical approaches are particularly striking, ranging from approximately four-fold for m=3 and m=4 to roughly twenty-fold at m=8.

10. Summary

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 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 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.[24] 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.

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 6 suggest that even under best-case circumstances and for a modest number of bootstrap replications such exercises will likely be time-intensive.

---

[23] 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.

[24] 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. 2005. "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 Intenstive 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 $S \in \{0,\ldots,m \times (g-1)\}$, m=2,3,4,5, g=3,4,5

		S																				
g	**m**	**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

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

Table 3
Simulation Results: m=3

	Probabilities: GHK t=			**Marginal Effects**						
				Analytical:	**Analytical: GHK t=**			**Numerical: GHK t=**		
y	**100**	**1,000**	**10,000**	***binormal***	**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
Sum	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 t=			Marginal Effects					
				Analytical: GHK t=			Numerical: GHK t=		
	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
Sum	.998691	.998875	.999516	3.20E-17	-1.09E-17	1.58E-17	-.003702	-.000359	-.000317

Table 5
Simulation Results, Summary: m=8

	Probabilities: GHK t=			Marginal Effects					
				Analytical: GHK t=			Numerical: GHK t=		
	100	**1,000**	**10,000**	**100**	**1,000**	**10,000**	**100**	**1,000**	**10,000**
Sum	.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(t,xb,r)* with xb 1x(m-1) and R (m-1)x(m-1);
the Numerical Marginal Effects are computed using *ghk(t,xb,r,d1,d2)* with xb and d1 1xm, R mxm, and d2 1x(.5*m*(m-1)).
(In all cases the xb are generated as the successive rows from a uniform(N,m) matrix defined in Mata)

		Marginal Effects					
		Analytical: GHK t=			**Numerical: GHK t=**		
m	**N**	**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