HHS Public Access
Author manuscript
Empir Econ. Author manuscript; available in PMC 2019 February 25.

Published in final edited form as:
Empir Econ. 2017 March ; 53(2): 447–461. doi:10.1007/s00181-016-1090-8.

Marginal Effects in Multivariate Probit Models

John Mullahy

University of Wisconsin-Madison, NUI Galway, and NBER, jmullahy@wisc.edu, +1-608-265-5410 (phone)

Abstract

Estimation of marginal or partial effects of covariates **x** on various conditional parameters or functionals is often a 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 results covering the case of quadrant probability marginal effects in bivariate probit models for jointly distributed outcomes **y** have previously been described in the literature. This paper's goals are to extend Greene's results to encompass the general M≥2 multivariate probit (MVP) context for arbitrary orthant probabilities and to extended these results to models that condition on subvectors of **y** and to multivariate ordered probit data structures. It is suggested that such partial effects are broadly useful in situations wherein multivariate outcomes are of concern.

Keywords

multivariate probit; marginal effects; C30; C35

1. Introduction and Motivation

Given M-dimensioned multivariate outcomes **y** and K-dimensioned covariates **x** whose relationship can often be usefully cast in terms of the conditional distribution $F(\mathbf{y}|\mathbf{x})$ analysts will often focus attention on estimation of and inferences about functionals of the M univariate conditional marginals of $F(.)$, i.e $V\left(F_j\left(y_j|\mathbf{x}\right)\right)$. Such univariate focus may of course owe to the particular nature of the question(s) at hand, but such focus may also be restrictive. Since estimation and interpretation of marginal or partial effects of covariates **x** on outcomes is often a central feature of applied microeconometric analysis, functionals $V(.)$ defined on the full multivariate joint distribution $F(\mathbf{y}|\mathbf{x})$ and how these vary with **x** may rightly be of interest in particular settings. [1] That is, in general one might consider functions defined on **y**, $h(\mathbf{y}) = h\left(y_1, \ldots, y_M\right)$ their moments, and their corresponding marginal effects $\partial E[h(\mathbf{y})|\mathbf{x}]/\partial\mathbf{x}$.

[1]One obvious example is that of conditional product moments $E\left[\prod_{j=1}^{M} y_j^{b_j}|\mathbf{x}\right]$ of which conditional covariances may be the most familiar example. In such cases how $\sigma_{i,j}(\mathbf{x})$ varies with conditioning sets **x** may be of interest in applications (consider, e.g., GARCH and related literatures).

The main focus in this paper is on joint probabilities and how they vary with $\mathbf{x}$, i.e. where $h(\mathbf{y})$ is $1(\mathbf{y} = \mathbf{k})$ for integer- or binary- valued M-vectors $\mathbf{k}$ and the focus is on $\partial \Pr(\mathbf{y} = \mathbf{k} | \mathbf{x}) / \partial \mathbf{x}$. Such considerations arise in substantive applied problems that focus on various aspects or patterns involving the intrinsic jointness of multivariate discrete outcomes and, by extension, on the marginal effects of $\mathbf{x}$ on such quantities, for instance when an analysis concerns understanding a set of outcomes observed for subjects at one point in time, a single outcome observed for subjects over time, or some combination thereof.

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 perse for all or some k_j, i.e.

$$\mathrm{APE}_p = \mathrm{Avg}_{\mathbf{x}}\left(\frac{\partial \Pr\left(\mathbf{y} = \mathbf{k}_p \middle| \mathbf{x}\right)}{\partial \mathbf{x}}\right),$$

for some p or set of p’s in $\mathbb{P}$ (where $\mathbb{P}$ is the set of all possible values of $\mathbf{y}$). In practice, a variety of situations arise where understanding how a $\Delta\mathbf{x}$ intervention affects conditional joint or orthant probabilities of various outcome patterns is per se of central interest. Beyond this, consider an evaluation context where focus is on how a change in some xj (intervention, policy, etc.) affects expected utility through impacting the distribution outcomes $\mathbf{y}$ over which welfare is defined. Let utility be $V\left(y_1, \ldots, y_m\right) = V(\mathbf{y})$**.** Expected utility given $\mathbf{x}$ is then

$$E[V(\mathbf{y}) | \mathbf{x}] = \sum_{k_M = 0}^{1} \cdots \sum_{k_1 = 0}^{1} \left\{V\left(y_1 = k_1, \ldots, y_M = k_M\right) \times \Pr\left(y_1 = k_1, \ldots, y_M = k_M \middle| \mathbf{x}\right)\right\}.$$

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

$$\frac{\partial E(V(\mathbf{y}) | \mathbf{x})}{\partial \mathbf{x}} = \sum_{k_M = 0}^{1} \cdots \sum_{k_1 = 0}^{1} \left\{V\left(y_1 = k_1, \ldots, y_M = k_M\right) \times \frac{\Pr\left(y_1 = k_1, \ldots, y_M = k_M \middle| \mathbf{x}\right)}{\partial \mathbf{x}}\right\}.$$

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. Generally, given consistent estimates of the conditional probability structure $\Pr(\mathbf{y} = \mathbf{k} | \mathbf{x})$ for all $\mathbf{k}$, then one can use the approach described below to address questions involving the role of varying $\mathbf{x}$ on outcomes defined by $\Pr(\mathbf{y} = \mathbf{k} | \mathbf{x})$ as well as aggregates over or differences between such probabilities for different $\mathbf{k}$ of interest. While the conceptualization of marginal effects in such contexts is no different from that arising in univariate contexts -- i.e. how conditional functionals vary with $\mathbf{x}$ - the computation of such marginal effects may present challenges in both computation and interpretation beyond those arising in univariate-outcome settings.

In the specific context of binary or ordered probit models -- the main subject of this paper -- estimation of partial effects like $\partial \Pr(\mathbf{y} \in A | \mathbf{x}) / \partial \mathbf{x}$ are often a central focus (here, A is some outcome set of interest). Such estimation is straightforward in standard univariate models for

$\partial \Pr(\mathbf{y} \in 1|\mathbf{x})/\partial \mathbf{x}$, and Greene (1996, 1998)[2] has extended these calculations to quadrant probability marginal effects $\partial \Pr(y_1 = k_1, y_2 = k_2|\mathbf{x})/\partial \mathbf{x}, k_j \in \{0, 1\}$, in bivariate probit models.

This paper extends these results to the general m≥2 multivariate probit (MVP) case for arbitrary orthant probabilities. Specifically the paper derives analytical representations of $\partial \Pr(y_1 = k_1, y_M = k_M|\mathbf{x})/\partial \mathbf{x}$ or, in shorthand, $\partial \Pr(\mathbf{y} = \mathbf{k}|\mathbf{x})/\partial \mathbf{x}$ where, for the binary probit case, $\mathbf{y} = [y_j]$ is the M-variate binary outcome vector, $\mathbf{k} = [k_j]$ is an m-vector of zeros or ones indicating any of the any of the 2^M possible outcomes, $\mathbf{x}$ are conditioning covariates,[3] and Pr(…) is a joint or orthant probability from a multivariate normal distribution.[4] While Greene's results are well established, the analytical formulae describing the general orthant probability result are not evident in the literature. This paper derives such results, which contain Greene's bivariate result as a special case, and suggests their potential applicability in applied contexts.

The remainder of the paper is organized as follows. Section 2 derives the main analytical results for arbitrary joint distributions. Section 3 describes the nature of the data in probit contexts, discusses estimation of multivariate probit models, and obtains the specific marginal effect formulae for multivariate probit models. Building on these results, Section 4 derives the marginal effects of probabilities that are conditioned on subvectors of **y** and Section 5 derives marginal effects for multivariate orderedprobit models. Section 6 reports an empirical exercise in a model of the determinants of multiple chronichealth conditions. Section 7 summarizes. Detailed derviations of the main resultsare presented in appendixes.

2. Results for Arbitrary Multivariate Distributions

The paper first establishes the main results on marginal effects for arbitrary joint distributions and then proceeds in the next section to obtain the particular results for the multivariate probit (MVP) model.

Let $\mathbf{u} = [u_1, \ldots, u_M]$ be continuously measured random variables with population joint distribution function $F = (u_1, \ldots, u_M)$. A standard result (or definition) is

$$\left.\frac{\partial^M F(u_1, \ldots, u_M)}{\partial u_1 \cdots \partial u_M}\right|_{\mathbf{u} = \mathbf{v}} = f(v_1, \ldots, v_M), \quad (1)$$

[2]See also Christofides et al. (1997, 1998).

[3]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, count measures, etc.) can be accommodated straightforwardly with the understanding that discrete differences in $\Pr(y_1 = k_1, \ldots, y_M = k_M|\mathbf{x})$ due to $\Delta x_j = 1$ will be of interest; these can be computed by evaluating $\Pr(y_1 = k_1, \ldots, y_M = k_M|\mathbf{x})$ at two different values of x_j and then differencing.

[4]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 the observed binary **y** (see (4) 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 $2^M \times M$ matrix whose rows (arranged arbitrarily) are the 2^M possible outcome configurations **k**. Let $\mathbb{P}$ be a 2^M-element set indexing rows of **K** having typical indexing element p, so that $\mathbf{k}_p = \mathbf{k}_{p\bullet}$ will denote a particular (p-th) outcome configuration.

where f(…) is the joint density and v_j are specific values of u_j. Note that (1) can be expressed as:

$$f_j(u_j) \times \left. \frac{\partial^{M-1} F\left(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_M \middle| u_j\right)}{\partial u_1 \cdots \partial u_{j-1} \partial u_{j+1} \cdots \partial u_M} \right|_{\mathbf{u} = \mathbf{v}} = f(v_1, \ldots, v_M), \text{ for any } j = 1, \ldots, M.$$

The partial derivative of F(**u**) with respect to u_j satisfies (see Appendix A for detailed derivations):

$$\left. \frac{\partial F(\mathbf{u})}{\partial u_j} \right|_{\mathbf{u} = \mathbf{v}} = f_j(v_j) \times F_{-j}\left(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_M \middle| v_j\right), \ j = 1, \ldots, M. \quad (2)$$

Suppose F(**u**) is evaluated at $\mathbf{u} = \mathbf{c}(\boldsymbol{\theta}) = \left[c_1(\boldsymbol{\theta}), \ldots, c_M(\boldsymbol{\theta})\right]$ where θ is a parameter (scalar or vector) shared across the M margins of F(**u**), and where all c_j (θ) are differentiable in θ. Thus with $F(\mathbf{c}(\boldsymbol{\theta})) = F\left(c_1(\boldsymbol{\theta}), \ldots, c_m(\theta)\right)$ a Standard chain rule for differentiation along with (2) yields:

$$\frac{\partial F\left(c_1(\boldsymbol{\theta}), \ldots, c_M(\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{M} \left\{ \frac{\partial F\left(c_1(\boldsymbol{\theta}), \ldots, c_M(\boldsymbol{\theta})\right)}{\partial c_j(\boldsymbol{\theta})} \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\}$$
$$= \sum_{j=1}^{M} \left\{ f_j\left(c_j(\boldsymbol{\theta})\right) \times F_{-j}\left(c_1(\boldsymbol{\theta}), \ldots, c_{j-1}(\boldsymbol{\theta}), c_{j+1}(\boldsymbol{\theta}), \ldots, c_M(\boldsymbol{\theta}) \middle| c_j(\boldsymbol{\theta})\right) \times \frac{dc_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\}$$

(3)

3. Results for the Multivariate Probit Model

In the binary probit context the outcomes $\mathbf{y} = [y_m]$ can be thought of as arising in the standard probit context as binary indicators of threshold crossings of latent marginal normal variates:[5]

$$y_j^* = \mathbf{x}\,\boldsymbol{\beta}_j + \varepsilon_j, \ j = 1, \ldots, M \quad (4)$$

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

[5]This stochastic structure allows for but does not appeal specifically to a common factor error structure for $\boldsymbol{\varepsilon}$ in (4). It may be that such an assumption would simplify estimation and, ultimately, computation of the marginal effects.

$$\boldsymbol{\varepsilon} = \left[\varepsilon_1, \ldots, \varepsilon_M\right] \sim \mathrm{MVN}(\mathbf{0}, \mathbf{R})$$

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

The parameters $\mathbf{B} = \left[\ \beta_1^T, \ldots, \ \beta_M^T\right]$ and $\mathbf{R}$ can be estimated using algorithms like Stata's *mvprobit* (Cappellari and Jenkins 2003) that uses a full-information approach (i.e. estimating all elements of **B** and **R** simultaneously) with simulated ML. Alternatively **B** and **R** can be estimated consistently using a computationally less demanding approach suggested by Mullahy (2016) that uses a chain of bivariate probit estimators to estimate the MVP model. For present purposes the method of estimation is not of concern so long as consistent estimates of **B** and **R** are available.

Recall that in the binary-outcome case there are 2^M possible outcome configurations. For each configuration $\mathbf{k}_p$, $p \in \mathbb{P}$, one has a corresponding conditional outcome probability $\Pr\left(y_1 = k_{1p}, \ldots, y_M = k_{Mp} | \mathbf{x}\right)$ The derivations obtained in Appendix B give the MVP marginal effects as

$$\frac{\partial \Pr\left(y_1 = k_{1p}, \ldots, y_M = k_{Mp} | \mathbf{x}\right)}{\partial \mathbf{x}} = \sum_{j=1}^{M} \left\{ \phi\left(\alpha_{jp}\right) \times \Phi_{z, V_{jp}}\left(L_{jp}\right) \times \left(s_{jp}\, \boldsymbol{\beta}_j\right)^T \right\}, \quad (5)$$

where all relevant notation is defined in Appendix B. For M=2 this is the result obtained by Greene (1998, p. 298). Greene's result in his notation,

$$\frac{\partial \mathrm{BVN}\Phi\left(\beta' \mathbf{x}_1 + \gamma,\ \alpha' \mathbf{x}_2, \rho\right)}{\partial z_k} = \left\{ \phi\left(\beta' \mathbf{x}_1 + \gamma\right) \Phi\left[\left(\alpha' \mathbf{x}_2 - \rho\left(\beta' \mathbf{x}_1 + \gamma\right)\right)\right] / \sqrt{1 - \rho^2} \right\} \beta_z$$
$$+ \left\{ \phi\left(\alpha' \mathbf{x}_2\right) \Phi\left[\left(\left(\beta' x_1 + \gamma\right) - \rho\left(\alpha' \mathbf{x}_2\right)\right)\right] / \sqrt{1 - \rho^2} \right\} \alpha_z$$

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

$$\frac{\partial\, \Phi_{\mathbf{Q}_{p*}}\left(\alpha_{1p*}, \alpha_{2p*}\right)}{\partial \mathbf{x}} = \sum_{j=1}^{2} \left\{ \phi\left(\alpha_{jp*}\right) \times \Phi_{z,\mathbf{V}_{jp*}}\left(\frac{\alpha_{(3-j)p*} - \alpha_{jp*}\tau_{(3-j)jp*}}{\sqrt{1-\tau_{12p*}^{2}}}\right) \times \left(s_{jp*}\, \boldsymbol{\beta}_j\right)^{T} \right\}, \quad (6)$$

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

In closing this section, three points merit consideration. First, only an (M-1)-dimension cumulative normal must be evaluated to obtain the marginal effects. So, e.g., for M=3 bivariate cumulative functions like Stata's *binormal(…)* can be used in lieu of simulation procedures. Second, note that the familiar panel probit model (Greene 2004) is a special case $\left(\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_M\right)$ of (4) when period-specific covariates $\mathbf{x}_t$ are time-invariant. As such, its orthant marginal effects can be computed exactly as above. Even when the $\mathbf{x}_t$ vary over time in the panel probit context, the relevant marginal effects can be obtained as a straightforward modification of (6). Finally, in practice it will typically be the case that estimation of the empirical counterparts to the APEs discussed in Section 1 will be the main objective. The approach to such estimation would in general follow that used for computation of marginal effects in most regression contexts, i.e.

$$\hat{APE}_p = \sum_{i=1}^{N} Wgt_i \left(\frac{\partial \hat{Pr}\left(y_{i1} = k_{1p}, \ldots, y_{iM} = k_{Mp} \middle| \mathbf{x}_i\right)}{\partial \mathbf{x}_i}\right),$$

where wgt_i might be N^{-1} or some other quantity reflecting sampling or other weighting schemes.

4. Marginal Effects of Orthant 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. $\partial \Pr\left(y_1 \mid y_2, \mathbf{x}\right) / \partial \mathbf{x}$ may be of interest in some instances.[6] Using the approach developed above, this idea can be extended straightforwardly to the general multivariate probit context as follows.

[6] In applied studies an explicit formulation of the model of interest as $\Pr\left(\mathbf{y}_a = \mathbf{k}_{p,a} \mid \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}\right)$ is often absent, and this conditional probability may or may not be the parameterwhose marginal effects are of interest. See Greene (1996) for conceptual discussion.

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 M*-vectors and $\mathbf{y}_b$ and $\mathbf{k}_{p,b}$ are (M-M*)-vectors. Suppose interest is in the quantities $\Pr(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x})$ and $\partial \Pr(\mathbf{y}_a = \mathbf{k}_{p,a} | \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}) / \partial \mathbf{x}$. Note that

$$\Pr\left(\mathbf{y}_a = \mathbf{k}_{p,a} \middle| \mathbf{y}_b = \mathbf{k}_{p,b}, \mathbf{x}\right) = \frac{\Pr\left(\mathbf{y} = \mathbf{k}_p \middle| \mathbf{x}\right)}{\Pr\left(\mathbf{y}_b = \mathbf{k}_{p,b} \middle| \mathbf{x}\right)} = = \frac{\Phi_{\mathbf{Q}_p}\left(\alpha_{1_P}, \ldots, \alpha_{M_p}\right)}{\Phi_{\mathbf{Q}_{p,b}}\left(\alpha_{(M^*+1)_P}, \ldots, \alpha_{M_p}\right)},$$

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

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

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. 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. Assume that (4) holds but now for j=1, …,M each observed y_j assumes one of g possible values, $y_j \in \{0, \ldots, (g-1)\}$ [7] with the mapping given by:

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

and with $-\infty = \mu_{0j} < \mu_{1j} < \ldots < \mu_{gj} = +\infty$. As such for each j there are g-1 free threshold parameters $\{\mu_{1j}, \ldots, \mu_{(g-1)j}\}$. Let $\mu_j = [\mu_{1j}, \ldots, \mu_{(g-1)j}]^T$, $\mathbf{M} = [\mu_1, \ldots, \mu_M]$ and $\omega_{cj} = \mu_{cj} - \mathbf{x}\,\boldsymbol{\beta}_j$ for all j. It follows that

[7] Allowing the y_j to have different numbers of outcomes is straightforward; the assumption of equal numbers of categories across j is made solelyto keep notation from becoming unwieldy.

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

where $\varphi(u_j)$ is a univariate N(0,1) density.[8]

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 $\mathbb{P}$ be a g^M -element set indexing columns of **C** having typical indexing element r, so that $\mathbf{c}_r = \mathbf{C}_{\bullet r}$ will denote a particular (r-th) outcome configuration. Thus

$$\Pr\left(\mathbf{y} = \mathbf{c}_r\middle|\mathbf{x}\right) = \int_{\omega_{\left(c_{rM}-1\right)}}^{\omega_{c_{rM}}} \cdots \int_{\omega_{\left(c_{r1}-1\right)}}^{\omega_{c_{r1}}} \phi_{\mathbf{R}}\left(u_1, \ldots, u_M\right)du_1 \ldots du_M, \quad r \in \mathbb{C}. \quad (7)$$

Note that (7) will be a sum of signed multivariate normal cdfs including(zeros and ones at lower and upper integration limits), so that marginal effects at any $\mathbf{c}_r$ are the corresponding signed sum of the components' marginals. For example, for the trivariate ordered probit model, (7) is

$$\begin{aligned}\int_{\omega_{\left(c_{r3}-1\right)}}^{\omega_{c_{r3}}} \int_{\omega_{\left(c_{r2}-1\right)}}^{\omega_{c_{r2}}} \int_{\omega_{\left(c_{r1}-1\right)}}^{\omega_{c_{r1}}} \phi_{\mathbf{R}}\left(u_1, u_2, u_3\right)du_1du_2du_3 = \\ \Phi_{\mathbf{R}}\left(\omega_{c_{r1}}, \omega_{c_{r2}}, \omega_{c_{r3}}\right) - \Phi_{\mathbf{R}}\left(\omega_{c_{r1}}, \omega_{c_{r2}}, \omega_{\left(c_{r3}-1\right)}\right) - \\ \Phi_{\mathbf{R}}\left(\omega_{c_{r1}}, \omega_{\left(c_{r2}-1\right)}, \omega_{c_{r3}}\right) - \Phi_{\mathbf{R}}\left(\omega_{\left(c_{r1}-1\right)}, \omega_{c_{r2}}, \omega_{c_{r3}}\right) + \\ \Phi_{\mathbf{R}}\left(\omega_{c_{r1}}, \omega_{\left(c_{r2}-1\right)}, \omega_{\left(c_{r3}-1\right)}\right) + \Phi_{\mathbf{R}}\left(\omega_{\left(c_{r1}-1\right)}, \omega_{c_{r2}}, \omega_{\left(c_{r3}-1\right)}\right) + \\ \Phi_{\mathbf{R}}\left(\omega_{\left(c_{r1}-1\right)}, \omega_{\left(c_{r2}-1\right)}, \omega_{c_{r3}}\right) + \Phi_{\mathbf{R}}\left(\omega_{\left(c_{r1}-1\right)}, \omega_{\left(c_{r2}-1\right)}, \omega_{\left(c_{r3}-1\right)}\right)\end{aligned}$$

6. Empirical Illustration

This section provides an empirical illustration of the application of MVP marginal effects. The example is drawn from a larger project studying the determinants of multiple chronic conditions in U.S. adult populations. The data are from the 2010 Medical Expenditure Panel Survey (MEPS) Household Component, with a focus on adults aged 18+. Five common

[8]Estimation of the M-variate multivariate ordered probit model can be approached using the methods spelled out in Mullahy (2016).

chronic conditions, measured as binary outcomes, are of interest: Hypertension, Asthma, Depression, Hyperlipidemia, and Diabetes. As such, there are 2^5=32 possible outcome patterns. An MVP model is estimated in which covariates are age (in years), schooling (in years), and gender. The estimation sample size is N=23,328. Sample average marginal effectswith respect to age and schooling are computed using the methods described in Section 3, with the results reported in Table 1.

Table 1 indicates that most patterns of poorer health conditions except some in which asthma is prominent are estimated to be increasing with age, with the probability of the perfect health (i.e. $\mathbf{y} = \mathbf{0}$) outcome decreasing by roughly one percentage point per additional year of age. Schooling's estimated marginal effects are positive for the perfect health outcome, negative for most outcomes that involve the presence of diabetes, and somewhat mixed otherwise.[9]

7. Summary

This paper has derived the marginal effects for multivariate probit models of arbitrary dimension M≥2, thus generalizing a result obtained by Greene (1996, 1998) for the bivariate probit case. 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.[10] 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 for purposes of inference would seem to provide a more straightforward approach

Finally, a potentially interesting extension of the results obtained here would be to situations involving conditionally heteroskedastic multivariate probit models. In this instance, $\mathrm{vcov}(\varepsilon \mid \mathbf{x}) = \mathrm{V}(\mathbf{x})$ in (4). While estimation of such models itself may be challenging, the additional richness afforded by such parameterizations in describing multivariate probit data structures may be important to exploit in particular applications.

Acknowledgments:

Thanks are owed to Kevin Denny, Bill Greene, Jeff Hoch, Alberto Holly, Stephen Jenkins, Mari Palta, Ron Thisted, participants in seminars at University College Dublin, Minnesota, AHRQ, and Wisconsin, and an anonymous referee for their helpful comments, and to Katherine Mullahy for valuable editorial assistance. This work has been supported in part by the Health & Society Scholars Program, by NICHD grant P2C HD047873 to the Center for Demography and Ecology, and by the Robert Wood Johnson FoundationEvidence for Action Program (grant 73336), all at the University of Wisconsin-Madison. An earlier, longer draft was circulated as NBER W.P. 17588.

[9]Of course, for each covariate the sum of the marginal effects across all 32 patterns must be zero.
[10]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.

Appendix A:: Detailed Derivations for the General Case

For an intuition for (2), note that in the M=2 case the partial derivative w.r.t. u_1 of the function $g(u_1, u_2) \equiv \partial F(u_1, u_2)/\partial u_2$ evaluated at **u**=**v** must in light of (1) yield 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 (2); this follows since, at **u** = **v**,

$$\frac{\partial f_2(v_2) \times F(v_1 | v_2)}{\partial v_1} = f_2(v_1) \times \frac{\partial F(v_1 | v_2)}{\partial v_1} = f_2(v_2) \times f(v_1 | v_2) = f(v_1, v_2)\,. \quad \text{(A.1)}$$

By recursion, this result generalizes to M>2 by working backwards from the M-th cross partial derivative. The general sequence of partial derivatives of F(…) is (differentiating w.o.l.o.g. in the order j=1,2,…,M):

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

$$\frac{\partial^r F(\mathbf{v})}{\partial v_1 \cdots \partial v_r} = f_1(v_1) \times \left\{ \prod_{k=2}^{r} f(v_k | v_1, \ldots, v_{k-1}) \right\} \times F_{-\{1, \ldots, r\}}(v_{r+1}, \ldots, v_M | v_1, \ldots, v_r), \ r = 2, \ldots, M\text{-}1$$

$$\frac{\partial^M F(\mathbf{v})}{\partial v_1 \cdots \partial v_M} = f(\mathbf{v})\,.$$

This result is trivial when the v_j are mutually independent, in which case all the conditioning arguments are irrelevant.

Alternatively (2) can be obtained directly using Leibniz's rule for differentiation of integrals whose limits depend on the variable of differentiation. Since $F(\mathbf{v}) = \int_{-\infty}^{v_M} \cdots \int_{-\infty}^{v_1} f(\mathbf{u}) du_1 \cdots du_M$ 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

$$\frac{\partial}{\partial v_j}\left(\int_{-\infty}^{v_M}\cdots\int_{-\infty}^{v_1} f(\mathbf{u})du_1\cdots du_M\right) = \int_{-\infty}^{v_M}\cdots\int_{-\infty}^{v_{j+1}}\int_{-\infty}^{v_{j-1}}$$

$$\cdots\int_{-\infty}^{v_1}\left(\frac{\partial}{\partial v_j}\int_{-\infty}^{v_1} f\left(u_1,\dots,u_M\right)du_j\right)du_1\cdots du_{j-1}du_{j+1}\cdots du_M$$

$$= \int_{-\infty}^{v_M}\cdots\int_{-\infty}^{v_{j+1}}\int_{-\infty}^{v_{j-1}}$$

$$\cdots\int_{-\infty}^{v_1}\left(f\left(u_1,\dots,u_{j-1},v_j,u_{j+1},\dots,u_M\right)\right)du_1\cdots du_{j-1}du_{j+1}\cdots du_M$$

$$= \int_{-\infty}^{v_M}\cdots\int_{-\infty}^{v_{j+1}}\int_{-\infty}^{v_{j-1}}$$

$$\cdots\int_{-\infty}^{v_1}\left(f\left(u_1,\dots,u_{j-1},u_{j+1},\dots,u_M\middle|v_j\right)\times f\left(v_j\right)\right)du_1\cdots du_{j-1}du_{j+1}\cdots du_M$$

$$= f\left(v_j\right)\times\int_{-\infty}^{v_M}\cdots\int_{-\infty}^{v_{j+1}}\int_{-\infty}^{v_{j-1}}$$

$$\cdots\int_{-\infty}^{v_1}\left(f\left(u_1,\dots,u_{j-1},u_{j+1},\dots,u_M\middle|v_j\right)\right)du_1\cdots du_{j-1}du_{j+1}\cdots du_M$$

$$= f\left(v_j\right)\times F\left(v_1,\dots,v_{j-1},v_{j+1},\dots,v_M\middle|v_j\right)$$

Analogous results appear in the literature on copula joint distribution functions (Frees and Valdez (1998); Trivedi and Zimmer (2005)) in which the joint distribution of **y** is represented in copula form as

$$C\left(F_1\left(y_1\right),\dots,F_M\left(y_M\right)\right) = C\left(u_1,\dots,u_M\right) = F(\mathbf{u})$$

with F_j denoting the marginal distribution function of y_j, with the u_j being marginally uniform variates. A familiar result in the bivariate copula literature is that $\partial C\left(u_1,u_2\right)/\partial u_1 = F\left(u_2 | u_1\right)$ This is essentially equivalent to (A.1) since uniform marginal densities satisfy $f_j\ (u_j)=1$. Note, however, that there are instances in the copula literature in which results like $\partial F\left(u_1,u_2\right)/\partial u_1 = F\left(u_2 | u_1\right)$ stated. In light of (A.1), this result in general does not hold unless $f_j(u_j)=1$

Appendix B:: Detailed Derivations for the Multivariate Probit Model

Let $s_{jp}=2k_{jp}-1$ so $s_{jp} \in \{-1,1\}$ and define correspondingly the M×M diagonal transformation matrixes $\mathbf{T}_p = \text{diag}\left[s_{jp}\right], p = 1,\dots,2^M, j = 1,\dots,M$ Define for each p the 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

$$\mathbf{Q}_p = \begin{bmatrix} 1 & s_{1p}s_{2p}\rho_{12} & \cdots & s_{1p}s_{Mp}\rho_{1M} \\ s_{1p}s_{2p}\rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ s_{1p}s_{Mp}\rho_{1M} & \cdots & & 1 \end{bmatrix} = \begin{bmatrix} 1 & \tau_{12p} & \cdots & \tau_{1Mp} \\ \tau_{12p} & 1 & & \vdots \\ \vdots & & \ddots & \\ \tau_{1Mp} & \cdots & & 1 \end{bmatrix}.$$

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

$$\Pr\left(y_1 = k_{1p}, \ldots, y_M = k_{Mp} \middle| \mathbf{x}\right) = \Phi_{\mathbf{Q}_p}\left(s_{1p}\mathbf{x}\,\beta_1, \ldots, s_{Mp}\mathbf{x}\,\boldsymbol{\beta}_M\right) = \Phi_{\mathbf{Q}_p}\left(\alpha_{1p}, \ldots, \alpha_{Mp}\right),$$

where $\Phi_{\mathbf{Q}}$ is the cumulative of an MVN($\mathbf{0}$,$\mathbf{Q}$) distribution with density $\varphi_{\mathbf{Q}}(\ldots)$ and $\alpha_{jp} = s_{jp}\mathbf{x}\boldsymbol{\beta}_j$. Using the transformed matrixes $\mathbf{Q}$ in place of the original correlation matrixes $\mathbf{R}$ streamlines the exposition since for each configuration p the outcome orthant probability can be described by a joint cumulative rather than by a notationally 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 the distribution of $\boldsymbol{\varepsilon}$ around the origin.

To obtain the MVP's marginal effects it thus suffices to obtain the particular expressions corresponding to the second line in (3). $f_j\left(c_j(\boldsymbol{\theta})\right)$ is a univariate N(0,1) density and $F_{-j}\left(c_1(\boldsymbol{\theta}), \ldots, c_{j-1}(\boldsymbol{\theta}), c_{j+1}(\boldsymbol{\theta}), \ldots, c_m(\boldsymbol{\theta}) | c_j(\boldsymbol{\theta})\right)$ is the cumulative of a conditional (M-1)-variate Multivariate normal distribution. The $c_j(\boldsymbol{\theta})$ in (3) 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 $d\left(s_j\mathbf{x}\,\boldsymbol{\beta}_j\right)/d\mathbf{x} = s_j\,\boldsymbol{\beta}_j$. Substituting into (14) $\phi(\ldots)$ for $f(\ldots)$, $\Phi(\ldots)$ for $F(\ldots)$, and α_{jp} for $c_j(\boldsymbol{\theta})$ gives:

$$\frac{\partial\,\Phi_{\mathbf{Q}_p}\left(\alpha_{1p}, \ldots, \alpha_{Mp}\right)}{\partial\mathbf{x}} = \sum_{j=1}^{M}\left\{\left(\frac{\partial\,\Phi_{\mathbf{Q}_p}\left(\alpha_{1p}, \ldots, \alpha_{Mp}\right)}{\partial\alpha_{jp}}\right) \times \left(\frac{\partial\alpha_{jp}}{\partial\mathbf{x}}\right)\right\}$$

$$= \sum_{j=1}^{M}\left\{\left(\phi\left(\alpha_{jp}\right) \times \Phi_{\mathbf{Q}_{p\{-j\}}}\left(\alpha_{1p}, \ldots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \ldots, \alpha_{Mp}\middle|\alpha_{jp}\right)\right) \times \left(s_{jp}\,\boldsymbol{\beta}_j\right)^T\right\}.$$

Given consistent estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{Q}}$, estimation of $\partial\,\Phi_{\mathbf{Q}_p}\left(\alpha_{1p}, \ldots, \alpha_{Mp}\right)/\partial\mathbf{x}$ is complicated only by evaluation of the term $\Phi_{\mathbf{Q}_{p\{-j\}}}\left(\alpha_{1p}, \ldots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \ldots, \alpha_{Mp} | \alpha_{jp}\right)$ The following result provides a basis for this calculation:

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

Suppose $\mathbf{z} = [z_1, \ldots, z_M] \sim MVN(\mathbf{0}, \boldsymbol{\Omega})$ Partition $\boldsymbol{\Omega}$ as $\begin{bmatrix} \omega_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}$ with ω_{11} scalar. Then $\mathbf{z}_{-1} = [z_2, \ldots, z_M]$ conditional on z1 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 $\mathbf{z}_{-j} = [z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_M]$ 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

$$\mathbf{z}_{-1} | z_1 \sim MVN\left(\begin{bmatrix} z_1\tau_{12p} \\ \vdots \\ z_1\tau_{1Mp} \end{bmatrix}, \begin{bmatrix} 1-\tau_{12p}^2 & \tau_{23p}-\tau_{12p}\tau_{13p} & \cdots & \tau_{2Mp}-\tau_{12p}\tau_{1Mp} \\ \tau_{23p}-\tau_{12p}\tau_{13p} & 1-\tau_{13p}^2 & & \vdots \\ \vdots & & \ddots & \\ \tau_{2Mp}-\tau_{12p}\tau_{1Mp} & \cdots & & 1-\tau_{1Mp}^2 \end{bmatrix}\right), \quad (8)$$

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

To obtain $\Phi_{\mathbf{Q}_{p\{-j\}}}\left(\alpha_{1p}, \ldots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \ldots, \alpha_{Mp} | \alpha_{jp}\right)$, define the (M-1)-vector of differences

$$\Delta_{-j,p} = \left[\left(\alpha_{1p} - \alpha_{jp}\tau_{1jp}\right), \ldots, \left(\alpha_{(j-1)p} - \alpha_{jp}\tau_{(j-1)jp}\right), \left(\alpha_{(j+1)p} - \alpha_{jp}\tau_{(j+1)jp}\right), \ldots, \left(\alpha_{Mp} - \alpha_{jp}\tau_{Mjp}\right)\right]^T, \quad (9)$$

and an $(M-1) \times (M-1)$ diagonal transformation matrix $\mathbf{H}_{jp} = diag_{k \neq j}\left[\left(\sqrt{1 - \tau_{jkp}^2}\right)^{-1}\right]$ Let $\mathbf{L}_{jp} = \mathbf{H}_{jp}\boldsymbol{\Delta}_{-j,p}$ be the corresponding (M-1)-vector of normalized differences. Then $\Phi_{\mathbf{Q}_{p\{-j\}}}\left(\alpha_{1p}, \ldots, \alpha_{(j-1)p}, \alpha_{(j+1)p}, \ldots, \alpha_{Mp} | \alpha_{jp}\right)$ can be computed by referring $\mathbf{L}_{jp}$ to $\Phi_{\mathbf{Z}, \boldsymbol{\Sigma}}(\ldots)$, which is the cumulative of an (M-1)-variate MVN($\mathbf{0}$,Σ) 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

$\left(\tau_{rcp} - \tau_{jrp}\tau_{jcp}\right)/\sqrt{\left(1 - \tau_{jrp}^2\right)\left(1 - \tau_{jcp}^2\right)}$ Let this matrix be denoted $\mathbf{V}_{jp}$. The results derived in this appendix now provide the basis for computing the quantities of interest in (5).

References

Cappellari L, Jenkins SP (2003) Multivariate probit regression using simulated maximum likelihood. The Stata Journal 3(3):278–294

Christofides LN, Stengos T, Swidinsky R (1997) On the calculation of marginal effects in the bivariate probit model. Economics Letters 54(3):203–208

Christofides LN, Stengos T, Swidinsky R (2000) Corrigendum. Economics Letters 68(3):339

Frees EW, Valdez EA (1998) Understanding relationships using copulas. North American Actuarial Journal 2(1):1–25

Greene WH (1996) Marginal effects in the bivariate probit model. NYU Stern School of Business Working Paper EC-96–11

Greene WH (1998) Gender economics courses in liberal arts colleges: further results. Journal of Economic Education 29(4):291–300

Greene WH (2004) Convenient estimators for the panel probit model: further results. Empirical Economics 29(1):21–47

Greene WH, Hensher DA (2010) Modeling ordered choices: A primer Cambridge University Press, Cambridge

Hajivassiliou V, McFadden D, Ruud P (1996) Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. Journal of Econometrics 72(1–2):85–134

Huguenin J, Pelgrin F, Holly A (2009) Estimation of multivariate probit models by exact maximum likelihood. University of Lausanne, IEMS Working Paper 09–02

Mullahy J (2016) Estimation of multivariate probit models via bivariate probit. Stata Journal (in press)

Rao CR (1973) Linear statistical inference and its applications, 2nd Edition. Wiley, New York

Trivedi PK, Zimmer DM (2005) Copula modeling: an introduction for practitioners. Foundations and Trends in Econometrics 1(1):1–111

Table 1

Multiple Chronic Conditions Model: Estimated Sample Average Marginal Effects

Outcomes (y=k)						Avg. Marginal Effects	
Hypertens.	Asthma	Depression	Hyperlipid.	Diabetes	Sample Freq.	Age	Schooling
0	0	0	0	0	.6133	−.011600	.003796
0	0	0	0	1	.0331	.000115	−.001320
0	0	0	1	0	.0141	.001100	.001213
0	0	0	1	1	.0036	.000203	−.000313
0	0	1	0	0	.0147	−.000483	.000031
0	0	1	0	1	.0012	.000005	−.000132
0	0	1	1	0	.0012	.000103	.000126
0	0	1	1	1	.0003	.000027	−.000049
0	1	0	0	0	.0386	−.000405	.000302
0	1	0	0	1	.0050	.000001	−.000068
0	1	0	1	0	.0052	.000055	.000105
0	1	0	1	1	.0015	.000013	−.000019
0	1	1	0	0	.0079	−.000054	.000028
0	1	1	0	1	.0015	−.0000003	−.000015
0	1	1	1	0	.0016	.000012	.000023
0	1	1	1	1	.0007	.000003	−.000006
1	0	0	0	0	.0782	.002439	−.001045
1	0	0	0	1	.0081	.000435	−.001145
1	0	0	1	0	.0107	.003227	.001019
1	0	0	1	1	.0026	.002522	−.001802
1	0	1	0	0	.0141	.000205	−.000123
1	0	1	0	1	.0019	.000050	−.000152
1	0	1	1	0	.0030	.000468	.000140
1	0	1	1	1	.0009	.000525	−.000380
1	1	0	0	0	.0566	.000141	−.000027
1	1	0	0	1	.0076	.000032	−.000090
1	1	0	1	0	.0156	.000301	.000155
1	1	0	1	1	.0033	.000307	−.000182
1	1	1	0	0	.0363	.000027	−.000010
1	1	1	0	1	.0059	.000007	−.000025
1	1	1	1	0	.0087	.000089	.000044
1	1	1	1	1	.0031	.000129	−.000078