NBER WORKING PAPER SERIES

ESTIMATION OF MULTIVARIATE PROBIT MODELS VIA BIVARIATE PROBIT

John Mullahy

Working Paper 21593
http://www.nber.org/papers/w21593

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015

I would like to thank Bill Greene, Stephen Jenkins, João Santos Silva, and an anonymous reviewer for helpful comments on earlier drafts. Support has been provided by NICHD Grant P2CHD047873 to UW-Madison's Center for Demography and Ecology and by the RWJF Health & Society Scholars Program at UW-Madison. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimation of Multivariate Probit Models via Bivariate Probit
John Mullahy
NBER Working Paper No. 21593
September 2015
JEL No. C3,I1

ABSTRACT

Models having multivariate probit and related structures arise often in applied health economics. When the outcome dimensions of such models are large, however, estimation can be challenging owing to numerical computation constraints and/or speed. This paper suggests the utility of estimating multivariate probit (MVP) models using a chain of bivariate probit estimators. The proposed approach offers two potential advantages over standard multivariate probit estimation procedures: significant reductions in computation time; and essentially unlimited dimensionality of the outcome set. The time savings arise because the proposed approach does not rely simulation methods; the dimension advantage arises because only pairs of outcomes are considered at each estimation stage. Importantly, the proposed approach provides a consistent estimator of all the MVP model's parameters under the same assumptions required for consistent estimation based on standard methods, and simulation exercises suggest no loss of estimator precision.

John Mullahy
University of Wisconsin-Madison
Dept. of Population Health Sciences
787 WARF, 610 N. Walnut Street
Madison, WI 53726
and NBER
jmullahy@facstaff.wisc.edu

A data appendix is available at:
http://www.nber.org/data-appendix/w21593

1. Introduction

Models having multivariate probit and related structures arise often in applied health economics (see Mullahy, 2011, for references). When the outcome dimensions of such models are large, however, estimation can be challenging owing to numerical computation constraints and/or speed.

This paper suggests the utility of estimating multivariate probit (MVP) models using a chain of bivariate probit estimators. It will be seen that the proposed approach, based on Stata's *biprobit* and *suest* procedures and driven by a Mata function *bvpmvp(...)*, affords two potential advantages over Stata's *mvprobit* procedure: significant reductions in computation time; and essentially unlimited dimensionality of the outcome set (*mvprobit*'s limit is M=20 outcomes).[1] The time savings arise because, unlike *mvprobit*, *bvpmvp(...)* does not rely simulation methods; the dimension advantage arises because only pairs of outcomes are considered at each estimation stage. Importantly, the proposed *bvpmvp(...)* approach provides a consistent estimator of all the MVP model's parameters under the same assumptions required for consistent estimation via *mvprobit*, and simulation exercises reported below suggest no loss of estimator precision relative to *mvprobit*.

The approach suggested here was inspired by the goal of embedding MVP estimation in a large-replication bootstrap exercise. The simulation results presented in Section 5 suggest that the computation time savings afforded by the *bvpmvp(...)* method relative to *mvprobit* can be significant while numerical differences in the respective point

[1] Stata SE's restriction that *matsize* cannot exceed 11,000 ultimately places a limit on the size of the parameter vector that can be estimated. All references to Stata herein are to Stata/SE, Version 13.1. Whether the results obtained here using Stata generalize to other statistical packages is an open question.

estimates and estimated standard errors are trivial. Since the potential applicability of MVP models is broad, it is valuable in practice that such potential not be thwarted by computational challenges.

The plan for the remainder of the paper is as follows. Section 2 describes the MVP model. Section 3 describes the *bvpmvp(...)* method. Section 4 describes the comparison empirical exercises. Section 5 presents the comparative results. Section 6 considers parallel issues involved in estimation of multivariate ordered probit models. Section 7 summarizes.

2. The Multivariate Probit Model

The multivariate probit model as typically specified is:

$$y^*_{ij} = \mathbf{x}_i \boldsymbol{\beta}_j + u_{ij} \tag{1}$$

$$y_{ij} = 1\left(y^*_{ij} > 0\right) \tag{2}$$

$$\mathbf{u}_i = \left[u_{i1}, \ldots, u_{iM}\right] \sim \mathrm{MVN}\left(\mathbf{0}, \mathbf{R}\right) \quad \text{or} \quad \mathbf{y}^*_i = \left[y_{i1}, \ldots, y_{iM}\right] \sim \mathrm{MVN}\left(\mathbf{x}_i \mathbf{B}, \mathbf{R}\right) \tag{3}$$

where i=1,...,N indexes observations, j=1,...,M indexes outcomes, $\mathbf{x}_i$ is a K-vector of exogenous covariates, the $\mathbf{u}_i$ are assumed to be iid independent across i but correlated across j for any i, and "MVN" denotes the multivariate normal distribution. (Henceforth the "i" subscripts will be suppressed.) The standard normalization sets the diagonal elements of **R** equal to 1 so that **R** is a correlation matrix with off-diagonal elements ρ_{pq},

$\{p,q\} \in \{1,...,M\}$, $p \neq q$.[2] With standard full rank conditions on the $\mathbf{x}$'s and each $|\rho_{pq}| < 1$ then $\mathbf{B} = [\beta_1,...,\beta_M]$ and $\mathbf{R}$ will be identified and estimable with sufficient sample variation in the $\mathbf{x}$'s.

3. Estimation and Inference

Estimation of the M-outcome multivariate probit model using *mvprobit* requires simulation of the MVN probabilities (Cappellari and Jenkins, 2003), with *mvprobit* computation time increasing in M, K, N, and simulation draws (D).[3] It turns out, however, that all the parameters (**B**,**R**) can be estimated consistently using bivariate probit -- implemented as Stata's *biprobit* procedure -- while consistent inferences about all these parameters are afforded via Stata's *suest* procedure. Since the proposed approach will be seen to be significantly faster in terms of computation time with no obvious disadvantages, this strategy may merit consideration in applied work.

The key result for the proposed estimation strategy is that the multivariate normal distribution is fully characterized by the mean vector **xB** and correlation matrix **R**. For present purposes, the key feature of the multivariate (conditional) normal distribution

[2] This normalization rules out cases like heteroskedastic errors (Wooldridge, 2010, section 15.7.4). While this normalization is common -- normalizing each univariate marginal to be a standard probit, for instance -- it is not the only possible normalization of the covariance matrix.

[3] Specifically, in the empirical exercises reported below as well as in some other simulations not reported here, it is found that *mvprobit* computation time increases: trivially in K; essentially proportionately in D; slightly more than proportionately in N; and at a rate between 2^M and 3^M in M. Greene and Hensher, 2010, suggest that MVP computation time would increase with 2^M but the results obtained in the simulations undertaken here suggest a somewhat greater rate of increase.

$F\left(y_1^*,\ldots,y_M^*|\mathbf{x}\right)$ is that all its bivariate marginals $F\left(y_j^*,y_m^*|\mathbf{x}\right)$ are bivariate normal with mean vectors and correlation matrixes corresponding to the respective submatrixes of $\mathbf{xB}$ and $\mathbf{R}$ (Rao, 1973, 8a.2.10).

Under the normalization that the diagonal elements of $\mathbf{R}$ are all one, the $\mathbf{B}$ parameters are identified based on knowledge of all M (conditional) univariate marginals $F\left(y_j^*|\mathbf{x}\right)$; there is no need to appeal to the multivariate features of $F\left(y_1^*,\ldots,y_M^*|\mathbf{x}\right)$ to identify **B.** The .5M(M-1) bivariate marginals provide the additional information about the ρ_{pq} parameters. As such, identification of the parameters of all the bivariate marginals implies identification[4] of the parameters of the full multivariate joint distribution so that consistent estimation of all the bivariate marginal probit models $\Pr\left(y_p=t_p,y_q=t_q|\mathbf{x}\right)$ provides consistent estimates of all the parameters $\left(\mathbf{B},\mathbf{R}\right)$ of the full multivariate probit model $\Pr\left(y_1=t_1,\ldots,y_M=t_M|\mathbf{x}\right)$ for $t_j\in\left\{0,1\right\}$, j=1,...,M.

Estimation via Bivariate Probit

The proposed approach, which can be implemented using the Mata function *bvpmvp(...)* described below, is as follows. First, corresponding to each possible outcome pair, $.5M\left(M-1\right)$ bivariate probit models are estimated using *biprobit* yielding a single

[4] As discussed below, identification of all the bivariate marginals implies overidentification of **B**.

estimate[5] of each ρ_{pq} and M-1 estimates of each β_j, j=1,...,M. Each of the M-1 estimates of β_j is itself consistent since each *biprobit* specification uses the same normalization on the relevant submatrixes of **R**. Each of these estimates $\left(\widehat{\beta_p},\widehat{\beta_q},\widehat{\rho_{pq}}\right)_b$, b=1,...,.5M$\left(M-1\right)$, is stored and then combined using Stata's *suest* procedure, which provides a consistent estimate of the joint variance-covariance matrix of all $M\left(M-1\right)\left(.5+K\right)$ parameters estimated with the $.5M\left(M-1\right)$ *biprobit* estimates. Denote this vector of parameter estimates and its estimated variance-covariance matrix as $\hat{\alpha}$ and $\hat{\Omega}$, respectively.[6]

Second, the simple averages $\hat{\beta}_{jA}=\left(\frac{1}{M-1}\right)\sum_{\substack{m=1\\m\neq j}}^{M}\hat{\beta}_{jm}$ are computed. This gives a $k\times M$ matrix of estimated averaged coefficients, denoted $\widehat{\mathbf{B}_A}=\left[\widehat{\beta_{1A}},...,\widehat{\beta_{MA}}\right]$. Since a weighted average of consistent estimators is in general a consistent estimator, the resulting $\widehat{\mathbf{B}_A}$ will itself be consistent for **B**. This averaging arises because the **B** parameters in the proposed approach are overidentified, i.e. there are M-1 consistent estimates of each β_j, j=1,...,M. One could use some other rule to compute a single consistent estimate of each β_j from among the M-1 candidates, but unless alternative strategies could boast significant precision gains, computational simplicity recommends the simple average as an obvious

---

[5] *biprobit* estimates directly the inverse hyperbolic tangent of ρ_{pq} or $.5\ln\left(\left(1+\rho_{pq}\right)/\left(1-\rho_{pq}\right)\right)$.

[6] $\hat{\alpha}$ and $\hat{\Omega}$ are the *suest* stored matrix results e(b) (a row vector) and e(V), respectively.

solution. See the Appendix for further discussion.

Finally, let $\mathbf{Q}$ denote the $.5M(M-1)$ vector of the $\tanh^{-1}(\rho_{jk})$ estimated in each *biprobit* specification, and define the $M(.5(M-1)+K)\times 1$ vector $\widehat{\boldsymbol{\Theta}}=\left[\operatorname{vec}\left(\widehat{\mathbf{B}_A}\right)^T, \widehat{\mathbf{Q}}^T\right]^T$.

Define $\mathbf{H}$ as the $M(.5(M-1)+K)\times M(M-1)(.5+K)$ averaging and selection matrix that maps $\widehat{\boldsymbol{\alpha}}$ to $\widehat{\boldsymbol{\Theta}}$, i.e. $\widehat{\boldsymbol{\Theta}}=\mathbf{H}\widehat{\boldsymbol{\alpha}}^T$; the elements of $\mathbf{H}$ are 1/(M-1), one, or zero.[7] The estimated variance-covariance matrix of $\widehat{\boldsymbol{\Theta}}$, useful for inference, is given by $\widehat{\operatorname{var}}\left(\widehat{\boldsymbol{\Theta}}\right)=\mathbf{H}\widehat{\boldsymbol{\Omega}}\mathbf{H}^T$.

bvpmvp(...)*: A Mata Function to Implement the Proposed Estimation Approach*

The function *bvpmvp(...)* returns the $M(k+.5(M-1))\times(M(k+.5(M-1))+1)$ matrix whose first column is $\widehat{\boldsymbol{\Theta}}^T$ and whose remaining elements are the $M(k+.5(M-1))$-dimension symmetric square matrix $\widehat{\operatorname{var}}\left(\widehat{\boldsymbol{\Theta}}\right)$. *bvpmvp(...)* takes six arguments: (1) a string containing the names of the M outcomes; (2) a string containing the

[7] A general form of the $\mathbf{H}$ matrix is complicated to express concisely. As an example, for M=3 and K=2 the 9×15 $\mathbf{H}$ matrix, computed internally by *bvpmvp(...)*, is

$$\mathbf{H}=\begin{bmatrix} .5 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

names of the K-1 non-constant covariates; (3) a (possibly null) string containing any "if" conditions for estimation; (4) a scalar indicating whether or not to display the interim estimation results; (5) a scalar indicating the rounding level of presented results; and (6) a scalar indicating whether or not to display the final results. For example:

```
bv1 = bvpmvp("y1 y2 y3 y4","x1 x2 x3 x4","if _n<=10000",0,.001,1)
bv2 = bvpmvp(yn,xn,ic,0,.001,1)
```

bvpmvp(...)'s summary report displays the $\widehat{\mathbf{B}_{\mathrm{A}}}$ estimates, their estimated standard errors, and the estimated correlation matrix $\widehat{\mathbf{R}}$; an example is provided in Exhibit 1. Of course, suppression of these results may be useful, for instance, in simulation or bootstrapping exercises. The do file containing the Mata code for *bvpmvp(...)* is available with this paper's supplementary materials.

4. Simulation Exercises

To assess the relative performance of the proposed approach and the approach based on *mvprobit* a simulation exercise was conducted. Three sample sizes (N=2,000, N=10,000, N=50,000) are considered. The data structure corresponding to (1)-(3) has either K=5 or K=9 covariates **x** (four or eight independently distributed uniform variates plus a constant) and M=8 binary outcomes y_{ij} (only four of which are used in some specifications) corresponding to latent y^*_{ij} having cross-outcome correlations ρ_{jk} variously in $\left\{.2,\ 1/\sqrt{10},\ .5\right\}$ for all $j \neq k$, specifically

$$
\mathbf{R} = \begin{bmatrix}
1 & & & & & & & \\
10^{-.5} & 1 & & & & & & \\
.5 & 10^{-.5} & 1 & & & \text{(symm.)} & & \\
10^{-.5} & .2 & 10^{-.5} & 1 & & & & \\
.5 & 10^{-.5} & .5 & 10^{-.5} & 1 & & & \\
10^{-.5} & .2 & 10^{-.5} & .2 & 10^{-.5} & 1 & & \\
.5 & 10^{-.5} & .5 & 10^{-.5} & .5 & 10^{-.5} & 1 & \\
10^{-.5} & .2 & 10^{-.5} & .2 & 10^{-.5} & .2 & 10^{-.5} & 1
\end{bmatrix}
$$

For *mvprobit*, the *draws(.)* option was set both at 10 and 20. The simulations are performed using Stata/SE Version 13.1 on an iMac 3.4GHz Intel Core i7 processor and OS X v10.8.[8].

5. Simulation Results

Key results of the simulations are summarized in Tables 1-3. Table 1 displays the absolute and relative computation times for *mvprobit* and *bvpmvp(...)* estimation across the various combinations of the N, M, K, and D parameters. Enormous differences in computation time are seen between the two estimation methods across all the different parameter combinations (for reference, it may be useful to recall that there are 86,400 seconds in one day). Tables 2 and 3 present a side-by-side comparison of the point estimates of **B** and **R** obtained in one select specification (N=10,000, M=4, K=5). For both **B** and **R** the differences between the *mvprobit* and *bvpmvp(...)* point estimates and

[8] The simulations set Stata's *matsize* parameter at 600 for all specifications. In some preliminary investigation, it was observed that computation time for *bvpmvp(...)* increased significantly when *matsize* was set much larger than necessary; this was not the case for *mvprobit*.

corresponding estimated standard errors are trivial.

In light of these results, use of methods like *bvpmvp(...)* to estimate MVP models merits consideration when computation time is an important consideration.[9]

6. Multivariate Ordered Probit Models

Analogous conceptual considerations arise in the context of multivariate ordered probit (MVOP) models in which the observed ordered outcomes are $yo_j \in \{0,...,G_j\}$ for finite integers $G_j \geq 1$. MVOP modeling involves estimation of and inference about the parameters **B** and **R** as well as the vector of category cutpoints, **C** (for each outcome yo_j there are G_j cutpoints that delineate the G_j+1 categories).[10]

---

[9] It should be noted that these simulations paint what is in some sense a "worst-case" picture for *mvprobit* estimation. The simulations use *mvprobit* "out of the box," i.e. without specifying any options that might enhance estimation speed (see the Stata "help" file for *mvprobit* and also Cappellari and Jenkins, 2003 and 2006). For instance, specifying a smaller number of draws (e.g. *draws(3)* or *draws(5)*) would clearly result in faster estimation times; any diminished performance of the *mvprobit* estimator relative to the performance at greater number of draws would be a potential consideration, however. Alternatively, using good starting values for **R** via *mvprobit*'s *atrho0(.)* option might also be expected to result in faster estimation times. One such approach would involve two stages: (1) estimate the full model using *mvprobit* with a small number of draws, e.g. *draws(1)* or *draws(2)*; and (2) use the estimate of **R** thus obtained to provide starting values for a second *mvprobit* estimation with a larger number of draws (e.g. *draws(10)* or *draws(20)*) being specified. This approach -- with *draws(1)* specified initially, followed by *draws(10)* -- was examined in some simulations. It was observed in this instance that the two-stage approach resulted in roughly a 10% reduction in overall estimation time, due mainly to a smaller number of iterations (three vs. four) required for convergence in the second stage.

This paper also has not considered how estimation using Stata's *cmp* procedure to estimate the MVP model would compare with the *bvpmvp(...)* approach.

I would like to thank Stephen Jenkins and an anonymous referee for their insights and suggestions on these matters.

[10] For the MVOP model **B** will not contain a parameter for the constant term since this is absorbed into the cutpoints **C**.

An estimation strategy fully analogous to *bvpmvp(...)* is not available since the *bioprobit* procedure (Sajaia, 2008) does not permit postestimation prediction with the *score* option, as required by *suest*. However, an alternative, fully consistent, and computationally efficient approach is available, as follows. First, estimate M univariate ordered probit models using Stata's *oprobit* procedure and store these results using *estimates store*. This provides consistent estimates of the **B** and **C** parameters. Second, estimate a chain of bivariate binary probit models using *biprobit* -- as with *bvpmvp(...)* -- and store these estimates using *estimates store*. This provides a consistent estimate of **R**.[11] Note that any thresholds used to map the ordered yo_{ij} to their corresponding coarsened binary outcomes should result in consistent estimates of **R**. *biprobit* uses the rule that a nonbinary outcome is treated as zero for zero values and one otherwise; this is a convenient mapping that minimizes programming burden. Third, combine all the estimates stored in these two steps using *suest*. The estimates from *suest* can then be used for inference. The do file containing the Mata code for the function *bvopmvop(...)* that implements this approach is available with this paper's supplementary materials.[12] An example of *bvopmvop(...)* output is presented in Exhibit 2.[13]

---

[11] Note that this also provides consistent estimates of **B**, but these are unnecessary given those obtained in the first step.

[12] *bvopmvop(...)* accommodates ordered outcomes having different numbers of cutpoints, including mixed ordered and binary outcomes. The single cutpoint estimated in *oprobit* for binary outcomes is -1 times the corresponding constant term that would be estimated using *probit*.

[13] The outcomes in this example are ordered versions yo_j of the y_j used in the earlier simulations in which the outcome value 2 is assigned if $1 \le y_j^* \le 2$ and 3 is assigned if $y_j^* > 2$. Then y_2 combines the top two categories and y_3 combines the top three categories (i.e. y_3 is the original binary measure). Thus, the numbers of categories are G_1=4, G_2=3, G_3=2, and G_4=4.

7. Summary

This paper has presented a novel estimation strategy for consistent estimation of and inference about the parameters of MVP and MVOP models. The straightforward implementation of these approaches using available Mata programs recommends their consideration in applied work, particularly in situations involving large numbers of outcomes (M), large sample sizes (N), or in situations requiring repeated MVP estimation like bootstrapping exercises.

In closing, it should be noted that the methods suggested here may prove useful in many but not all applications of multivariate probit models. Ultimately the methods proposed here -- as well as the *mvprobit* method -- permit estimation of the joint conditional probability model $\Pr\left(\mathbf{y}=\mathbf{k}\middle|\mathbf{x}\right)$ for the M-vectors of outcomes $\mathbf{y}$, all possible 2^M vectors $\mathbf{k}=[k_m]$, $k_m \in \left\{0,1\right\}$, and exogenous covariates $\mathbf{x}$. As such, when these joint conditional probabilities are *per se* the estimands of interest, when they are instrumentally of interest in the estimation of other quantitites (see Mullahy, 2011, for discussion), or when reduced forms of structural models are of interest, the approach suggested here may prove useful. However in other MVN contexts with binary outcomes -- e.g. where endogenous y_m are RHS variables in the structural models for other latent y_j^* -- consistent estimation of the structural parameters will typically demand attention to the full joint probability structure, not just its bivariate marginals.[14]

[14] Thanks are owed to an anonymous reviewer for emphasizing these points.

References

Cappellari, L. and S.P. Jenkins. 2003. "Multivariate Probit Regression Using Simulated Maximum Likelihood." *Stata Journal* 3: 278–294.

Cappellari, L. and S.P. Jenkins. 2006. "Calculation of Multivariate Normal Probabilities by Simulation, with Applications to Maximum Simulated Likelihood Estimation." *Stata Journal* 6: 156-189.

Greene, W.H. and D.A. Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.

Mullahy, J. 2011. "Marginal Effects in Multivariate Probit and Kindred Discrete and Count Outcome Models, with Applications in Health Economics." NBER W.P. 17588.

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

Sajaia, Z. 2008. *BIOPROBIT: Stata Module for Bivariate Ordered Probit Regression*. Boston College Department of Economics, Statistical Software Components, No. S456920.

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

Appendix: Additional Remarks on Combining *biprobit* Estimates

In general, the optimal approach to combining such multiple estimates in the overidentified case is to use a minimum-distance estimator with an optimal weight matrix (Wooldridge, 2010, section 14.5). In the present context this would amount to computing a weighted average for each point estimate, i.e. $\hat{\beta}_{jkw} = \sum_{\substack{m=1 \\ m \neq j}}^{M} w_{jkm} \hat{\beta}_{jkm}$, j=1,...,M, k=1,...,K. Implementing the minimum-distance approach can be computationally challenging, however. For example, consider the simplest case, M=3. The optimal (variance-minimizing) weights even in this instance are complicated functions of the estimates' variances and covariances; suppressing the j,k subscripts, for $\{p,q,r\} \in \{1,2,3\}$, $p \neq q \neq r$ these optimal weights are:

$$w_r = \frac{\sigma_{pp}\sigma_{qq} - \sigma_{pq}^2 - \sigma_{qq}\sigma_{pr} - \sigma_{pp}\sigma_{rq} - \sigma_{pr}\sigma_{pq} - \sigma_{pq}\sigma_{rq}}{\sigma_{pp}\sigma_{rr} + \sigma_{rr}\sigma_{qq} + \sigma_{pp}\sigma_{qq} - \sigma_{pr}^2 - \sigma_{pq}^2 - \sigma_{rq}^2 + 2\left(\sigma_{pr}\sigma_{pq} + \sigma_{pq}\sigma_{rq} + \sigma_{pr}\sigma_{rq} - \sigma_{pp}\sigma_{rq} - \sigma_{rr}\sigma_{pq} - \sigma_{qq}\sigma_{pr}\right)}$$

where $\sigma_{\bullet\bullet}$ are variances and covariances of the parameter estimates (the empirical counterpart, $\widehat{w_r}$, would use $\widehat{\sigma_{\bullet\bullet}}$). The algebraic complexity of these weights increases rapidly as M increases.

The considerable additional computational complexity involved in implementing such a minimum-distance approach is unlikely to provide much benefit (in terms of precision) unless the optimal w_{jkm} were to diverge dramatically from 1/(M-1). The simulations undertaken here suggest this is unlikely to be the case. In general the optimal weights will diverge from the equi-weighted case of 1/(M-1) to the extent that the variances and covariances of and between the parameter point estimates differ substantively across the (M-1) estimates.[15]

---

[15] Bill Greene suggested to me that a computationally straightforward middle-ground weighting strategy would be to, in essence, ignore the cross-estimator covariances and compute the variance-matrix-weighted quantities:

$$\widehat{\beta_{jv}} = \left[\sum_{\substack{m=1 \\ m \neq j}}^{M} \left[\widehat{\mathrm{var}}\left(\widehat{\beta_m}\right) \right]^{-1} \right]^{-1} \times \sum_{\substack{m=1 \\ m \neq j}}^{M} \left[\widehat{\mathrm{var}}\left(\widehat{\beta_m}\right) \right]^{-1} \widehat{\beta_m}, \quad j=1,\ldots,M.$$

For illustrative purposes, selecting arbitrarily the (M-1) point estimates corresponding to the parameter β_{11} (outcome y_1, covariate x_1) for the N=10,000, M=8 and K=5 specification, the range of the seven point estimates $\widehat{\beta_{11}}$ is [.3266, .3288], the range of the corresponding seven estimated point estimate variances is [.001983, .001995], and the range of the 28 estimated point estimate covariances is [.001983, .001993]. It is thus unlikely that the optimal weights would diverge much from 1/(M-1).

The ultimately important result is that at least insofar as the simulations conducted for this paper are concerned, the differences between the *mvprobit* and *bvpmvp(...)* point estimates and estimated standard errors are inconsequentially small (see Tables 2 and 3).

Table 1 Estimation Time Comparisons (in Seconds)						
Parameters				Computation Time		Relative Difference (Ratio)
N	M	K	D	*mvprobit*	*bvpmvp(...)*	
2,000	4	5	10	29	1	29
			20	53		53
		9	10	28	1	28
			20	54		54
	8	5	10	1,219	5	244
			20	2,041		408
		9	10	1,036	8	130
			20	2,044		256
10,000	4	5	10	142	2	71
			20	263		132
		9	10	137	3	46
			20	258		86
	8	5	10	4,628	14	331
			20	10,469		748
		9	10	4,669	19	246
			20	9,833		518
50,000	4	5	10	986	12	82
			20	1,937		161
		9	10	995	18	55
			20	1,970		109
	8	5	10	35,833	65	551
			20	72,406		1114
		9	10	36,647	86	426
			20	73,204		851
Legend N: Number of sample observations M: Number of outcomes K: Number of covariates (including constant term) D: Number of draws for *mvprobit*						
Note: Stata's *matsize* parameter is set at 600 for all specifications.						

Table 2 *mvprobit* and *bvpmvp(...)* Comparison: $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{B}_{\mathbf{A}}}$ Point Estimates, One Example (N=10,000, M=4, K=5; Estimated Standard Errors in Parentheses)			
Outcome	Covariate	*mvprobit* (draws=20)	*bvpmvp(...)*
y_1	x_1	0.3265 (.0448)	.3279 (.0446)
	x_2	-0.3301 (.0447)	-.3314 (.0447)
	x_3	0.3184 (.0447)	.3198 (.0449)
	x_4	-0.3902 (.0448)	-.3916 (.0447)
	Constant	0.3901 (.0466)	.3909 (.0464)
y_2	x_1	-0.4487 (.0456)	-.4487 (.0455)
	x_2	0.5624 (.0458)	.5620 (.0456)
	x_3	-0.3998 (.0457)	-.3977 (.0457)
	x_4	0.4000 (.0456)	.3961 (.0457)
	Constant	-0.5086 (.0474)	-.5079 (.0474)
y_3	x_1	0.3102 (.0445)	.3151 (.0446)
	x_2	0.3846 (.0445)	.3875 (.0449)
	x_3	-0.3188 (.0446)	-.3206 (.0447)
	x_4	-0.3462 (.0446)	-.3496 (.0447)
	Constant	0.3230 (.0463)	.3210 (.0463)
y_4	x_1	0.4567 (.0455)	.4573 (.0457)
	x_2	-0.4438 (.0455)	-.4408 (.0457)
	x_3	-0.4489 (.0456)	-.4516 (.0457)
	x_4	0.4555 (.0456)	.4499 (.0453)
	Constant	-0.4552 (.0472)	-.4524 (.0472)

Table 3

mvprobit and *bvpmvp(…)* Comparison: $\widehat{\mathbf{R}}$ Point Estimates, One Example

(N=10,000, M=4, K=5; Estimated Standard Errors in Parentheses)

R	*mvprobit* (draws=20)	*bvpmvp(…)*
ρ_{12}	.3190 (.0158)	.3308 (.0159)
ρ_{13}	.4942 (.0134)	.5073 (.0134)
ρ_{14}	.2766 (.0160)	.2872 (.0161)
ρ_{23}	.3356 (.0156)	.3424 (.0158)
ρ_{24}	.2000 (.0163)	.2034 (.0167)
ρ_{34}	.3059 (.0157)	.3086 (.0160)

Exhibit 1: Sample Output from *bvpmvp(…)* (N=10,000, M=4, K=5)

```
. mata
---------------------------------- mata (type end to exit) ----------
----------------
: yn="y1 y2 y3 y4"
: xn="x1 x2 x3 x4"
: ic="if _n<=10000"
: bv1=bvpmvp(yn,xn,ic,1,.001,1)

  ********************************************
  *                                          *
  *       Multivariate Probit: Results       *
  *                                          *
  ********************************************

  N. of Observations (from suest): 10000

  Estimation Sample: if _n<=10000

  Averaged Beta-Hat Point Estimates and Estimated Standard Errors

             1        2        3        4        5
     +-----------------------------------------------+
   1 |                y1       y2       y3       y4  |
   2 |                                               |
   3 |      x1     .328    -.449     .315     .457   |
   4 |           (.045)   (.046)   (.045)   (.046)   |
   5 |                                               |
   6 |      x2    -.331     .562     .388    -.441   |
   7 |           (.045)   (.046)   (.045)   (.046)   |
   8 |                                               |
   9 |      x3      .32    -.398    -.321    -.452   |
  10 |           (.045)   (.046)   (.045)   (.046)   |
  11 |                                               |
  12 |      x4    -.392     .396     -.35      .45   |
  13 |           (.045)   (.046)   (.045)   (.045)   |
  14 |                                               |
  15 |   _cons     .391    -.508     .321    -.452   |
  16 |           (.046)   (.047)   (.046)   (.047)   |
  17 |                                               |
     +-----------------------------------------------+
```

(continued)

Exhibit 1 (continued)

```
Estimated Correlation (Rho) Matrix and Estimated Standard Errors

           1         2         3         4         5
   +-----------------------------------------------+
 1 |                  y1        y2        y3        y4  |
 2 |                                                |
 3 |        y1         1      .331      .507      .287  |
 4 |                        (.016)    (.013)    (.016)  |
 5 |                                                |
 6 |        y2      .331         1      .342      .203  |
 7 |              (.016)              (.016)    (.017)  |
 8 |                                                |
 9 |        y3      .507      .342         1      .309  |
10 |              (.013)    (.016)              (.016)  |
11 |                                                |
12 |        y4      .287      .203      .309         1  |
13 |              (.016)    (.017)    (.016)            |
14 |                                                |
   +-----------------------------------------------+
```

```
Cut & Paste Matrix, Averaged Beta-Hat Point Estimates

(.328 , -.449 , .315 , .457) \
(-.331 , .562 , .388 , -.441) \
(.32 , -.398 , -.321 , -.452) \
(-.392 , .396 , -.35 , .45) \
(.391 , -.508 , .321 , -.452)
```

```
Cut & Paste Matrix, Estimated Correlation Matrix

(1 , .331 , .507 , .287) \
(.331 , 1 , .342 , .203) \
(.507 , .342 , 1 , .309) \
(.287 , .203 , .309 , 1)
```

Exhibit 2: Sample Output from *bvopmvop(...)* (N=10,000, M=4, K=5)

```
. mata
------------------------------- mata (type end to exit) ----------
----------------
: yn="y1o y2o y3o y4o"
: xn="x1 x2 x3 x4"
: ic="if _n<=10000"
: bv2=bvopmvop(yn,xn,ic,1,.001,1)

  ******************************************************
  *                                                    *
  *       Multivariate Ordered Probit: Results         *
  *                                                    *
  ******************************************************

  N. of Observations (from suest): 10000

  Estimation Sample: if _n<=10000

  Beta-Hat and Cutpoint Point Estimates and Estimated Standard Errors
     (Note: SEs are from suest ests.)

                1         2         3         4         5
     +-----------------------------------------------+
   1 |                  y1o       y2o       y3o       y4o  |
   2 |                                                 |
   3 |        x1      .379     -.457      .316      .464  |
   4 |               (.038)    (.043)    (.045)    (.043)  |
   5 |                                                 |
   6 |        x2     -.325       .53      .388      -.44  |
   7 |               (.038)    (.044)    (.045)    (.043)  |
   8 |                                                 |
   9 |        x3      .338     -.404     -.321     -.471  |
  10 |               (.038)    (.043)    (.045)    (.043)  |
  11 |                                                 |
  12 |        x4     -.393      .397     -.348       .45  |
  13 |               (.038)    (.043)    (.045)    (.043)  |
  14 |                                                 |
  15 |      cut1     -.354      .485     -.319      .447  |
  16 |                (.04)    (.045)    (.046)    (.045)  |
  17 |                                                 |
  18 |      cut2      .356     1.379        --     1.305  |
  19 |                (.04)    (.047)              (.047)  |
  20 |                                                 |
  21 |      cut3     1.079        --        --      2.18  |
  22 |               (.041)                        (.054)  |
  23 |                                                 |
     +-----------------------------------------------+
```

(continued)

Exhibit 2 (continued)

```
Estimated Correlation (Rho) Matrix and Estimated Standard Errors

            1         2         3         4         5
    +----------------------------------------------+
  1 |                y1o       y2o       y3o       y4o |
  2 |                                                  |
  3 |      y1o         1      .331      .507      .287 |
  4 |                       (.016)    (.013)    (.016) |
  5 |                                                  |
  6 |      y2o      .331         1      .342      .203 |
  7 |             (.016)              (.016)    (.017) |
  8 |                                                  |
  9 |      y3o      .507      .342         1      .309 |
 10 |             (.013)    (.016)              (.016) |
 11 |                                                  |
 12 |      y4o      .287      .203      .309         1 |
 13 |             (.016)    (.017)    (.016)           |
 14 |                                                  |
    +----------------------------------------------+

Cut & Paste Matrix, Beta-Hat and Cutpoint Point Estimates

(.379 , -.457 , .316 , .464) \
(-.325 , .53 , .388 , -.44) \
(.338 , -.404 , -.321 , -.471) \
(-.393 , .397 , -.348 , .45) \
(-.354 , .485 , -.319 , .447) \
(.356 , 1.379 , . , 1.305) \
(1.079 , . , . , 2.18)

Cut & Paste Matrix, Estimated Correlation Matrix

(1 , .331 , .507 , .287) \
(.331 , 1 , .342 , .203) \
(.507 , .342 , 1 , .309) \
(.287 , .203 , .309 , 1)
```