Multivariate Versus Multinomial Probit:
When are Binary Decisions Made Separately also Jointly Optimal?

Dale J. Poirier and Deven Kapadia

University of California, Irvine

March 10, 2012

Abstract

We provide an analysis of the question in the title in terms of a bivariate probit framework representing two (possibly correlated) separate decisions, and a multinomial probit framework representing the four possible outcomes viewed as one joint decision. We offer a Bayesian treatment that builds on Weeks and Orme (1998) and Di Tommaso and Weeks (2000) who showed that the bivariate probit corresponds to a *singular* four-dimensional multinomial probit under testable restrictions. We also discuss extensions to trivariate and quadrivariate probit.

Key words: Bayesian, posterior odds, multivariate probit, multinomial probit

1. Introduction

The observation that a multivariate binomial model can be recast as a multinomial model has been long known [e.g., Cox (1972, p. 115), Amemiya (1981, p. 1525), Velandia et al. (2009)]. But all of the implications of the nesting are not widely understood. For example, in the case of ***J-dimensional multivariate probit (J-MVP)*** nested within a ***M-dimensional multinomial probit (M-MNP)***, where $M = 2^J$, Amemiya (1981, p. 1525) states: "Therefore, the theory of statistical inference I discussed in regard to a multi-response model in Section 3.A is valid for a multivariate model without modification."

In under-appreciated contributions, Weeks and Orme (1998) and Di Tommaso and Weeks (2000) added an important qualifier that necessitates a modification to Amemiya's remark. Consider the case of ***bivariate probit (2-MVP)*** in which the four outcomes are viewed as a ***4-dimensional multinomial probit (4-MNP)***. Weeks and Orme (1998) point out that 2-MVP corresponds to a

singular multinomial model which lies on the boundary of the permissible parameter space. This irregularity can affect asymptotic distribution theory for likelihood ratio and Wald tests [Chernoff (1954)]. In contrast, Moran (1971) and Chant (1974) showed that the asymptotic properties of the score/LM test under such "non-standard" conditions are unaffected. This led Weeks and Orme (1998) to propose the score test as a specification test for 2-MVP. Di Tomasso and Weeks (2000) implement this test for labor force participation and fertility decisions. The present paper offers a alternative Bayesian posterior odds framework and extends it to higher dimensional caes.

The intuition behind Weeks and Orme is (1998) and Di Tommaso and Weeks (2000) is straightforward. 2-MVP consists of two equations involving two shocks. 4-MNP involves three shocks (its identified structure is trinomial). Hence, a singularity is required in the trinomial specification in order to equate to a binary structure. This singularity provides a specification test for BVP within the more general 4-MNP.

While nesting 2-MVP inside 4-MNP is of general interest, generalization of 2-MVP in the direction of 4-MNP is particularly attractive when the two discrete outcomes of the 2-MVP correspond to two decisions by the ***same*** individual. In 2-MVP the two decisions are made separately, but not necessarily independently. If the 2-MVP can be shown to be equivalent to 4-MNP, then the two univariate decisions are also ***jointly*** optimal.

2. Underlying Latent Variable Models

Consider n = 1, ..., N independent observations on the latent multivariate normal regression

$$y_n^* = (I_J \otimes x_n')\beta + u_n, \quad u_n|X_n \sim \mathbb{N}_J(0_J, \Omega), \tag{1}$$

where $y_n^* = [y_{n1}^*, ..., y_{nJ}^*]'$, x_n is a K×1 vector of covariates, $X_n = (I_J \otimes x_n')$ is a J×JK matrix, $\beta = [\beta_1', ..., \beta_J']'$, Ω is a J×J covariance matrix, and $\mathbb{N}_J(0_J, \Omega)$ denotes a J-dimensional normal distribution with zero mean and covariance $\Omega = [\omega_{ij}]$. (1) is the latent model underlying J-dimensional MVP model in which y_n^* (n = 1, ..., N) are not observed, but only their component signs

$$y_{nj} = \begin{Bmatrix} 1, & \text{if } y_{nj}^* > 0 \\ 0, & \text{if } y_{nj}^* \leq 0 \end{Bmatrix}. \tag{2}$$

Next consider n = 1, ..., N independent observations on the M-dimensional latent multivariate normal regression

$$z_n^* = (I_M \otimes x_n')\gamma + \varepsilon_n, \quad \varepsilon_n|X_n \sim \mathbb{N}_M(0_M, \Sigma), \tag{3}$$

where $z_n^* = [z_{n1}^*, ..., z_{nM}^*]'$, $\gamma = [\gamma_1', ..., \gamma_M']'$, and Σ is a M×M covariance matrix. This is the latent variable model underlying M-MNP in which agents construct latent Gaussian utilities and select the category that corresponds to the largest utility, i.e., z_n^* (n = 1, ..., N) are not observed, but rather only which is the maximum:

$$z_n = m, \text{ if } z_{nm}^* = \max\{z_{n1}^*, ..., z_{nM}^*\} \quad (n = 1, ..., N). \tag{4}$$

3. Identification and Priors

The latent variable models (1) and (3) underlying the J-MVP and the M-MNP models, respectively, are identified if the latent data are observable, but they are unidentified if only the discrete choices (2) and (4) are observed. Priors (dogmatic or proper) provide vehicles for adding additional information. Two broad approaches have been suggested: (i) add dogmatic restrictions to some parameters in order to yield the remaining parameters identified [e.g., McCulloch et al. (2000) and Nobile (2000)], and (ii) ignore the identifiability problem, perform the analysis on the unidentified model with proper priors, and post-process samples by scaling to obtain draws on identified parameters [e.g., McCulloch and Rossi (1994), Nobile (1988), Imai and van Dyk (2005a), Edwards and Allenby (2003)]. The first approach is common in the case of J-MVP [e.g., Chib and Greenberg (1998)]. In the case of M-MNP, both approaches have been used - even by the same authors [e.g., McCulloch and Rossi (1994), McCulloch et al. (2000)].

Approach (i) is used in non-Bayesian analyses. For example, the scale identification problem for J-MVP can be eliminated by setting all the diagonal elements of Ω equal to one, denoting the restricted matrix as $\tilde{\Omega} = [\tilde{\omega}_{ij}]$. This identifies β but changes the interpretation of the off-diagonal elements of $\tilde{\Omega}$ to correlations rather than covariances. Thus, in a Bayesian setting, a prior on these off-diagonal elements needs to be adjusted before imposing the diagonal restrictions. Let $\psi = [\psi_1, ..., \psi_J]'$, $\psi_j \in \{0, 1\}$(J = 1, ..., J), denote the support of $y_n = [y_{n1}, ..., y_{nJ}]'$. The J-MVP choice probability is

$$\pi_{n\psi_1 \cdots \psi_J} = \text{Prob}(y_n = \psi \,|\, \beta, \tilde{\Omega}) = \int_{\Psi(\psi)} \phi_J(y_n^* \,|\, (I_J \otimes x_n')\beta, \tilde{\Omega}) \; dy_n^*, \tag{5}$$

where $\phi_J(\cdot)$ is the J-dimension multivariate normal density, and the region of integration is $\Psi(\psi) = \Psi(\psi_1) \times \ldots \times \Psi(\psi_J)$ with

$$\Psi(\psi_j) = \begin{Bmatrix} (-\infty, 0], & \text{if } \psi_j = 0 \\ (0, \infty), & \text{if } \psi_j = 1 \end{Bmatrix} \;\; (j = 1, \ldots, J). \tag{6}$$

The likelihood function is $\mathfrak{L}^{MVP}(\beta, \tilde{\Omega}; y) = \prod_{n=1}^{N} \pi_{n\psi_1 \cdots \psi_J}$, where $y = [y_1', \ldots, y_J']'$.

M-MNP suffers from both location and scale identification problems. The location identification problem is addressed by differencing and measuring effects relative to one of the choices, in effect setting the location of one choice outcome to zero. For example, subtracting the M^{th} equation from the first $M - 1$ equations leads to

$$\tilde{z}_n^* = (I_{M-1} \otimes x_n')\tilde{\gamma} + \tilde{\varepsilon}_n, \quad \tilde{\varepsilon}_n | X_n \sim \mathbb{N}_{M-1}(0_{M-1}, \tilde{\Sigma}), \tag{7}$$

where $\tilde{z}_n^* = [z_{n1}^* - z_{nM}^*, \ldots, z_{n(M-1)}^* - z_{nM}^*]'$, $\tilde{\gamma} = [(\gamma_1 - \gamma_M)', \ldots, (\gamma_{M-1} - \gamma_M)']'$, $\tilde{\varepsilon}_n = [\varepsilon_{n1} - \varepsilon_{nM}, \ldots, \varepsilon_{n(M-1)} - \varepsilon_{nM}]'$, $\tilde{\Sigma} = [\tilde{\sigma}_{ij}] = [I_{M-1}, -\iota_{M-1}]\,\Sigma\,[I_{M-1}, -\iota_{M-1}]'$, and $\iota_{M-1} = [1, \ldots, 1]'$. Since the ordering of the utilities is invariant to additive shift or multiplicative rescaling, identifying assumptions on the location and scale are needed. The normalization $\gamma_M = 0_K$ and $\tilde{\sigma}_{11} = 1$ identifies the remaining parameters. The M-MNP choice probability is

$$p_{nj} = \text{Prob}(z_n = j \,|\, \gamma, \tilde{\Sigma}) = \int_{A_j} \phi_{M-1}(\tilde{z}_n^* \,|\, (I_{M-1} \otimes x_n')\tilde{\gamma}, \tilde{\Sigma}) \; d\,\tilde{z}_n^*, \tag{8}$$

where the region of integration is given by

$$A_j = \begin{Bmatrix} \left\{ \tilde{z}_n^* \,|\, \tilde{z}_{nj}^* > \max(\tilde{z}_{-nj}^*, 0) \right\}, & \text{if } j < M \\ \left\{ \tilde{z}_n^* \,|\, \tilde{z}_{nj}^* < 0 \right\}, & \text{if } j = M \end{Bmatrix} \;\; (j = 1, \ldots, J), \tag{9}$$

and z_{-nj}^* denotes all element in z_n^* except the j^{th}. The likelihood function is $\mathfrak{L}^{MNP}(\gamma; \tilde{\Sigma}; z) = \prod_{n=1}^{N} p_{n z_n}$.

For M-MNP Burgette and Nordheim (2010) showed that the choice of which element you fix can have a meaningful impact on posterior predictions in the case of the prior used by Imai and van Dyk (2005a): $\tilde{\gamma} \sim \mathbb{N}(\underset{\sim}{g}, \underline{A}^{-1})$ independent of $\tilde{\Sigma}$ with the latter having an inverted Wishart distribution with degrees of freedom $\underline{\nu} > M - 1$, positive definite scale matrix $\underline{S} = [\underline{s}_{ij}]$, and marginal density

$$f(\tilde{\Sigma}|\underline{S}, \underline{\nu}) \propto |\tilde{\Sigma}|^{-(\underline{\nu}+M)/2} \exp[\mathrm{tr}(\underline{S}\tilde{\Sigma}^{-1})]^{-\underline{\nu}(M-1)/2} \tag{10}$$

subject to $\tilde{\sigma}_{11} = 1$ and $\underline{s}_{11} = 1$. Basically, the choice of prior for the identified parameters should not be invariant to how the scale problem is solved. To avoid this problem, Burgette and Nordheim (2010) proposed a model that identifies the scale of the model by fixing the trace of the covariance matrix, which makes the prior covariance invariant to joint permutations of the rows and columns. Burgette and Nordheim (2010) also showed that Bayesian M-MNP predictions can be sensitive to the specification of the base choice, here, the M^{th}. This problem exists because, instead of specifying a prior for the original utilities and inducing a prior on the base-subtracted utilities, it has been customary to specify a prior directly on base-subtracted utilities.

Burgette and Nordheim (2011) propose two models that do not have this base category problem. They describe a Bayesian M-MNP whose prior and likelihood (and therefore posterior) can be symmetric with respect to relabeling the outcome categories, unless prior knowledge suggests a desired deviation from that symmetry. Rather than selecting a reference category whose utility is assumed to be equal to zero, they enforce a series of "sum-to-zero" restrictions, both on the unit-level errors and on the regression coefficients. If respondents choose from M categories, other MNP methods transform the utilities to (M-1)-dimensional space. Instead, Burgette and Nordheim constrain utilities to exist in a (M-1)-dimensional hyperplane in M-space.

4. Nesting MVP in MNP

To see how to nest a J-MVP model in a M-MNP model, first consider the case J = 2 and M = 4 considered by Weeks and Orme (1998). For n = 1, ..., N suppose

$$\tilde{z}^*_{n3} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2}. \tag{11}$$

Singularity (11) implies the $K+3$ restrictions

$$\tilde{\gamma}_3 = \tilde{\gamma}_1 + \tilde{\gamma}_2, \quad \tilde{\sigma}_{i3} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2} \ (i = 1, 2, 3). \tag{12}$$

In addition we impose the testable restriction

$$\tilde{\sigma}_{22} = 1. \tag{13}$$

Under (12) and (13), 4-MNP reduces to 2-MVP. To see this, suppose $z_n = 1$. Then (6) implies $\tilde{z}^*_{n1} > \tilde{z}^*_{n2}$, $\tilde{z}^*_{n1} > \tilde{z}^*_{n3}$, and $\tilde{z}^*_{n1} > 0$. Substituting (10) into the second inequality implies $0 > \tilde{z}^*_{n2}$. These last two inequalities imply the first two. Proceeding similarly, $z_n = 2$ implies $\tilde{z}^*_{n2} > 0$ and $0 > \tilde{z}^*_{n1}$, $z_n = 3$ implies $\tilde{z}^*_{n1} > 0$ and $\tilde{z}^*_{n2} > 0$, and $z_n = 4$ implies $0 > \tilde{z}^*_{n1}$ and $0 > \tilde{z}^*_{n2}$. This shows the singular 4-MNP reduces to 2-MVP with

$$p_{n1} = \pi_{n10}, \quad p_{n2} = \pi_{n01}, \quad p_{n3} = \pi_{n11}, \quad p_{n4} = \pi_{n00}. \tag{14}$$

In other words, the two binary decisions are also optimal in the quadnomial context. The parameters of the 2-MVP are given by $\beta_1 = \tilde{\gamma}_1$, $\beta_2 = \tilde{\gamma}_2$, and $\tilde{\omega}_{12} = \tilde{\sigma}_{12}$.

In the case J = 3 and M = 8 the number of singularities required increases noticeably over the previous case because of the need to reduce the seven dimensions of 8-MNP down to the three dimensions of trivariate probit (3-MVP):

$$\tilde{z}^*_{n4} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2}, \quad \tilde{z}^*_{n5} = \tilde{z}^*_{n1} + \tilde{z}^*_{n3}, \quad \tilde{z}^*_{n6} = \tilde{z}^*_{n2} + \tilde{z}^*_{n3}, \quad \tilde{z}^*_{n7} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2} + \tilde{z}^*_{n3}, \tag{15}$$

for n = 1, ..., N. The singularities in (15) imply the $4K + 28$ restrictions

$$\tilde{\gamma}_4 = \tilde{\gamma}_1 + \tilde{\gamma}_2, \quad \tilde{\gamma}_5 = \tilde{\gamma}_1 + \tilde{\gamma}_3, \quad \tilde{\gamma}_6 = \tilde{\gamma}_2 + \tilde{\gamma}_3, \quad \tilde{\gamma}_7 = \tilde{\gamma}_1 + \tilde{\gamma}_3 + \tilde{\gamma}_3, \tag{16}$$

$$\tilde{\sigma}_{i4} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2}, \quad \tilde{\sigma}_{i5} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i3}, \quad \tilde{\sigma}_{i6} = \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3}, \quad \tilde{\sigma}_{i7} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3} \tag{17}$$

for i = 1, ..., 7. Analogous to (13), we also impose the testable restrictions

$$\tilde{\sigma}_{22} = \tilde{\sigma}_{33} = 1. \tag{18}$$

Then it can be shown that (16) - (18) imply the singular 8-MNP reduces to 3-MVP with

$$p_{n1} = \pi_{n100}, \quad p_{n2} = \pi_{n010}, \quad p_{n3} = \pi_{n001}, \ p_{n4} = \pi_{n110}, \tag{19}$$

$$p_{n5} = \pi_{n101}, \quad p_{n6} = \pi_{n011}, \quad p_{n7} = \pi_{n111}, \ p_{n8} = \pi_{n000}. \tag{20}$$

The parameters of the 3-MVP are $\beta_1 = \tilde{\gamma}_1$, $\beta_2 = \tilde{\gamma}_2$, $\beta_3 = \tilde{\gamma}_3$, $\omega_{12} = \tilde{\sigma}_{12}$, $\omega_{13} = \tilde{\sigma}_{13}$, and $\omega_{23} = \tilde{\sigma}_{23}$.

Finally, the case J = 4 and M = 16 involves the eleven singularities.

$$\tilde{z}^*_{n5} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2}, \quad \tilde{z}^*_{n6} = \tilde{z}^*_{n1} + \tilde{z}^*_{n3}, \quad \tilde{z}^*_{n7} = \tilde{z}^*_{n1} + \tilde{z}^*_{n4}, \quad \tilde{z}^*_{n8} = \tilde{z}^*_{n2} + \tilde{z}^*_{n3}, \tag{21}$$

$$\tilde{z}^*_{n9} = \tilde{z}^*_{n2} + \tilde{z}^*_{n4}, \quad \tilde{z}^*_{n10} = \tilde{z}^*_{n3} + \tilde{z}^*_{n4}, \quad \tilde{z}^*_{n11} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2} + \tilde{z}^*_{n3}, \quad \tilde{z}^*_{n12} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2} + \tilde{z}^*_{n4}, \tag{22}$$

$$\tilde{z}^*_{n13} = \tilde{z}^*_{n1} + \tilde{z}^*_{n3} + \tilde{z}^*_{n4}, \quad \tilde{z}^*_{n14} = \tilde{z}^*_{n2} + \tilde{z}^*_{n3} + \tilde{z}^*_{n4}, \quad \tilde{z}^*_{n15} = \tilde{z}^*_{n1} + \tilde{z}^*_{n2} + \tilde{z}^*_{n3} + \tilde{z}^*_{n4}, \tag{23}$$

$$\tilde{\gamma}_5 = \tilde{\gamma}_1 + \tilde{\gamma}_2, \quad \tilde{\gamma}_6 = \tilde{\gamma}_1 + \tilde{\gamma}_3, \quad \tilde{\gamma}_7 = \tilde{\gamma}_1 + \tilde{\gamma}_4, \quad \tilde{\gamma}_8 = \tilde{\gamma}_2 + \tilde{\gamma}_3, \quad \tilde{\gamma}_9 = \tilde{\gamma}_2 + \tilde{\gamma}_4, \quad \tilde{\gamma}_{10} = \tilde{\gamma}_3 + \tilde{\gamma}_4, \tag{24}$$

$$\tilde{\gamma}_{11} = \beta_1 + \beta_2 + \beta_3, \quad \tilde{\gamma}_{12} = \beta_1 + \beta_2 + \beta_4, \quad \tilde{\gamma}_{13} = \beta_1 + \beta_3 + \beta_4, \quad \tilde{\gamma}_{14} = \beta_2 + \beta_3 + \beta_4, \tag{25}$$

$$\tilde{\gamma}_{15} = \beta_1 + \beta_2 + \beta_3 + \beta_4, \tag{26}$$

$$\tilde{\sigma}_{i5} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2}, \quad \tilde{\sigma}_{i6} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i3}, \quad \tilde{\sigma}_{i7} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i4}, \quad \tilde{\sigma}_{i8} = \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3}, \tag{27}$$

$$\tilde{\sigma}_{i9} = \tilde{\sigma}_{i2} + \tilde{\sigma}_{i4}, \quad \tilde{\sigma}_{i10} = \tilde{\sigma}_{i3} + \tilde{\sigma}_{i4}, \quad \tilde{\sigma}_{i11} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3}, \quad \tilde{\sigma}_{i12} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2} + \tilde{\sigma}_{i4}, \tag{28}$$

$$\tilde{\sigma}_{i13} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i3} + \tilde{\sigma}_{i4}, \quad \tilde{\sigma}_{i14} = \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3} + \tilde{\sigma}_{i4}, \quad \tilde{\sigma}_{i15} = \tilde{\sigma}_{i1} + \tilde{\sigma}_{i2} + \tilde{\sigma}_{i3} + \tilde{\sigma}_{i4}, \tag{29}$$

for n = 1, ..., N. Singularities (21) - (23) imply the restrictions for (i = 1, ..., 15). Analogous to (13) and (18), we also impose the testable restrictions

$$\tilde{\sigma}_{22} = \tilde{\sigma}_{33} = \tilde{\sigma}_{44} = 1. \tag{30}$$

Then it can be shown that (24) - (30) imply the singular 16-MNP reduces to 4-MVP. The parameters of the 4-MVP are given by $\beta_1 = \tilde{\gamma}_1$, $\beta_2 = \tilde{\gamma}_2$, $\beta_3 = \tilde{\gamma}_3$, $\beta_4 = \tilde{\gamma}_4$, $\omega_{12} = \tilde{\sigma}_{12}$, $\omega_{13} = \tilde{\sigma}_{13}$, $\omega_{14} = \tilde{\sigma}_{14}$, $\omega_{23} = \tilde{\sigma}_{23}$, $\omega_{24} = \tilde{\sigma}_{24}$, and $\omega_{34} = \tilde{\sigma}_{34}$. It can be shown that (16) - (18) imply the singular 8-MNP reduces to 3-MVP with

$$p_{n1} = \pi_{n100}, \quad p_{n2} = \pi_{n010}, \quad p_{n3} = \pi_{n001}, \quad p_{n4} = \pi_{n110}, \tag{31}$$

$$p_{n5} = \pi_{n101}, \quad p_{n6} = \pi_{n011}, \quad p_{n7} = \pi_{n111}, \quad p_{n8} = \pi_{n000}. \tag{32}$$

5. Estimation

The development of MCMC techniques led to an explosion of Bayesian analysis of limited dependent variable models in the 1990s. Early Albert and Chib (1993) recognized the data augmentation introduced by Tanner and Wong (1987) had major implications for computation. Such ideas were expanded upon by Meng and van Dyk (1999) and van Dyk and Meng (2001). Rather than struggle with computing multivariate normal integrals appearing in the likelihood function, instead it was possible to generate samples from the posterior *without* computing such integrals [also see Geweke et al. (1994) for an early contribution]. Imai and van Dyk (2005) provide a nice summary discussion of the literature in the context of MNP.

6. Specification Tests

Let $H_{0,J}$ denote the hypothesis that a 2^J-MNP model reduces to a J-MVP model and $H_{1,J}$ denote the hypothesis that it does not. We denote the prior densities for all unknown parameters by $f(\beta,\Omega|H_{0,J})$ and $f(\tilde{\gamma},\tilde{\Sigma}_{-11}|H_{1,J})$, where $\tilde{\Sigma}_{-11}$ denotes all the elements in $\tilde{\Sigma}$ excluding the identifying restriction $\tilde{\sigma}_{11}=1$. Recall we have imposed the normalization $\gamma_M = 0_K$. The Bayes factor in favor of $H_{1,J}$, $B_{10} = f(y|H_{1,J})/f(z|H_{1,J})$, where

$$f(y|H_0) = \iint f(\beta,\Omega)\,\mathcal{L}^{MVP}(\beta,\Omega;y)\ d\beta\, d\Omega, \tag{33}$$

$$f(z|H_1) = \iint f(\tilde{\gamma},\tilde{\Sigma}_{-11})\,\mathcal{L}^{MNP}(\tilde{\gamma},\tilde{\Sigma}_{-11};z)\ d\tilde{\gamma}\, d\Sigma_{-11}. \tag{34}$$

Note that it is also of interest to consider the case starting with M-MNP model and asking whether it has a singular covariance matrix giving rise to a sub-model which has a MVP structure. But it is not necessary that $M = 2^J$. For example, Figures 1-3 display the tree structure of 2-MVP, 4-MNP, and a mixed model with M = 6 where the last four categories have a binary structure that renders 2-MVP plausible for them. From (7),

$$\tilde{z}_n^* = (I_5 \otimes x_n{}')\tilde{\gamma} + \tilde{\varepsilon}_n, \quad \tilde{\varepsilon}_n|X_n \sim \mathbb{N}_5(0_5,\tilde{\Sigma}), \tag{35}$$

where $\tilde{z}_n^* = [z_{n1}^* - z_{n6}^*, ..., z_{n5}^* - z_{n6}^*]'$, $\tilde{\gamma} = [(\gamma_1-\gamma_6)', ..., (\gamma_5-\gamma_6)']'$, $\tilde{\varepsilon}_n = [\varepsilon_{n1}-\varepsilon_{n6}, ..., \varepsilon_{n5}-\varepsilon_{n6}]'$, $\tilde{\Sigma} = [\tilde{\sigma}_{ij}] = [I_5, -\iota_5]\Omega[I_5, -\iota_5]'$. Proceeding analogous to (11) - (13), the singularity

$$\tilde{z}_{n5}^* = \tilde{z}_{n3}^* + \tilde{z}_{n4}^*, \tag{36}$$

implies the $K+3$ restrictions

$$\tilde{\gamma}_5 = \tilde{\gamma}_3 + \tilde{\gamma}_4, \quad \tilde{\sigma}_{i5} = \tilde{\sigma}_{i3} + \tilde{\sigma}_{i4}\ (i = 3,4,5). \tag{37}$$

In addition we impose the testable restriction

$$\tilde{\sigma}_{44} = 1. \tag{38}$$

Under (37) and (38), the structure for z_{nm} (m = 3, 4, 5, 6) reduces to a 2-MVP. The parameters of this 2-MVP are given by $\beta_3 = \tilde{\gamma}_3$, $\beta_4 = \tilde{\gamma}_4$, and $\omega_{34} = \tilde{\sigma}_{34}$. The 6-MNP model has 5K + 14 parameters: $\tilde{\gamma}_m$ (m = 1, ..., 5) and $\tilde{\sigma}_{12}, ..., \tilde{\sigma}_{55}$. The model subject to the K + 4 restrictions in (37) and (38) has the unrestricted parameters $\tilde{\gamma}_m$ (m = 1, ..., 4) and $\tilde{\sigma}_{12}, ..., \tilde{\sigma}_{34}, \tilde{\sigma}_{15}$, and $\tilde{\sigma}_{25}$.

6. Empirical example

7. Discussion

Identification can be fragile in MNP unless there are exclusion restrictions on the $\tilde{\boldsymbol{\gamma}}_m$ (m = 1, ..., M) across choices [Keane (1992)]. Researchers have tried to add further structure and eliminate some parameters by putting restrictions on Ω and $\tilde{\boldsymbol{\Sigma}}$, e.g., exchangeability across equations or a factor structure. The singularity restrictions discussed here provide alternative restrictions that may be attractive when a subset of the multinomial choices have a binary structure. Furthermore, the restrictions are numerous and increase rapidly as M increases.

References

Amemiya, T. 1981. Qualitative response models: a survey. Journal of Economic Literature XIX: 1483-1536.

Burgette LF, Hahn PR. 2011. Symmetric Bayesian multinomial probit models. Duke University, unpublished manuscript.

Burgette LF, Nordheim EV. 2010. The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model. Duke University, unpublished manuscript.

Chant D. 1974. On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika* 61: 291-298.

Chernoff H. 1954. On the distribution of the likelihood ratio. Annals of Mathematical Statistics 25: 573-578.

Cox DR. 1972. Analysis of multivariate binary data. Journal of the Royal Statistical Society, Series

C 21: 113-120.

Di Tomasso ML,Weeks M. 2000. Decision structures and discrete choices: an application to labour market participation and fertility. unpublished manuscript.

Geweke J, Keane M, Runkle D. 1994. Alternative computational approaches to inference in the multinomial probit model. Review of Economics and Statistics 76: 609–632.

Godfrey LG. 1991. *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*. Cambridge University Press: Cambridge.

Imai K, van Dyk DA. 2005. A Bayesian analysis of the multinomial probit model using marginal data augmentation. Journal of Econometrics 124: 311 – 334.

McCulloch R, Rossi P. 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207–240.

McCulloch R, Rossi P. 2000. Reply to Nobile. Journal of Econometrics 99, 347–348.

McCulloch R, Polson NG, Rossi P. 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. Journal of Econometrics 99, 173–193.

Meng X-L, van Dyk DA. 1999. Seeking efficient data augmentation schemes via conditional and marginal augmentation. Biometrika 86: 301–320.

Moran PAP. 1971. Maximum likelihood estimation in non-standard conditions. *Proceedings of the Cambridge Philosophical Society* 70: 441-450.

Nobile A. 1998. A hybrid Markov chain for the Bayesian analysis of the multinomial probit model.

Statistics and Computing 8: 229–242.

Nobile A. 2000. Comment: Bayesian multinomial probit models with normalization constraint. Journal of Econometrics 99: 335–345.

Tanner MA, Wong WH. 1987. The calculation of posterior distributions by data augmentation (with discussion). Journal of the American Statistical Association 82: 528–550.

van Dyk DA, Meng X-L. 2001. The art of data augmentation (with discussions). Journal of Computational and Graphical Statistics 10: 1–111.

Velandia MR, Rejesus M, Knight TO, Sherrick BJ. 2009. Factors affecting farmers' utilization of agricultural risk management tools: the case of crop insurance, forward contracting, and spreading sales. *Journal of Agricultural and Applied Economics* 41: 107-123.

Weeks M, Orme C. 1998. The statistical relationship between bivariate and multinomial choice models. Working Paper, Department of Applied Economics, University of Cambridge.

Figure 1: Tree diagram for 2-MVP

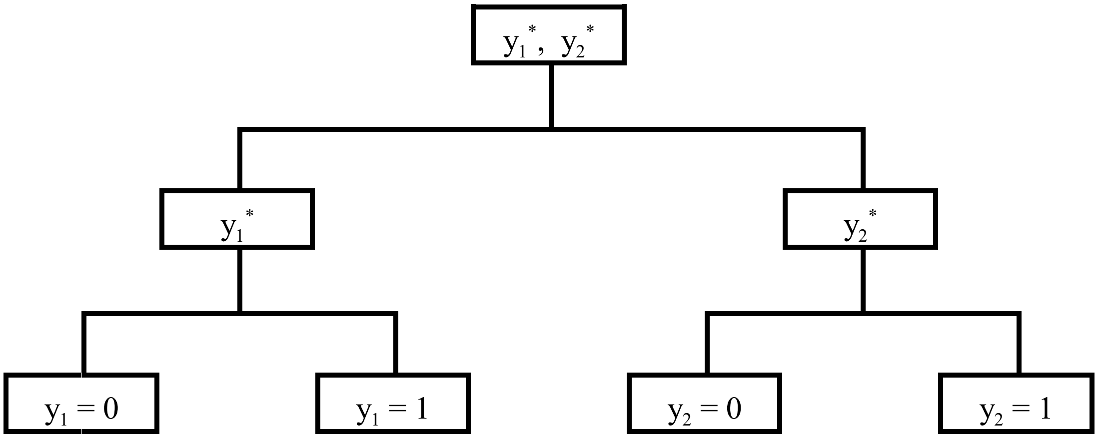

Figure 2: Tree diagram for 4-MNP

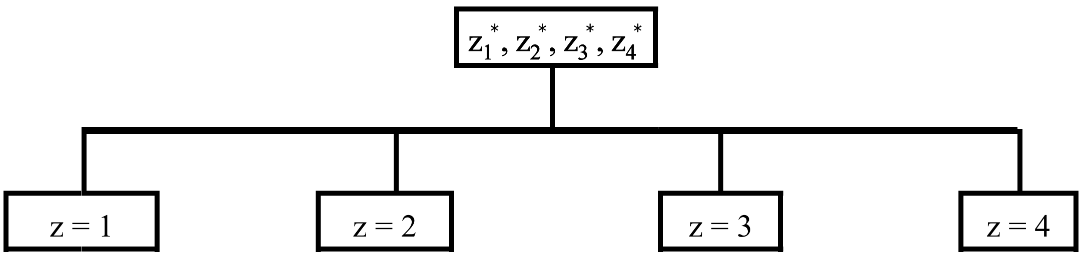

Figure 3: Tree diagram for Mixed Case with M = 6