The Statistical Relationship Between Bivariate and Multinomial Choice Models

MELVYN WEEKS* and Chris Orme†

January 1998

Abstract

We demonstrate the conditions under which the bivariate probit model can be considered a special case of the more general multinomial probit (MNP) model. Since the attendant parameter restrictions produce a singular covariance matrix, the subsequent problems of testing on the boundary of the parameter space are circumvented by the construction of a score test.

Acknowledgements: We are grateful to Paul Ruud and Christian Gourieroux for valuable comments and suggestions.

*Department of Applied Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DE, United Kingdom. Tel: +44 (1)223 335260; EMail: Melvyn.Weeks@econ.cam.ac.uk.

†School of Economic Studies, University of Manchester, Manchester M13 9PL.

1. Introduction

The bivariate and multinomial probit models can both be thought of as natural extensions of the simple (binomial) probit model. As Greene (1993) points out, in formulating the bivariate probit (BVP) model two 'probit', or binary, regression equations are defined with correlated disturbances, yielding a discrete analogue of the seemingly unrelated regression model in which the dependent variable is a two dimensional vector of binary outcomes. The multinomial probit (MNP) model, on the other hand, might appear to be appropriate for a rather different problem in which the dependent variable of interest is the unique choice made from a set of distinct alternatives. In this note, we point out that the BVP model might usefully be interpreted as a special case of a MNP model. This is achieved by recognising that the two binary equations (or decisions), as specified in the BVP approach, generates *four* mutually exclusive outcomes (or alternatives), which can be modelled in an MNP framework. Formally, it can then be shown that the MNP model nests the BVP model, via parametric restrictions, and an appreciation of this may be of use to applied workers in formulating appropriate parametric specifications since, in many published studies involving discrete choice, both the BVP and multinomial models (such as the MNP and multinomial logit (MNL)) have been employed without apparent consideration of the statistical and behavioural relationship that exist between the two different model structures; see, however, some discussion in Pudney (1989) (pp.122-3). This is of some interest, since a commonly used behavioural/economic interpretation of the MNP model would be that the observed 'joint' decision is made on the basis of utility maximisation over the four possibilities that the bivariate problem can generate. Consequently, the BVP model is itself, consistent with such utility maximising behaviour when the appropriate parametric restrictions, imposed on the more general MNP specification, are satisfied. This suggests the availability of a formal test of the BVP specification which might also be used to assess whether or not this behavioural interpretation, accommodated by the MNP model, is data consistent.

The outline of the paper is as follows: In section 2, the basic parametric BVP and MNP models are outlined and contrasted. In section 3, the set of parametric restrictions are derived, for a MNP model, which yields the BVP as a special case and an appropriate score test derived. In constructing such a test care must be taken since the restrictions produce a singular covariance matrix in the implied MNP specification. The subsequent problems of testing on the boundary of the parameter space are circumvented by the construction of a score test. Section 4 concludes.

2. Bivariate and Multinomial Models

Cox (1972) describes a number of approaches for analysing multivariate binary outcomes. As the number of binary random variables (h) under consideration increases, a simplistic method is to treat each of the variables as independent. In economics, assuming a parametric mean and additive error, we are then faced with h independent binary probit (or logit) regressions. With $h = 2$, dependence is easily introduced via the BVP model and has been used extensively in a number of fields, including economics and bioassay. A different approach is to treat the 2^h outcomes as being generated from a multinomial distribution, which gives the multinomial counterparts of the simple probit and logit models.

The bivariate model is popular in biological assay and was first introduced by Ashford and Sowden (1970) in recognition of the dependence between distinct physiological systems. The authors, following Mantel (1966), are careful to point out the distinction between a model which recognises dependence over two distinct systems which determine 4 mutually exclusive outcomes, and a single system providing four outcomes with an arbitrary dependence structure. (The multivariate and multinomial models extend this to consider dependence over h systems and the 2^h mutually exclusive outcomes, respectively.) Apart from the comments by McFadden (1981), it appears that the precise nature of this distinction is not widely appreciated in economics. Greene (1993) (p.913) notes that the multivariate models are distinct from the multinomial choice models in that for the former, the focus is upon the modelling of two or more decisions, with each decision involving two alternatives, whereas in the latter case there is a single decision among two or more alternatives. Notwithstanding this behavioural distinction, in the next section it is demonstrated how the MNP model nests BVP model.

To begin with, however, let us re-examine the basic statistical relationship between bivariate and multinomial choice models. Let $(y^*, z^*)'$ be a vector of scalar random variables, each with support on the real line, and define $y = \mathbf{1}(y^* > 0)$ and $z = \mathbf{1}(z^* > 0)$, in which $\mathbf{1}(.)$ is the indicator function; i.e., $y = 1$ if $y^* > 0$, and is zero otherwise. Based upon these relationships we may enumerate four mutually exclusive outcomes for the pair (y, z). Let the outcome be denoted by the discrete random variable s, taking on values $j = 1, \ldots, 4$, to which a value measure (or utility level), denoted v^*, is assigned. This is illustrated in Table 1:

Table 1:

y	z	s	v^*
1	1	1	v_1^*
1	0	2	v_2^*
0	1	3	v_3^*
0	0	4	v_4^*

Although the above table represents a highly abstract system, there exist many examples from inter alia bioassay and economics where the relationship between two (or more) Bernoulli random variables and the associated states is important.

Example 1

In economics y might represent the discrete decision (D_1) whether or not to work. This decision might be based upon the respective utilities of being in or out of the labour force, denoted y_1^* and y_0^*, with the implication that $y^* = (y_1^* - y_0^*)$.

Further, z might denote the decision (D_2) whether or not to take up formal (i.e., paid) child-care (again based upon a comparison of underlying utilities, denoted z_1^* and z_0^* with $z^* = z_1^* - z_0^*$). The two decisions may be correlated and, in this instance, the discrete random variable s enumerates the four possible combinations of: work, not work $\left(W, \bar{W}\right)$, child-care, not child-care $\left(C, \bar{C}\right)$ outcomes. This is illustrated in Figure 1(a). Here the probability that an individual both works and utilises paid child-care is given by the probability $\Pr(y^* > 0 \cap z^* > 0)$. In Figure 1(b) we consider a somewhat different structural framework, for the same problem, in which a single decision (D) yields one of 4 possible outcomes, but based upon a ranking of the utilities (or, the implied value measure). In this multinomial model, and with reference to Table 1, the probability of working *and* using paid child-care is given by the probability $\Pr(\bigcap\limits_{j=2}^{4} v_1^* - v_j^* > 0)$.

Example 2

In bioassay y might indicate the discrete response of a subject after taking a particular drug. However, since it is known that subjects may exhibit both a main effect and a number of other possibly related side effects, it might be necessary to monitor a secondary response, z. The observed response is then the joint realisation of the two underlying physiological systems. For example, Fay (1957) documents research into the respiratory systems of coalminers. Two symptoms were examined: breathlessness and wheeze. For each symptom we may consider a continuum of exposure (say y^* and z^*), where beyond certain thresholds, the subject is said to be suffering from the condition. Since each response function has two levels (afflicted, not afflicted), the two binary variables (say y and z) define four possible states.

Figure 1 (a)
Bivariate Model

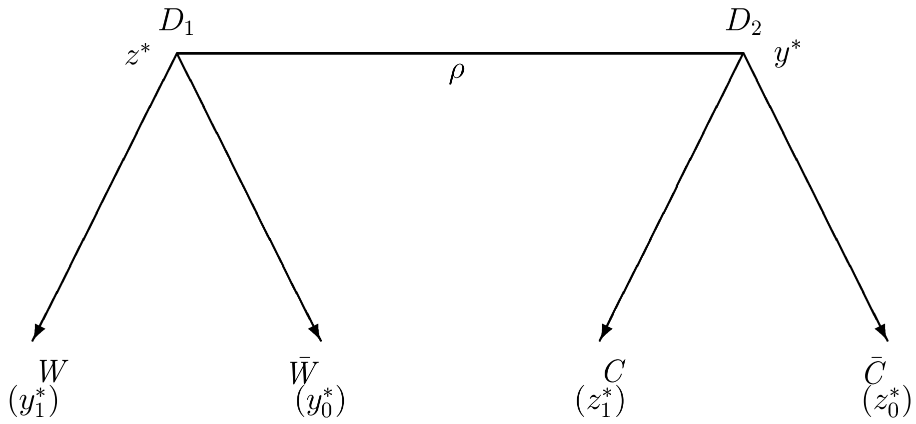
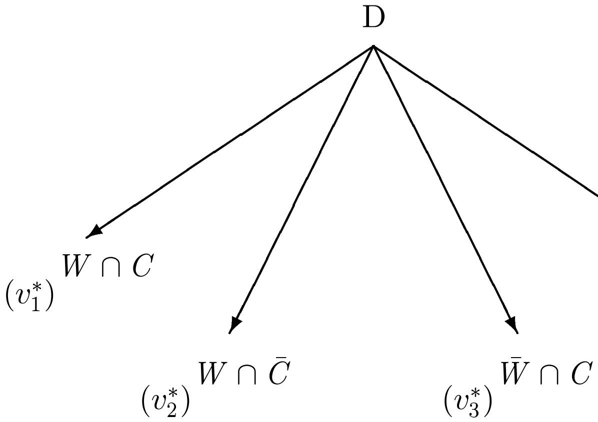


Figure 1 (b)
Multinomial Model



Example 3

As another example, McCullagh and Nelder (1989) report a study of mortality due to radiation. Exposed and non-exposed individuals were classified at the end of the study as dead or alive, with mortality further classified according to deaths due to cancer or other causes. Cancer deaths were further differentiated according to leukemia deaths and deaths from other cancers. The four mutually exclusive response categories are therefore: alive, death from causes other than cancer, deaths from cancers other than leukemia and deaths from leukemia. Again the critical issue is how do we model the observed data. In this case it seems appropriate to model the data in terms of a number of distinct dichotomies, and therefore proceed by making a separate study of total mortality, cancer mortality and leukemia mortality, rather than entertain a BVP structure.

However, there may arise other situations in which the appropriate model is less clear. In such cases, a fundamental issue is under what circumstances we might use the information contained in the random variables y and z (perhaps allowing for correlation), and when we should use the random variable s. In addition it would be instructive to know under what conditions the two are equivalent and this question is addressed in the next section.

2.1. Bivariate Probit

The BVP model is derived from the latent variables y^* and z^* introduced above. Specifically, we assume that y^* and z^* are distributed bivariate normal with a simple parametric structure given by

$$\begin{aligned} y^* &= \mathbf{x}'\boldsymbol{\alpha} + u \\ z^* &= \mathbf{x}'\boldsymbol{\beta} + w \end{aligned} \tag{2.1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameter vectors, $\mathbf{x}$ is a $(k \times 1)$ vector of regressors and the joint distribution of u and w is bivariate normal with correlation ρ:

$$\begin{pmatrix} u \\ w \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{2.2}$$

As discussed previously, (2.1) generates the four possible outcomes of the pair (y, z); see Table 1. The corresponding four probabilities will be denoted π_j, $j = 1, \ldots, 4$ and are relatively straightforward to derive. For example, $s = 1$ if and only if $y = 1$ *and* $z = 1$. Thus

$$\begin{aligned} \pi_1 &= \Pr(s = 1) = \Pr(y^* > 0 \cap z^* > 0) \\ &= \Pr(u > -\mathbf{x}'\boldsymbol{\alpha} \quad \cap \quad v > -\mathbf{x}'\boldsymbol{\beta}). \end{aligned}$$

The other three probabilities are obtained in an obvious manner. Formally, if $\Phi_2(a, b; \rho)$ denotes the bivariate normal distribution function of (u, w) evaluated at the point (a, b) and $\Phi(c)$ the standard normal distribution function evaluated at the point c, then it is easy to see that

$$\begin{aligned} \pi_1 &= \Phi_2(\mathbf{x}'\boldsymbol{\alpha}, \mathbf{x}'\boldsymbol{\beta};\rho) \\ \pi_2 &= \Phi_2(\mathbf{x}'\boldsymbol{\alpha}, -\mathbf{x}'\boldsymbol{\beta}; -\rho) = \Phi(\mathbf{x}'\boldsymbol{\alpha}) - \pi_1 \\ \pi_3 &= \Phi_2(-\mathbf{x}'\boldsymbol{\alpha}, \mathbf{x}'\boldsymbol{\beta}; -\rho) = \Phi(\mathbf{x}'\boldsymbol{\beta}) - \pi_1 \\ \pi_4 &= \Phi_2(-\mathbf{x}'\boldsymbol{\alpha}, -\mathbf{x}'\boldsymbol{\beta};\rho) = 1 - \pi_1 - \pi_2 - \pi_3. \end{aligned} \tag{2.3}$$

The key point to emphasise in the above is that the probabilities for the *four* states have been calculated using the *two* random variables y^* and z^*. As such, the required probabilities involve only the evaluation of the bivariate distribution, $\Phi_2(a, b; \rho)$, and two univariate distributions, $\Phi(a)$ and $\Phi(b)$.

Finally, introduce the mutually exclusive indicators $c_j = \mathbf{1}(s = j)$, $j = 1, \ldots, 4$; i.e., $c_j = 1$ if and only if $s = j$. Then contributions to a log-likelihood function of an arbitrary individual take the form $\sum_{j=1}^{4} c_j \ln(\pi_j)$.

2.2. Multinomial Probit

As shown above, although the BVP model can provide estimates of the probabilities for four mutually exclusive states, this is done using combinations of the two underlying latent random variables. Therefore, the model implies a particular statistical structure. In a MNP model of four states (or alternatives), denoted by the random variable s, the value (or utility) of each state is represented explicitly by the random variable v^*, which can take on values v_j^*, $j = 1, \ldots, 4$, (see Table 1) and, in applications common in economics and biology, it is understood that the outcome with the highest 'value' attached is observed. As a result, the calculation of probabilities for each of the outcome is somewhat different, than in the BVP model. Here, the probability that $s = j$ (or, equivalently, $c_j = 1$, $j = 1, \ldots, 4$) may be written

$$p_j = \Pr(c_j = 1) = \Pr(\bigcap_{l \neq j}^{4} v_j^* - v_l^* > 0) \tag{2.4}$$

and, in general, requires the evaluation of a *trivariate* integral. Contributions to the log-likelihood are thus $\sum_{j=1}^{4} c_j \ln(p_j)$, and the form of the p_j are given below together with some comments on identification issues.

In the MNP model, the natural specification would be

$$v_j^* = \mathbf{x}'\boldsymbol{\gamma}_j + \varepsilon_j; \; j = 1, \ldots, 4, \tag{2.5}$$

in which the $\boldsymbol{\gamma}_j$ are vectors of unknown parameters and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)'$ is distributed multivariate normal with zero mean vector and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{jl}\}$; i.e., $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. In order to obtain the probabilities defined in (2.4), introduce $v_{jl}^* = v_j^* - v_l^*$ so that $v_{jl}^* = v_{1l}^* - v_{1j}^*$, say. Some straightforward calculation then yields:

$$\begin{aligned}
p_1 = &\; \Pr\left(v_{12}^* > 0 \cap v_{13}^* > 0 \cap v_{14}^* > 0\right), \\
p_2 = &\; \Pr\left(v_{12}^* < 0 \cap (v_{13}^* - v_{12}^*) > 0 \cap (v_{14}^* - v_{12}^*) > 0\right), \\
p_3 = &\; \Pr\left(v_{13}^* < 0 \cap (v_{13}^* - v_{12}^*) < 0 \cap (v_{14}^* - v_{13}^*) > 0\right), \\
p_4 = &\; \Pr\left(v_{14}^* < 0 \cap (v_{14}^* - v_{12}^*) < 0 \cap (v_{14}^* - v_{13}^*) < 0\right),
\end{aligned} \tag{2.6}$$

so that all the required probabilities can be defined in terms of the three random variables, v_{1j}^{*}, $j = 2, \ldots, 4$, and differences thereof.[1]

Furthermore, by defining $\boldsymbol{\delta}_j = \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_j$ and $\eta_j = \varepsilon_1 - \varepsilon_j$, the preceding probabilities in (2.6) become

$$
\begin{aligned}
p_1 = &\ \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\delta}_2 \cap \eta_3 > -\mathbf{x}'\boldsymbol{\delta}_3 \cap \eta_4 > -\mathbf{x}'\boldsymbol{\delta}_4\right), \\
p_2 = &\ \Pr\left(\eta_2 < -\mathbf{x}'\boldsymbol{\delta}_2 \cap (\eta_3 - \eta_2) > -\mathbf{x}'\left(\boldsymbol{\delta}_3 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_2) > -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_2\right)\right), \\
p_3 = &\ \Pr\left(\eta_3 < -\mathbf{x}'\boldsymbol{\delta}_3 \cap (\eta_3 - \eta_2) < -\mathbf{x}'\left(\boldsymbol{\delta}_3 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_3) > -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_3\right)\right), \\
p_4 = &\ \Pr\left(\eta_4 < -\mathbf{x}'\boldsymbol{\delta}_4 \cap (\eta_4 - \eta_2) < -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_3) < -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_3\right)\right),
\end{aligned}
\tag{2.7}
$$

which will be determined by the joint distribution of $\boldsymbol{\eta} = (\eta_2, \eta_3, \eta_4)'$, being $MVN(\mathbf{0}, \boldsymbol{\Omega})$. Note that exactly the same probabilities obtain by substituting $\eta_j^{\dagger} = \omega\eta_j$ and $\boldsymbol{\delta}_j^{\dagger} = \omega\boldsymbol{\delta}_j$ throughout, $j = 2, \ldots 4$. It is therefore clear that not only are the individual $\boldsymbol{\gamma}_j$ in (2.5) not identifiable but that it is only possible to identify $\boldsymbol{\delta}_j$ and $\boldsymbol{\Omega}$ up to a factor of proportionality. To overcome this a normalisation is required on one of the elements of $\boldsymbol{\Omega}$; e.g., $var\left(\eta_2\right) = 1$, would be sufficient to identify $\boldsymbol{\delta}_j$ and the remaining 5 distinct parameters in $\boldsymbol{\Omega}$. These arguments can be generalised to a MNP model, defined by (2.5), in which there are J possible outcomes/choices, so that only $J(J-1)/2 - 1$ covariance parameters in $\boldsymbol{\Omega} = var(\boldsymbol{\eta})$, $(J - 1 \times J - 1)$ are identified together with the $\boldsymbol{\delta}_j$, $(k \times 1)$, $j = 2, \ldots, J$; see, for example, the discussion in Pudney (1989).

In the next section, conditions are derived under which $p_j = \pi_j$, $j = 1, \ldots, 4$, implying that the BVP model is a special case of the MNP model.

3. A Restricted Multinomial Probit

It was shown above that the MNP model requires a restriction on one of the elements of $\boldsymbol{\Omega}$, in order for parameters to be identified. The BVP model obtains

[1] This demonstrates that the dimensionality of the multinomial problem is reduced in that the requisite probability calculations involve the estimation of 3-fold multivariate integrals.

from the MNP model, with $J = 4$, by imposing further restrictions on $\mathbf{\Omega}$, and the $\boldsymbol{\delta}_j$, as detailed below.

3.1. Parametric restrictions

The way in which the MNP model formally nests the BVP model is described by the following proposition:

Proposition 3.1. *Let* (i) $\begin{pmatrix} \eta_2 \\ \eta_3 \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$; (ii) $\eta_4 = \eta_2 + \eta_3$; (iii) $\boldsymbol{\delta}_2 = \boldsymbol{\alpha}$, $\boldsymbol{\delta}_3 = \boldsymbol{\beta}$, $\boldsymbol{\delta}_4 = \boldsymbol{\alpha} + \boldsymbol{\beta}$. *Then* $\pi_j = p_j$, $j = 1, \ldots, 4$.

The result is readily established and we demonstrate it only for p_1. Upon substitution of $(i) - (iii)$, in the expression for p_1, we can write

$$\begin{aligned} p_1 &= \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\alpha} \cap \eta_3 > -\mathbf{x}'\boldsymbol{\beta} \cap (\eta_2 + \eta_3) > -\mathbf{x}'\left(\boldsymbol{\alpha} + \boldsymbol{\beta}\right)\right), \\ &= \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\alpha} \cap \eta_3 > -\mathbf{x}'\boldsymbol{\beta}\right), \\ &= \Phi_2\left(\mathbf{x}'\boldsymbol{\alpha}, \mathbf{x}'\boldsymbol{\beta}; \rho\right) = \pi_1. \end{aligned}$$

It is clear that the conditions identified in the above proposition lead to a constrained MNP model; i.e., the MNP model is algebraically equivalent to the BVP model if (2.5) is subject to parametric restrictions, which are now investigated.

Firstly, note that condition (ii) in the above proposition implies that

$$\begin{pmatrix} \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & \rho & 1+\rho \\ \rho & 1 & 1+\rho \\ 1+\rho & 1+\rho & 2(1+\rho) \end{bmatrix}\right) \tag{3.1}$$

in which the covariance matrix is singular. The question we wish to address now is, given (3.1), what can we say about the distribution of the vector $\boldsymbol{\varepsilon}$? Since there are only 5 identifiable covariance parameters, in this case, we therefore begin by imposing a number of rather innocuous restrictions upon the first three variances - namely, $var(\varepsilon_l) = 1$, $l = 1, \ldots, 3$. It then follows that

$$
\begin{aligned}
1 &= var(\eta_2) = var(\varepsilon_1 - \varepsilon_2) \\
&= 2\left(1 - \sigma_{12}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
1 &= var(\eta_3) = var(\varepsilon_1 - \varepsilon_3) \\
&= 2\left(1 - \sigma_{13}\right),
\end{aligned}
$$

giving $\sigma_{12} = \sigma_{13} = \frac{1}{2}$. Continuing in this fashion the restricted covariance matrix of the vector $\boldsymbol{\varepsilon}$ emerges as

$$
\boldsymbol{\Sigma}_R = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & \rho & \frac{1}{2} + \rho \\ \frac{1}{2} & \rho & 1 & \frac{1}{2} + \rho \\ 0 & \frac{1}{2} + \rho & \frac{1}{2} + \rho & 1 + 2\rho \end{bmatrix}, \tag{3.2}
$$

noting that $\rho = E(\eta_2\eta_3) = E((\varepsilon_1 - \varepsilon_2)(\varepsilon_1 - \varepsilon_3)) = \sigma_{23}$. It then immediately follows that the BVP model (2.1) is obtained from the MNP model (2.5) under the following set of parameter restrictions:

1. $var(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_R$;
2. $\boldsymbol{\delta}_4 = \boldsymbol{\delta}_2 + \boldsymbol{\delta}_3$, with $\boldsymbol{\delta}_2 = \boldsymbol{\alpha}$ and $\boldsymbol{\delta}_3 = \boldsymbol{\beta}$ free.

Given that in the unrestricted form of the MNP model there would only be $5+3k$ free parameters, this corresponds to $(5 + 3k) - (1 + 2k) = 4 + k$ parametric restrictions. In Section 3.3 we discuss a procedure for testing such restrictions. Before doing so, however, it is shown how the constrained MNP model (or BVP model) described above is, in fact, consistent with a rather restrictive form of utility maximisation behaviour.

3.2. A Behavioural Interpretation

Consider again Table 1 and assume that, in the context of choices (y, z), the joint decision, s, is indeed made on the basis of utility maximisation, but by considering the utility obtained from y and z individually rather than *jointly* (as would be the case in the MNP model). Specifically, write

$$\begin{aligned} y_y^* &= \mathbf{x}'\boldsymbol{\alpha}_y + u_y, \quad y = 0, 1, \\ z_z^* &= \mathbf{x}'\boldsymbol{\beta}_z + v_z, \quad z = 0, 1, \end{aligned} \tag{3.3}$$

where y_0^* is the utility derived from the decision $y = 0$ and y_1^* is that derived from $y = 1$, with a similar interpretation for z_0^* and z_1^*. Under the assumption of utility maximising behaviour on individual decisions, $y = 0$ (respectively, $y = 1$) is observed if and only if $y_0^* - y_1^* > 0$ (respectively, $y_1^* - y_0^* > 0$), and similarly for z. This characterisation gives the BVP model of (2.1) in which: $y^* = y_1^* - y_0^*$, $z^* = z_1^* - z_0^*$, $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0$, $\boldsymbol{\beta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$, $u = u_1 - u_0$ and $v = v_1 - v_0$, with distributional assumption (2.2) in order to ensure identification.

Now assume that the utilities, denoted v_j^*, $j = 1, \ldots, 4$, and derived from the joint outcome (y, z), are formed additively as $y_y^* + z_z^*$. Thus,

$$\begin{aligned} v_1^* &= y_1^* + z_1^*, \\ v_2^* &= y_1^* + z_0^*, \\ v_3^* &= y_0^* + z_1^*, \\ v_4^* &= y_0^* + z_0^*. \end{aligned}$$

This is a strong assumption; it says that the utilities, v_j^*, are *additively separable* and, given (3.3) also implies that (2.5) correctly expresses the utility derived from the joint decision. Note that, given (3.3), additive separability is not only sufficient, but it is also necessary for this representation of v_j^*.[2] Although restrictive,

[2] Of course, any linear combination of y_y^* and z_z^* will generate linear v_j^*, but arbitrary scale and

let us pursue this for the moment and write these utilities as $v_j^* = \mathbf{x}'\boldsymbol{\gamma}_j + \varepsilon_j$, with $\boldsymbol{\delta}_j = \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_j$ and $\eta_j = \varepsilon_1 - \varepsilon_j$, as in Section 2.2. Then, from (3.3), assumption of (2.2) and the implied restriction of additive separability, it immediately follows that the conditions of Proposition 3.1 are satisfied.

From the above discussion the following conclusions may be drawn:

1. if (2.5) is an adequate representation for v_j^*, and the parametric restrictions of section 3.1 are true, then not only is v_j^* additively separable in the bivariate utilities, but the approach offered by the BVP model *is* consistent with utility maximisation over the 4 distinct alternatives;

2. if (2.5) is an adequate representation for v_j^* and the parametric restrictions are not true, then the BVP model can *not* be consistent with such utility maximising behaviour, since v_j^* is not additively separable;

3. if (2.5) is not an adequate representation for v_j^* (e.g., non-linearities exist in the regression specification), then the BVP model *may be* consistent with utility maximisation over the 4 implied alternatives, but v_j^* is not additively separable in the bivariate utilities.

These conclusions suggest a possible testing strategy in order to shed some light on whether, in the context of a bivariate decision problem, the BVP model is consistent with utility maximisation over the implied four alternative states. Firstly, test the adequacy of modelling v_j^* as in (2.5), in which the regressor vector is the same as that which appears in the individual probit equations of the bivariate decision problem; e.g. test for non-linearities in the regression specification.. If (2.5) appears data consistent, then proceed to test the parametric restrictions as described in Section 3.3, below. If the restrictions pass this test, then there is evidence that the BVP model is consistent with utility maximising

location parameters in any such linear combination will not be indentifiable from the bivariate data.

behaviour over the four distinct alternatives and that the utility derived from the joint outcome is additively separable in the individual utilities.. If the test rejects the restrictions, then there is evidence that the BVP model is not consistent with such behaviour. Finally, if (2.5) is an incorrect model for v_j^*, then BVP model may be consistent with utility maximisation over the four distinct alternatives but v_j^* can not be additively separable in y_y^* and z_z^*.

3.3. A Score Test

Since the parametric restrictions, described in Section 3.1, imply a singular covariance matrix in the MNP model, the most natural likelihood ratio procedure is not strictly available due to the problem of testing on the boundary of the parameter space. A score test procedure is therefore outlined, whose asymptotic validity is unaffected by such a problem. First we define an appropriate $(m \times 1)$ vector of unrestricted parameters $\boldsymbol{\theta}' = (\boldsymbol{\delta}', \boldsymbol{\sigma}')$ for the MNP model, (2.5), which accommodates the BVP model as a special case:

$$\boldsymbol{\delta}' = (\boldsymbol{\delta}_2', \boldsymbol{\delta}_3', \boldsymbol{\delta}_4'), \qquad \boldsymbol{\sigma}' = (\sigma_{23}, \sigma_{14}, \sigma_{24}, \sigma_{34}, \sigma_{44}),$$

where $\boldsymbol{\Sigma} = \{\sigma_{jl}\}$, $j, l = 1, \ldots, 4$, and $m = 5+3k$. Conditional on $\mathbf{x}$, let the relevant four probabilities, (2.7), be expressed as functions of $\boldsymbol{\theta}$; i.e., $p_k(\boldsymbol{\theta}|\mathbf{x})$. Then based on N independent realisations of the indicator c_j, denoted c_{ij}, $i = 1, \ldots, N$, the log-likelihood is $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))$. From this the $(m \times 1)$ score vector and $(m \times m)$ Hessian matrix are

$$\begin{aligned} \mathbf{g}(\boldsymbol{\theta}) &= \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \frac{\partial \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))}{\partial \boldsymbol{\theta}}, \\ \mathbf{H}(\boldsymbol{\theta}) &= \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \frac{\partial^2 \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \end{aligned}$$

Now let $\left(\tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\beta}}', \tilde{\rho}\right)$ denote (restricted) maximum likelihood estimates of the BVP model, (2.1), and define accordingly the restricted MNP maximum likelihood estimates as

$$\begin{aligned}\tilde{\boldsymbol{\delta}}' &= \left(\tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\alpha}}' + \tilde{\boldsymbol{\beta}}'\right) \\ \tilde{\boldsymbol{\sigma}}' &= \left(\tilde{\rho}, 0, \tfrac{1}{2} + \tilde{\rho}, \tfrac{1}{2} + \tilde{\rho}, 1 + 2\tilde{\rho}\right).\end{aligned}$$

An asymptotically valid test of the implied $q = 4+k$ parameter restrictions can be based on the score test statistic given by $\mathcal{S} = N^{-1}\mathbf{g}(\tilde{\boldsymbol{\theta}})' \left\{\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right)\right\}^{-1} \mathbf{g}(\tilde{\boldsymbol{\theta}})$, where $\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right)$ is any consistent estimator for the average information matrix, under the assumption that the restrictions under test are valid; e.g., $\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right) = -\frac{1}{N}\mathbf{H}\left(\tilde{\boldsymbol{\theta}}\right)$. In large samples, $\mathcal{S}$ is distributed as a chi-square random variable with q (number of restrictions) degrees of freedom when the parametric restriction imposed on the MNP model are correct. Significantly large values of $\mathcal{S}$ would be provide evidence against the BVP model and, as a consequence, it would also suggest that the BVP specification is inconsistent with utility maximisation over the implied four possible outcomes.

4. Concluding Remarks

In this note we have pointed out the nature of the statistical relationship between the bivariate probit (BVP) model and multinomial probit (MNP) model. Specifically, a general MNP specification could be used to model the four mutually exclusive outcomes, which a bivariate decision process can generate, and a restricted version of this yields the BVP model. This suggests a way of testing the BVP model by nesting it within the MNP model and a score test procedure has been outlined for this purpose. (A score test is used since the implied parametric restrictions yield a singular covariance matrix in the MNP specification). Such a test might also be used by the applied worker as a way of assessing whether, in the bivariate decision problem, individuals maximise utility over the four possible outcomes that can be generated.

References

Ashford, J., and R. Sowden (1970): "Multivariate Probit Analysis," *Biometrics*, 26, 535–546.

Cox, D. R. (1972): "The Analysis of Multivariate Binary Data," *Applied Statistics*, vol. 21, 113–126.

Fay, J. W. (1957): "The National Coal Board's Pneumoconisos Research," Discussion paper.

Greene, W. H. (1993): *Econometric Analysis*. Macmillan Publisher Company, 2nd edn.

Mantel, N. (1966): "Models for Complex Contingency Tables and Polychotomous Dosage Response Curves," *Biometrics*, 22, 83–95.

McFadden, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, pp. 198–272. Cambridge MA:MIT Press.

Pudney, S. (1989): *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes.* Blackwell.