

What can we learn about correlations from multinomial probit estimates?

Chiara Monfardini* J.M.C. Santos Silva†

February 2006

Abstract

It is well known that, in a multinomial probit, only the covariance matrix of the location and scale normalized utilities are identified. In this study, we explore the relation between these identifiable parameters and the original elements of the covariance matrix, to find out what can be learnt about the correlations between the stochastic components of the non-normalized utilities. We show that, in certain circumstances, it is possible to obtain information on these behavioural parameters and define appropriate tools for inference. We illustrate the usefulness of our results in applied settings using an example.

JEL classification code: C25.

Key words: Correlations; Equicorrelation; Identification; Inequality restrictions.

*Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2 - Bologna, Italia. Fax: +39 051221968. E-mail: chiaram@economia.unibo.it.

†ISEG/Universidade Técnica de Lisboa. R. do Quelhas 6, 1200-781 Lisboa, Portugal. Fax: +351 213922781. E-mail: jmcass@iseg.utl.pt. Santos Silva gratefully acknowledges the partial financial support from Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER

1. INTRODUCTION

In the context of discrete choice modelling, the multinomial probit model (MNP) is often adopted as a way to avoid the well known limitations of the simpler multinomial logit (MNL), viz., the independence from irrelevant alternatives. This property of the MNL follows from the assumption that the stochastic components of the utilities are independent and identically distributed as type 1 extreme value variates. The inadequacy of such assumption, for many cases where it is realistic to assume that some alternatives are more similar to each other for the individual performing the choice, has been thoroughly noted in the literature.

Although the MNP does not impose any restrictions on the covariance matrix of the stochastic components of the utilities, its elements are not identified. Indeed, due to the fact that in any random utility model the utility functions are only identified up to scale and location (Dansie, 1985), all that is possible to identify are the parameters in the covariance matrix of the normalized utilities. These parameters are functions of the original elements of the covariance matrix and are unfit to be given an economic or behavioural interpretation.

In this study, we explore the relation between the original elements of the covariance matrix and their functions which are identified after normalization, to find out what can be learnt about the correlations between the stochastic components of the non-normalized utilities. We show that, in certain circumstances, it is possible to obtain information on these behavioural parameters and define appropriate tools for inference. The results we obtain are very simple, but they appear not to be currently available in the specialized literature.

Although we focus on the MNP model, some of our results are easy to extend to other models, like the mixed multinomial logit of McFadden and Train (2000), or the heterogeneity adjusted logit of Chesher and Santos Silva (2002). This issue, however, is not pursued in the current version of this paper and is the subject of ongoing research.

The remainder of the paper is organized as follows. Section 2 briefly presents the MNP and introduces the necessary notation. Section 3 discusses the information content of the identified elements in the normalized covariance matrix of the MNP. In particular, we show that it is not possible to distinguish a model with homoskedastic¹ and independent errors from a model with homoskedastic and equicorrelated disturbances. Moreover, we show how the space of the identified covariance matrix parameters in a MNP can be partitioned into two regions. One region corresponds to the existence of some non-zero covariances between the errors of the original (non-normalized) utilities. The points in the second region may or may not entail non-null covariances between the original stochastic utilities. In section 4 we present an empirical illustration of our main results and, finally, section 5 concludes.

2. THE MULTINOMIAL PROBIT

For simplicity, we present the MNP for the three alternative case ($J = 3$). The model assumes that individuals select one of three mutually exclusive alternatives. The random utility of individual i , $i = 1, \dots, N$, for choice j , $j = 1, 2, 3$, is formulated as

$$u_{ij} = \alpha_j + x_i' \beta_j + \varepsilon_{ij} \tag{1}$$

where: x_i is a $(k \times 1)$ vector of explanatory variables for individual i , which may contain both individual specific characteristics and alternative specific attributes faced by individual i ;² $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})'$ is a vector of stochastic terms which is assumed to be distributed as a trivariate normal, identically and independently across the N individuals, with zero

¹Throughout the paper, homoskedasticity is interpreted as meaning that the stochastic component of the utility has the same variance for all alternatives.

²Keane (1992) shows that in multinomial probit models identification is *tenuous* unless exclusion restrictions are present, i.e., there are some alternative-specific attributes which enter as regressors only the utility function associated with one alternative and not the others.

mean and covariance matrix

$$\Sigma = Cov(\varepsilon_i) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

with $\sigma_{jj} > 0, \forall j$ (positive definiteness).

Arranging the parameters in (1) as $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$, $\beta = (\beta'_1, \beta'_2, \beta'_3)'$, the log-likelihood function associated with the model is

$$L(\alpha, \beta, \Sigma) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 y_{ij} \ln P_{ij}(\alpha, \beta, \Sigma) \quad (2)$$

where $y_{ij} = 1$ if individual i chooses alternative j and $y_{ij} = 0$ otherwise, while $P_{ij}(\alpha, \beta, \Sigma) = \Pr(u_{ij} > u_{ik}, k \neq j = 1, 2, 3)$ represents the probability that individual i chooses alternative j and involves the evaluation of a bivariate integral in this three alternative case. Unfortunately, it is not possible to get unique maximum likelihood estimates of the parameters α, β, Σ in the above model, as they are not identified. Dansie (1985) gives the first systematic explanation of the identification problem in multinomial probit models and of its two sources. The first source of the identification problem is that the observed choices are only informative on the differences of the utilities and not on the utilities themselves. This means that, in (2), all the probabilities of selection $P_{ij}(\alpha, \beta, \Sigma)$ can be rewritten in terms of differenced utilities, without altering the value of the log-likelihood function. In what follows, we discuss the identification taking differences with respect to the utilities associated with $j = 3$, i.e., we take the third alternative as the reference state used to normalize location of the latent variable. This leads to

$$u_{il}^* = u_{il} - u_{i3} = \alpha_l^* + x'_i \beta_l^* + \varepsilon_{il}^* \quad (3)$$

where, $\alpha_l^* = \alpha_l - \alpha_3$, $\beta_l^* = \beta_l - \beta_3$, $\varepsilon_{il}^* = \varepsilon_{il} - \varepsilon_{i3}$, $l = 1, 2, 3$. As a consequence, $u_{i3}^* = 0$ and the relevant distribution of the disturbances is not the above-mentioned trivariate one, but the bivariate distribution of $\varepsilon_i^* = (\varepsilon_{i1}^*, \varepsilon_{i2}^*)'$, which is normal with zero mean and covariance matrix

$$\Sigma^* = Cov(\varepsilon_i^*) = \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{22}^* \end{pmatrix} \quad (4)$$

with $\sigma_{lk}^* = E(\varepsilon_{il} - \varepsilon_{i3})(\varepsilon_{ik} - \varepsilon_{i3})$, $l, k = 1, 2$. The second source of the identification problem concerns the lack of information on the scale in the available data, i.e., the utilities can be multiplied by an arbitrary constant without changing the value of the log-likelihood function. Therefore, in order to achieve identification, it is necessary to impose a restriction on Σ^* , and only two out of the three parameters of the bivariate covariance matrix are identified. The usual way of imposing this identification restriction is to standardize in order to have the first utility disturbance with unit variance, i.e., the utilities become

$$u_{il}^{**} = \frac{u_{il}^*}{\sqrt{\sigma_{11}^*}} = \alpha_l^{**} + x_i' \beta_l^{**} + \varepsilon_{il}^{**} \quad (5)$$

with $\alpha_l^{**} = \alpha_l^* / \sqrt{\sigma_{11}^*}$, $\beta_l^{**} = \beta_l^* / \sqrt{\sigma_{11}^*}$, $\varepsilon_i^{**} = \varepsilon_i^* / \sqrt{\sigma_{11}^*}$, $l = 1, 2$, $u_{i3}^{**} = 0$ and

$$\Sigma^{**} = Cov(\varepsilon_i^{**}) = \begin{pmatrix} 1 & \sigma_{12}^{**} \\ \sigma_{21}^{**} & \sigma_{22}^{**} \end{pmatrix} \quad (6)$$

with $\sigma_{lk}^{**} = \sigma_{lk}^* / \sigma_{11}^*$, $l, k = 1, 2$.

The log-likelihood function of the identified model can now be written as

$$L(\alpha^{**}, \beta^{**}, \Sigma^{**}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 y_{ij} \ln P_{ij}^{**}(\alpha^{**}, \beta^{**}, \Sigma^{**}) \quad (7)$$

where, $P_{ij}^{**}(\alpha^{**}, \beta^{**}, \Sigma^{**}) = \Pr(u_{ij}^{**} > u_{ik}^{**}, k \neq j = 1, 2, 3)$. For example

$$\begin{aligned} P_{i1}^{**}(\alpha^{**}, \beta^{**}, \Sigma^{**}) &= \Pr(u_{i1}^{**} > u_{i2}^{**}, u_{i1}^{**} > 0) \\ &= \int_{-\infty}^{\frac{(\alpha_1^{**} + x_i' \beta_1^{**}) - (\alpha_2^{**} + x_i' \beta_2^{**})}{\sqrt{1 + \sigma_{22}^{**} - 2\sigma_{12}^{**}}}} \int_{-\infty}^{\alpha_1^{**} + x_i' \beta_1^{**}} \varphi(z_1, z_2; \rho_1) dz_1 dz_2 \end{aligned}$$

where $\varphi(z_1, z_2; \rho_1)$ is the bivariate normal density function of two random variables having zero mean, unit variance and correlation coefficient $\rho_1 = (1 - \sigma_{12}^{**}) / \sqrt{1 + \sigma_{22}^{**} - 2\sigma_{12}^{**}}$. Similarly, P_{i2}^{**} and P_{i3}^{**} can be derived.

This approach, consisting of working directly in a $J - 1$ space, is the more general one. Alternatively, identification can be achieved by imposing arbitrary normalizations and identification restrictions in the J space, for example, by setting some of the covariances equal to zero. Bunch (1991) points out that this practice has been adopted in some studies without a clear recognition of the number of restrictions required for identification and

shows that it can lead to invalid (i.e., non positive definite) covariance matrixes in the $J - 1$ space. Note that this is the approach of the routine implemented in STATA 9 (StataCorp, 2005). Imposing a structure on the covariance matrix makes the estimated coefficients directly interpretable as covariances between the errors of the original utilities, but this interpretation rests on strong and untestable assumptions. We consider the case in which no structure is imposed on the covariance matrix.

3. INFERENCE ON BEHAVIOURAL PROPERTIES

3.1. The identified covariance elements

We add to the notation above a superscript (j) denoting the chosen reference state. When $j = 3$, the elements of the (2×2) matrix Σ^* in (4) are expressed as the following functions of the elements of the original (3×3) covariance matrix Σ

$$\begin{aligned}\sigma_{11}^{*(3)} &= E(\varepsilon_1 - \varepsilon_3)^2 = \sigma_{11} + \sigma_{33} - 2\sigma_{13} \\ \sigma_{22}^{*(3)} &= E(\varepsilon_2 - \varepsilon_3)^2 = \sigma_{22} + \sigma_{33} - 2\sigma_{23} \\ \sigma_{12}^{*(3)} &= E(\varepsilon_1 - \varepsilon_3)(\varepsilon_2 - \varepsilon_3) = \sigma_{12} - \sigma_{13} - \sigma_{23} + \sigma_{33}.\end{aligned}$$

The covariance element in the identified covariance matrix $\Sigma^{**(3)}$ in (6) is given by $\sigma_{12}^{**(3)} = \sigma_{12}^{*(3)} / \sigma_{11}^{*(3)}$.

Notice that changing the reference state adopted for estimation allows the identification of two other covariance elements in the corresponding (2×2) matrix $\Sigma^{**(j)}$, namely $\sigma_{12}^{**(1)} = \sigma_{12}^{*(1)} / \sigma_{11}^{*(1)}$ and $\sigma_{12}^{**(2)} = \sigma_{12}^{*(2)} / \sigma_{11}^{*(2)}$ where

$$\begin{aligned}\sigma_{12}^{*(1)} &= E(\varepsilon_2 - \varepsilon_1)(\varepsilon_3 - \varepsilon_1) = \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} \\ \sigma_{12}^{*(2)} &= E(\varepsilon_1 - \varepsilon_2)(\varepsilon_3 - \varepsilon_2) = \sigma_{13} - \sigma_{23} - \sigma_{12} + \sigma_{22} \\ \sigma_{11}^{*(1)} &= E(\varepsilon_2 - \varepsilon_1)^2 = E(\varepsilon_1 - \varepsilon_2)^2 = \sigma_{11}^{*(2)}.\end{aligned}$$

3.2. Independence vs. equicorrelation

The so-called independent multinomial probit (IMP) is often presented in the literature as the MNP that more closely approximates the MNL. Indeed, the IMP, characterized by independent and homoskedastic errors, has been used by Horowitz (1981) and Terza (1998) as the basis of tests of the independence from irrelevant alternatives property. However, it is perhaps not widely appreciated that the identified covariance structure of the IMP is indistinguishable from the identified covariance structure corresponding to the homoskedastic and equicorrelated probit (HEP). This applies to any homoskedastic discrete choice model whose identification requires normalization of location and scale.

Proposition 1 *Equicorrelation and non-correlation are observationally equivalent in any homoskedastic discrete choice model identified through normalization of location and scale.*

Proof: *Without loss of generality, consider an equicorrelated and homoskedastic probit, with original covariance matrix Σ^{HEP} characterized by diagonal elements $\sigma_{jj} = \tau$, $\tau > 0$ for all $j = 1, \dots, J$ and off-diagonal elements $\sigma_{js} = \gamma$. After normalizing the location with respect to alternative J , the errors are of the type $\varepsilon_{il}^* = \varepsilon_{il} - \varepsilon_{iJ}$ and $\sigma_{lk}^{*(J)} = E(\varepsilon_{il} - \varepsilon_{iJ})(\varepsilon_{ik} - \varepsilon_{iJ}) = \tau - \gamma$ for all l, k . That is, all the covariances of the location normalized errors ε_{il}^* , $l = 1, \dots, J-1$, must be equal to each other and to $\tau - \gamma$. The variance elements are all equal to $\sigma_{ll}^{*(J)} = E(\varepsilon_{il} - \varepsilon_{iJ})^2 = 2(\tau - \gamma)$. Therefore, after normalization of scale is performed, we obtain the following identified covariance matrix for the HEP*

$$\Sigma_{(J-1) \times (J-1)}^{**HEP} = \begin{bmatrix} 1 & 0.5 & \cdots & 0.5 \\ 0.5 & 1 & 0.5 & \vdots \\ \vdots & 0.5 & \ddots & 0.5 \\ 0.5 & \cdots & 0.5 & 1 \end{bmatrix}.$$

The particular case of the IMP is obtained setting $\gamma = 0$ and it is straightforward to see that this restriction has no impact on the structure of the identified covariance matrix (cf. Terza, 1998, p. 7). ■

This result states that models like the IMP and MNL may be an adequate representation of behaviour when all the alternatives exhibit the same degree of similarity to each other. Independence is obtained in particular homoskedastic cases for which the common correlation coefficient is equal to zero. However, that cannot be distinguished from an homoskedastic and equicorrelated error structure. The result has an intuitive appeal in terms of behavioural interpretation since independence is often used just as a way to express the idea that, for a given individual, all alternatives are equally dissimilar.

The results above suggest that the procedures of Horowitz (1981) and Terza (1998) actually have as the null hypothesis the homoskedasticity and equicorrelation of the errors, rather than their homoskedasticity and independence. Within the MNP estimation, the equicorrelation hypothesis

$$\begin{aligned} H_0 & : \Sigma^{**} = \Sigma^{**HEP} \\ H_1 & : \Sigma^{**} \neq \Sigma^{**HEP} \end{aligned}$$

can be easily checked using the classical likelihood ratio or Wald tests.

3.3 Dependence

It has been noted in empirical applications of the MNP that the estimation results on the identified parameters of the covariance matrix are not useful for inferring individual preferences and are difficult to interpret from an economic perspective. This is a direct consequence of the identification problem of the econometric model, which only permits the estimation of particular functions of the parameters of interest. The following proposition shows that there is a region of the identified covariance matrix parameter space of the MNP which corresponds to the existence of at least one non-zero covariance between the errors of the non-normalized utilities.

Proposition 2 (*case $J = 3$*) *Non positiveness of one of the identified covariance elements across the reference states is a sufficient condition for the existence of a non null covariance among the error terms of the original stochastic utilities. That is $\sigma_{12}^{**^{(j)}} \leq 0$ for some j ($j = 1, 2, 3$) $\implies \sigma_{lm} \neq 0$ for some l, m ($l, m = 1, 2, 3; l \neq m$).*

Proof: Consider the identified covariance elements obtained with reference state 3 $\sigma_{12}^{**(3)} \leq 0 \Leftrightarrow \sigma_{12}^{*(3)} = \sigma_{12} - \sigma_{13} - \sigma_{23} + \sigma_{33} \leq 0 \Leftrightarrow \sigma_{33} \leq -\sigma_{12} + \sigma_{13} + \sigma_{23}$. As $\sigma_{33} > 0$, the case of all σ_{lm} being simultaneously equal to zero must be ruled out. The same applies to the identified covariance elements $\sigma_{12}^{**(2)}$ and $\sigma_{12}^{**(1)}$, obtained with alternative reference states. ■

It is straightforward to generalize the proposition above to the case of $J > 3$, where the number of identified covariance elements for a given reference state j is $M = (J-1) \times (J-2)/2$. The result is obtained replacing $\sigma_{12}^{**(j)}$ with $\sigma_{rs}^{**(j)}$, ($r = 1, \dots, J-1; s = 2, \dots, J; r \neq s$). For example, with $J = 4$, reference state 4, we have $\sigma_{12}^{*(4)} = E(\varepsilon_1 - \varepsilon_4)(\varepsilon_2 - \varepsilon_4)$, $\sigma_{13}^{*(4)} = E(\varepsilon_1 - \varepsilon_4)(\varepsilon_3 - \varepsilon_4)$, $\sigma_{23}^{*(4)} = (\varepsilon_2 - \varepsilon_4)(\varepsilon_3 - \varepsilon_4)$. The following proposition applies in the general case of J alternatives:

Proposition 3 (general case) *Non positiveness of one of the identified covariance elements across the reference states is a sufficient condition for the existence of a non null covariance among the error terms of the original stochastic utilities. That is $\sigma_{rs}^{**(j)} \leq 0$ for some j ($j = 1, \dots, J$) and some r, s ($r = 1, \dots, J-1; s = 2, \dots, J; r \neq s$) $\implies \sigma_{lm} \neq 0$ for some l, m ($l, m = 1, \dots, J; l \neq m$).*

Proof: Trivial, analogous to $J = 3$ case. ■

The proposition above states that the sign of the identified covariance element has an informational content about the covariance structure of the stochastic utilities in the original space. Indeed, if, across all possible reference states, it is possible to find an identified covariance element which is equal to zero or negative, this implies that the original covariances can not all be equal to zero. Not surprisingly, given the identification problem, this is a weak result in the sense that it gives no information on how many and which covariances are different from zero. Moreover, finding that the identified covariance elements are positive across all possible reference states is inconclusive about the existence of non-zero covariances between the stochastic components of the non-normalized utility functions.

This result sheds some light on MNP applications in which identified correlation elements are found to be negative and high in absolute value (see for example Monfardini, 2003). We are now in the position to argue that these results can be interpreted as due to the non validity of the non-correlation/independence assumption, therefore supporting the use of models that do not impose a priori the zero covariance pattern.

The proposition above can be made operational without having to estimate the MNP with different reference states. For illustrative purposes, consider the case $J = 3$.³ According to Proposition 1, some dependence exists as long as an identified covariance element across the 3 possible reference states is zero or negative. To exclude this case, all identified covariance elements across the 3 possible reference states have to be strictly positive. We can then define the two following mutually exclusive sets of conditions:

$$\begin{aligned} D^{**} &: \sigma_{12}^{**(1)} \leq 0 \quad \vee \quad \sigma_{12}^{**(2)} \leq 0 \quad \vee \quad \sigma_{12}^{**(3)} \leq 0 \\ \overline{D}^{**} &: \sigma_{12}^{**(1)} > 0 \quad \wedge \quad \sigma_{12}^{**(2)} > 0 \quad \wedge \quad \sigma_{12}^{**(3)} > 0, \end{aligned}$$

where the set \overline{D}^{**} is inconclusive, being consistent with both dependence or independence patterns, while set D^{**} corresponds to the existence of some dependence.

The inequalities defining D^{**} can be transformed in order to make them dependent only on the identified covariance matrix elements corresponding to a chosen reference state. That is, it is possible to transform the inequalities defined over the different reference states to different restrictions applied to the identified covariance matrix elements from a single reference state. To show this, it is useful to go one step back to the one-star notation denoting only normalization of location, but not of scale, and write

$$\begin{aligned} \sigma_{12}^{*(1)} &= E(\varepsilon_2 - \varepsilon_1)(\varepsilon_3 - \varepsilon_1) = E[(\varepsilon_2 - \varepsilon_3) + (\varepsilon_3 - \varepsilon_1)](\varepsilon_3 - \varepsilon_1) = -\sigma_{12}^{*(3)} + \sigma_{11}^{*(3)} \\ \sigma_{12}^{*(2)} &= E(\varepsilon_1 - \varepsilon_2)(\varepsilon_3 - \varepsilon_2) = E[(\varepsilon_1 - \varepsilon_3) + (\varepsilon_3 - \varepsilon_2)](\varepsilon_3 - \varepsilon_2) = -\sigma_{12}^{*(3)} + \sigma_{22}^{*(3)} \\ \sigma_{12}^{*(3)} &= E(\varepsilon_1 - \varepsilon_3)(\varepsilon_2 - \varepsilon_3). \end{aligned}$$

These two relations allow us to express the three restrictions defining D^{**} as functions of $\sigma_{12}^{*(3)}$ and $\sigma_{22}^{*(3)}$, i.e., the parameters of the identified covariance matrix after location

³In the appendix we illustrate how the same manipulation can be performed for $J > 3$ and explicitly consider the case of $J = 4$.

and scale normalization. Indeed,

$$\begin{aligned}\sigma_{12}^{**(3)} &\leq 0 \\ \sigma_{12}^{**(1)} &\leq 0 \Leftrightarrow \frac{\sigma_{12}^{*(1)}}{\sigma_{11}^{*(1)}} \leq 0 \Leftrightarrow \frac{-\sigma_{12}^{*(3)} + \sigma_{11}^{*(3)} \frac{\sigma_{11}^{*(3)}}{\sigma_{11}^{*(1)}}}{\sigma_{11}^{*(3)}} \leq 0 \Leftrightarrow 1 - \sigma_{12}^{**(3)} \leq 0 \\ \sigma_{12}^{**(2)} &\leq 0 \Leftrightarrow \frac{\sigma_{12}^{*(2)}}{\sigma_{11}^{*(2)}} \leq 0 \Leftrightarrow \frac{-\sigma_{12}^{*(3)} + \sigma_{22}^{*(3)} \frac{\sigma_{11}^{*(3)}}{\sigma_{11}^{*(2)}}}{\sigma_{11}^{*(3)}} \leq 0 \Leftrightarrow \sigma_{22}^{**(3)} - \sigma_{12}^{**(3)} \leq 0.\end{aligned}$$

Let $S^{(3)}$ be the admissible region of the identified covariance matrix elements with reference state 3.⁴ The result above implies the following partition of $S^{(3)} = D^{(3)} \cup \overline{D}^{(3)}$, with $D^{(3)} \cap \overline{D}^{(3)} = \emptyset$ and

$$\begin{aligned}D^{(3)} &: \sigma_{12}^{**(3)} \leq 0 \vee \sigma_{12}^{**(3)} \geq 1 \vee \sigma_{12}^{**(3)} \geq \sigma_{22}^{**(3)} \\ \overline{D}^{(3)} &: 0 < \sigma_{12}^{**(3)} < 1 \wedge \sigma_{12}^{**(3)} < \sigma_{22}^{**(3)},\end{aligned}$$

where $\overline{D}^{(3)}$ is the inconclusive region, while points in $D^{(3)}$ correspond to the existence of some dependence.

Figure 1 illustrates the partition of $S^{(3)}$. Notice that the inconclusive region contains the point $(\sigma_{12}^{**(3)} = 0.5, \sigma_{22}^{**(3)} = 1)$ which, as discussed above, is compatible both with independence and equicorrelation.

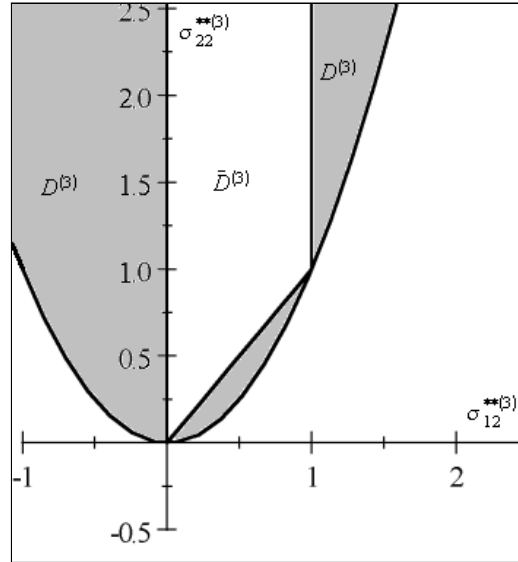


Fig. 1 - The partition of $S^{(3)}$

⁴Notice that the region of the two dimensional space with $(\sigma_{12}^{**(3)})^2 > \sigma_{22}^{**(3)}$ is not admissible.

In empirical applications, it may be interesting to check whether the estimation results allow the researcher to reach any conclusion concerning the existence of dependence among the stochastic components of the non-normalized utilities.

The set of inequalities defining $D^{(3)}$ can be formally tested using the procedures initially developed by Perlman (1969),⁵ or the alternative method proposed by Hansen (2003). However, these methods are not particularly attractive for routine use by practitioners. Fortunately, it is possible to use much simpler tools to check whether the data are compatible with a matrix $\Sigma^{**(3)}$ whose identified elements belong to $D^{(3)}$.

Let $\hat{\sigma}_{12}^{**(3)}$ and $\hat{\sigma}_{22}^{**(3)}$ denote the maximum likelihood estimates of the identified elements of $\Sigma^{**(3)}$ and let \hat{V} denote an estimator of their covariance matrix. Defining $\delta = (\delta_1, \delta_2)'$, with $\delta_l = \hat{\sigma}_{1l}^{**(3)} - \sigma_{1l}^{**(3)}$, $l = 1, 2$, standard asymptotic results imply that $\delta' \hat{V}^{-1} \delta \sim \chi_{(2)}^2$.

If $q_{1-\alpha}^k$ denotes the $1 - \alpha$ quantile of the $\chi_{(k)}^2$ distribution, then the inequality $\delta' \hat{V}^{-1} \delta < q_{1-\alpha}^2$ defines a confidence region of level $100(1 - \alpha)\%$ for the vector $(\sigma_{12}^{**(3)}, \sigma_{22}^{**(3)})'$. If this confidence region is contained within $D^{(3)}$, this provides evidence supporting the existence of at least one non-zero covariance between the errors of the non-normalized utilities.

In case $J > 3$, visual inspection is not appropriate to check whether the confidence ellipsoid is within the dependence region. In such cases, one can just minimize the quadratic form defining the confidence region with respect to the elements of $\Sigma^{**(J)}$, subject to the restriction that these belong to the inconclusive region. If the minimum obtained is smaller than $q_{1-\alpha}^v$, where v is the number of identified variance and covariance elements of $\Sigma^{**(J)}$,⁶ then the confidence region is not contained within the region of the parameter space that implies dependence.

4. AN EMPIRICAL ILLUSTRATION

In this section, we use the data on the choice of fishing mode studied by Herriges and Kling (1999) to illustrate the application of our results.⁷ These data consist of a sample of

⁵See also Wolak (1989, 1991).

⁶The total number of distinct identified elements of the matrix is $v = [(J - 1)J/2] - 1$.

⁷These data are made available through Cameron and Trivedi (2005).

1182 individuals for which we have information on the choice of fishing mode (beach, pier, private boat or charter boat), as well as on income (which is individual specific and does not vary across alternatives), catch rate and price (both of which vary across individuals and fishing mode). For our illustration, we use a trinomial model obtained by collapsing in a non-boat option the choices pier and beach, which have the same price.⁸ Descriptive statistics for these variables are given in Table 1.

Table 1: Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
Modes				
Noboat	0.26	0.44	0	1
Private	0.35	0.48	0	1
Charter	0.38	0.49	0	1
Income (\$1000)	4.10	2.46	0.42	12.50
Prices (\$)				
Pnoboat	103.42	103.64	1.29	843.19
Pprivate	55.26	62.71	2.29	666.11
Pcharter	84.38	63.54	27.29	691.11
Catch rates				
Qnoboat	0.20	0.17	0.04	0.49
Qprivate	0.17	0.21	0.00	0.74
Qcharter	0.63	0.71	0.00	2.31

For this particular example, the following specification is adopted for the stochastic utilities associated with the three alternatives

$$\begin{aligned}
 u_{i1} &= \alpha_1 + \beta_{11}\text{Income}_i + \beta_{12}\text{Pnoboat}_i + \beta_{13}\text{Qnoboat}_i + \varepsilon_{i1}, \\
 u_{i2} &= \alpha_2 + \beta_{21}\text{Income}_i + \beta_{22}\text{Pprivate}_i + \beta_{23}\text{Qprivate}_i + \varepsilon_{i2}, \\
 u_{i3} &= \alpha_3 + \beta_{31}\text{Income}_i + \beta_{32}\text{Pcharter}_i + \beta_{33}\text{Qcharter}_i + \varepsilon_{i3}.
 \end{aligned} \tag{8}$$

This specification implies that each utility depends only on its own price and catch rate, and we let the parameters of these regressors to differ across the three alternatives. The

⁸The catch rate for the non-boat alternative is obtained as the average of the catch rates for the beach and pier options.

identified model used for estimation is subsequently obtained by normalizing with respect to the third choice (charter boat). Therefore, after location and scale normalization, the estimated model is (cf. with 5):

$$u_{i1}^{**} = \alpha_1^{**} + \beta_{11}^{**}\text{Income}_i + \beta_{12}^{**}\text{Pnoboat}_i - \beta_{32}^{**}\text{Pcharter}_i + \beta_{13}^{**}\text{Qnoboat}_i - \beta_{33}^{**}\text{Qcharter}_i + \varepsilon_{i1}^{**}$$

$$u_{i2}^{**} = \alpha_2^{**} + \beta_{21}^{**}\text{Income}_i + \beta_{22}^{**}\text{Pprivate}_i - \beta_{32}^{**}\text{Pcharter}_i + \beta_{23}^{**}\text{Qprivate}_i - \beta_{33}^{**}\text{Qcharter}_i + \varepsilon_{i2}^{**}$$

with $\alpha_l^{**} = (\alpha_l - \alpha_3)/\sqrt{\sigma_{11}^{**}}$, $\beta_{l1}^{**} = (\beta_{l1} - \beta_{31})/\sqrt{\sigma_{11}^{**}}$, $\varepsilon_{l1}^{**} = (\varepsilon_{l1} - \varepsilon_{31})/\sqrt{\sigma_{11}^{**}}$, $l = 1, 2$ and $\beta_{jk}^{**} = \beta_{jk}/\sqrt{\sigma_{11}^{**}}$, $j = 1, 2, 3; k = 2, 3$.

Table 2 presents the main estimation results. Since $1 < \hat{\sigma}_{12}^{**(3)} < \hat{\sigma}_{22}^{**(3)}$, the pair $(\hat{\sigma}_{12}^{**(3)}, \hat{\sigma}_{22}^{**(3)})$ falls into $D^{(3)}$, which is compatible with the existence of some dependence between the error terms of the stochastic utilities in (8). In order to check the strength of this evidence, we can follow the procedure described in the previous section, to check whether a 95% confidence region for $(\sigma_{12}^{**(3)}, \sigma_{22}^{**(3)})$ is contained within $D^{(3)}$. Minimization of $\delta'\hat{V}^{-1}\delta$ with respect to $\sigma_{12}^{**(3)}$ and $\sigma_{22}^{**(3)}$, subject to the restriction that these parameters belong to $\overline{D}^{(3)}$, leads to $\delta'\hat{V}^{-1}\delta = 0.0971$, for $\sigma_{12}^{**(3)} = 0.9999$ and $\sigma_{22}^{**(3)} = 2.3029$. Since $0.097173 < q_{0.95}^2 = 5.99$, the 95% confidence region is not contained within $D^{(3)}$.

Table 2: Trinomial Probit results

Parameter	Estimate	Std. Error
α_1^{**}	-0.1789	0.2094
β_{11}^{**}	0.1346	0.0258
β_{12}^{**}	-0.0168	0.0019
β_{32}^{**}	-0.0064	0.0012
β_{13}^{**}	2.9222	0.4608
β_{33}^{**}	0.4105	0.0737
α_2^{**}	-0.7006	0.2300
β_{21}^{**}	0.1538	0.0504
β_{22}^{**}	-0.0106	0.0020
β_{23}^{**}	1.1316	0.3650
$\sigma_{12}^{**(3)}$	1.1775	0.5696
$\sigma_{22}^{**(3)}$	2.7785	1.5939
$Cov(\sigma_{12}^{**(3)}, \sigma_{22}^{**(3)})$	0.8690	
Log-likelihood	-0.8077	

In this case, because $J = 3$, this result can be confirmed by visual inspection of Figure 2, which plots the 95% confidence region for $(\sigma_{12}^{**(3)}, \sigma_{22}^{**(3)})$, as well as the partition of $S^{(3)}$ into $D^{(3)}$ and $\overline{D}^{(3)}$. Clearly, there is an overlap between this confidence region and $\overline{D}^{(3)}$, and therefore the evidence in favour of the existence of dependence between the errors in (8) is weak. From this picture, it is also clear that the point $(\sigma_{12}^{**(3)} = 0.5, \sigma_{22}^{**(3)} = 1)$ belongs to this confidence region and therefore the equicorrelated homoskedastic probit cannot be rejected.⁹ Given the relatively small size of the sample used in this illustration, these results do not come as a surprise. However, in many applications the samples used are much larger and in those cases more conclusive results may be obtained.

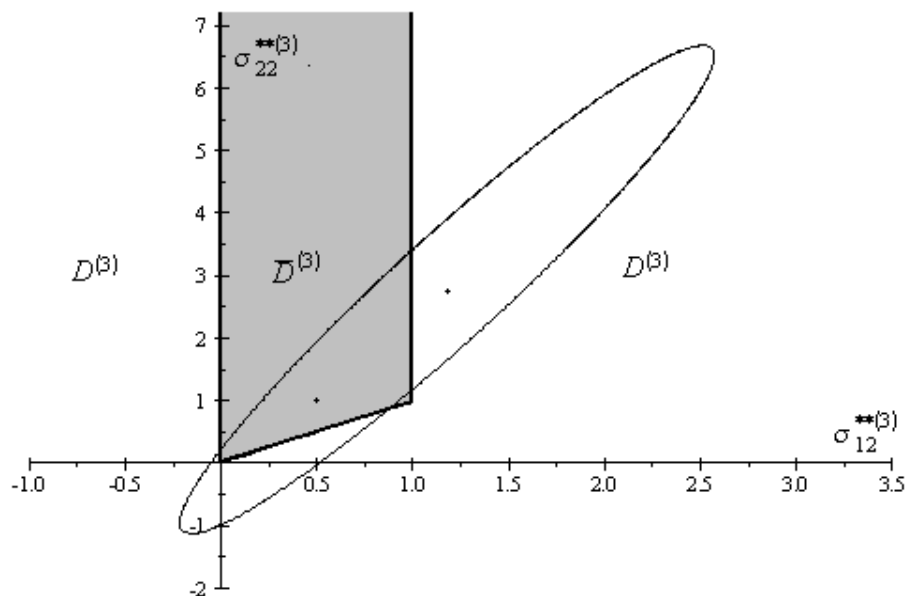


Fig. 2 - Confidence region for $\sigma_{12}^{**(3)}, \sigma_{22}^{**(3)}$ and the partition of $S^{(3)}$.

5. CONCLUSIONS

In this paper, we study what can be learnt about correlations between the errors of the original (non-normalized) stochastic utilities from multinomial probit estimates, which only identify the covariance parameters for the errors of the normalized utilities. Two results emerge.

⁹The Wald test statistic for $H_0 : \Sigma^{**} = \Sigma^{**HEP}$ is 1.421, to which corresponds a p-value of 0.491.

First, we show that, with homoskedastic errors, it is not possible to distinguish an independent probit from a probit with equicorrelated errors. Second, more interestingly, we show that certain combinations of the identified parameters from the normalized covariance matrix of the errors of the stochastic utilities, imply the existence of correlations between the errors of the original (non-normalized) utilities. This result provides additional information to practitioners, which generally believe that the estimation results on the identified parameters of the covariance matrix are not useful for inferring about the more interesting parameters of the original utilities. For this result to be useful in practice, it is necessary to have simple inference tools, to check the compatibility of the data with the existence of correlations between the errors of the original utilities. We show that this can easily be done using confidence regions, whose use is illustrated in an application.

APPENDIX

To illustrate what happens in the general case of a number of alternatives J greater than 3, consider the case of $J = 4$ alternatives. The number of distinct identified covariance elements is $M = 3$. According to Proposition 2, some dependence exists as long as an identified covariance element across the 4 possible reference state is zero or negative. To exclude this case, all identified covariance elements across the 4 possible reference state have to be strictly positive. This delivers a set of $M \times J$ inequalities. Applying the same manipulation as above, such inequalities can be written as a function of the elements corresponding to a fixed reference state, to define the inconclusive region. Taking the fourth alternative as the reference state, the restrictions defining the inconclusive region $\overline{D}^{(4)}$ can be derived as follows:

1. $\sigma_{12}^{**(4)} = \frac{\sigma_{12}^{*(4)}}{\sigma_{11}^{*(4)}} = \frac{E(\varepsilon_1 - \varepsilon_4)(\varepsilon_2 - \varepsilon_4)}{E(\varepsilon_1 - \varepsilon_4)^2} > 0$
2. $\sigma_{13}^{**(4)} = \frac{E(\varepsilon_1 - \varepsilon_4)(\varepsilon_3 - \varepsilon_4)}{\sigma_{11}^{*(4)}} > 0$
3. $\sigma_{23}^{**(4)} = \frac{E(\varepsilon_2 - \varepsilon_4)(\varepsilon_3 - \varepsilon_4)}{\sigma_{11}^{*(4)}} > 0$

4. $\sigma_{12}^{**(3)} > 0 \Leftrightarrow \frac{\sigma_{12}^{*(3)}}{\sigma_{11}^{*(3)}} > 0 \Leftrightarrow \frac{E(\varepsilon_1 - \varepsilon_3)(\varepsilon_2 - \varepsilon_3)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(3)}} > 0 \Leftrightarrow \sigma_{12}^{**(4)} - \sigma_{13}^{**(4)} - \sigma_{23}^{**(4)} + \sigma_{33}^{**(4)} > 0$
5. $\sigma_{13}^{**(3)} > 0 \Leftrightarrow \frac{E(\varepsilon_1 - \varepsilon_3)(\varepsilon_4 - \varepsilon_3)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(3)}} > 0 \Leftrightarrow -\sigma_{13}^{**(4)} + \sigma_{33}^{**(4)} > 0$
6. $\sigma_{23}^{**(3)} > 0 \Leftrightarrow \frac{E(\varepsilon_2 - \varepsilon_3)(\varepsilon_4 - \varepsilon_3)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(3)}} > 0 \Leftrightarrow -\sigma_{23}^{**(4)} + \sigma_{33}^{**(4)} > 0$
7. $\sigma_{12}^{**(2)} > 0 \Leftrightarrow \frac{E(\varepsilon_1 - \varepsilon_2)(\varepsilon_3 - \varepsilon_2)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(2)}} > 0 \Leftrightarrow \sigma_{13}^{**(4)} - \sigma_{12}^{**(4)} - \sigma_{23}^{**(4)} + \sigma_{22}^{**(4)} > 0$
8. $\sigma_{13}^{**(2)} > 0 \Leftrightarrow \frac{E(\varepsilon_1 - \varepsilon_2)(\varepsilon_4 - \varepsilon_2)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(2)}} > 0 \Leftrightarrow -\sigma_{12}^{**(4)} + \sigma_{22}^{**(4)} > 0$
9. $\sigma_{23}^{**(2)} > 0 \Leftrightarrow \frac{E(\varepsilon_3 - \varepsilon_2)(\varepsilon_4 - \varepsilon_2)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(2)}} > 0 \Leftrightarrow -\sigma_{23}^{**(4)} + \sigma_{22}^{**(4)} > 0$
10. $\sigma_{12}^{**(1)} > 0 \Leftrightarrow \frac{E(\varepsilon_2 - \varepsilon_1)(\varepsilon_3 - \varepsilon_1)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(1)}} > 0 \Leftrightarrow \sigma_{23}^{**(4)} - \sigma_{12}^{**(4)} - \sigma_{13}^{**(4)} + 1 > 0$
11. $\sigma_{13}^{**(1)} > 0 \Leftrightarrow \frac{E(\varepsilon_2 - \varepsilon_1)(\varepsilon_4 - \varepsilon_1)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(1)}} > 0 \Leftrightarrow -\sigma_{12}^{**(4)} + 1 > 0$
12. $\sigma_{23}^{**(1)} > 0 \Leftrightarrow \frac{E(\varepsilon_3 - \varepsilon_1)(\varepsilon_4 - \varepsilon_1)}{\sigma_{11}^{*(4)}} \frac{\sigma_{11}^{*(4)}}{\sigma_{11}^{*(1)}} > 0 \Leftrightarrow -\sigma_{13}^{**(4)} + 1 > 0$

REFERENCES

- Bunch, D.S. (1991). "Estimability in the Multinomial Probit Model," *Transportation Research B*, 25B, 1-12.
- Cameron, C. and Trivedi, P (2005). *Microeconometrics, Methods and Applications*, Cambridge: Cambridge University Press.
- Chesher, A. and Santos Silva, J.M.C. (2002). "Taste Variation in Discrete Choice Models," *The Review of Economic Studies*, 69, 147-168.
- Dansie, B.R. (1985). "Parameter Estimability in the Multinomial Probit Model," *Transportation Research B*, 19B, 526-528.
- Hansen, P.R. (2003). "Asymptotic Tests of Composite Hypotheses," Working Papers 2003-09, Brown University, Department of Economics.
- Herriges, J.A. and Kling, C.L. (1999). "Non linear Income Effects in Random Utility Models," *Review of Economics and Statistics*, 81, 62-72.
- Horowitz, J.L. (1981). "Testing the Multinomial Logit Model against the Multinomial Probit Model without Estimating the Probit Parameters," *Transportation Science*, 15, 153-163.
- Keane, M.P. (1992). "A Note on Identification in the Multinomial Probit Model," *Journal of Business and Economic Statistics*, 10, 193-200.
- McFadden, D. and Train, K. (2000). "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447-470.
- Monfardini, C. (2003). "An Illustration of Cox's Non-nested Testing Procedure for Logit and Probit Models," *Computational Statistics and Data Analysis*, 42, 425-444.
- Perlman, M. D. (1969). "One-Sided Testing Problems in Multivariate Analysis," *The Annals of Mathematical Statistics*, 40, 549-567.
- StataCorp (2005), *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP.

- Terza J.V. (1998), "Independence in Multinomial Choice Models: the Horowitz Test Revisited," *mimeo*, available at <http://people.musc.edu/~terza/>.
- Wolak, F. A. (1989). "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models," *Econometric Theory*, 5, 1-35.
- Wolak, F. A. (1991). "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, 59, 981-995.