Economics Bulletin

What can we learn about correlations from multinomial probit estimates?

Chiara Monfardini
Dipartimento di Scienze Economiche, Università di Bologna

J.M.C. Santos Silva
Department of Economics, University of Essex and CEMAPRE

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 note, 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 are grateful to John Mullahy for kindly suggesting the example and providing the dataset used in the empirical illustration. Santos Silva gratefully acknowledges the partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).
Citation: Monfardini, Chiara and J.M.C. Santos Silva, (2008) "What can we learn about correlations from multinomial probit estimates?." *Economics Bulletin*, Vol. 3, No. 28 pp. 1-9
Submitted: March 19, 2008. **Accepted:** May 21, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C20028A.pdf

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. Unlike the multinomial logit, the MNP does not impose any restrictions on the covariance matrix of the stochastic components of the utilities. However, due to the fact that in any random utility model the utility functions are only identified up to scale and location (Dansie, 1985), in a MNP it is not possible to identify the elements of the covariance matrix. Indeed, 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 but are unfit to be given an economic or behavioural interpretation.

In this note, we show that certain combinations of the identified parameters from the covariance matrix of the errors of the normalized stochastic utilities imply the existence of correlations between the errors of the original (non-normalized) utilities. Although we focus on the MNP, our result is 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).

2. Identified parameters in the MNP

For simplicity, we present the MNP for the three alternative case ($J = 3$), but all the results can be generalized to models with larger choice-sets. The model assumes that individuals select one of three mutually exclusive alternatives. The random utility of individual i, $i = 1, ..., N$, for choice j, $j = 1, 2, 3$, is formulated as

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

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, and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})'$ 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 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 $\sigma_{jj} > 0$, $\forall j$. Unfortunately, it is not possible to get unique maximum likelihood estimates of the parameters $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$, $\beta = (\beta_1', \beta_2', \beta_3')'$ and Σ, as they are not identified (Dansie, 1985).

One source of the identification problem is that the observed choices are only informative on the differences of the utilities. For example, taking the third alternative as the reference state and using the superscript (j) to denote the chosen reference state, we obtain

$$u_{ij}^{*(3)} = u_{ij} - u_{i3} = \alpha_j^{*(3)} + x_i'\beta_j^{*(3)} + \varepsilon_{ij}^{*(3)}$$

where, $\alpha_j^{*(3)} = \alpha_j - \alpha_3$, $\beta_j^{*(3)} = \beta_j - \beta_3$, $\varepsilon_{ij}^{*(3)} = \varepsilon_{ij} - \varepsilon_{i3}$, $j = 1, 2, 3$. As a consequence, $u_{i3}^{*(3)} = 0$ and the relevant distribution of the disturbances is the bivariate distribution of $\varepsilon_i^{*(3)} = (\varepsilon_{i1}^{*(3)}, \varepsilon_{i2}^{*(3)})'$, which is normal with zero mean and covariance matrix

$$\Sigma^{*(3)} = Cov(\varepsilon_i^{*(3)}) = \begin{pmatrix} \sigma_{11}^{*(3)} & \sigma_{12}^{*(3)} \\ \sigma_{21}^{*(3)} & \sigma_{22}^{*(3)} \end{pmatrix}$$

with $\sigma_{jk}^{*(3)} = E(\varepsilon_{ij} - \varepsilon_{i3})(\varepsilon_{ik} - \varepsilon_{i3})$, $j, k = 1, 2$.

The other source of the identification problem is that the data provides no information on the scale of the utilities. To achieve identification, it is necessary to impose a restriction on $\Sigma^{*(3)}$, 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

$$u_{ij}^{**(3)} = \frac{u_{ij}^{*(3)}}{\sqrt{\sigma_{11}^{*(3)}}} = \alpha_j^{**(3)} + x_i'\beta_j^{**(3)} + \varepsilon_{ij}^{**(3)}$$

with $\alpha_j^{**(3)} = \alpha_j^{*(3)} \Big/ \sqrt{\sigma_{11}^{*(3)}}$, $\beta_j^{**(3)} = \beta_j^{*(3)} \Big/ \sqrt{\sigma_{11}^{*(3)}}$, $\varepsilon_i^{**(3)} = \varepsilon_i^{*(3)} \Big/ \sqrt{\sigma_{11}^{*(3)}}$, $j = 1, 2$, $u_{i3}^{**(3)} = 0$ and

$$\Sigma^{**(3)} = Cov(\varepsilon_i^{**(3)}) = \begin{pmatrix} 1 & \sigma_{12}^{**(3)} \\ \sigma_{21}^{**(3)} & \sigma_{22}^{**(3)} \end{pmatrix}$$

with $\sigma_{jk}^{**(3)} = \sigma_{jk}^{*(3)} \Big/ \sigma_{11}^{*(3)}$, $j, k = 1, 2$. 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 $\sigma_{12}^{**(1)} = \sigma_{12}^{*(1)} \Big/ \sigma_{11}^{*(1)}$ and $\sigma_{12}^{**(2)} = \sigma_{12}^{*(2)} \Big/ \sigma_{11}^{*(2)}$ where

$$\begin{aligned}
\sigma_{12}^{*(1)} &= E(\varepsilon_{i2} - \varepsilon_{i1})(\varepsilon_{i3} - \varepsilon_{i1}) = \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} \\
\sigma_{12}^{*(2)} &= E(\varepsilon_{i1} - \varepsilon_{i2})(\varepsilon_{i3} - \varepsilon_{i2}) = \sigma_{13} - \sigma_{23} - \sigma_{12} + \sigma_{22} \\
\sigma_{11}^{*(1)} &= E(\varepsilon_{i2} - \varepsilon_{i1})^2 = E(\varepsilon_{i1} - \varepsilon_{i2})^2 = \sigma_{11}^{*(2)}.
\end{aligned}$$

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.

3. Inference on covariance patterns

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 1 *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 $(j = 1,2,3) \Longrightarrow \sigma_{kl} \neq 0$ *for some* k, l $(k, l = 1,2,3; k \neq l)$.

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_{jk} *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.* ■

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: 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.

Fortunately, this proposition can be made operational without having to estimate the MNP with different reference states. The inequalities $\sigma_{12}^{**(j)} \leq 0$ $(j = 1,2,3)$ can be transformed in order to make them dependent only on the identified covariance matrix elements corresponding to a chosen reference state. To show this, it is useful to go back to the one-star notation denoting only normalization of location, but not of scale, and write

$$\sigma_{12}^{*(1)} = E(\varepsilon_{i2} - \varepsilon_{i1})(\varepsilon_{i3} - \varepsilon_{i1}) = E\left[(\varepsilon_{i2} - \varepsilon_{i3}) + (\varepsilon_{i3} - \varepsilon_{i1})\right](\varepsilon_{i3} - \varepsilon_{i1}) = -\sigma_{12}^{*(3)} + \sigma_{11}^{*(3)}$$
$$\sigma_{12}^{*(2)} = E(\varepsilon_{i1} - \varepsilon_{i2})(\varepsilon_{i3} - \varepsilon_{i2}) = E\left[(\varepsilon_{i1} - \varepsilon_{i3}) + (\varepsilon_{i3} - \varepsilon_{i2})\right](\varepsilon_{i3} - \varepsilon_{i2}) = -\sigma_{12}^{*(3)} + \sigma_{22}^{*(3)}.$$

These two relations allow us to express the three inequalities above as functions of $\sigma_{12}^{**(3)}$ and $\sigma_{22}^{**(3)}$, i.e., the parameters of the identified covariance matrix after location and scale normalization, taking the third alternative as reference state. Indeed,

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

Let $S^{(3)}$ be the admissible region of the identified covariance matrix elements with reference state 3, which is defined by the condition $\sigma_{22}^{**(3)} - \left(\sigma_{12}^{**(3)}\right)^2 > 0$. Then, using the symbols $\vee$ and $\wedge$ to denote respectively the OR and AND logical operators, the result above implies the following partition of $S^{(3)} = D^{(3)} \cup \overline{D}^{(3)}$, with $D^{(3)} \cap \overline{D}^{(3)} = \varnothing$ and

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

where $\overline{D}^{(3)}$ is an inconclusive region, being consistent with both dependence or independence patterns, while points in $D^{(3)}$ correspond to the existence of some dependence. Figure 1 illustrates the partition of $S^{(3)}$.

FIGURE 1 ABOUT HERE

The inconclusive region in Figure 1 contains the point $\left(\sigma_{12}^{**(3)} = 0.5, \sigma_{22}^{**(3)} = 1\right)$, which defines the so-called independent multinomial probit (IMP), which is characterized by homoskedastic and independent errors.[1] Therefore, it appears that this point in the inconclusive region actually implies absence of correlation among the errors of the non-normalized utilities. However, it is perhaps not widely appreciated that the identified covariance structure of the IMP is indistinguishable from the identified covariance structure corresponding to a MNP with equicorrelated and homoskedastic errors. Actually, as the following proposition shows, this applies to any homoskedastic discrete choice model whose identification requires normalization of location and scale.

Proposition 2 *Equicorrelation and non-correlation are observationally equivalent in any homoskedastic discrete choice model identified through normalization of location and scale.*

Proof: *Consider a random utility model whose errors are equicorrelated and homoskedastic, with original covariance matrix* Σ^{HE} *characterized by diagonal elements* $\sigma_{jj} = \tau$, $\tau > 0$ *for all* $j = 1, ..., J$ *and off-diagonal elements* $\sigma_{jk} = \gamma$*. After normalizing the location with respect to alternative* J*, the errors are of the type* $\varepsilon_{ij}^{*(J)} = \varepsilon_{ij} - \varepsilon_{iJ}$ *and* $\sigma_{jk}^{*(J)} = E(\varepsilon_{ij} - \varepsilon_{iJ})(\varepsilon_{ik} - \varepsilon_{iJ}) = \tau - \gamma$ *for all* j, k*. The variance elements are all equal to* $\sigma_{jj}^{*(J)} = E(\varepsilon_{ij} - \varepsilon_{iJ})^2 = 2(\tau - \gamma)$*. Therefore, after normalization of scale is performed, we obtain the following identified covariance matrix*

$$\underset{(J-1)\times(J-1)}{\Sigma^{**HE}} = \begin{bmatrix} 1 & 0.5 \cdots & 0.5 \\ 0.5 & 1 \;\; 0.5 & \vdots \\ \vdots & 0.5 \;\; \ddots & 0.5 \\ 0.5 \cdots & 0.5 & 1 \end{bmatrix}.$$

The particular case of independence 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).* ■

The set of inequalities defining $D^{(3)}$ can be formally tested using the procedures initially developed by Perlman (1969), 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 $D^{(3)}$. It is important to notice, however, that the procedure proposed below is not a specification test, nor is it designed to assess the validity of the IMP. Rather, its purpose is simply to provide the researcher with some additional information about the structure of the covariance matrix of the non-normalized utilities.

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 $\delta_j = \hat{\sigma}_{j2}^{**(3)} - \sigma_{j2}^{**(3)}$, $j = 1, 2$, standard asymptotic results imply that

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

$\delta'\hat{V}^{-1}\delta \mathrel{\dot\sim} \chi^2_{(2)}$. Then, using $q^{(v)}_{1-\alpha}$ to denote the $1-\alpha$ quantile of the $\chi^2_{(v)}$ distribution, the inequality $\delta'\hat{V}^{-1}\delta < q^{(2)}_{1-\alpha}$ defines a confidence region of level $100\,(1-\alpha)\,\%$ for the vector $\left(\sigma^{**(3)}_{12}, \sigma^{**(3)}_{22}\right)'$. If this confidence region does not overlap with $\overline{D}^{(3)}$, this provides evidence supporting the existence of at least one non-zero covariance between the errors of the non-normalized utilities.

The inconclusive region is defined by a set of $J(J-1)(J-2)/2$ inequalities and, for $J > 3$, visual inspection is not appropriate to check whether it overlaps with the confidence ellipsoid. In this case, information on the possible overlapping of the two regions can be obtained by minimizing 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 $q^{(v)}_{1-\alpha}$, with $v = [(J-1)J/2] - 1$ being the number of identified elements of $\Sigma^{**(J)}$, one concludes that there is an overlapping of the two regions.

4. An empirical illustration

To illustrate the application of our results, we use data from the 2002 U.S. Medical Expenditure Panel Survey (MEPS) and estimate a model of health insurance choice. Our dataset consist of a sample of 22,283 individuals aged between 18 and 64, for which we have information on insurance status (any private insurance, only public insurance, no insurance), as well as on a set of standard covariates. Descriptive statistics for these variables, and the dataset itself, can be obtained from the authors upon request.

In the specification we adopt, the stochastic utilities associated with the three alternatives depend on all the available regressors, whose parameters vary across the alternatives. Table 1 presents the main estimation results obtained using "no insurance" as the reference state for normalization. The names of the covariates are self-explanatory, and we notice that the omitted geographic region is West.

TABLE 1 ABOUT HERE

The pair $\left(\hat{\sigma}^{**(3)}_{12}, \hat{\sigma}^{**(3)}_{22}\right)$ falls into $D^{(3)}$, which is compatible with the existence of some dependence between the error terms of the non-normalized stochastic utilities. In order to check the strength of this evidence, we follow the procedure described in the previous section to check whether a 95% confidence region for $\left(\sigma^{**(3)}_{12}, \sigma^{**(3)}_{22}\right)$ overlaps with $\bar{D}^{(3)}$. Minimization of $\delta'\hat{V}^{-1}\delta$ with respect to $\sigma^{**(3)}_{12}$ and $\sigma^{**(3)}_{22}$, subject to the restriction that these parameters belong to $\overline{D}^{(3)}$, leads to $\delta'\hat{V}^{-1}\delta = 9.254$, for $\sigma^{**(3)}_{12} \cong 1$ and $\sigma^{**(3)}_{22} \cong 1$. Since $9.254 > q^{(2)}_{0.95} = 5.99$, the 95% confidence region does not overlap with $\bar{D}^{(3)}$. This provides evidence in favour of the existence of non-zero covariances between the errors of the original utilities attached with each insurance status.

5. Concluding remarks

Horowitz (1981) and Terza (1998) proposed specification tests which allow the researcher to choose between the IMP and MNP. Our results complement this sort of tests. While Proposition 2 clarifies the meaning of their null hypothesis, Proposition 1 provides additional information on the structure of the covariance matrix of the errors of the non-normalized utilities when the null is rejected. Naturally, the procedure we propose and illustrate for multinomial probit model has a broader scope of application and can be useful in the context of other multinomial choice models, like the mixed multinomial logit.

References

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.

Horowitz, J.L. (1981). "Testing the Multinomial Logit Model against the Multinomial Probit Model without Estimating the Probit Parameters," *Transportation Science*, 15, 153-163.

McFadden, D. and Train, K. (2000). "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447-470.

Perlman, M. D. (1969). "One-Sided Testing Problems in Multivariate Analysis," *The Annals of Mathematical Statistics*, 40, 549-567.

Terza J.V. (1998), "Independence in Multinomial Choice Models: the Horowitz Test Revisited," *mimeo*, available at http://people.musc.edu/~terza/.

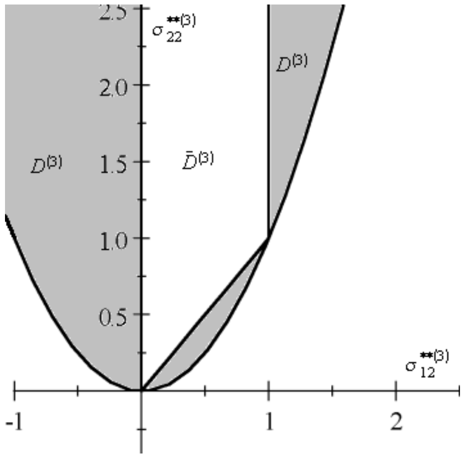

Fig. 1 - The partition of $S^{(3)}$.

Table 1. Trinomial probit results for health insurance choice

	Any private		Only public	
Regressors	Coef.	St. err.	Coef.	St. err.
Intercept	−0.79305	0.07489	−0.68841	0.07473
Family size	−0.01488	0.00606	−0.02104	0.00800
North East	0.25538	0.03474	0.26283	0.03856
Mid West	0.30107	0.03253	0.23135	0.05630
South	−0.07510	0.02572	−0.14101	0.05249
Age	0.00592	0.00083	0.00686	0.00109
Female	0.26078	0.02110	0.31775	0.04561
Black	0.06038	0.03045	0.18853	0.09215
Years of education	0.07909	0.00437	0.05456	0.01532
Personal income	0.01406	0.00665	0.00466	0.00665
Covariance parameters		Estimate	St. err.	
$\sigma_{12}^{**(3)}$		1.12512	0.09264	
$\sigma_{22}^{**(3)}$		1.32076	0.28646	
$Cov\left(\widehat{\sigma}_{12}^{**(3)}, \widehat{\sigma}_{22}^{**(3)}\right)$	0.02645			
Log-likelihood	−15260.6			