 Journal of Business & Economic Statistics, April 1992, Vol. 10, No. 2

A Note on Identification in the Multinomial Probit Model

Michael P. Keane
Industrial Relations Center and Department of Economics, University of Minnesota, Minneapolis, MN 55455, and Research Department, Federal Reserve Bank of Minneapolis, Minneapolis, MN 55480

Although formal conditions for identification in the multinomial probit (MNP) model are now clearly established, little is known about how various estimable MNP specifications perform in practice. This article shows that parameter identification in the MNP model is extremely tenuous in the absence of exclusion restrictions. This previously unnoticed fact is important because formal identification of MNP models does *not* require exclusion restrictions, and many potential economic applications of MNP are to situations in which exclusion restrictions are not readily available. Thus, failure to be aware of the difficulties present in such situations may lead to reporting of unreliable results.

KEY WORDS: Discrete choice; Latent variables; Parameter estimability.

The multinomial probit (MNP) model has rarely been used as a model of choice in applied work, despite its well-known advantages over the popular logit model (i.e., its relaxation of the restrictive independence of irrelevant alternatives assumption). The lack of use of MNP stems from the computational burden involved in its estimation. The model generates choice probabilities that are multivariate integrals of order $M - 1$, where M is the number of alternatives. Thus, even when $M = 3$, estimation by maximum likelihood (ML) is expensive given large data sets.

The recent development of a computationally practical method of simulated moments (MSM) estimator for the MNP model (see McFadden 1989; Pakes and Pollard 1989), along with the development of practical parameterization methods that avoid proliferation of covariance matrix elements (see Ben-Akiva and Bolduc 1989; Bolduc 1991; Elrod and Keane 1991), has raised renewed interest in MNP as a model of choice. Given the lack of practical experience with MNP models, however, there is a need to develop a "folklore" concerning the conditions under which the model performs well. An important step in that direction is the recent paper by Bunch and Kitamura (1991).

The purpose of this article is to demonstrate that parameter identification in MNP models is extremely tenuous in the absence of *exclusion restrictions*. By exclusion restrictions I mean restrictions that certain exogenous variables in the model do not affect the utility levels of certain alternatives. This fact is important, because formal identification of MNP models does *not* require exclusion restrictions, and (to my knowledge) the fact that identification is tenuous in their absence has not been previously noted in the literature. Hence, given the new MSM technology, researchers are likely to attempt simulation estimation of unrestricted multinomial probits. This may, in turn, lead to the reporting of unreliable results.

For clarity of exposition, it is important to note that there are two types of identification problems. For a formally nonidentified model, there is a range of parameter values that generate the maximized value of the objective function. In addition, a model may be formally identified yet exhibit very small variation in the objective function from its maximum over a wide range of parameter values. I refer to identification in such cases as being *tenuous* or *fragile*. Common symptoms of this problem in practice are a close-to-singular Hessian, large standard errors, and inability of optimization algorithms to find steps that improve the objective or to achieve convergence. As we shall see, these symptoms of fragile identification are quite severe in MNP models without exclusion restrictions.

Note that in complex nonlinear models such as MNP, formal identification or nonidentification is often very difficult to prove due to the complexity of the analytical Hessian. Thus, in this article, rather than working with the analytical Hessian, I use a series of trial estimations of MNP models on Monte Carlo and actual data to illustrate the nature of the identification problem. Because of the difficulties in proving identification in nonlinear models, a common practice in applied work is to simply attempt to estimate a model and see if the Hessian is singular (or nearly singular). Such practice is dangerous when using simulation estimators of the type likely to be applied to MNP models, because simulation error will generate contours where the true objective function is flat and will generate a nonsingular Hessian when the true Hessian is singular. An illustration of this danger was provided by Horowitz, Sparmann, and Daganzo (1982). They were testing the accuracy of simulated ML based on the Clark approximation for purposes of estimating the MNP model. One of the models they used for the tests was actually nonidentified, but approximation errors introduced contours in the objective function that masked the problem and allowed them to obtain estimates. (Their other tests did use identified models, however, and they indicate that the Clark approximation does not work well for purposes of estimating the MNP model.) Because of this problem, all of the estimation in this article is done using ML estimation based on accurate numerical evaluations of the MNP choice probabilities.

1. THE MULTINOMIAL PROBIT MODEL

In this section, I describe the trinomial probit model (TNP). Extension to multinomial probit is obvious. In the TNP model there are three alternatives with random utilities given by

$$\begin{aligned} U_{1i} &= \alpha_1 + \beta_1 X_i + \varepsilon_{1i} \\ U_{2i} &= \alpha_2 + \beta_2 X_i + \varepsilon_{2i} \\ U_{3i} &= \alpha_3 + \beta_3 X_i + \varepsilon_{3i}. \end{aligned} \tag{1}$$

Here, X_i is a vector of regressors for person i, and β_j is the corresponding coefficient vector for alternative j. In this notation, there are exclusion restrictions if we have $\beta_{jk} = 0$ for some element k of X_i but $\beta_{j'k} \neq 0$ for some $j' \neq j$. For example, if X_{ik} is an attribute of alternative 1, a common exclusion restriction would be $\beta_{jk} = 0$ for $j \neq 1$. Such exclusion restrictions arise naturally in problems of transportation-mode choice—the area where MNP has usually been applied—because mode attributes are readily available in data. Then, it is natural to assume that attributes of mode j (such as price or travel time) affect only the utility generated by mode j and do not affect the utilities generated by other modes. In many economic applications, however (e.g., choice of industry or occupation), the data available to the econometrician will typically include only attributes of survey respondents and not of respondents' alternative choices.

The key feature separating probit from other choice models (i.e., logit or conditional logit) is the assumption that the stochastic terms $(\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i})$ have a multivariate normal distribution with covariance matrix

$$\operatorname{cov}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}. \tag{2}$$

In this model, the probability of individual i choosing alternative 1 is given by

$$\Pr_{1i} = \int_{w=-\infty}^{(\overline{U}_1 - \overline{U}_2)/(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})^{1/2}} \phi(w) \times \Phi[(\overline{U}_1 - \overline{U}_3)/[(\sigma_1^2 + \sigma_3^2 - 2\sigma_{13})(1 - r_1^2)]^{1/2} - wr_1/(1 - r_1^2)^{1/2}]dw, \tag{3}$$

where $\overline{U}_j \equiv \alpha_j + \beta_j X_i$, $r_1 \equiv (\sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23})/[(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})(\sigma_1^2 + \sigma_3^2 - 2\sigma_{13})]^{1/2}$, and $\phi(\,)$ and $\Phi(\,)$ are the normal density and distribution function, respectively (see Hausman and Wise 1978). Here r_1 is the correlation between $\varepsilon_1 - \varepsilon_2$ and $\varepsilon_1 - \varepsilon_3$. The probabilities of choosing alternatives 2 and 3 are expressed symmetrically.

This unrestricted model is not identified for two obvious reasons. First, a proportional change in all elements of the covariance matrix and of the (α_j, β_j) for $j = 1, 3$ leaves all probabilities unaffected. Second, addition of a constant to all of the α_j leaves all probabilities unchanged, because choice depends only on utility differences.

Thus, as described by Bunch (1991), Dansie (1985), and Albright, Lerman, and Manski (1977), identification may be achieved by normalizing the utility of one alternative to 0 and by restricting one covariance matrix element. (Other normalizations are possible.) Setting $U_3 = 0$ and $\sigma_1 = 1$, we have the model

$$\begin{aligned} U_{1i} &= \alpha_1 + \beta_1 X_i + \varepsilon_{1i} \\ U_{2i} &= \alpha_2 + \beta_2 X_i + \varepsilon_{2i} \\ U_{3i} &= 0 \end{aligned} \tag{1$'$}$$

with error covariance matrix

$$\operatorname{cov}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} 1 & \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}. \tag{2$'$}$$

Then we have

$$\Pr_{1i} = \int_{w=-\infty}^{(\overline{U}_1 - \overline{U}_2)/(1 + \sigma_2^2 - 2\sigma_{12})^{1/2}} \phi(w) \times \Phi[\overline{U}_1/(1 - r_1^2)^{1/2} - wr_1/(1 - r_1^2)^{1/2}]dw, \tag{3$'$}$$

where $r_1 = (1 - \sigma_{12})/(1 + \sigma_2 - 2\sigma_{12})^{1/2}$, the correlation between $\varepsilon_1 - \varepsilon_2$ and ε_1. Again, $\Pr_{2i}$ and $\Pr_{3i}$ are expressed symmetrically.

Note that here, for any *particular* X vector, we can always find alternative values for (α_j, β_j), $j = 1, 2$, and (σ_{12}, σ_2) that give the same values for the choice probabilities. We cannot, however, find alternative values for these parameters that give the same values for the choice probabilities for *all* X. Thus, as was pointed out by Heckman and Sedlacek (1985), the TNP model is identified so long as X contains a single regressor that varies over individuals. No exclusion restrictions are required for formal identification. Unfortunately, even the proper conditions for formal identification of MNP models appear to be not widely known. Bunch and Kitamura (1991) pointed out that nearly half of the existing applications of MNP have used formally nonidentified models.

2. IDENTIFICATION PROBLEMS IN THE TRINOMIAL PROBIT MODEL

This section presents evidence that identification in the TNP model of Equations (1′) and (2′) is extremely fragile in the absence of exclusion restrictions. This evidence is from both Monte Carlo and actual data. Consider first the evidence from Monte Carlo data. A data set of size 8,000 was constructed by drawing values for a single regressor X_i (i = 1), 8,000 from the $N(6, 5)$ distribution, and by drawing $(\varepsilon_{1i}, \varepsilon_{2i})$ for i = 1, 8,000 with covariance matrix given by (2′).

TNP model estimates were obtained by ML using the algorithm of Berndt, Hall, Hall, and Hausman (1974), which will be referred to as the BHHH algorithm. Let θ denote the vector of all parameters of the model, and let θ^k denote the estimate of θ at iteration k. The log-likelihood function L for the TNP model is given by

$$L(\theta^k) = \sum_{i=1}^{N} \sum_{j=1}^{3} d_{ij} \ln \Pr_{ji}(\theta^k), \tag{4}$$

where d_{ij} is an indicator equal to 1 if i chooses alternative j and 0 otherwise. The TNP choice probabilities $\Pr_{ji}(\theta^k)$ were evaluated using 100 term tetrachoric expansions. A modified Newton–Raphson step is given by

$$\theta^{k+1} = \theta^k - \lambda^k H^{-1} \left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta^k}, \tag{5}$$

where

$$H \equiv \left.\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\right|_{\theta^k}$$

is the Hessian evaluated at θ^k and λ^k is a step size that is 1 on the first step of an iteration but that is reduced if a unit step does not improve the log-likelihood function. Of course, numerical evaluation of the Hessian is quite computationally expensive in this model. In the BHHH algorithm, the fact that the expected value of the Hessian is (at the optimum) equal to the negative of the expectation of the outer product of the gradient vectors is invoked to justify approximation of the Hessian by

$$H \approx - \left.\frac{\partial L(\theta)}{\partial \theta} \frac{\partial L(\theta)}{\partial \theta'}\right|_{\theta^k}.$$

In this article, this approximation to the Hessian was used to obtain steps and in covariance matrix calculations. When using Monte Carlo data, the outer product of the gradients was evaluated at the θ vector. Since in the Monte Carlo data we know the true model and since the sample size of 8,000 is rather large, this approximation to the Hessian should be a good one.

The first set of results using Monte Carlo data is presented in Table 1. The true values of the parameters are reported in the column headed "True value" [here ρ is defined as corr$(\varepsilon_1, \varepsilon_2)$]. The columns headed (1)–(4) report estimates obtained with ρ and σ_2 pegged at various values. The $\chi^2(2)$ tests are for the null that the hypothesized constraints are valid [note that the optimized value of the log-likelihood for the unrestricted model is −7,953.57 (Table 2, col. (1))]. The true values of ρ and σ_2 used in constructing the data were .60 and 1.50, respectively.

In column (2), ρ is restricted to 0 (whereas σ_2 is restricted to the true value of 1.5). The deterioration of the log-likelihood resulting from this restriction is only .57 of a point. Similarly, in Column (3), σ_2 is restricted to 1.00 whereas ρ is pegged at the true value of .60. The deterioration of the log-likelihood resulting from this restriction is only one point. In column (4), ρ is restricted to 0, and σ_2 is restricted to 1.00. The deterioration in the log-likelihood is 2.25 points, giving a $\chi^2(2)$ value of 4.5 that is not significant at the 10% level (critical value = 4.605).

It is apparent that restrictions on ρ and σ_2 produce some slight deterioration in fit of the TNP model without exclusion restrictions—as they must because these parameters are formally identified. The improvement in fit obtained by introducing these parameters is so minor, however, as to render their identification in practice problematic. The problem is illustrated in Table 2. Here TNP models were estimated with ρ and σ_2 free. The starting values used for the runs in columns (1)–(4) of Table 2 are, respectively, the estimates in columns (1)–(4) of Table 1. Classic symptoms of fragile identification were found in all runs in Table 2. First, the Hessian was so close to singular that, to obtain a sensible direction vector for use in BHHH iterations, the Marquadt (1963) procedure of adding positive diagonal elements to the (approximate) Hessian had to be applied (note, however, that this was not done when cal-

Table 1. Trinomial Probit Model—ML Estimates With ρ and σ_2 Restricted

Parameter	True value	(1)	(2)	(3)	(4)
ρ	.60	.60	.00	.60	.00
σ_2	1.50	1.50	1.50	1.00	1.00
α_1	−.80	−.7593	−.7411	−.7148	−.6875*
		(.0503)	(.0494)	(.0476)	(.0471)
β_1	.20	.1943	.1260**	.1633**	.1029**
		(.0093)	(.0089)	(.0086)	(.0083)
α_2	−2.00	−1.9234	−2.2000*	−1.2490**	−1.4811**
		(.0734)	(.0781)	(.0524)	(.0545)
β_2	.40	.3933	.3927	.2782**	.2759**
		(.0124)	(.0129)	(.0091)	(.0092)
Log-likelihood		−7,953.57	−7,954.14	−7,954.57	−7,955.82
$\chi^2(2)$		.00	1.14	2.00	4.50

NOTE: Standard errors are in parentheses. Double asterisks indicate that an estimated parameter differs from the true value at the 1% level. An asterisk indicates the 5% level. The $\chi^2(2)$ statistic is for the null that the restrictions on ρ and σ_2 are valid (the 10% critical value is 4.605). Sample size is 8,000.

Table 2. Trinomial Probit Model—ML Estimates With ρ and σ_2 Unrestricted

Parameter	True value	(1)	(2)	(3)	(4)
ρ	.60	.60007 (1.6274)	.0228 (1.0462)	.6121 (1.3209)	.0176 (1.1423)
σ	1.50	1.5065 (2.4405)	1.5106 (1.5031)	1.0133 (.6053)	1.0122 (.7544)
α_1	−.80	−.7590 (.1797)	−.7435 (.0701)	−.7161 (.0966)	−.6902 (.0711)
β_1	.20	.1945 (.1073)	.1288 (.0766)	.1658 (.1036)	.1052 (.0747)
α_2	−2.00	−1.9303 (4.4203)	−2.2069 (2.3958)	−1.2582 (1.5697)	−1.4932 (1.3220)
β_2	.40	.3944 (.5830)	.3954 (.3169)	.2810 (.1719)	.2790 (.1609)
Log-likelihood		−7,953.57	−7,954.05	−7,954.34	−7,955.49

NOTE: Standard errors are in parentheses. Sample size is 8,000.

culating the covariance matrices). Second, despite using the Marquadt procedure, it was extremely difficult for the algorithm to find improving steps, and in all four instances the parameters did not move far from the starting values. Third, the estimated standard errors of the ρ and σ_2 estimates are very large, and the standard errors of all other parameters increase dramatically when ρ and σ_2 are unrestricted. For example, compare column (1) of Table 2, in which all estimates are close to the true values, with column (1) of Table 1, in which ρ and σ_2 are fixed at their true values. In column (1), the estimated standard error of $\hat{\rho}$ is 1.6274, so all points in the -1 to 1 range are within one standard error of the estimate. The standard error of $\hat{\sigma}_2$ is 2.4405, so σ_2 is only .62 of a standard error above 0. Furthermore, freeing ρ and σ_2 to be estimated increases the standard errors on the regressor coefficients by up to 6,000% (the standard error on $\hat{\alpha}_1$ increases from .05 to .18, that on $\hat{\beta}_1$ from .009 to .107, that on $\hat{\alpha}_2$ from .07 to 4.42, and that on $\hat{\beta}_2$ from .012 to .58).

These results are not isolated to this particular data set but were consistently present in many experiments (and also in actual data, as we shall see). The source of the fragility of identification in the TNP model is that movements in the regressor coefficients can effectively mimic the effects of changes in the covariance parameters. Thus, restricting the covariance parameters has little effect on the fit of the model, and the values of the covariance parameters are difficult to disentangle from the values of the regressor coefficients.

This is illustrated in Figure 1, in which the true $\overline{U}_1$ and $\overline{U}_2$ are plotted against those estimated in Table 1, column (2). Here the true ρ is .60, but ρ is fixed at 0. The effect of an increase in ρ is twofold: (1) It increases the probability of alternative 2 when $\overline{U}_2 > \overline{U}_1$ and reduces the probability of alternative 2 when $\overline{U}_2 < \overline{U}_1$, and (2) since *both* ε_1 and ε_2 must be less than certain thresholds to generate choice 3, it increases the probability of alternative 3. To see why the first effect is present, observe that as $\rho \to 1$ we approach a deterministic rule whereby the alternative that gives the largest $\overline{U}$ is always chosen. Concerning the second effect, observe that as $\rho \to 1$, ε_1 and ε_2 collapse into a single random variable, so the probability of them both being less than a certain threshold increases. When ρ is incorrectly fixed at 0, both the $\overline{U}_1$ and $\overline{U}_2$ lines shift to mimic these two effects of an increase in ρ. Looking at Figure 1, we see that the $\overline{U}_1$ line flattens, thereby increasing the distance between $\overline{U}_1$ and $\overline{U}_2$ at any given distance to the right or left of the $\overline{U}_1 = \overline{U}_2$ point. This increases the probability of alternative 2 when $\overline{U}_2 > \overline{U}_1$ (and vice versa), thus mimicking effect 1. Moreover, we see that both the $\overline{U}_1$ and $\overline{U}_2$ lines shift downward. This increases the probability of alternative 3, thus mimicking effect 2.

In Figure 2, the true $\overline{U}_1$ and $\overline{U}_2$ are plotted against those estimated in Table 1, column (3). Here the true σ_2 is 1.5, but σ_2 is fixed at 1.0. The effect of an increase in σ_2 is twofold: (1) It reduces the probability of alternative 2 when $\overline{U}_2 > \overline{U}_1$ and increases the probability of alternative 2 when $\overline{U}_2 < \overline{U}_1$, and (2) it increases the probability of alternative 3 when $\overline{U}_2 > 0$ and reduces the probability of alternative 3 when $\overline{U}_2 < 0$. In Figure 2, observe that the $\overline{U}_2$ line flattens relative to the $\overline{U}_1$ line, thus reducing the distance between $\overline{U}_1$ and $\overline{U}_2$ at any given distance to the right or left of the $\overline{U}_1 = \overline{U}_2$ point. This reduces the probability of alternative 2 when $\overline{U}_2 > \overline{U}_1$ (and vice versa), thus mimicking effect 1. The pronounced flattening of the $\overline{U}_2$ curve also mimics effect 2.

Note that changes in ρ and σ_2 both affect the relative probabilities of alternatives 1 and 2 to the right and left of the $\overline{U}_1(X) = \overline{U}_2(X)$ point. Increases in ρ increase the probability that alternative j will be chosen if $\overline{U}_j(X) > \overline{U}_{j'}(X)$. Increases in σ_2 reduce the probability that alternative 2 will be chosen when $\overline{U}_2(X) > \overline{U}_1(X)$. These effects can be mimicked by changes in the relative slopes of the $\overline{U}_1(X)$ and $\overline{U}_2(X)$ lines that either increase or reduce the distance $|\overline{U}_1(X) - \overline{U}_2(X)|$ at any given distance to the right or left of the $\overline{U}_1(X) = \overline{U}_2(X)$ point (i.e., for *all* X). This suggests that different effects of ρ, σ_2, and the regressors may be disentangled by introducing exclusion restrictions. Then there are multiple $\overline{U}_1(X) = \overline{U}_2(X)$ points and rotations of the $\overline{U}_1(X)$ and $\overline{U}_2(X)$ lines that either increase or reduce the distance $|\overline{U}_1(X) - \overline{U}_2(X)|$ for *all* X are impossible.

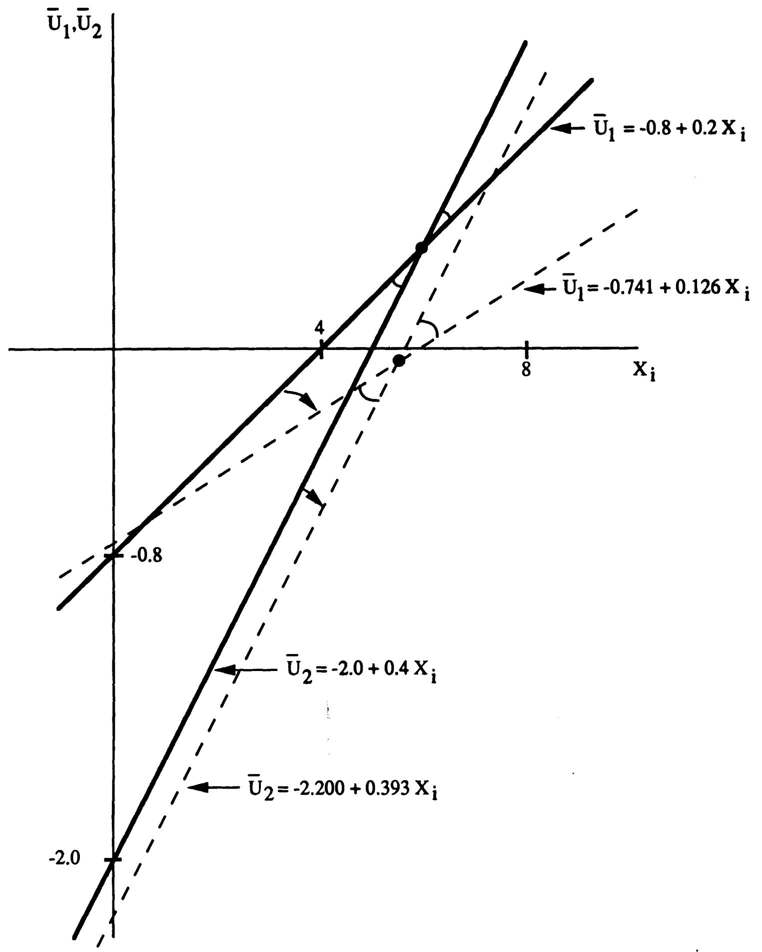

Figure 1. Effect of Fixing ρ at .0 When True ρ = .6. Note that the solid lines are the true $\overline{U}_1$ and $\overline{U}_2$ lines. The dashed lines are the $\overline{U}_1$ and $\overline{U}_2$ lines estimated with ρ fixed at .0 when the true ρ is .60 [see Table 1, col. (2)].

A TNP model with exclusion restrictions is estimated in Table 3 (p. 199). This model has the form

$$\overline{U}_{1i} = \alpha_1 + \beta_{11}X_{1i} + \beta_{12}X_{2i}$$

$$\overline{U}_{2i} = \alpha_2 + \beta_{21}X_{1i} + \beta_{23}X_{3i}, \qquad (6)$$

where X_{1i} is constructed as before, but X_{2i} and X_{3i} are dummy variables equal to 1 with probability .50. Thus there are four $\overline{U}_1(X_{1i}) = \overline{U}_2(X_{1i})$ points corresponding to the four possible values of (X_{1i}, X_{2i}), and any pivoting to increase or reduce the distance $|\overline{U}_1(X_{1i}) - \overline{U}_2(X_{1i})|$ for all X_{1i} is impossible.

The true values of the parameters of the model are reported in the column headed "True value." All values are the same as before except that the new parameters β_{12} and β_{23} equal $-.60$. The unrestricted model results are reported in column (4). In contrast to the results in Table 2, all of the parameters of the model are estimated with precision. Columns (1)–(3) report estimates obtained with ρ and σ_2 restricted as in Table 1, columns (2)–(4). Here, all of the restrictions are overwhelmingly rejected. For example, the $\chi^2(2)$ statistic for the restriction $\rho = .0$, $\sigma_2 = 1.0$ is 44.78, compared to a 1% significance level of 9.21. Clearly, the problem of fragile identification in the TNP model is solved by introducing exclusion restrictions.

Experimentation with a wide range of specifications revealed that in MNP models it is necessary to have one exclusion from each utility index to avoid identification problems. Simply introducing additional regressors, without introducing exclusion restrictions, does not solve the problem. This is illustrated by the results in column (5) of Table 3, where estimates were obtained with the $\beta_{13} = \beta_{22} = 0$ restrictions removed [the results from col. (4) were used as starting values]. The symptoms of fragile identification, including increases in the standard errors of up to 700%, are again apparent here.

It is also apparent that the identification problems found in the preceding Monte Carlo data do not result from the particular distributional assumptions on the X_i. This is best seen by considering a TNP model estimated on actual data. In Table 4, I report results of estimating a model of industry choice on data from the National Longitudinal Survey of Young Men (NLS). This application is motivated by the fact that the only application of MNP in labor economics is that of Heckman and Sedlacek (1985), who considered industry choice in Current Population Survey data. As in the work of Heckman and Sedlacek, industries are grouped into three alternatives—manufacturing (M), nonmanufacturing (NM), and unemployment. The NLS sample used

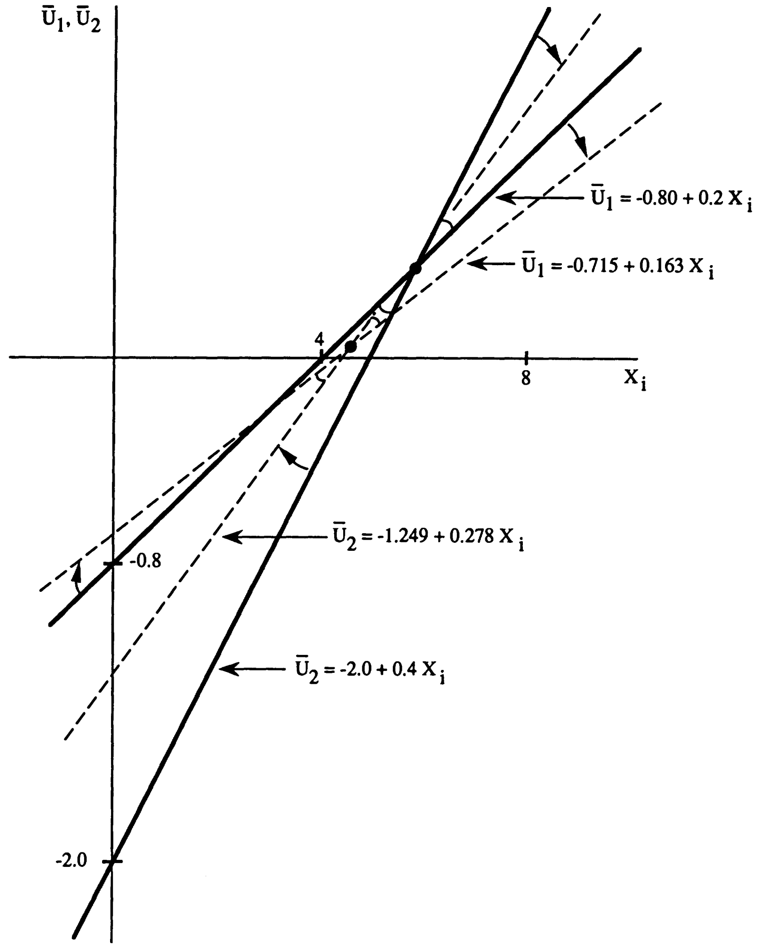

Figure 2. Effect of Fixing σ_2 at 1.0 When True σ_2 = 1.5. Note that the solid lines are the true $\bar{U}_1$ and $\bar{U}_2$ lines. The dashed lines are the $\bar{U}_1$ and $\bar{U}_2$ lines estimated with σ_2 fixed at 1.0 when the true σ_2 is 1.5 [see Table 1, col. (3)].

here has 11,886 person-year observations from 1966–1981. For a complete description of the data, see Keane, Moffitt, and Runkle (1988). The person-specific regressors are the national unemployment rate (U-RATE), a time trend (TREND), years of education (EDUC), labor market experience and its square (EXPER, EXPER^2), a dummy to indicate if respondent is white (WHITE), a dummy to indicate if respondent is married (WIFE), and the number of child dependents of the respondent (KIDS).

Following Heckman and Sedlacek (1985), an alternative specific variable—sector-specific nonlabor income (NLINC)—is also included in the model. The rationale for this variable is that it captures differences by sector in unemployment compensation and other benefits. Since nonlabor income is only observed for the chosen sector, instruments of the type used by Heckman and Sedlacek are constructed as follows: Nonlabor income for all workers in a particular sector is regressed on a set of instruments. Instruments are respondent's age, years of education, a Standard Metropolitan Statistical Area (SMSA) resident dummy, a South regional dummy, local labor-force size, TREND, EXPER, EXPER^2, WHITE, WIFE, KIDS, and the interactions of age with the SMSA and South dummies and of the SMSA dummy with the South dummy. Fitted values of sector-specific nonlabor income are then constructed for each sector for each worker. NLINC for a particular sector only enters the utility index for that sector, so the estimated model has exclusion restrictions. The other restrictions are that the utility index for the unemployment alternative is normalized to 0 and the manufacturing-utility-index error is normalized to have unit variance. ρ is the correlation of the manufacturing and nonmanufacturing utility function errors, and σ_2 is the standard deviation of the nonmanufacturing error.

The first column of Table 4 reports results obtained with ρ pegged at 0 and σ_2 pegged at 1. This model gives a log-likelihood value of $-10{,}300.710$. Unfortunately, the coefficient on nonlabor income in nonmanufacturing is estimated to be negative. The column (2) estimates are obtained with ρ and σ_2 free, using the column (1) estimates as starting values. As with the Monte Carlo data, this again resulted in a close-to-singular (approximate) Hessian, so that the Marquadt procedure again had to be used to obtain a reasonable step matrix for the BHHH algorithm. (Again, however, the procedure

Table 3. Trinomial Probit Model With Exclusion Restrictions—ML Estimates

Parameter	True value	(1)	(2)	(3)	(4)	(5)
ρ	.60	.00	.60	.00	.6368 (.0786)	.7190 (.2846)
σ	1.50	1.50	1.00	1.00	1.5215 (.2099)	2.5786 (1.4086)
α_1	−.80	−.7336 (.0516)	−.7649 (.0489)	−.6811 (.0497)	−.8110 (.0522)	−.7961 (.0839)
β_{11}	.20	.1338** (.0088)	.1673** (.0083)	.1125** (.0083)	.2023 (.0151)	.2295 (.0234)
β_{12}	−.60	−.6501 (.0350)	−.5063** (.0269)	−.6336 (.0339)	−.5324 (.0451)	−4786 (.1284)
β_{13}	0					−.0855 (.1039)
α_2	−2.00	−2.1767* (.0812)	−1.3103** (.0534)	−1.4698 (.0564)	−2.0146 (.2903)	−3.4868 (2.2865)
β_{21}	.40	.4041 (.0128)	.2832** (.0088)	.2814** (.0091)	.4086 (.0492)	.6431 (.3343)
β_{22}	0					.2376 (.4283)
β_{23}	−.60	−.7235** (.0475)	−.4120** (.0262)	−.4848 (.0382)	−.5984 (.0947)	−1.0880 (.7287)
Log-likelihood		−7,675.53	−7,661.55	−7,678.72	−7,656.33	−7,655.17
$\chi^2(2)$		38.40**	10.40**	44.78**		2.32

NOTE: Standard errors are in parentheses. Double asterisks indicate that an estimated parameter differs from the true value at the 1% level. An asterisk indicates the 5% level. In columns (1)–(3), the $\chi^2(2)$ statistic is for the null that the restrictions on ρ and σ_2 are valid. In column (5), the $\chi^2(2)$ statistic is for the null that $\beta_{13} = \beta_{22} = 0$. Sample size is 8,000.

was not used when calculating covariance matrices.) All of the estimates remain very close to their column (1) values, and the standard errors for several (in particular ρ and σ_2) are very large. The log-likelihood value obtained is −10,299.700, which is an insignificant 1.01 improvement over the constrained model. The column (3) estimates are also obtained with ρ and σ_2 free, using as starting values estimates from a model including only

Table 4. Trinomial Probit Model of Industry Choice—ML Estimates

	(1)		(2)		(3)	
Parameter	M	NM	M	NM	M	NM
Regressor coefficients						
NLINC	.0077** (.0033)	−.0471** (.0061)	.0081** (.0036)	−.0473** (.0229)	.0029 (.0028)	−.0367 (.0255)
U-RATE	−.0760** (.0145)	−.0484** (.0135)	−.0752** (.0176)	−.0484** (.0179)	−.0978** (.0202)	−.0804** (.0222)
TREND	−.0231** (.0070)	.0462** (.0089)	−.0233** (.0078)	.0461* (.0270)	−.0116 (.0078)	.0370 (.0314)
EDUC	.0121* (.0071)	.1083** (.0074)	.0138 (.0140)	.1106** (.0496)	.0328** (.0152)	.1024* (.0552)
EXPER	.0252** (.0109)	−.0290** (.0110)	.0269** (.0111)	−.0277 (.0205)	.0142 (.0115)	−.0229 (.0249)
EXPER^2	−.0017** (.0005)	.0005 (.0005)	−.0018** (.0005)	.0005 (.0007)	−.0015** (.0005)	−.0000 (.0008)
WHITE	.1047** (.0495)	.0865* (.0479)	.1117** (.0565)	.0976* (.0567)	.1455** (.0585)	.1356* (.0695)
WIFE	.4711** (.0390)	.9468** (.0922)	.4782** (.0639)	.9599** (.3610)	.5080** (.0678)	.9110** (.3911)
KIDS	.1164** (.0225)	−.1777** (.0321)	.1174** (.0276)	−.1798* (.1053)	.0984** (.0259)	−.1092 (.1184)
CONSTANT	−.0585 (.1434)	−.1268 (.1156)	−.0741 (.1730)	−.1346 (.3379)	.4552 (.1814)	.3131 (.3499)
Covariance matrix						
ρ	.0000		.0315 (.4093)		.6419* (.3682)	
σ_2	1.0000		1.0506** (.5087)		1.1596** (.5843)	
Log-likelihood	−10,300.710		−10,299.700		−10,299.645	

NOTE: M denotes manufacturing utility index coefficients. NM denotes nonmanufacturing. Standard errors are in parentheses. Double asterisks indicate significance at the 5% level. An asterisk indicates significance at the 10% level. The estimates in column (1) were obtained by fixing ρ and σ_2 at 0 and 1, respectively. The estimates in column (2) were obtained using the column (1) estimates as starting values. The estimates in column (3) were obtained using as starting values the estimates from a model that included only CONSTANT, U-RATE, TREND, and NLINC as regressors and that held ρ and σ_2 fixed at 0 and 1, respectively.

constants, U-RATE, TREND, and NLINC and with ρ and σ_2 pegged at 0 and 1, respectively. (The Marquadt procedure was again used to obtain a step matrix.) This produced a value of ρ very far from 0 (the estimate is .6419) and several coefficient estimates far from the column (1) and column (2) values. The log-likelihood function value is $-10{,}299.645$, however, which is not significantly different from the column (1) and column (2) values, and the standard errors for a number of parameters are again quite large. Thus the parameters ρ and σ_2 do not seem to be well identified despite the inclusion of sector-specific regressors. The reason that identification problems remain is apparently that NLINC does not have sufficient correlation with sectoral choices in these data. In Heckman and Sedlacek's (1985) data, the sector-specific nonlabor income variable was very highly significant in both the manufacturing and nonmanufacturing utility indexes.

3. CONCLUSION

This article has illustrated, via Monte Carlo tests and an application to actual data, that identification of multinomial probit models in the absence of exclusion restrictions is extremely fragile, despite the fact that exclusion restrictions are *not* necessary for these models to be formally identified. A straightforward geometric intuition suggests that this fragility arises because it is difficult to disentangle covariance parameters from regressor coefficients in such models.

The reason that the fragility of identification in MNP models without exclusion restrictions has not been previously noted is that these models have rarely been applied, and almost all existing applications are to transportation-mode choice. There, exclusion restrictions arise naturally because there are mode-specific attributes that only affect the utility derived from choosing the specific mode. In other applications, such as in labor economics for example, exclusion restrictions do not arise naturally. Microeconomic data sets usually contain only attributes of survey respondents themselves. If one attempted to model, say, choice of industry or occupation, attributes of the alternatives would not usually be available in the data. This renders the application of MNP to such choice problems problematic. Because of the recent advent of practical simulation estimators for the MNP model, many more empirical researchers are likely to attempt applications of the model to many previously untried problems—in many of which exclusion restrictions do not naturally arise. Thus it is important that the practical limitations of MNP be realized.

ACKNOWLEDGMENTS

I thank the Alfred P. Sloan Foundation for its support and the two anonymous referees for their comments. Any views expressed here are mine and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[*Received August 1990. Revised August 1991.*]

REFERENCES

Albright, R., Lerman, S., and Manski, C. (1977). "Report on the Development of an Estimation Program for the Multinomial Probit Model," report prepared by Cambridge Systematics for the Federal Highway Administration.

Ben-Akiva, M., and Bolduc, D. (1989), "Multinomial Probit With Taste Variation and Generalized Autoregressive Alternatives," paper presented at the 69th Annual Meeting, Transportation Research Board, Washington.

Berndt, E., Hall, R., and Hausman, J. (1974), "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3, 653–666.

Bolduc, D. (1991), "Generalized Autoregressive Alternatives in the Multinomial Probit Model," unpublished manuscript, Université Laval, Département d'Économique. Submitted to *Transportation Research B*.

Bunch, D. S. (1991), "Estimability in the Multinomial Probit Model," *Transportation Research B*, 25, 1–12.

Bunch, D. S., and Kitamura, R. (1991), "Multinomial Probit Estimation Revisited: Testing Estimable Model Specifications, Maximum Likelihood Algorithms, and Probit Integral Approximations for Trinomial Models of Household Car Ownership," working paper, University of California, Davis, Graduate School of Management.

Dansie, B. (1985), "Parameter Estimability in the Multinomial Probit Model," *Transportation Research B*, 19, 526–528.

Elrod, T., and Keane, M. (1991), "Modelling Heterogeneity in Consumer Choice Behavior," unpublished manuscript, University of Alberta, Faculty of Business.

Hausman, J., and Wise, D. (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

Heckman, J., and Sedlacek, G. (1985), "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077–1125.

Horowitz, J. L., Sparmann, J. M., and Daganzo, C. F. (1982), "An Investigation of the Clark Approximation of the Multinomial Probit Model," *Transportation Science*, 16, 382–401.

Keane, M., Moffitt, R., and Runkle, D. (1988), "Real Wages Over the Business Cycle: Estimating the Impact of Heterogeneity With Micro Data," *Journal of Political Economy*, 96, 1232–1266.

Marquadt, D. (1963), "An Algorithm for the Estimation of Non-Linear Parameters," *Society for Industrial and Applied Mathematics Journal*, 11, 431–441.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

Pakes, A., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1058.