E C O N O M I C S B U L L E T I N

A Note on the Relationship of the Ordered and Sequential Probit Models to the Multinomial Probit Model

Daisuke Nagakura
University of Washington

Abstract

In this note, I show that the ordered and sequential probit models are special cases of the multinomial probit model where the disturbance terms in the latent variables degenerate or those variances converge to zero at a certain rate.

Citation: Nagakura, Daisuke, (2004) "A Note on the Relationship of the Ordered and Sequential Probit Models to the Multinomial Probit Model." *Economics Bulletin,* Vol. 3, No. 40 pp. 1−7
Submitted: September 16, 2004. **Accepted:** October 27, 2004.
URL: http://www.economicsbulletin.com/2004/volume3/EB−04C40007A.pdf

1. Introduction

The ordered, sequential, and multinomial probit models are widely used in analyzing discrete choice problems (See Davidson and MacKinnon, 2004, and Amemiya, 1985, for details of these models). These three models have been considered as different types of models, i.e., it is not believed that one of them nests other two as special cases. In this note, I show that the ordered and sequential probit models are obtained as special cases of the multinomial probit model where the disturbance terms in the latent variables are degenerated or those variances converge to zero at a certain rate.

2. Definitions of the three models

First, let us state the definitions of these three models. Let the number of alternatives be J. Define the unobserved latent variable as $y_t^* = \mathbf{x}_t'\gamma + \eta_t$, $\eta_t \sim NID(0, \sigma_\eta)$, where $\mathbf{x}_t$ is a $k \times 1$ observable explanatory vector without a constant term, and γ is a $k \times 1$ coefficient vector. Let y_t be a discrete random variable whose value ranges from 1 to J. Then, the ordered probit model with J alternatives is defined as follows:

Definition 1 (ordered probit model)

$$y_t = \begin{cases} 1 & \text{if } \ y_t^* < d_1 \\ 2 & \text{if } \ d_1 \le y_t^* < d_2 \\ \vdots & \\ J-1 & \text{if } \ d_{J-2} \le y_t^* < d_{J-1} \\ J & \text{if } \ d_{J-1} \le y_t^* \end{cases}$$

where $d_1 < d_2 < \cdots < d_{J-1}$.

The parameter d_k is called "threshold parameter". It is customary to set $\sigma_\eta = 1$ for identification of parameters. The estimation can be easily done by the maximum likelihood estimation.

Second, I shall define the sequential probit model. The $J-1$ latent variables for the model are given by $y_{tj}^+ = \phi_j + \mathbf{x}_t'\boldsymbol{\delta}_j + \zeta_{tj}$, $\zeta_{tj} \sim NID(0,1)$ for $j = 1, \ldots, J-1$, where ϕ_j is a scalar constant term, and $\boldsymbol{\delta}_j$ is a $k \times 1$ coefficient vector. Then, the sequential probit model is defined as follows:

Definition 2 (sequential probit model)

$$y_t = \begin{cases} 1 & \text{if } y_{t1}^+ \ge 0 \\ 2 & \text{if } y_{t1}^+ < 0, \text{ and } y_{t2}^+ \ge 0 \\ \vdots & \\ J-1 & \text{if } y_{t1}^+ < 0, y_{t2}^+ < 0, \ldots, y_{tJ-2}^+ < 0, \text{ and } y_{tJ-1}^+ \ge 0 \\ J & \text{if } y_{t1}^+ < 0, y_{t2}^+ < 0, \ldots, y_{tJ-2}^+ < 0, \text{ and } y_{tJ-1}^+ < 0 \end{cases}$$

In the sequential probit model, each decision is made sequentially according to a binary probit model. Whether or not an alternative j is selected is determined before an alternative k ($>j$) is considered. See Amemiya (1985, p310) for details of this model.

Third, I shall give the definition of the multinomial probit model. The J latent variables for the model is defined as $y_{tj}^{\circ}=\alpha_j+\mathbf{x}_t'\boldsymbol{\beta}_j+\varepsilon_{tj}$, $\boldsymbol{\varepsilon}_t \sim N(0,\boldsymbol{\Omega})$ for $j=1,\ldots, J$, where α_j is a scalar constant term, $\boldsymbol{\beta}_j$ is a $k \times 1$ coefficient vector, and $\boldsymbol{\varepsilon}_t$ is a $J \times 1$ normal random vector with typical element ε_{tj} and the covariance matrix

$$\boldsymbol{\Omega}=\begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1J}\sigma_1\sigma_J \\ & \sigma_2^2 & \cdots & \rho_{2J}\sigma_2\sigma_J \\ & & \ddots & \vdots \\ & & & \sigma_J^2 \end{bmatrix},$$

where σ_j is the standard deviation of ε_j and ρ_{kj} $(=\rho_{jk})$ is the correlation coefficient between ε_j and ε_k. The definition of the multinomial probit model is given as follows:

Definition 3 (multinomial probit model)

$$y_t = j \quad \text{when} \quad y_{tj}^{\circ} > y_{ti}^{\circ} \quad \forall i, \quad i \neq j.$$

We usually set $\alpha_r=0$, $\boldsymbol{\beta}_r=\mathbf{0}$, $\sigma_r=0$, $\sigma_s=1$ for some r and s, for parameter identification (Note that $\boldsymbol{\varepsilon}_t$ becomes $J-1 \times 1$ normal random vector by this normalization).

2. Main results

Hereafter, the subscript t is suppressed for the sake of notational simplicity.

Proposition 1: *The multinomial probit model reduces to the ordered probit mode with* $d_j=\dfrac{\sigma_k(\alpha_{j+1}-\alpha_j)}{\rho_{kj}\sigma_j-\rho_{k\,j+1}\sigma_{j+1}}$ *for* $j=1,\ldots, J-1$, *where* $\rho_{kk}=1$, *if the following conditions are satisfied for some k:*

(A) $\rho_{kj}=1$ or -1 $\forall j$[1], (B) $\rho_{kj}\sigma_k^{-1}\sigma_j\boldsymbol{\gamma}=\boldsymbol{\beta}_j$ $\forall j$,

(C) $\rho_{k1}\sigma_1<\rho_{k2}\sigma_2<\cdots<\sigma_k<\cdots<\rho_{kJ}\sigma_J$,

and (D) $\dfrac{\alpha_r-\alpha_j}{\rho_{kj}\sigma_j-\rho_{kr}\sigma_r}<\dfrac{\alpha_s-\alpha_j}{\rho_{kj}\sigma_j-\rho_{ks}\sigma_s}$, for $r<s$, $r, s, j=1,\ldots, J$, $r, s\neq j$ [2]

Proof. It can be easily shown that the conditional distribution of ε_j conditioned on ε_k is $N(\rho_{kj}\sigma_k^{-1}\sigma_j\varepsilon_k, \sigma_j^2(1-\rho_{kj}^2))$. Under the condition (A), the conditional variance becomes zero and ε_j is perfectly correlated with ε_k, i.e.,

---

[1] When this is the case, other correlation coefficients also become positive or negative one.

[2] It can be shown that this condition is equivalent to $\dfrac{\alpha_r-\alpha_j}{\rho_{kj}\sigma_j-\rho_{kr}\sigma_r}<\dfrac{\alpha_s-\alpha_j}{\rho_{kj}\sigma_j-\rho_{ks}\sigma_s}$ for $r<j<s$.

$$\varepsilon_j = \rho_{kj}\sigma_k^{-1}\sigma_j\varepsilon_k \quad (1)$$

Additionally, if the condition (B) is satisfied, the condition for $y = j$ in the multinomial probit model becomes as follows:

$$y_j^\circ > y_i^\circ \quad \forall i,\ i \neq j \quad \Leftrightarrow \alpha_j + \mathbf{x}'\boldsymbol{\beta}_j + \varepsilon_j > \alpha_i + \mathbf{x}'\boldsymbol{\beta}_i + \varepsilon_i \quad \forall i,\ i \neq j$$

$$\Leftrightarrow \quad \mathbf{x}'(\boldsymbol{\beta}_j - \boldsymbol{\beta}_i) + \varepsilon_j - \varepsilon_i > \alpha_i - \alpha_j \quad \forall i,\ i \neq j$$

$$\Leftrightarrow \quad \sigma_k^{-1}(\rho_{kj}\sigma_j - \rho_{ki}\sigma_i)(\mathbf{x}'\boldsymbol{\gamma} + \varepsilon_k) > \alpha_i - \alpha_j \quad \forall i,\ i \neq j \quad \text{(From (1) and (B))},$$

where the notation $\Leftrightarrow$ means that the conditions on the both sides are equivalent. Furthermore, from the condition (C), the above is equivalent to

$$\mathbf{x}'\boldsymbol{\gamma} + \varepsilon_k > c_{ji}^{(k)} \ \text{ for } i < j \quad \text{and} \quad \mathbf{x}'\boldsymbol{\gamma} + \varepsilon_k < c_{ji}^{(k)} \ \text{ for } i > j, \quad (2)$$

where $c_{ji}^{(k)} \equiv \dfrac{\sigma_k(\alpha_i - \alpha_j)}{\rho_{kj}\sigma_j - \rho_{ki}\sigma_i}$. Note that $c_{ji}^{(k)} = c_{ij}^{(k)}$ clearly by its definition. From the condition (D), when j is fixed, the maximum of $c_{ji}^{(k)}$ for $i < j$ is $c_{j\,j-1}^{(k)}$, while the minimum of $c_{ji}^{(k)}$ for $i > j$ is $c_{j\,j+1}^{(k)}$. Thus, (2) is rewritten as[3]:

$$c_{j\,j-1}^{(k)} < \mathbf{x}'\boldsymbol{\gamma} + \varepsilon_k < c_{j\,j+1}^{(k)} \ (= c_{j+1\,j}^{(k)}). \quad (3)$$

Note that the inequalities in (3) reduce to only one side for $j = 1$ and J.

Setting $\varepsilon_k = \eta$, it have been shown that when the conditions (A)~(D) are satisfied, the multinomial probit model reduces to the ordered probit model with $d_j = \dfrac{\sigma_k(\alpha_{j+1} - \alpha_j)}{\rho_{kj}\sigma_j - \rho_{k\,j+1}\sigma_{j+1}}$ for $j = 1, \ldots, J-1$.[4] □

The conditions for $j = 1$, 2, and 3 are illustrated in figure 1 (a), (b), and (c), respectively. The shaded area is the area which satisfies the all required conditions, i.e., $y = j$ when $\mathbf{x}'\boldsymbol{\gamma} + \eta$ falls into this area. Note that ρ_{kj} can be negative one for $j < k$, but it must be positive one for $k < j$. For example if we set $k = 1$, $\rho_{1j} = 1$ for all j.

Example: Set $J = 4$, $\rho_{2j} = 1$ for all j, $\sigma_1 = 0$ $\sigma_2 = 1$, $\sigma_3 = 2$, $\sigma_4 = 3$, $\alpha_1 = 0$, $\alpha_2 = 2$, $\alpha_3 = 2.6$, $\alpha_4 = 3$, $\boldsymbol{\beta}_1 = 0$, $\boldsymbol{\beta}_2 = 1$, $\boldsymbol{\beta}_3 = 2$, and $\boldsymbol{\beta}_4 = 3$ in the multinomial probit model[5]. Notice that the conditions (A), (C) and (D) in Proposition 1 are satisfied. Then this reduces to the ordered probit model with $\boldsymbol{\gamma} = 1$, $\sigma_\eta = 1$, $d_1 = -2$, $d_2 = -0.6$, and $d_3 = -0.4$.

---

[3] $c_{j\,j-1}^{(k)} < c_{j\,j+1}^{(k)}$ is obvious from the condition (D).

[4] The equality in the definition 1 (and 2) is not essentially important because the latent variable is a continuous random variable, hence the equality holds with probability zero.

[5] This multinomial probit model is a normalized one by the previously mentioned way.

The next proposition states the relationship between the sequential and the multinomial probit models.

Proposition 2: *The sequential probit model can be obtained as a limiting case of the multinomial probit model if the following conditions are satisfied:*

(E) $\sigma_1 = 1$, $\alpha_1 = \phi_1$, $\boldsymbol{\delta}_1 = \boldsymbol{\beta}_1$, $\alpha_k \to 0$, $\boldsymbol{\beta}_k \to 0$, $\sigma_k \to 0$ *so that*

$$\lim_{\substack{\sigma_k \to 0 \\ \alpha_k \to 0}} \sigma_k^{-1}\alpha_k = \phi_k, \quad \lim_{\substack{\sigma_k \to 0 \\ \boldsymbol{\beta}_k \to \mathbf{0}}} \sigma_k^{-1}\boldsymbol{\beta}_k = \boldsymbol{\delta}_k, \quad \lim_{\substack{\sigma_k \to 0 \\ \alpha_l \to 0}} \sigma_k^{-1}\alpha_l = 0,$$

$$\lim_{\substack{\sigma_k \to 0 \\ \boldsymbol{\beta}_l \to \mathbf{0}}} \sigma_k^{-1}\boldsymbol{\beta}_l = \mathbf{0}, \quad \lim_{\substack{\sigma_k \to 0 \\ \sigma_l \to 0}} \sigma_k^{-1}\sigma_l = 0 \quad \textit{for} \quad k = 2,\ldots, J, \quad \textit{and} \quad l > k,$$

and (F) $\rho_{ij} = 0 \quad \forall i, j,\ i \neq j$.

Proof. Multiplying the both sides by σ_j or σ_i, the condition for $y = j$ in the multinomial probit model can be rewritten as

$$y_j^\circ > y_i^\circ \quad \forall i, \quad i \neq j$$

$$\Leftrightarrow \quad \sigma_i^{-1}\alpha_j + \mathbf{x}\sigma_i^{-1}\boldsymbol{\beta}_j + \sigma_i^{-1}\varepsilon_j > \sigma_i^{-1}\alpha_i + \mathbf{x}\sigma_i^{-1}\boldsymbol{\beta}_i + \sigma_i^{-1}\varepsilon_i, \text{ for } i < j \text{ and}$$

$$\sigma_j^{-1}\alpha_j + \mathbf{x}\sigma_j^{-1}\boldsymbol{\beta}_j + \sigma_j^{-1}\varepsilon_j > \sigma_j^{-1}\alpha_i + \mathbf{x}'\sigma_j^{-1}\boldsymbol{\beta}_i + \sigma_j^{-1}\varepsilon_i \quad \text{for } i > j.$$

Under the condition (E), the above reduces to

$$0 > \phi_i + \mathbf{x}'\boldsymbol{\delta}_i + \sigma_i^{-1}\varepsilon_i \quad \text{for } i < j, \quad \text{and} \quad \phi_j + \mathbf{x}'\boldsymbol{\delta}_j + \sigma_j^{-1}\varepsilon_j > 0 \quad \text{for } i > j$$

because $\sigma_k^{-1}\alpha_l$, $\sigma_k^{-1}\boldsymbol{\beta}_l$, and the variance of $\sigma_k^{-1}\varepsilon_l$ ($l > k$) converge to zero. Set $\zeta_k = \sigma_k^{-1}\varepsilon_k$, then the condition (F) ensures that $\zeta_k \sim NID(0,1)$. This completes the proof.
□

Intuitively, this is because the terms which converge to zero can be neglected compared with other terms which also converge to zero but at less rapid rate so that the condition (E) holds. To confirm the analytical proof of Proposition 2, I conduct a small Monte Carlo experiment. The number of choices is set at three. Samples for a multinomial probit model, $y_t^{(m)}$ t =1,…,30, are generated following the definition 3 with $y_{t1}^\circ = \varepsilon_{t1}$, $y_{t2}^\circ = 0.5/\sqrt{s} + \varepsilon_{t2}/\sqrt{s}$, and $y_{t3}^\circ = 1/s + \varepsilon_{t3}/s$, where ε_{tj} j =1,2,3 are random draws from *NID*(0,1) and *s* is some constant. Samples for a sequential probit model, $y_t^{(s)}$ t =1,…,30, are generated following the definition 2 with $y_{t1}^+ = y_{t1}^\circ$, and $y_{t2}^+ = \sqrt{s}\,y_{t2}^\circ$. According to Proposition 2, $y_t^{(m)}$ should converge to $y_t^{(s)}$ as *s* becomes large. Figure 2 (a), (b), (c) shows the actual values of $y_t^{(m)}$ for *s* =1, 10, and 1000, respectively, and those of $y_t^{(s)}$ are shown in Figure 2(d). The same ε_{tj} for *t* =1,…,30, *j* =1,2,3 are used for each value of *s* to make the result clearer. Observe that $y_t^{(m)}$ actually converges to $y_t^{(s)}$ as *s* becomes large, as insisted in Proposition 2.

3. Conclusion

I have shown that the ordered and sequential probit models are special cases of the multinomial probit model where the disturbance terms in the latent variables degenerate or those variances converge to zero at a certain rate. Which probit model should be used is an important problem in an empirical analysis. The results obtained in this note are useful for comparing these three models; specification tests for these models will be developed based on the results. For example, Kobayashi (2001) have proposed a test for ordered probit models against multinomial probit models in the case of three alternatives. The results of this note can be used for extending the test to the case of more than three alternatives. Also, a test for sequential probit models in this context has not been proposed as far as I know. These are the subjects of further research. Finally, I note that the conditions I have shown in this note are sufficient conditions, so the same results may be obtained under weaker conditions.

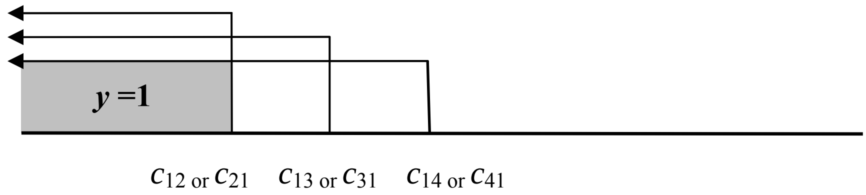
Reference

Amemiya, T. (1985) *Advanced Econometrics*, Cambridege, Mass., Harvard University Press.

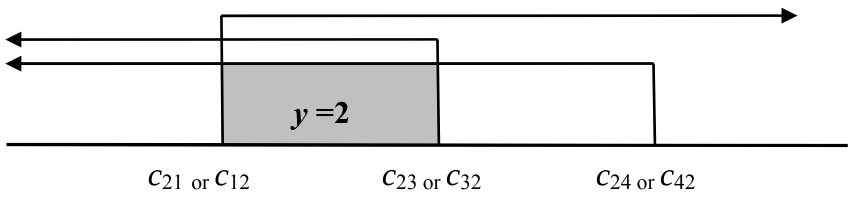
Kobayashi M. (2001) "A Test for Ordered Probit Models against Multinomial Probit Models," mimeo.

Davidson, R., and J. G.. MacKinnon (2004) *Econometric Theory and Methods*, New York, Oxford University press.

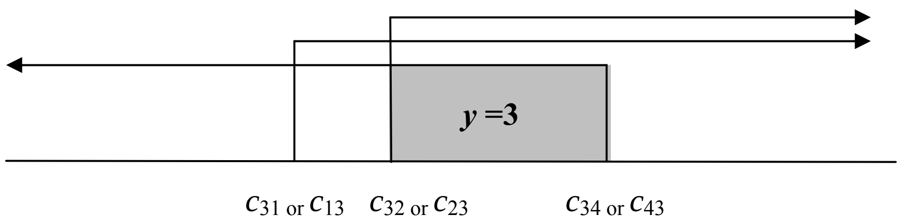
Figure 1. Conditions of the degenerated multinomial probit model



(a) conditions for y =1

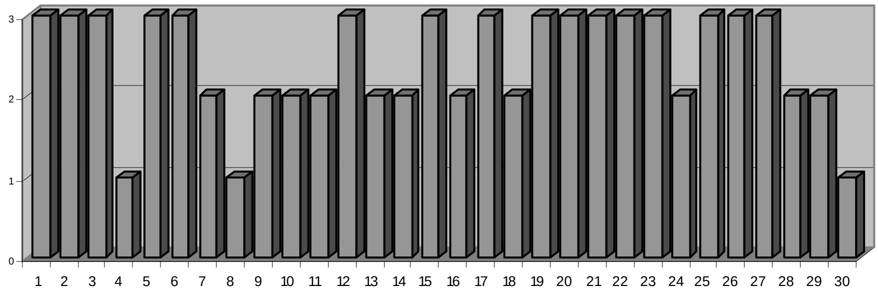


(b) conditions for y = 2

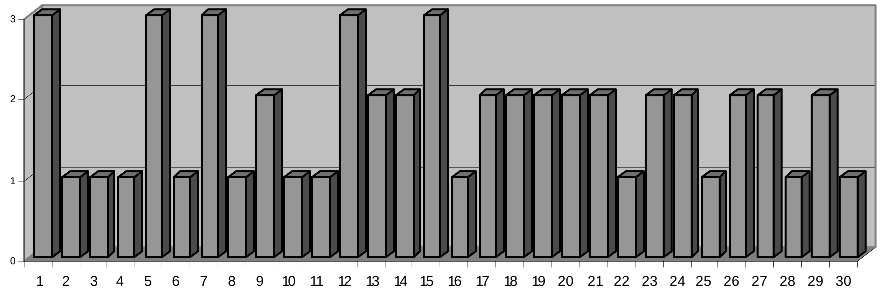


(c) conditions for y = 3

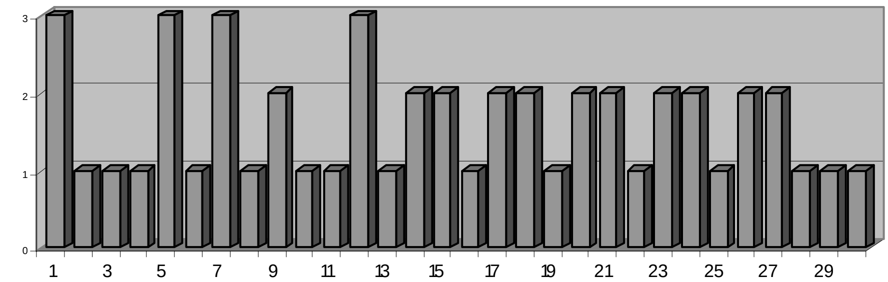
Figure 2. Actual values of the samples from the multinomial and sequential probit models



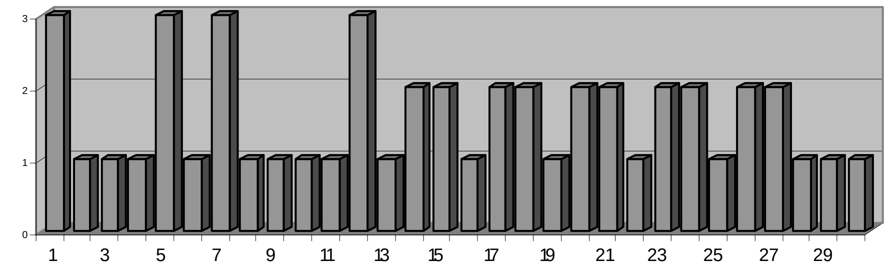
(a) samples from the *multinomial* probit model when s =1.



(b) samples from the *multinomial* probit model when s =10.



(c) samples from the *multinomial* probit model when s =1000.



(d) samples from the *sequential* probit model when s =1, 10 ,1000

Note: x axis indicates the no. of a sample, and y axis is for an actual value of a sample**.**