AN APPLICATION OF NESTED LOGIT MODELS TO
THE LABOR SUPPLY OF THE ELDERLY

TAKESHI AMEMIYA
KEIKO SHIMONO

TECHNICAL REPORT NO. 22
DECEMBER 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT NO. SES 85-20313

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD CALIFORNIA
AN APPLICATION OF NESTED LOGIT MODELS TO
THE LABOR SUPPLY OF THE ELDERLY

Takeshi Amemiya and Keiko Shimono

TECHNICAL REPORT NO. 22
December 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT NO. SES 85-20313

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN APPLICATION OF NESTED LOGIT MODELS TO
THE LABOR SUPPLY OF THE ELDERLY*

by
Takeshi Amemiya
Keiko Shimono

I. Introduction

In this paper we apply nested logit models to a study of the labor supply of the elderly. We assume that an individual chooses among the following four alternatives: (1) to retire (to be denoted by the symbol "R"), (2) to be self-employed ("S"), (3) to be employed full-time ("F"), and (4) to be employed part-time ("P"). We assume that a utility is associated with each alternative and the individual chooses that alternative for which the utility is greatest. The utility is completely known to the individual but is a random variable for the econometrician; hence, the assumption of utility maximization leads to a probabilistic model of choice (or, a qualitative response model). A specific assumption on the joint distribution of the utilities determines a specific qualitative response model.

Tachibanaki and Shimono [1985] used a multinomial logit model to analyze the same problem. A multinomial logit model is essentially equivalent to the assumption that the utilities for different alternatives are independent, each distributed according to the Type I extreme-value distribution (see McFadden

* A preliminary draft of this paper was presented at a regular seminar of Economic Research Center, Nagoya University. The authors have benefited from the comments of the participants of the seminar. Amemiya's research was partially supported by NSF grant SES-8520313 and Stanford University CEAS Faculty Fellowship.
We use nested logit models, which arise as we relax the assumption of independence in a specific manner, because we suspect that some of these alternatives (say, "F" and "P") are more similar to each other than to the others and, therefore, their utilities are likely to be correlated. At the least, there is no strong theoretical reason to assume independence. Moreover, since a multinomial logit model is a special case of a nested logit model, we can test it as a statistical hypothesis.

As we noted above, we think that some alternatives (such as "F" and "P") are more similar than others. However, economic theory does not clearly indicate an optimal way of nesting the four alternatives. Therefore, we have decided to try every possible way of nesting the four alternatives and choose the best model on the basis of both theoretical consideration and a statistical fit. It has generally turned out that the more theoretically plausible the model is, the better fit it exhibited.

Some people might object to our basic assumption of utility maximization on the grounds that the choice of the work alternatives for the elderly may often be determined more by the job availability and the decision of the potential employer (that is, by the demand side) than by the preference of the individual worker (that is, by the supply side). We should point out that the utility used in qualitative response models is the indirect utility which reflects the demand side as well as the supply side consideration. We can even do away with the term utility altogether and merely state that the choice is made so as to maximize a certain stochastic index associated with each alternative.

The order of the presentation of this paper is as follows: In Section 2
we briefly explain our data and variables. In Section 3 we define the nested logit models we estimate and discuss the estimation procedure. The results are presented in Section 4, and a brief conclusion given in Section 5.

2. Data and variables

We use the same data (except for the definition of the part-time employee as noted below) and the same independent variables as in Tachibanaki and Shimono [1985], so we shall briefly discuss them, referring the reader who want to examine them more closely to that article. We use the data collected by the "Survey on Labor Force Status of the Elderly" conducted by the Ministry of Labor in 1980. The ages covered are between 55 and 69. We use only the male samples who were full-time employees at age 55 and are living principally with their income. The employment status of the 4,101 individuals in our sample is divided into the four categories mentioned earlier. Their brief descriptions with their respective sample sizes are given below:

(1) Retired: Those who are not working and do not indicate a desire for working. 423 people.

(2) Self-employed: 389 people.

(3) Full-time employees: Those who are working as regular employees every day. 3048 people.

(4) Part-time employees: Those who are either working everyday but for a limited duration or employed regularly but not working every day. 241 people. (The Tachibanaki-Shimono study included, in addition, those who did not work everyday and were not employed regularly.)
We have excluded from our sample the "unemployed" individuals, those who are not working and have expressed their desire for working.

The independent variables are as follows: age, amount of savings, amount of private pension, amount of public pension, other family members' income, dummy for health status, and dummy for whether an individual receives income other than wage and pension.

3. Model Definitions and Estimation Methods

Let $U_{ij}$ be the utility of the $i$th individual choosing the $j$th alternative. We assume that it can be written as the sum of a nonstochastic term $\mu_{ij}$ and a stochastic term $\varepsilon_{ij}$:

$$U_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \ldots, n,$$
$$j = 1, 2, 3, 4.$$

We further assume that $\mu_{ij}$ is a linear function of independent variables:

$$\mu_{ij} = x_i' \beta_j.$$

Note that we let the independent variables depend on the individual's characteristics but not on the alternative's characteristics, whereas we let the coefficients depend on the alternative $j$. In some qualitative response models, notably those involving the choice of the transportation mode (see Domencich and McFadden [1975]), it is customary to let $x$ depend both on $i$ and $j$ and make $\beta$ invariant of $i$ or $j$. In our study the limitation of the data more or less forces us to adopt the specification (3.2).

The random term $\varepsilon_{ij}$ may be regarded as a function of all the independent
variables which are observable to the individual but not to the econometrician. Thus, it is a random variable in the eyes of the econometrician. We assume that $\varepsilon_{ij}$ are independent and identically distributed across $i$ but may be correlated across $j$. Different correlation structures imposed on $\varepsilon_{ij}$ lead to different qualitative response models.

Let $y_i$ denote the discrete dependent variable signifying the response of the $i$th individual. Then, the assumption of utility maximization implies that

$$P(y_i = j) = P(U_j > U_k \text{ for all } k \neq j).$$

Thus, the model is completed by specifying the joint distribution of $\varepsilon_{i1}$, $\varepsilon_{i2}$, $\varepsilon_{i3}$, and $\varepsilon_{i4}$.

We consider two basic type of nested logit models, denoted Type I and Type II. They are schematically described in Figure 1. In Type I, we let $\varepsilon_1$ and $\varepsilon_2$ be correlated with each other, let $\varepsilon_3$ be correlated with both $\varepsilon_1$ and $\varepsilon_2$, and let $\varepsilon_4$ be independent of the other three. In Type II, we let $\varepsilon_1$ and $\varepsilon_2$ be correlated with each other, let $\varepsilon_3$ and $\varepsilon_4$ be correlated with each other, and assume the two pairs to be independent. In other words, in Type I, 1 and 2 are the most similar alternatives, 3 is somewhat akin to 1 and 2, and 4 is entirely different. In Type II, there is a similarity between 1 and 2 or between 3 and 4 but not between the two pairs.

Next, we give the joint distribution of the $\varepsilon$'s and the ensuing probabilities for each type of the nested logit models. To simplify the notation, we have omitted the subscript $i$ from the variables $\varepsilon$ and $\mu$ in the subsequent derivation. We use the symbol $\Lambda(\cdot)$ to denote the logistic distribution function $\Lambda(x) = e^x/(1+e^x)$.
Figure 1: A Schematic Representation of Four Types of Nested Logit Models

Type I

Type II

Type III

Type IV
Type I

\[ F(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = \exp\{-[(e^{-\varepsilon_1/\rho} + e^{-\varepsilon_2/\rho}) \rho/\sigma + e^{-\varepsilon_3/\sigma} - e^{-\varepsilon_4/\sigma}] - e^{-\varepsilon_4/\sigma}, \ 0 < \sigma, \ \rho \leq 1. \]

\[ P(y=1 \mid y=1 \text{ or } 2) = \Lambda(x'\alpha_1), \text{ where } \alpha_1 = (\beta_1 - \beta_2)/\rho. \]

\[ P(y=1 \mid y=1 \text{ or } 2 \text{ or } 3) = \Lambda[x'(\beta_1 - \beta_3)/\sigma + (\rho/\sigma) \log(1+e^{-x'\alpha_1})]. \]

\[ P(y=1 \text{ or } 2 \mid y=1 \text{ or } 2 \text{ or } 3) = \Lambda[x'(\beta_1 - \beta_4) \rho/\sigma + e^{-x'\alpha_1}]. \]

Type II

\[ F(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = \exp\{-[(e^{-\varepsilon_1/\rho_1} + e^{-\varepsilon_2/\rho_1}) \rho_1 - (e^{-\varepsilon_3/\rho_2} + e^{-\varepsilon_4/\rho_2}) \rho_2], \ 0 < \rho_1, \rho_2 \leq 1. \]

\[ P(y=1 \mid y=1 \text{ or } 2) = \Lambda(x'\alpha_1), \text{ where } \alpha_1 = (\beta_1 - \beta_2)/\rho_1. \]

\[ P(y=1 \mid y=3 \text{ or } 4) = \Lambda(x'\alpha_2), \text{ where } \alpha_2 = (\beta_3 - \beta_4)/\rho_2. \]

\[ P(y=1 \text{ or } 2) = \Lambda[x'(\beta_2 - \beta_4) + \rho_1 \log(1+e^{-x'\alpha_1}) - \rho_2 \log(1+e^{-x'\alpha_2})]. \]

In estimating the parameters of these models, we use a three-step method, a simple generalization of the two-step method originally proposed by Domencich and McFadden [1975] for a nested logit model of three responses.

In each step one needs only to use a standard binary logit program. For Type I the method works as follows: (1). Estimate \( \alpha_1 \) using the first probability. (2). Insert that value into the second probability and estimate \( (\beta_1 - \beta_3)/\sigma \) and \( \rho/\sigma \). (3). Insert the estimates obtained in step (2) into the third probability and estimate \( \beta_1 - \beta_4 \) and \( \sigma \). Thus, eventually, we can identify the difference of every pair of \( \beta \)'s as well as \( \rho \) and \( \sigma \). For Type II the method works as follows: (1). Estimate \( \alpha_1 \) using the first probability. (2). Estimate \( \alpha_2 \) using the second probability. (3). Insert these estimates into the third
probability and estimate $\beta_2 - \beta_4$, $\rho_1$, and $\rho_2$. Thus, eventually, we can also identify $\beta_1 - \beta_2$ and $\beta_3 - \beta_4$.

The process of inserting the estimates obtained from one step to the next invalidates the formula for the asymptotic variance-covariance matrix used in a standard binary logit program. The formula must be theoretically derived and evaluated. However, the derivation of such a formula is straightforward (see Manski and McFadden [1981]).

Two additional types of models arise as special cases of Type I or II. Type III arises as a special case of Type I if we put $\sigma=1$ or as a special case of Type II if we put $\rho_2=1$. Type IV arises as a special case of Type I if we put $\sigma=\rho$. Their schematic representations are given in Figure 1.

By putting the four alternatives $R$, $S$, $F$, and $P$ into the four positions in every conceivable way, we can create 12 distinct models for Type I and 3 distinct models for Type II. They then yield 6 distinct models for Type III and 4 distinct models for Type IV. All the 25 models are schematically described in Figure 2. The numbering of models under each type is given in such a way that the same number means the same sequence of the four alternatives. We estimate all of these models, even though we do have some idea as to which models are theoretically more plausible than others. The results of estimation are reported and discussed in the next section.

In addition to the nested logit models, we also estimate a sequential logit model (see Amemiya [1985]) for the sake of a comparison. We only consider the following sequential decision: (1) The choice between "R" and "Not R" is done according to a binary logit model. (2) Given "Not R", one chooses between "S" and "Not S", also by a logit model. (3) Given "Not R and Not S",
Figure 2: A Schematic Representation of the 25 Nested Logit Models

Type I
(1) \[ \begin{array}{cccc} F & P & S & R \end{array} \]
(2) \[ \begin{array}{cccc} F & S & P & R \end{array} \]
(3) \[ \begin{array}{cccc} P & S & F & R \end{array} \]
(4) \[ \begin{array}{cccc} F & P & R & S \end{array} \]
(5) \[ \begin{array}{cccc} F & R & P & S \end{array} \]
(6) \[ \begin{array}{cccc} P & R & F & S \end{array} \]
(7) \[ \begin{array}{cccc} P & S & R & F \end{array} \]
(8) \[ \begin{array}{cccc} P & R & S & F \end{array} \]
(9) \[ \begin{array}{cccc} S & R & P & F \end{array} \]
(10) \[ \begin{array}{cccc} F & S & R & P \end{array} \]
(11) \[ \begin{array}{cccc} F & R & S & P \end{array} \]
(12) \[ \begin{array}{cccc} S & R & F & P \end{array} \]

Type II
(1) \[ \begin{array}{cccc} F & P & S & R \end{array} \]
(2) \[ \begin{array}{cccc} F & S & P & R \end{array} \]
(3) \[ \begin{array}{cccc} P & S & F & R \end{array} \]

Type III
(1) \[ \begin{array}{cccc} F & P & S & R \end{array} \]
(2) \[ \begin{array}{cccc} F & S & P & R \end{array} \]
(3) \[ \begin{array}{cccc} P & S & F & R \end{array} \]
(4) \[ \begin{array}{cccc} F & R & P & S \end{array} \]
(5) \[ \begin{array}{cccc} F & R & P & S \end{array} \]
(6) \[ \begin{array}{cccc} P & R & F & S \end{array} \]
(12) \[ \begin{array}{cccc} S & R & F & P \end{array} \]

Type IV
(1) \[ \begin{array}{cccc} F & P & S & R \end{array} \]
(4) \[ \begin{array}{cccc} F & P & R & S \end{array} \]
(7) \[ \begin{array}{cccc} P & S & R & F \end{array} \]
(10) \[ \begin{array}{cccc} F & S & R & P \end{array} \]
one chooses between "F" and "P" again by a logit model.

4. Results

The results of estimation are summarized in table 1. The maximum likelihood method is used to estimate the parameters of the multinomial logit model and the sequential logit model, and the three-step estimation method described in Section 3 is used to estimate the parameters of the nested logit models. For each model we report \((-1)\) times the value of the log likelihood function evaluated at the respective estimates of the parameters \((-\log L\) and the number of correct predictions \((\text{NCP})\). The NCP is computed as follows: For each individual we predict his job status to be the one having the highest estimated probability and calculate the number of the individuals for whom the predicted status is equal to the actual status. For the nested logit models we also report the estimates of \(\rho\) and \(\sigma\) and their standard deviations.

Let us first look at the column for \(-\log L\). At first we should point out that the multinomial logit model and the sequential logit model have the same and smallest number of the parameters among all the models, Type III and IV nested logit models have one additional parameter, and Type I and II have two additional parameters. Therefore, it is meaningful to compare \(-\log L\) only within each of the three groups having the same number of parameters. The more parameters there are in a model, the higher we expect \(\log L\) to be, and, unfortunately, we do not know how much \(\log L\) should increase on the average as an additional parameter is included in a model. The use of the Akaike information criterion (Akaike [1973]) is not warranted here because we do not
always use maximum likelihood estimation.

First, we look at \(-\log L\) for Type I and II nested logit models. It is gratifying to find that the more theoretically plausible the model is, generally the higher the \(\log L\) is. For example, it is theoretically most reasonable to couple "F" and "P" as the closest alternatives, and the three models of this kind, namely I(1), I(4), and II(1), all had high \(\log L\). Especially, the two theoretically most plausible models I(1) and I(4) had the highest values of \(\log L\). The two most distant alternatives are "F" and "R", and the three models coupling these alternatives, namely I(5), I(11), and II(3), had the three lowest values of \(\log L\).

Second, we consider NCP. The variation of this criterion across models is negligibly small, and whatever small variation there is seems unrelated either to the magnitude of \(\log L\) or to theoretical plausibility. Thus, we conclude that the discriminating power of this criterion is extremely small for our models.

Third, we consider the estimates of \(\rho\) and \(\sigma\). Suppose we choose Model I(1) as the most plausible among Type I and II nested logit models both from theoretical and statistical considerations and test the hypothesis \(\rho=\sigma=1\), which implies the multinomial logit model. Then, since their estimates both exceed unity and their theoretical values are bounded between zero and unity, we must accept the hypothesis at whatever significance level. The fact that the estimates of \(\rho\) and \(\sigma\) exceed unity when their true values are bounded between zero and one is not necessarily a matter of concern. If \(\rho=\sigma=1\), the estimated values such as \(\hat{\rho}=6.41\) and \(\hat{\sigma}=2.08\) in Model I(1) can occur with a reasonably large probability, given their respective standard deviations.
<table>
<thead>
<tr>
<th></th>
<th>(-\log L)</th>
<th>NCP</th>
<th>Estimate of (\rho)</th>
<th>Estimate of (\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multinominal Logit</td>
<td>2659.7</td>
<td>3213 (78.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nested Logit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(1)</td>
<td>2650.1</td>
<td>3209</td>
<td>6.41 (6.49)</td>
<td>2.08 (1.25)</td>
</tr>
<tr>
<td>(2)</td>
<td>2653.1</td>
<td>3209</td>
<td>4.96 (8.44)</td>
<td>4.86 (2.85)</td>
</tr>
<tr>
<td>(3)</td>
<td>2653.9</td>
<td>3209</td>
<td>3.62 (11.85)</td>
<td>4.78 (5.11)</td>
</tr>
<tr>
<td>(4)</td>
<td>2650.5</td>
<td>3202</td>
<td>4.01 (.89)</td>
<td>.68 (.14)</td>
</tr>
<tr>
<td>(5)</td>
<td>2660.7</td>
<td>3209</td>
<td>1.10 (.47)</td>
<td>1.14 (.28)</td>
</tr>
<tr>
<td>(6)</td>
<td>2655.9</td>
<td>3208</td>
<td>.25 (.26)</td>
<td>1.40 (.28)</td>
</tr>
<tr>
<td>(7)</td>
<td>2654.4</td>
<td>3209</td>
<td>1.56 (2.40)</td>
<td>.29 (.25)</td>
</tr>
<tr>
<td>(8)</td>
<td>2655.6</td>
<td>3208</td>
<td>.27 (.16)</td>
<td>.42 (.24)</td>
</tr>
<tr>
<td>(9)</td>
<td>2656.4</td>
<td>3212</td>
<td>.26 (.23)</td>
<td>.31 (.22)</td>
</tr>
<tr>
<td>(10)</td>
<td>2658.8</td>
<td>3210</td>
<td>2.74 (1.02)</td>
<td>.70 (.19)</td>
</tr>
<tr>
<td>(11)</td>
<td>2660.8</td>
<td>3209</td>
<td>.98 (.37)</td>
<td>.92 (.38)</td>
</tr>
<tr>
<td>(12)</td>
<td>2658.0</td>
<td>3212</td>
<td>.50 (.24)</td>
<td>1.21 (.30)</td>
</tr>
<tr>
<td>II(1)</td>
<td>2653.7</td>
<td>3210</td>
<td>5.53* (2.26)</td>
<td>1.22** (.42)</td>
</tr>
<tr>
<td>(2)</td>
<td>2656.2</td>
<td>3210</td>
<td>-.14* (1.23)</td>
<td>.14** (.25)</td>
</tr>
<tr>
<td>(3)</td>
<td>2660.6</td>
<td>3209</td>
<td>1.04* (.16)</td>
<td>2.30** (1.88)</td>
</tr>
<tr>
<td>III(1)</td>
<td>2657.1</td>
<td>3207</td>
<td>3.08 (1.23)</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>2664.6</td>
<td>3206</td>
<td>1.02 (1.16)</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>2675.2</td>
<td>3204</td>
<td>5.43 (3.82)</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>2661.5</td>
<td>3209</td>
<td>.97 (.21)</td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>2659.3</td>
<td>3209</td>
<td>.64 (.28)</td>
<td></td>
</tr>
<tr>
<td>(12)</td>
<td>2658.6</td>
<td>3212</td>
<td>.41 (.16)</td>
<td></td>
</tr>
<tr>
<td>IV(1)</td>
<td>2653.0</td>
<td>3208</td>
<td>4.95 (1.21)</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>2660.0</td>
<td>3209</td>
<td>1.11 (1.22)</td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>2655.5</td>
<td>3209</td>
<td>.31 (.18)</td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>2660.1</td>
<td>3211</td>
<td>.98 (.23)</td>
<td></td>
</tr>
<tr>
<td>Sequential Logit</td>
<td>2666.4</td>
<td>3210</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (1) NCP: the number of correct predictions.

(2) * is estimate of \(\rho_1\) and ** is estimate of \(\rho_2\).

(3) Standard deviations are given in brackets.
Table 2: A Comparison of the Multinomial Logit and Type I(1)
Nested Logit Models

<table>
<thead>
<tr>
<th>Multinomial</th>
<th>Const.</th>
<th>Age</th>
<th>savings</th>
<th>Private Pension</th>
<th>Public Pension</th>
<th>Other Members' Income</th>
<th>Health</th>
<th>Self</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial \log(p_1/p_2)}{\partial x}$</td>
<td>8.231</td>
<td>-0.094</td>
<td>0.051</td>
<td>-0.104</td>
<td>-0.116</td>
<td>0.009</td>
<td>0.745</td>
<td>-0.862</td>
</tr>
<tr>
<td>$\frac{\partial \log(p_1/p_2)}{\partial x}$</td>
<td>11.519</td>
<td>-0.154</td>
<td>-0.243</td>
<td>-0.060</td>
<td>-0.050</td>
<td>-0.005</td>
<td>0.422</td>
<td>-0.638</td>
</tr>
<tr>
<td>$\frac{\partial \log(p_1/p_4)}{\partial x}$</td>
<td>18.999</td>
<td>-0.255</td>
<td>0.205</td>
<td>-0.205</td>
<td>-0.266</td>
<td>-0.034</td>
<td>2.209</td>
<td>-1.473</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nested Logit I(1)</th>
<th>Const.</th>
<th>Age</th>
<th>savings</th>
<th>Private Pension</th>
<th>Public Pension</th>
<th>Other Members' Income</th>
<th>Health</th>
<th>Self</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial \log(p_1/p_2)}{\partial x}$</td>
<td>7.918</td>
<td>-0.089</td>
<td>0.066</td>
<td>-0.096</td>
<td>-0.110</td>
<td>0.010</td>
<td>0.742</td>
<td>-0.821</td>
</tr>
<tr>
<td>$\frac{\partial \log(p_1/p_3)}{\partial x}$</td>
<td>11.055</td>
<td>-0.156</td>
<td>-0.239</td>
<td>-0.062</td>
<td>-0.061</td>
<td>-0.004</td>
<td>0.524</td>
<td>-0.738</td>
</tr>
<tr>
<td>$\frac{\partial \log(p_1/p_4)}{\partial x}$</td>
<td>21.5888</td>
<td>-0.320</td>
<td>0.208</td>
<td>-0.272</td>
<td>-0.338</td>
<td>-0.028</td>
<td>2.770</td>
<td>-2.233</td>
</tr>
</tbody>
</table>

Notes: (1) Health: Dummy for health status.
(2) Self: Dummy whether a respondent receives income rather than wage and pension.
6.49 and 1.25.

So far, we have not paid any attention at the estimates of the regression coefficients obtained for each model. We now do so only for the multinomial logit model and Type I(1) model, the one we judged to be the best among the nested logit models. First, we should point out that \( \beta_j \) cannot be separately identified in these models and that only the difference \( \beta_j - \beta_k \) can be identified. Second, it is not meaningful to compare \( \beta_j - \beta_k \) for different models because their interpretations are totally different for different models. The meaningful quantity we should look at is that which expresses the degree of the effect of the independent variables upon the relative probabilities of the four alternatives. For this purpose we compare \( \partial \log(P_j/P_k) / \partial x \) for the aforementioned two models. For the multinomial logit model, the derivatives do not depend on \( x \). For the nested logit model, the derivatives are evaluated at the mean value of each independent variable. As can be seen from Table 2, the signs of these derivatives are identical for the two models, and, moreover, there is little appreciable difference in their magnitudes between the two models.

5. Conclusions

In this paper we took up the multinomial logit model of Tachibanaki and Shimono [1985], which analyzed the choice of the job status of the elderly, and fit the essentially same data to various nested logit models. The one in which the alternatives "Full-time employed" and "Part-time employed" are most closely nested, "Self-employed" given the next closest position, and "Retired"
set apart from the rest, turned out to be the most plausible model from both theoretical and statistical considerations. However, the hypothesis of the multinomial logit model was accepted under the framework of this most plausible nested logit model. Moreover, there was little appreciable difference in the estimates of the probabilities of the four alternatives between the two models. Thus, we conclude that the assumption of the multinomial logit model adopted by the Tachibanaki-Shimono paper is a reasonable one.
References


TECHNICAL REPORTS
NATIONAL SCIENCE FOUNDATION
GRANTS DMS 82-19748, DMS 86-03779, and SES 82-08180
Econometric Workshop


