NUMBER 28

# A NOTE ON MAXIMUM LIKELIHOOD ESTIMATION OF DISCRETE CHOICE MODELS FROM THE 1978 SURVEY OF DISABILITY AND WORK

Barry V. Bye and Sal Gallicchio

Division of Disability Studies

November 1982

Social Security Administration
Office of Policy
Office of Research and Statistics

## INTRODUCTION

The purpose of this paper is to inform users of the 1978 Survey of Disability and Work that the usual maximum likelihood procedures for estimating discrete choice models where the dependent variable is self reported disability status, beneficiary status or application status are not appropriate. A more complex approach is needed because the 1978 Survey sample is stratified on these factors.

Such sample designs have been referred to in the literature as retrospective, choice based or endogenous sampling designs. The likelihood functions for such designs involve the sample design parameters. This is not the case for simple random samples or samples which are stratified on only explanatory variables. The methodology discussed in this paper applies not only to socio-economic structural equation models, but also to the usual kinds of linear logistic models that researchers have used in post survey reports to analyze disability, beneficiary and application status. Although the methodology presented is not new, its application to the 1978 Survey design is more complex than that discussed in the literature thus requiring special attention. In this paper we suggest an estimation stratagey and discuss it in the context of an analysis of the effect of local labor market factors on the disability process.
The paper is organized as follows: We first give a brief overview of discrete choice analysis in retrospective sample designs which serves to identify the issues at hand. We then develop the likelihood for a simple design which has the basic structural features of the 1978 Survey design. Next we describe the 1978 Survey design and the local labor market analysis. This is followed by the derivation of conditional likelihoods for that analysis. Finally, an estimation strategy is suggested.

## AN OVERVIEW OF ANALYSIS IN RETROSPECTIVE SAMPLES

In a prospective study of the incidence of the occurence of some random response variable, Y , a sample of individuals is drawn from a general population and risk factors and other explanatory (regression) variables, X, are recorded and considered as fixed. For each sample case, the dependent variable, Y, is measured and viewed as a random event. Since the probability of the occurence of the event under study is often small, it is necessary to take large samples in order to estimate, precisely, the relationship between Y and X. Usually we are interested in the estimation of the parameters of a probabilistic model, $P(Y|X,\theta)$ by maximum likelihood methods. Since the dependent variable is the random variable, the maximum likelihood estimator (mle) of $\theta$ is just the solution of $\max_{\theta} \left[\prod_{obs} P(Y|X,\theta)\right]$.

Because of the need for large samples in prospective studies, it is often more practical to use retrospective techniques. In a retrospective study, the response variable, Y, is regarded as fixed, and separate samples of individuals are selected from the various categories of the response variable. The explanatory variables are regarded as random, conditional on response status. Because retrospective studies draw separate samples for each category of the response variable, Y, a smaller total sample size is usually required and data collection can often be completed quickly. 1/

In retrospective studies, the estimation of the probabilistic model $P(Y|X)$ is no longer straight forward since it is necessary to condition estimation on being in the sample. The likelihood of the sample data is given by

$$\lambda(Y,X) = \prod_{obs} P(X|Y).$$

The mle of $\theta$ for the probabilistic model $P(Y|X,\theta)$ can be obtained from

$$\max_{\theta} \left(\prod_{obs} \frac{P(Y|X,\theta)\ P(X)}{P(Y)}\right)$$

by applying Bayes rule to $P(X|Y)$ and choosing $\theta$ subject to the constraint

$$P(Y) = \int_X P(Y|Z,\theta)\ P(Z)\partial Z.$$

If $P(X)$ and/or $P(Y)$ are not known, then other methods must be used to estimate $\theta$.

---

1/ In addition to the 1978 Survey, retrospective sampling approaches have been used in other studies conducted by the Division of Disability Studies. Both the Bellmon Review (see the Technical Appendix to Implementation of Section 304(g) of Public Law 96-265, "Social Security Disability Amendments of 1980," Report to the Congress by the Secretary of Health and Human Services, January 1982) and the Study of Trial Work Patterns of Disability Insurance Beneficiaries (see Division of Disability Studies Workplan No. 07302) used retrospective sampling plans.

An even more complicated situation arises when the sample is stratified by both dependent and independent variables. As example of what can happen, consider the estimation of the odds ratio from a simple four-fold table. In this case, we have a dichotomous dependent variable and one dichotomous explanatory variable. The estimation of the odds ratio is analogous to the estimation of the parameters of a dicrete choice model where the logistic functional form is assumed.

Let X=1,0 by the explanatory variable, Y=1,0 by the variable. If a, b, c, and d represent the observed frequencies of X and Y, jointly, we have the following table:

| | | Y | |
|---|---|---|---|
| | | 0 | 1 |
| X | 0 | a | b |
| | 1 | c | d |

In prospective studies the odds ratio is given by $\frac{b/a}{d/c} = \frac{bc}{ad} = \frac{c/a}{d/b}$ which is the odds ratio from retrospective studies.

Therefore the odds ratio can be estimated from either study design, even though P(Y) and P(X) are usually not known.

If one stratifies by both dependent and explanatory variables, the estimation of the odds ratio must be adjusted to account for the sampling fraction for each combination. As an illustration, suppose the entire population was observed in the study and we obtained the following summary table:

| | | Y | | |
|---|---|---|---|---|
| | | 0 | 1 | Total |
| X | 0 | 500 | 1,500 | 2,000 |
| | 1 | 2,500 | 5,500 | 8,000 |
| | Total | 3,000 | 7,000 | 10,000 |

The population odds ratio would be given by

$$OR = \frac{1,500 \times 2,500}{500 \times 5,500} = \frac{15}{11}$$

If a sample is drawn which is stratified on both X and Y and the sampling rates are disproportionate, then knowledge of the sampling rates is required to obtain an unbiased estimate of the population odds ratio.

For example, if the sampling rates are $w_{00}=1$, $w_{01}=1$, $w_{10}=.2$ and $w_{11}=.1$ and the following table based on the sample was obtained

| | | Y: 0 | Y: 1 |
|---|---|---|---|
| X | 0 | 500 | 1,500 |
| | 1 | 500 | 550 |

the sample odds ratio would be

$$OR_S = \frac{500 \times 1,500}{500 \times 550} = \frac{30}{11}$$

which is very far from the correctly weighted population estimate

$$OR = \frac{\left[(5)(500)\right]\left[(1)(1500)\right]}{\left[(1)(500)\right]\left[(10)(550)\right]} = \frac{15}{11}$$

The general problem of estimating associations among dependent and independent variables in retrospective sample designs has received quite extensive treatment in the literature. The problem of estimating the parameters of logistic choice functions has been treated by Farewell 2/, Beaslow and Powers 3/, Prentice 4/, and Prentice and Pyke 5/. Manski and Lerman 6/ and Manski and McFadden 7/ treat these issues in the context of general functional forms for $P(Y|X,\theta)$. Manski and

---

2/ Farewell, V. T., "Some Results on the Estimation of Logistic Models Based on Retrospective Data," Biometrika (1979), 66, 1, pp. 27-32.

3/ Breslow, N. and Powers, W., "Are There Two Logistic Regressions for Retrospective Studies," Biometrics 34, 100-105.

4/ Prentice, Ross, "Use of the Logistic Model in Retrospective Studies," Biometrics 32, 599-606.

5/ Prentice, R. L. and Pyke, R., "Logistic Disease Incidence Models and Case Control Studies," Biometrika (1979), 66, 3, pp. 403-11.

6/ Manski, Charles F., and Lerman, Steven R., "The Estimation of Choice Probabilities From Choice Based Samples," Econometrica, Vol. 45, No. 8 (November, 1977).

7/ Manski, Charles F., and McFadden Daniel, "Alternative Estimators and Sample Design for Discrete Choice Analysis," unpublished Manuscript, January 1977.

McFadden also discuss the general estimation problem when samples are stratified on both dependent and independent variables considering all possible combinations of knowledge or the lack thereof of the marginal densitites P(X) and P(Y).

## A MORE COMPLICATED SAMPLE DESIGN

The 1978 Survey sample design is somewhat more complex than those that have been discussed in the previously cited literature. Before proceeding to the development of the likelihoods for this design, a simple example of the basic structure of the 1978 Survey problem might be helpful.

Suppose we have a set of observations (AP,X) 8/ with probability density function f(AP,X) which is specified up to a set of unknown parameters, $\theta$, such that

$$f(AP,X) = P(AP|X,\theta)\ P(X)$$

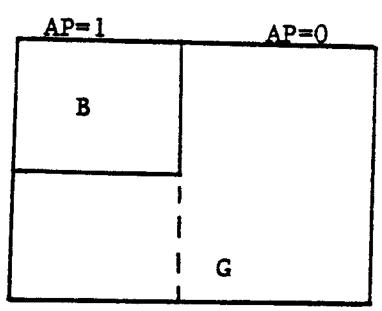
where

$P(AP|X,\theta)$ is the discrete choice function.

P(x) is the marginal density over X.

Let us further suppose that the set of (AP,X) is divided into two groups as shown in the following diagram:

Set of (AP,X)




8/ We can think of AP as the response variable taking on values of 1 and 0, and X as the set of regressors.

Some (but not all) of the observations with AP=1 have been placed in B. Assume that the joint density h, of B and G with (AP,X) is also known up to some unknown parameter $\phi$ so that

$$h(B,AP,X) = P(B|AP,X,\phi)\ P(AP,X)$$
$$h(G,AP,X) = P(G|AP,X,\phi)\ P(AP,X)$$

where according to the diagram

$$P(B|AP=0,X,\phi)=0 \text{ and}$$
$$P(G|AP=0,X,\phi)=1$$

Let us now suppose that we have fixed a sample design of size n which specifies that we choose $n_B$ cases from B and $n_G$ from G. Let $\mu_B = \frac{n_B}{n}$ and $\mu_G = \frac{n_G}{n}$.

We can now exhibit the likelihood, $\lambda$, of observations (AP=1,X) and (AP=0,X) from this design as follows:

$$\lambda(AP=1,X) = \mu_B\ P(AP=1,X|B) + \mu_G P(AP=1,X|G) \text{ and}$$

$$\lambda(AP=0,X) = \mu_G\ P(AP=0,X|G)\ \underline{9}/.$$

Using Bayes theorem and the probabilities previously defined, we can rewrite these likelihoods as follows:

$$\lambda(AP=1,X) = \mu_B\left(\frac{P(B|AP=1,X,\phi)\ P(AP=1|X,\theta)\ P(X)}{P(B)}\right)$$

$$+ \mu_G\left(\frac{P(G|AP=1,X,\phi)\ P(AP=1|X,\theta)\ P(X)}{P(G)}\right)$$

$$= P(AP=1|X,\theta)\ P(X)\left(\frac{\mu_B}{P(B)}\ P(B|AP=1,X,\phi) + \frac{\mu_G}{P(G)}\ P(G|AP=1,X,\phi)\right)$$

$$\lambda(AP=0,X) = \mu_G\ \frac{P(G|AP=0,X,\phi)\ P(AP=0|X,\theta)\ P(X)}{P(G)}$$

$$= P(AP=0|X)\ P(X)\left(\frac{\mu_G}{P(G)}\right)$$

---

9/ For this design P(AP=0,X|B)=0.

where P(B) and P(G) are the marginal probabilities of B and G respectively and are related to the other quantities by

$$P(B) = \int_{(AP=1,X)} P(B|AP=1,z,\phi)\ P(AP=1,z,\theta)\ P(z)dz$$

$$P(G) = \int_{(AP=1,X)} P(G|AP=1,z,\phi)\ P(AP=1|z,\theta)\ P(z)dz + \int_{(AP=0,X)} P(AP=0|z,\theta)\ P(z)dz$$

If we wish to estimate θ and ϕ by maximizing the usual likelihood function, assuming P(X) to be known, the choice of θ, ϕ would have to satisfy the constraints shown above. If P(X) is not known, which is usually the case, Manski and McFadden have suggested that θ and ϕ be estimated by maximizing the conditional likelihood function which in this case would contain products of terms of the form

$$\lambda(AP=1|X) = \frac{\lambda(AP=1,X)}{\lambda(AP=1,X) + \lambda(AP=0,X)}$$

and

$$\lambda(AP=0|X) = \frac{\lambda(AP=0,X)}{\lambda(AP=1,X) + \lambda(AP=0,X)}$$

where λ(AP,X), for AP=0,1, are given above.

They have shown that under suitable regularity conditions, the unconstrained conditional mle is consistent and asymtotically normally distributed. The advantage of using the conditional likelihood is that the unknown P(X) does not appear in the formulation. The standard errors of the estimated parameters can be obtained from the diagonals of the inverted conditional likelihood information matrix.

## THE 1978 SURVEY SAMPLE DESIGN 10/

The 1978 Survey sample must be construed, for some analyses, as having been stratified on both dependent and independent variables. A two-frame sampling approach was used. The first frame was a general population frame of noninstitutionalized adults ages 18-64 as of June 1978, in the continental United States. This frame was supplemented by a second frame consisting of recent Social Security disability beneficiaries and denied applicants also noninstitutionalized and age 18-64.

The 1976 Health Interview Survey (HIS) sample was used for selection from the general population frame. The HIS sample is designed for use by the National Center for Health Statistics for gathering basic data on the health of the general population. The Social Security frame consisted of about 1.8 million persons obtained from SSA's Master Beneficiary Record (MBR). About 1.67 million of these persons represented Social Security Title II disability beneficiaries with date of current entitlement between September 1972 and September 1977. The remaining records represented persons who applied for disability benefits but had been denied where the administrative denial action had been recorded on the MBR between January 1 and September 15, 1977.

For the purposes of the subsequent discussion, we assume that the overlap between the two frames has been removed. 11/ Thus, we can consider that the 1978 Survey sample design stratifies the noninstitutionalized population into three groups:

B--Beneficiaries with date of current entitlement 9/15/72 - 9/15/77

D--Denied applicants 1/1/77 - 9/15/77

G--All other persons not in B or D.

---

10/ For a detailed description of the 1978 Survey Sample design, see 1978 Survey of Disability and Work--Technical Introduction, Office of Research and Statistics, SSA Publication No. 13-11745, January 1982.

11/ This has, in fact, been done for most 1978 Survey Analyses. The reasons for and methods of exclusion are discussed in the reference cited in footnote 10, pages 15-17.

The Social Security Beneficiary stratum (B) was further stratified into four age substrata: under 35, 35-44, 45-54, 55-64. The general population stratum was further stratified into five substrata designed to represent levels of disability in 1976 based on several data items obtained from the HIS interview schedule. The diagram in figure 1 represents the overall design 12/:

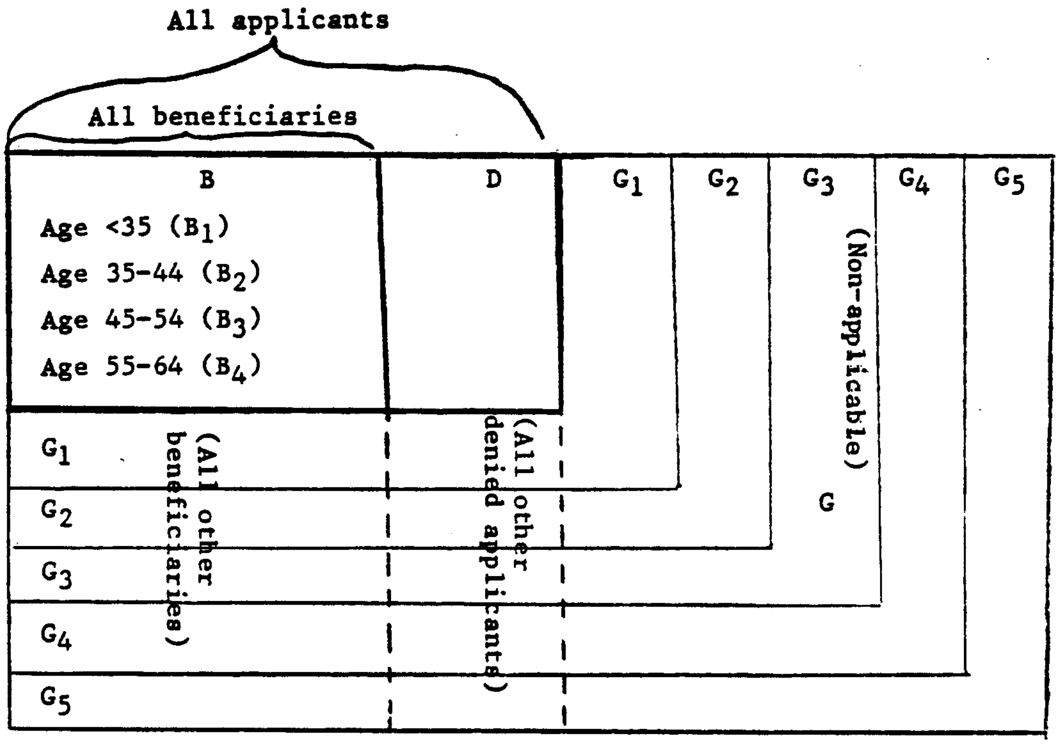



FIGURE 1.--Diagram of the 1978 Survey Sample stratification.

---

12/ The HIS sample was a stratified cluster sample similar to most Census-type designs. We assume that this geographic stratification and clustering has no bearing on the subsequent methodology proposed in this paper.

## AN EXAMPLE OF A DISCRETE CHOICE MODEL

A research design for a 1978 Survey analysis proposed by Jesse M. Levy provides a context for a discussion of the maximum likelihood methodology. 13/ Levy proposes to investigate the determinants of self-perceived disability, the decision to apply for benefits given that one perceives oneself to be disabled, and the award decision given application (and that the applicant is insured for disability). In his application and award analyses, Levy excludes those persons who have applied for benefits prior to January 1977. 14/ This leaves the 1978 Survey design with the following components:

B--Beneficiaries with date of entitlement 9/72-9/77 who had applied after 1/77.

D--Denied applicants 1/77-9/77 who had applied after 1/77.

G--General population who had not applied before 1/77; and are not in group B or D.

Levy assumes that the self perception of disability, the decision to apply for benefits on the part of the individual and the award decision by the Social Security Administration represent discrete choices determined by a vector of characteristics, X, which may include local labor market factors. For each of these events there is a choice probability function, $P(Y|X,\theta)$, which is known up to the unknown parameters $\theta$. 15/

In the remainder of this paper we will use application and award analyses as examples of the estimation approach. The development of the likelihood for the self-perceived disability models is almost identical to that of the application model.

---

13/ See Levy, Jesse M., "The Impact of Local Labor Market Characteristics on the Disability Process: Further Data," draft research proposal, 2/82, Division of Disability Studies, ORS.

14/ The reasons for this are beyond the scope of this paper.

15/ Levy proposes a random coefficient probit model.

## DERIVATION OF THE CONDITIONAL LIKELIHOOD--APPLICATION MODEL

The presentation of the likelihood of an observation from the 1978 Survey sample design follows closely the development in the last section. Let us define the following quantities:

Let

AP be the application indicator; that is AP=1 if the person applied for benefits after January 1977, and AP=0 otherwise.

X be the vector of exogenous variables upon which the application decision depends.

$B_i$ be the SSA beneficiary stratum indicators; i=1,...4.

$A_i$ be the indicator that the age for individual i is in the range specified by $B_i$.

D be the SSA denied applicant stratum indicator.

$G_j$ be the HIS stratum indicators; j=1,...5.

$P(B_i)$ be the proportion of the population in the SSA frame in stratum $B_i$; and similarly for P(D) and $P(G_j)$. 16/

$P(X,A_i)$ be the marginal density of $(X,A_i)$.

$\mu_{B_i}$ be the proportion of sample cases in stratum $B_i$; that is $\mu_{B_i} = \frac{n_{B_i}}{n}$ where $n_{B_i}$ is the number of sample cases in $B_i$ and n is the total sample size. Similarly define $\mu_D$ and $\mu_{G_j}$.

---

16/ After exclusions.

We will also assume that $P(AP|X,A_i,\theta)$ 17/ and $P(B_i|AP,X,A_i,\phi)$ 18/ are known functions up to the unknown parameters $\theta$ and $\phi$.

The likelihood of an observation $(AP=1,X,A_i)$ from the 1978 survey sample design is given by

$$\lambda(AP=1,X,A_i) = \mu_{B_i}\, P(AP=1,X,A_i|B_i) + \mu_D\, P(AP=1,X,A_i|D) + \sum_{j=1}^{5} \mu_{G_j}\, P(AP=1,X,A_i|G_j). \quad (1)$$

19/

Each term of this expression represents the joint probability of occurence of $(AP=1,X,A_i)$ and the indicated stratum; thus, for example, $\mu_{B_i}\, P(AP=1,X,A_i|B_i)$ is the probability of drawing a case from $B_i$ times the probability of drawing $(AP=1,X,A_i)$ given $B_i$ and so on.

Using Bayes theorem and collecting terms, equation (1) can be rewritten as:

$$\lambda(AP=1,X,A_i) = P(AP=1|X,A_i,\theta)\, P(X,A_i) \left( \frac{\mu_{B_i}}{P(B_i)}\, P(B_i|AP=1,X,A_i,\phi) + \frac{\mu_D}{P(D)}\, P(D|AP=1,X,A_i,\phi) + \sum_{j=1}^{5} \frac{\mu_{G_i}}{P(G_i)}\, P(G_j|AP=1,X,A_i,\phi) \right) \quad (2)$$

In a similar fashion, the likelihood of an observation $(AP=0,X,A_i)$ is given by

$$\lambda(AP=0,X,A_i) = P(AP=0|X,A_i,\theta)\, P(X,A_i) \left( \sum_{j=1}^{5} \frac{\mu_{G_j}}{P(G_j)}\, P(G_j|A_i=0,X,A_i,\phi) \right) \quad (3)$$

where $P(AP=0|X,A_i,\theta)=1-P(AP=1|X,A_i,\theta)$

The conditional likelihoods, $\lambda(AP=1|X,A_i)$ and $\lambda(AP=0|X,A_i)$ are shown on the next page.

---

17/ We assume implicitly that $P(AP|X,A_i,\theta)$ does not depend on $G_j$.
18/ And similarly for $P(D|AP,X,A_i,\phi)$, $P(G_j|AP,X,A_i,\phi)$.
19/ Note that $P(AP=1,X,A_i|B_j) = 0$ for $j \neq i$.

$$\lambda(AP=1|X,A_i) = \frac{P(AP=1|X,A_i,\theta)}{P(AP=1|X,A_i,\theta)\ M_1 + (1-P(AP=1|X,A_i,\theta))\ M_2}$$

$$\lambda(AP=0|X,A_i) = \frac{(1-P(AP=1|X,A_i,\theta))M_2}{P(AP=1|X,A_i,\theta)\ M_1 + (1-P(AP=1|X,A_i,\theta))M_2}$$

Where

$$M_1 = \frac{\mu_{Bi}}{P(B_i)}\ P(B_i|AP=1,X,A_i,\theta) + \frac{\mu_D}{P(D)}\ P(D|AP=1,X,A_i,\phi) + \sum_{j=1}^{5} \frac{\mu_{Gj}}{P(G_j)}\ P(G_j|AP=1,X,A_i,\phi)$$

$$M_2 = \sum_{j=1}^{5} \frac{\mu_{Gi}}{P(G_j)}\ P(G_j|AP=0,X,A_i,\phi)$$

Although we are carrying a lot of baggage in the formulae on page 13, the general nature of the sample design seems to leave little choice. Things could be much simplified if we could assume that $P(X|AP=1)$ was independent of the strata. The problem is, however, that the applicants have not been randomly distributed among the strata. In particular, given knowledge of SSA's disability program, it is unlikely that the distribution of X for beneficiaries is the same as that for denied applicants.

The complexity of the overall estimation problem depends to a large extent on the kinds of simplifying assumptions one is willing to make about the conditional probability functions involving $\theta$ and $\phi$. Some aspects of these issues will be discussed in a subsequent section.

Finally, it can be noted that if one assumes that the self-perception of disability is a necessary condition to application for benefits, then the conditional likelihood functions for analyses of self-perceived disability have the same form as those just presented, where the conditional probability functions take on meaning in that context.

## THE CONDITIONAL LIKELIHOOD--AWARD MODEL

The likelihoods for the award model shown below are similar to those developed for the applications model. We are dealing with award status conditional on application. The terms defined in the previous section should be taken in that context. Also note that allowed applicants can only come from the $B_i$ and $G_j$ strata whereas the denied applicants come from only the D and $G_j$ strata.

The award likelihoods are given by:

$$\lambda(AW=1,X,A_i) = P(AW=1|X,A_i,\theta)\,P(X,A_i)\left(\frac{\mu_{B_i}}{P(B_i)}P(B_i|AW=1,X,A_i,\phi)+\sum_{j=1}^{5}\frac{\mu_{G_j}}{P(G_j)}P(G_j|AW=1,X,A_i,\phi)\right)$$

$$\lambda(AW=0,X,A_i)=\left(1-P(AW=1|X,A_i,\theta)\right)P(X,A_i)\left(\frac{\mu_D}{P(D)}P(D|AW=0,X,A_i,\phi)+\sum_{j=1}^{5}\frac{\mu_{G_j}}{P(G_j)}P(G_j|AW=0,X,A_i,\phi)\right)$$

## A TWO STEP ESTIMATION APPROACH

We now turn to the practical problem of the estimation of $\theta$ and $\phi$ for the application model. We propose a two step approach where $\phi$ is estimated first. Then the expression $M_1$ and $M_2$ in (4) can be computed for each sample case and $\theta$ obtained by maximizing the conditional likelihood function. 20/ In all cases, the quantities $\mu_{B_i}$, $\mu_D$, $\mu_{G_j}$, $P(B_i)$, $P(D)$ and $P(G_j)$ are treated as known. 21/

To see how $\phi$ can be estimated, let S represent the sample strata $(B_i, D, G_j)$ with the following correspondence: $S=1\rightarrow B_1$, $S=2\rightarrow B_2$, $S=3\rightarrow B_3$, $S=4\rightarrow B_4$, $S=5\rightarrow D$, $S=6\rightarrow G_1, \ldots S=11\rightarrow G_5$. $M_1$ and $M_2$ are just linear functions of the conditional probabilities $P(S|AP,X,A_i,\phi)$ where $\phi$ is a vector of unknown parameters, AP is the application status and X is a vector of regressors.

Assume that $p(S|AP,X,A_i,\phi)$ can be represented by a probabilistic model with unknown parameters $\phi$. 22/ As before, the likelihood of observing $(S,X,A_i)$ is given by

$$\lambda(S,X,A_i) = \mu_s\, p(AP,X,A_i|S)$$

which by Bayes theorem

$$= \mu_s \frac{p(S|AP,X,A_i,\phi) \cdot p(X,A_i)}{P(S)}$$

---

20/ One could also leave $\phi$ unspecified in the likelihood and estimate $\theta$ and $\phi$ simultaneously. However, due to the large number of parameters and observations plus the complicated nature of the assumed probabilistic models, this becomes computationally impractical.

21/ In practice, the P's can be consistently estimated using the individual case weights available in the file. For Levy's analysis, this would be done only for persons who had not applied prior to January 1, 1977. The $\mu$'s for this subsample are also unknown but can be consistently estimated by counting the unweighted subsample cases which meet the criterion.

22/ Whether or not one chooses a separate set of parameters for each i and performs a separate mle on each partition, $A_i$, of the sample depends on whether one wants to assume that the coefficients of the regressor variables are constant across $A_i(i=1,4)$. Of course it is possible to test this hypotheses after one has completed the separate estimation of each $A_i$. As an intermediate position, assume that the coefficients of the regressor variables, X, are the same for all i and allow for different constants by making $A_i$ a dummy variable in the model.

whence the conditional likelihood is

$$\lambda(S|X,A_i) = \frac{P(S|AP,X,A_i,\phi)\ \frac{\mu_s}{P(S)}}{\sum_S P(S|AP,X,A_i,\phi)\ \frac{\mu_s}{P(S)}}$$

where $\mu_s$ is the probability of choosing strata S and P(S) is the marginal density of S.

An estimation of $\phi$ can then be obtained by maximizing the conditional likelihood function:

$$\prod_S \left[\prod_{obs\ in\ S} \lambda(S|X,A_i)\right].$$ 23/ 24/

Once $\phi$ has been obtained, the quantities M1 and M2 can be computed for each case and $\theta$ can be obtained by maximizing the conditional likelihood function:

$$\left(\prod_{AP=1} \lambda(AP=1|X,A_i)\right) \left(\prod_{AP=0} \lambda(AP=0|X,A_i)\right).$$ 25/

---

23/ If one chooses the logistic functional form for $P(S|AP,X,A_i,\phi)$; that is $P(S|X,A_i,\phi) = (1 + e^{-(\beta_0+X\beta+\gamma A_i)})^{-1}$ where $\phi=(\beta_0,\beta_1,\gamma)$ then $\beta$ and $\gamma$ can be estimated by the usual prospective analysis approach, and $\beta_0$ obtained after proper adjustment in terms of the $\mu_s$ and P(S). (See Manski and Lerman). These parameters would be estimated separately for AP=0 and AP=1.

24/ Note that

$$P(S|AP,X,A_i,\phi)=P\left(S|AW,X,A_i,\phi_{AW}\right)\ P\left(AW|X,A_i,\theta_{AW}\right)$$

The quantities $\phi_{AW}$ and $\theta_{AW}$ are obtained from the award model estimation. If the award model is estimated before the application model then $P(S|AP,X,A_i\phi)$ can be computed directly and $\phi$ (for the application model) need not be estimated.

25/ The maximization of either of these likelihoods can be handled quite manageably by a computer program, DFP, developed in the Division of Disability Studies. This program uses the Davidon-Fletcher-Powell method of optimization and supplies both estimates of the unknown parameters and their standard errors. This technique is described in G. R. Walsh, Methods of Optimigation, J. Wiley, 1975.

One further comment on the estimation is required. At the present time no data field is available either in house or on the public use tapes which identifies the HIS sample strata (the $G_j$). Without this item it appears as though the HIS sample will have to be treated as unstratified. 26/ While treating the HIS sample this way significantly reduces the number of parameters to be estimated in $\phi$, it also entails the very strong assumption that the joint distribution of $(AP,AW,X,A_i)$ does not depend on $G_j$. This assumption is much stronger than the earlier assumption that $P(AP|X,A_i)$ and $P(AW|X,A_i)$ do not depend on $G_j$. 27/ The direction and magnitude of potential biases in estimated parameters is not known.

## CONCLUSION

We have shown in this paper that the likelihood functions for certain discrete choice models for the 1978 Survey are quite a bit more complicated than those ordinarily encountered. We have suggested an estimation approach which maximizes the conditional likelihood in two steps.

We should note that not all discrete choice models need have this complicated formulation. An example which may occur frequently would be labor force status. If one can exhibit a regression vector which exhausts a model of labor force status then the stratification in the 1978 Survey is of no consequence to the estimation. Of course Social Security Disability program status may be included in such a regression. In this case, the sample design is exogenous and the kernal of the likelihood function contains only the regression model specification as usual.

---

26/ We are now obtaining these data from the Bureau of the Census.

27/ We also feel that this assumption is much stronger than the ignoring of the geographic stratification in the HIS sample.

**Also, there are alternative estimation strategies which might be entertained. If the number of regressor variables is small, a generalized least squares approach might be used on a summary crosstabulation of the dependent and independent variables** together with the sampling variance/covariance matrix for the table. 28/

Finally, as discussed in the reference in footnote 10, the HIS general population sample with overlap cases included constitutes an exogenous sample of the entire population and might be suitable in its own right for certain analyses without recourse to the complex methodology described above. It does however seem improbabl that such an analytic strategy could be very efficient after having excluded thousands of cases belonging to the SSA sample.

---

28/ This technique is described in Grizzle, Starmer, and Koch, "Analysis of Categorical Data by Linear Models," Biometrics 25, 1969, pp 489-504.