MONASH

8/90

MONASH
UNIVERSITY

AUSTRALIA

AN ANALYSIS OF THE EFFECT OF AN OFFENDER'S EMPLOYMENT STATUS ON THE TYPE OF SENTENCE CHOSEN BY THE MAGISTRATE

Nicola J. Crichton and Timothy R. L. Fry

Working Paper No. 8/90

July 1990

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 86746 962 5

AN ANALYSIS OF THE EFFECT OF AN OFFENDER'S EMPLOYMENT STATUS ON THE TYPE OF SENTENCE CHOSEN BY THE MAGISTRATE

Nicola J. Crichton and Timothy R. L. Fry

Working Paper No. 8/90

July 1990

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS, COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

An Analysis of the Effect of an Offender's Employment Status on the Type of Sentence Chosen by the Magistrate.

Nicola J. Crichton[1] and Timothy R.L. Fry[2]

1. Department of Mathematical Statistics and Operational Research, University of Exeter, Streatham Court, Rennes Drive, Exeter, UK.

2. Department of Econometrics, Monash University, Clayton, Victoria 3168, Australia.

Summary

In this paper we will consider the problem of estimating a multinomial logit model for choice behaviour, when the data has been collected using choice–based sampling. We illustrate the estimation process using a set of data collected by the National Association for the Care and Resettlement of Offenders to investigate whether an offender's employment status affected the sentence chosen by the magistrate.

Keywords: multinomial logit; choice–based sampling; choice model; unemployment; magistrates court.

1. Introduction

The purpose of this paper is to examine the problem of estimating a multinomial logit model when the data arises from choice–based sampling. Our interest in this estimation problem arose because we wished to investigate some data concerned with the sentence imposed on offenders who appeared in the magistrates court. The data was collected by The National Association for the Care and Resettlement of Offenders (NACRO) to answer a range of questions, in particular whether a relationship exists between unemployment and the type of sentence imposed. The data will be described briefly in section 2, a more detailed description is given in Crow and Simon (1987a).

Most estimation methods assume random sampling, which in the context of our problem would mean a random selection of offenders is drawn and we observe the sentence type imposed. The random sample may be simple or perhaps stratified on the basis of some explanatory variables. In a choice–based sampling process a sequence of offenders who have had a particular sentence type imposed is drawn and their characteristics are observed. That is we select a sample of individuals with a particular outcome variable and collect information about the variables we think influenced the choice of outcome variable. We could think of this as selecting a stratified random sample where the stratification is on the outcome variable. This procedure is discussed in more detail in Manski and McFadden (1981), and Manski and Lerman (1977).

This type of sampling is commonly used in economic surveys and in epidemiological case–control studies. There are often significant cost savings associated with using a choice–based sample. For example, if we are studying choice of mode of transport for getting to work, it is much cheaper to survey passengers at the station and car users at the car park than to interview commuters in their home. As far as epidemiological case–control studies are concerned, they provide economies of cost and study duration, especially for rare diseases.

In this paper we attempt to modify existing theoretical results to a form more suitable for practical use. Indeed most of the results are attainable with standard statistical packages. We hope this will encourage use of the more appropriate multinomial logit model for the analysis of quantal choice data in preference to using linear probability models, even when the data comes from choice–based sampling.

2. The NACRO data and analysis

Crow and Simon (1987a, b) give detailed information about how the data were collected, the twenty–one hypotheses that were stated for the research, and provide an analysis of the data. In discussing the background to the study Crow and Simon (1987a) state "Employment status is often believed to have some bearing on the way that offenders are dealt with, but has been subject to little detailed study. This, and the growth of unemployment in recent years, gave rise to the study reported here." We can take this as the primary motivation for the study.

Ideally the study would have included both crown courts and magistrates courts. However, resources were limited and it was felt it would be difficult to get access to crown courts. The study was therefore confined to a small number of magistrates courts in England and Wales. Six courts were chosen to represent different experiences of unemployment and sentencing traditions. Three areas of England and Wales were selected to represent different experiences of unemployment; one in which unemployment was relatively low; one in which unemployment had been high for many years; and one which had experienced a transition from low to high unemployment in recent years. Within each area two magistrates courts were selected, one which used custodial sentences more frequently than the area average and one which used such sentences less frequently than the area average, as shown in Table 1. A sample period was selected for each court, as indicated in Table 1, to provide at least 500 cases per court.

[Table 1 about here]

The sample design is a cluster design, based on area unemployment and custody use. Over the sampling period every male property offender, aged 17 or over, sentenced in the court was included in the data set, except for those offenders who were fined where every second case was included. This was a mixture of census and systematic sampling procedures. Notice that the sentence type was the basis for the "every other fines case" sampling. Sentence type is the variable whose frequency of occurrence it is hoped to explain in the subsequent analysis. Thus, sentence type is our outcome variable and the sample has been selected by a choice–based sampling procedure.

The data extracted from the court record consisted of age; ethnic origin; details of the offence and previous criminal record; employment status on day of sentence; social enquiry recommendation (SER), if any; and the sentence type imposed. We might have thought it desirable to collect information on other potential explanatory variables, for example, income, educational qualifications, employment status at the time of committing the crime, however, this was not done.

The data made available to us was a restricted subset of the information originally collected and consists of age; offending score; employment status on the day of sentence; SER recommendation, if any; and sentence type imposed. The offending score (OSCORE) is constructed from information about the offence committed and the offender's previous history; the variables included and their 'scores' are given in Table 2. An offender's OSCORE is obtained by adding up the score he obtains for each variable in Table 2. We were only provided with the final OSCORE, not the constituent variables which we would have preferred. For a first time offender OSCORE has a minimum value of 3 and a maximum of 10, whilst for an offender with a criminal history the minimum is 5 and the maximum is 19.

[Table 2 about here]

Although the study considers male property offenders over 17 years old, those offenders aged $17 - 20$ are treated differently by the courts than older men. Thus Crow and Simon (1987a, b) often consider separately men aged $17 - 20$ and those aged 21 and over. We consider only those aged 21 and over in our analysis. Crow and Simon (1987a,

b) consider nine sentence types: fine; discharge, compensation order only; probation; community service order; suspended imprisonment; immediate custody; deferred; attendence centre; and committed to crown court for sentence. Since attendence centre is only appropriate for $17 - 20$ year olds it does not occur in our data. In addition we have omitted from our analysis offenders with sentence deferred and those committed to crown court for sentence, because the numbers receiving such sentences are very small and their final sentence is unknown. In our analysis we consider only those male property offenders age 21 years or more who received one of the following sentences: fine; discharge; probation; community service order (CSO); suspended imprisonment (SI); and immediate custodial sentence (ICS), giving a total sample size of 1359 men.

Crow and Simon (1987a, b) report the results of fitting linear probability models to the data. In their analysis they treated the data as if they were a random sample. To overcome the underrepresentation of fines cases they simply 'double up' the fines category, that is they included each fine case twice. Such a procedure assumes that each fine case included had a missing identical twin.

Suppose the sentence categories are indexed by $j = 1, 2, \ldots, J$ and the offenders are indexed by $k = 1, 2, \ldots, K$. In order to fit the linear probability models J binary variables y_{jk} were constructed, such that

$$y_{jk} = \begin{cases} 1, & \text{if offender } k \text{ receives sentence } j; \\ 0, & \text{otherwise.} \end{cases}$$

Two explanatory variables were used in the model, offending score (x_{1k}) and employment status (x_{2k}), were x_{1k} is OSCORE for offender k and

$$x_{2k} = \begin{cases} 1, & \text{if offender } k \text{ is in employment on day of sentence;} \\ 0, & \text{otherwise.} \end{cases}$$

The model estimated can be written

$$y_{jk} = \beta_{0j} + \beta_{1j} x_{1k} + \beta_{2j} x_{2k} + e_{jk}$$

where e_{jk} is the error term. This is a model for the probability that $y_{jk} = 1$.

For each offender $y_{jk} = 1$ for a single sentence type and is zero for all others, so $\sum_j y_{jk} = 1$. This holds for all possible values of x_{1k} and x_{2k}, hence it follows that

$$\sum_j \beta_{0j} = 1; \qquad \sum_j \beta_{1j} = 0; \qquad \sum_j \beta_{2j} = 0; \qquad \sum_j e_{jk} = 0.$$

Thus we need only estimate the parameters for five of the six sentence types. The estimation can be carried out in any package able to do multiple regression.

Crow and Simon (1987b) present the results of such an analysis for each sentence type in each court separately. However, they only provide the values of the employment coefficient for the analyses in which all courts are considered together, so court effects are ignored. We have performed, for our data set with only six sentence types, the same type of analysis as Crow and Simon used, for each court separately, the results are shown in Table 3.

[Table 3 about here]

There are several objections to the analysis as performed by Crow and Simon (1987b). Firstly, the 'doubling up' procedure seems an unnecessary assumption. If we introduce an appropriate weighting variable to allow for the under sampling of fines cases, as discussed in section 4, we would have obtained the same parameter estimates but got larger standard errors. This reflects the fact that our data set is not really as large as that implied by the 'doubling up' approach.

Amemiya (1981, 1985) and Maddala (1983) give a number of objections to linear probability models, among these is that the predicted probabilities can be outside the interval $[0,1]$. We can observe this happening in the NACRO data set, for example, the probability of a fine for an employed individual in court 4 will be greater than one if their OSCORE is less than six. There are nine such cases in the data set, so it is not simply that the models give impossible probabilities for variable combinations that are rare.

Finally the linear probability model neglects the multi–alternative nature of the problem. The analysis fits separate (independent) regressions for each sentence type and assumes that the alternative to sentence type j is 'not sentence type j', where as the alternatives are the other sentence types. These objections lead us to our reanalysis of the data using a more appropriate model and making allowance for the choice–based sampling.

3. The Multinomial Logit Model

When considering response or choice probabilities, parametric models are often proposed. Logistic models are frequently used in the analysis of epidemiological studies (Breslow and Day 1980). In such studies there are generally only two outcome states, diseased or not diseased. The multinomial logit model generalises this analysis to situations in which the outcome variable has several states and has been used in the econometric literature to model choice behaviour.

Let there exist $j = 1, 2, \ldots, J$ mutually exclusive and exhaustive responses (e.g. sentence types, illness states) and $z = 1, 2, \ldots, Z$ mutually exclusive and exhaustive stimulus values. By stimulus values we mean possible combinations of the explanatory variables, for example if our explanatory variables are OSCORE and employment status there are 34 combinations so $Z = 34$. Let P_{jz} be the probability of response j conditional on stimulus z. Our interest is in inferring these conditional response probabilities from our sample.

A variety of parametric forms might be considered for P_{jz}, we have chosen to use the multinomial logit model. For a particular stimulus value, z, the response variable has J categories with category probabilities P_{jz}, $j = 1, 2, \ldots, J$. The probability that in a sample of n_z independent observations we obtain n_{jz} in the jth category is

$$P(n_{1z}, n_{2z}, \ldots, n_{Jz}) = \frac{n_z!}{n_{1z}!\, n_{2z}! \ldots n_{Jz}!} {P_{1z}}^{n_{1z}} {P_{2z}}^{n_{2z}} \ldots {P_{Jz}}^{n_{Jz}},$$

where $\sum_j P_{jz} = 1$, that is there are only $J - 1$ distinct probabilities.

The multinomial logit transformation of $P_{1z}, \ldots, P_{Jz}$ is the parameter set $\theta_{1z}, \ldots, \theta_{Jz}$ defined by

$$\theta_{jz} = \log(P_{jz}/P_{1z}) \qquad j = 1, 2, \ldots, J.$$

We can model the logits with a linear combination of the explanatory variables, so that $\theta_{jz} = \beta_{0j} + \beta'_{1j}\mathbf{x}_z$, where $\mathbf{x}_z$ is the vector of explanatory variables making up stimulus value z. The restriction $\sum_j P_{jz} = 1$ leads to the following expression for P_{jz},

$$P_{jz} = \frac{\exp(\beta_{0j} + \beta'_{1j}\mathbf{x}_z)}{\sum_{k=1}^{J} \exp(\beta_{0k} + \beta'_{1k}\mathbf{x}_z)} \qquad j = 1, 2, \ldots, J; \quad z = 1, 2, \ldots, Z. \tag{1}$$

However, $\theta_{1z} = \log(P_{1z}/P_{1z}) = 0$, so we can re-write (1) as

$$\begin{aligned} P_{1z} &= \frac{1}{1 + \sum_{k=2}^{J} \exp(\beta_{0k} + \beta'_{1k}\mathbf{x}_z)} & z = 1, 2, \ldots, Z, \\ P_{jz} &= \frac{\exp(\beta_{0j} + \beta'_{1j}\mathbf{x}_z)}{1 + \sum_{k=2}^{J} \exp(\beta_{0k} + \beta'_{1k}\mathbf{x}_z)} & j = 2, 3, \ldots, J. \end{aligned} \tag{2}$$

The multinomial logit model would seem to be appropriate for the problem described in section 2. It overcomes the objections to the linear probability models, since the predicted probabilities always lie in the interval [0, 1] and the model allows for the multi–alternative structure. Although the model structure is more complicated than the linear probability model it is computationally easier to deal with than probit or complementary log-log models. A major difference between the linear probability model and the multinomial logit model is that we consider the effect of explanatory variables to be additive in the linear model and multiplicative in the logit model.

An objection to the logit model is the assumption of independence of irrelevant alternatives, explained by Judge et al (1985, p770–771). The effect of this assumption is that the odds of a particular dichotomous choice are unaffected as additional alternatives are added. This seems implausible if some of the alternatives are close substitutes, for example two buses identical except for colour. For the sentencing data this is not such an unreasonable assumption since the sentence types are not close substitutes.

Having chosen a model we need to estimate the parameters. Assuming we have a random sample, this can be done in many packages, for example SPSSX, GLIM, GENSTAT. We have chosen to consider estimation in GENSTAT as this allows us to write procedures appropriate for use in later sections. Although the multinomial distribution is not one of the distributions explicitly included in GLIM and GENSTAT we can make use of the relationship between the multinomial distribution and the Poisson distribution to fit the models. This relatonship is shown in Aitkin et al (1989, p231–232) and McCullagh and Nelder (1989, p210).

For convenience in later sections we define a variable $x_{0z} = 1$ for all z, so $\beta_{0j} = \beta_{0j}x_{0z}$ and we can simply write $\beta_{0j} + \beta'_{1j}\mathbf{x}_z$ as $\beta'_j\mathbf{x}_z$.

To fit the model in GENSTAT or GLIM we need to define a response factor with J levels, corresponding to the multinomial response categories. We also need to define, for each explanatory variable, a factor with an appropriate number of levels, thus giving a complete set of classifying factors. The parameter β_j is the coefficient of the interaction between the jth level of the response factor and the explanatory variables $\mathbf{x}$. In addition the model must contain a parameter for each explanatory variable combination, that is each observed multinomial distribution. This requires the inclusion in the model of a nuisance factor with Z levels, which can be conveniently achieved by fitting all levels of interaction for the set of classifying factors. An example is given by McCullagh and Nelder (1989, p212) and a detailed example of fitting the multinomial logit model in GLIM is provided by Aitkin et al (1989, p235).

If the data arises as a random sample we will be able to use this methodology as presented, however, if we have a choice–based sample some modifications are necessary. These modifications are given in section 4.

4. Estimating the Multinomial logit model for choice–based data

Consider the population joint distribution of stimuli and responses,

$$\Pi_{jz} = P_{jz}p_z = Q_{zj}q_j, \tag{3}$$

where P_{jz} is the probability of response j conditional on stimulus z; p_z is the marginal probability of stimulus z; Q_{zj} is the probability of stimulus z conditional on response j; and q_j is the marginal probability of response j. Our interest is in P_{jz} for which a model was proposed in section 3. However, with choice–based data we know about Q_{zj} rather than P_{jz}. Notice that

$$q_j = P(\text{response } j) = \sum_{z=1}^{Z} P(\text{response } j|z)P(z) = \sum_{z=1}^{Z} P_{jz}p_z,$$

so that

$$Q_{zj} = \frac{P_{jz}p_z}{\sum\limits_{z=1}^{Z} P_{jz}p_z}.$$

If we have a choice–based sample we can only make statements about P_{jz} if we have information about q_j and p_z. Manski and Lerman (1977) and Manski and McFadden (1981) consider situations in which choice–based samples are combined with exact knowledge of q_j and/or p_z. The assumption of such knowledge is often unrealistic, so Hsieh et al (1985) consider the situation in which q_j are estimated from an auxillary random sample. We will consider both these situations and the implications for our model. In each situation we will consider two approaches to the estimation.

4.1. When the population proportions, q_j, are known.

This situation was first considered by Manski and Lerman (1977) who proposed and developed the weighted exogenous sampling maximum likelihood (WESML) estimator. Let h_j be the probability an observation in our choice based sample has response j; then the joint distribution for our sample is

$$\begin{aligned} f(j,z) = P(z| \text{ response } j)P(\text{response } j) &= Q_{zj}h_j \\ &= \frac{P_{jz}p_zh_j}{q_j} \end{aligned} \tag{4}$$

Suppose N is the total number of individuals in the sample and n_{jz} is the number of individuals with stimulus z who selected j. The exogenous sampling log-likelihood function would be

$$\ell = \sum_{n=1}^{N} \log P_{j_n z_n} = \sum_{z=1}^{Z} \sum_{j=1}^{J} n_{jz} \log P_{jz}$$

Manski and Lerman (1977) show that under choice–based sampling this yields inconsistent estimates. Instead they propose using the weighted exogenous likelihood function

$$W_N = \sum_{n=1}^{N} w(j_n) \log P_{j_n z_n} = \sum_{z=1}^{Z} \sum_{j=1}^{J} w_j n_{jz} \log P_{jz} \tag{5}$$

where $w_j = (q_j/h_j)$, q_j is assumed known and h_j is the sample proportion. Manski and Lerman (1977) show that maximisation of (5) gives estimates that are consistent and asymptotically normal with covariance matrix given by

$$V = \Omega^{-1} \Delta \Omega^{-1} \tag{6}$$

where

$$\Omega = \left[-E \left(\frac{\partial^2 w_j \log P_{jz}}{\partial\beta\partial\beta'} \right) \right] \qquad \text{and} \qquad \Delta = \left[E \left(\frac{\partial w_j \log P_{jz}}{\partial\beta} \right) \left(\frac{\partial w_j \log P_{jz}}{\partial\beta'} \right) \right]$$

It can be shown that these matrices reduce to

$$\Omega = N \sum_{z=1}^{Z} \sum_{j=1}^{J} P_{jz} \frac{\partial \log P_{jz}}{\partial\beta} \frac{\partial \log P_{jz}}{\partial\beta'} p_z = N \sum_{z=1}^{Z} \sum_{j=1}^{J} \frac{1}{P_{jz}} \frac{\partial P_{jz}}{\partial\beta} \frac{\partial P_{jz}}{\partial\beta'} p_z$$

$$\Delta = N \sum_{z=1}^{Z} \sum_{j=1}^{J} w_j P_{jz} \frac{\partial \log P_{jz}}{\partial\beta} \frac{\partial \log P_{jz}}{\partial\beta'} p_z = N \sum_{z=1}^{Z} \sum_{j=1}^{J} w_j \frac{1}{P_{jz}} \frac{\partial P_{jz}}{\partial\beta} \frac{\partial P_{jz}}{\partial\beta'} p_z$$

A proof is given in Amemiya (1985) and these are the forms used in Manski and McFadden (1981).

We could do the maximisation of expression (5) using GENSTAT, setting w_j in a weight vector and following the procedure described in section 3. To calculate estimates of P_{jz} we use the following form

$$P_{jz} = \frac{w_j \exp(\hat{\beta}_j' \mathbf{x}_z)}{\sum_{k=1}^{J} w_k \exp(\hat{\beta}_k' \mathbf{x}_z)} = \frac{\left(\frac{w_j}{w_1}\right) \exp(\hat{\beta}_j' \mathbf{x}_z)}{1 + \sum_{k=2}^{J} \left(\frac{w_k}{w_1}\right) \exp(\hat{\beta}_k' \mathbf{x}_z)}$$

We can simply include w_j/w_1 in the parameter β_{0j} given by GENSTAT, so the intercept parameter is $\beta_{0j}+\log(w_j/w_1)$. The parameters associated with the explanatory variables require no alteration, however, the standard errors will require some adjustment. We can calculate the correct standard errors using the expressions for Ω and Δ. We will also need an estimate of p_z to use these forms, this will simply be n_z/N where $n_z = \sum_j n_{jz}$.

Suppose we have H explanatory variables, then Ω and Δ appropriate for the multinomial logit model are as follows. We consider parameters in the order $\beta_{02},\ldots,\beta_{0J}$ then the $J-1$ parameters associated with explanatory variable 1, $\beta_{12},\ldots,\beta_{1J}$, then those for explanatory variable 2 and so on.

$$\Omega(i,i) = \sum_{z=1}^{Z} n_z P_{iz}(1-P_{iz})x_z x_z \qquad \text{and} \qquad \Omega(i,k) = -\sum_{z=1}^{Z} n_z P_{iz} P_{kz} x_z x_y$$

where,

$$\begin{aligned}
&x_z = 1 \text{ for all } z \text{ if } i \leq J,\\
&x_z = x_{1z},\ P_{iz} = P_{(i-J)z} \text{ if } J+1 \leq i \leq 2J,\\
&x_z = x_{2z},\ P_{iz} = P_{(i-2J)z} \text{ if } 2J+1 \leq i \leq 3J,\\
&\vdots\\
&x_z = x_{Hz},\ P_{iz} = P_{(i-HJ)z} \text{ if } HJ+1 \leq i \leq (H+1)J,\\
&x_y = 1 \text{ if } k \leq J,\\
&x_y = x_{1z},\ P_{kz} = P_{(k-J)z} \text{ if } J+1 \leq k \leq 2J,\\
&\vdots\\
&x_y = x_{Hz},\ P_{kz} = P_{(k-HJ)z} \text{ if } HJ+1 \leq i \leq (H+1)J.
\end{aligned}$$

Using similar notation

$$\Delta(i,i) = \sum_{z=1}^{Z} n_z \left(P_{iz}^2(w_1P_{1z} + w_2P_2 + \ldots + w_JP_{Jz} - 2w_i) + w_iP_{iz}\right) x_z x_z$$

$$\Delta(i,k) = \sum_{z=1}^{Z} n_z P_{iz} P_{kz}(w_1P_{1z} + w_2P_2 + \ldots + w_JP_{Jz} - w_i - w_k)x_z x_y.$$

A GENSTAT procedure could be written to calculate these matrices and hence the new covariance matrix.

An alternative approach is to generalise the method suggested by both Breslow and Day (1980, p203) and McCullagh and Nelder (1989, p113) for binary data. Consider $P(\text{Choice } j|\mathbf{x}_z) = P_{jz}$ as given by (2). Introduce a dummy variable Y to define whether an individual is sampled or not, and denote the sampling proportions by

$$\pi_j = P(Y=1| \text{ choice } j) = P(\text{in sample } | \text{ choice } j)$$

assume the sampling proportions depend only on choice j and not on the explanatory variables, so $P(Y=1|j, \mathbf{x}_z) = \pi_j$ for all z. We can now use Bayes' theorem to compute the choice frequency among sampled individuals with explanatory variable $\mathbf{x}_z$.

$$\begin{aligned} P(\text{choice } j|Y=1, \mathbf{x}_z) &= \frac{P(Y=1|\text{choice } j, \mathbf{x}_z)P_{jz}}{P(Y=1|\text{choice } 1, \mathbf{x}_z)P_{1z} + \sum_{k=2}^{J} P(Y=1|\text{choice } k, \mathbf{x}_z)P_{kz}} \\ &= \frac{\pi_j \exp(\beta_{0j} + \boldsymbol{\beta}_j'\mathbf{x}_z)}{\pi_1 + \sum_{k=2}^{J} \pi_k \exp(\beta_{0k} + \boldsymbol{\beta}_k'\mathbf{x}_z)}; \qquad (7) \\ &= \frac{\left(\frac{\pi_j}{\pi_1}\right) \exp(\beta_{0j} + \boldsymbol{\beta}_j'\mathbf{x}_z)}{1 + \sum_{k=2}^{J} \left(\frac{\pi_k}{\pi_1}\right) \exp(\beta_{0k} + \boldsymbol{\beta}_k'\mathbf{x}_z)} \\ &= \frac{\exp(\beta_{0j}^{\dagger} + \boldsymbol{\beta}_j'\mathbf{x}_z)}{1 + \sum_{k=2}^{J} \exp(\beta_{0k}^{\dagger} + \boldsymbol{\beta}_k'\mathbf{x}_z)}, \qquad (8) \end{aligned}$$

where

$$\beta_{0j}^{\dagger} = \beta_{0j} + \log\left(\frac{\pi_j}{\pi_1}\right) \qquad j = 1, 2, \ldots, J \qquad (9)$$

If we use GENSTAT as suggested in section 3, the estimates for the parameters associated with the explanatory variables are correct, but we will get estimates for $\beta_{0j}^{\dagger}$ rather than β_{0j}. We can easily recover the required estimates using equation (9), since

$$\pi_j = P(Y=1|j) = \frac{P(Y=1,j)}{P(j)} = \frac{P(j|Y=1)P(Y=1)}{P(j)} = \frac{(n_j/N)P(Y=1)}{q_j},$$

where q_j, the population proportion selecting j, is assumed known. Also n_j and N are assumed fixed. Thus $\text{Var}(\beta_{0j}) = \text{Var}(\beta_{0j}^{\dagger})$ so the standard errors produced by GENSTAT will be correct.

4.2 When the population proportions, q_j are estimated.

In many situations we will not know the proportion of the population selecting choice j, q_j. For example in a survey to determine a model of choice selection for breakfast cereals we would not know the proportion of the population selecting a particular brand. However, we could carry out a survey of a large random sample of the population and just ask what brand of breakfast cereal they eat in order to estimate the proportions q_j. If we estimate q_j with an auxillary sample, independent of our choice–based sample, both the estimation procedures discussed in section 4.1 are still appropriate. Both give estimators that are consistent and asymptotically normal, however, the variances of the estimators will be different to those given in section 4.1.

The WESML estimation is discussed by Hsieh et al (1985) who provide an expression for the covariance matrix. By using the appropriate expressions for $\frac{\partial \log P_{jz}}{\partial \beta}$, from our multinomial logit model, we can evaluate the covariance matrix given by Hsieh et al (1985). This could be done with a GENSTAT procedure.

For multiplicative intercept models such as ours, Hsieh et al (1985) prove that conditional maximum likelihood (CML) estimation and full–information concentrated likelihood equation (FICLE) estimation will give estimators that coincide. The second approach considered in section 4.1 is CML estimation. Hsieh et al use the form given in equation (7), so the relationship between the intercept parameters is

$$\exp(\beta^*_{0j}) = \pi_j \exp(\beta_{0j}) \qquad j = 1, 2, \ldots, J. \tag{10}$$

Suppose we estimate the population proportions, q_j, by a sample of size T in which t_j people selected j, so $\hat{q}_j = t_j/T$. Then we will have

$$\hat{\pi}_j = \frac{(n_j/N)P(Y=1)}{t_j/T}, \tag{11}$$

but, the term $P(Y = 1)$ will cancel in equation (7). So if we let $\hat{\pi}^*_j = \frac{n_j/N}{t_j/T}$, we can replace π_j in equation (10) by π^*_j. GENSTAT will give us β^*_{0j} rather than β_{0j}, however

$$\exp(\beta_{0j}) = \frac{\exp(\beta^*_{0j})}{\hat{\pi}^*_j} = \left[\exp(\beta^*_{0j})\left(\frac{N}{n_j}\right)\right]\left[\frac{t_j}{T}\right]. \tag{12}$$

Also GENSTAT gives variances for β^*_{0j}. Hsieh et al (1985) obtain variances for $\exp(\beta_{0j})$ by considering $\exp(\beta_{0j})$ as the product of two independent random variables, so

$$\begin{aligned}
V\left[\exp(\beta_{0j})\right] &= \left[E\left(\frac{t_j}{T}\right)\right]^2 V\left[\left(\frac{N}{n_j}\right)\exp(\beta^*_{0j})\right] + \left[E\left(\left(\frac{N}{n_j}\right)\exp(\beta^*_{0j})\right)\right]^2 V\left(\frac{t_j}{T}\right) \\
&\quad + V\left[\left(\frac{N}{n_j}\right)\exp(\beta^*_{0j})\right] V\left(\frac{t_j}{T}\right) \\
&= \left(\frac{t_j/T}{n_j/N}\right)^2 V\left(\exp(\beta^*_{0j})\right) + \left(\frac{\exp(\beta^*_{0j})}{n_j/N}\right)^2 \frac{1}{T}\left(\frac{t_j}{T}\right)\left(1-\frac{t_j}{T}\right) \\
&\quad + \frac{1}{T}\left(\frac{t_j}{T}\right)\left(1-\frac{t_j}{T}\right)\left(\frac{V\left(\exp(\beta^*_{0j})\right)}{n_j/N}\right).
\end{aligned} \tag{13}$$

However, Hsieh et al omit the last of the three terms since it will in general be small. This still leaves the problem of what is the variance of $\exp(\beta^*_{0j})$. Since the estimators are asymptotically normally distributed, $\exp(\beta^*_{0j})$ will be asymptotically lognormal, hence

$$V\left[\exp(\beta^*_{0j})\right] = \exp(2\beta^*_{0j})\exp(v^*_j)\left(\exp(v^*_j)-1\right) \tag{14}$$

where $v^*_j = \sqrt{V(\beta^*_{0j})}$ and is given by GENSTAT. This enables us to get $V\left(\exp(\beta^*_{0j})\right)$ which we can use in (13) to get $V\left(\exp(\beta_{0j})\right)$. We need to use an expression similar to (14) to recover $V(\beta_{0j})$.

An alternative and more straightforward method for getting $V(\beta_{0j})$ is to use a Taylor series approximation. Taking logs of (12) and rearranging gives

$$\beta_{0j} = \log\left(\frac{t_j}{T}\right) - \log\left(\frac{n_j}{N}\right) + \beta^*_{0j}$$

which, since n_j and N are fixed, gives

$$V(\beta_{0j}) = V\left[\log\left(\frac{t_j}{T}\right)\right] + V(\beta^*_{0j}).$$

A Taylor series expansion of $\log(t_j/T)$ about $E(t_j/T) = q_j$ gives

$$V\left[\log\left(\frac{t_j}{T}\right)\right] = \frac{1-q_j}{Tq_j} + \frac{(1-q_j)(3-q_j)}{2T^2q_j^2} + O\left(\frac{1}{T^3}\right). \tag{15}$$

To evaluate this we need to use $\hat{q}_j = t_j/T$. This approach will give $V(\beta_{0j})$ directly. In practice t_j will not be zero and T will be large. If we explicitly exclude the possibility that t_j is zero we need to divide expression (15) by

$$[1 - P(t_j = 0)]^2 = \left[1 - (1 - q_j)^T\right]^2$$

In most practical situations T will be large enough to make $(1 - q_j)^T$ very close to zero, so we will not generally need to make this adjustment to expression (15).

The Hsieh et al (1985) approach, working from equation (7), has resulted in $\beta_{01} \neq 0$. Usually we would have chosen $\beta_{01} = 0$ and adjusted the other parameters appropriately. If we now make such an adjustment we will need to make further alterations to the parameter standard errors. This will require knowledge of the covariance between β_{01} and β_{0j}.

Suppose instead we work from relationship (9), then

$$\begin{aligned}
\beta_{0j} &= \beta^{\dagger}_{0j} - \log\left(\frac{\pi_j}{\pi_1}\right) \qquad\qquad j = 2, \ldots, J \\
&= \beta^{\dagger}_{0j} - \log\left(\frac{n_j/N}{t_j/T} \times \frac{t_1/T}{n_1/N}\right) \\
&= \beta^{\dagger}_{0j} - \log\left(\frac{n_j(t_1/T)}{n_1(t_j/T)}\right) \\
&= \beta^{\dagger}_{0j} + \log\left(\frac{n_1}{n_j}\right) + \log\left(\frac{t_j}{T}\right) - \log\left(\frac{t_1}{T}\right)
\end{aligned}$$

Since n_1 and n_j are fixed we have, assuming the choice–based sample and the auxillary sample are independent, that

$$V(\beta_{0j}) = V(\beta^{\dagger}_{0j}) + V\left[\log\left(\frac{t_j}{T}\right)\right] + V\left[\log\left(\frac{t_1}{T}\right)\right] - 2\mathrm{Cov}\left[\log\left(\frac{t_j}{T}\right), \log\left(\frac{t_1}{T}\right)\right]. \quad (16)$$

Using Taylor series expansion about $E(t_j/T)$ and $E(t_1/T)$ gives

$$V(\beta_{0j}) = V(\beta^{\dagger}_{0j}) + V\left[\log\left(\frac{t_j}{T}\right)\right] + V\left[\log\left(\frac{t_1}{T}\right)\right] + 2\left(\frac{1}{T} - \frac{(q_j q_1 - q_1 - q_j)}{2T^2 q_j q_1}\right) + O\left(\frac{1}{T^3}\right) \quad (17)$$

where the second and third terms are given by (15). Again we should consider excluding the possibility that t_j is zero, as we did for expression (15).

5. Results for the NACRO data

In order to use the methodology described in sections 3 and 4 we will assume that the data set was collected as a choice-based sample and that the number of individuals of each sentence type was fixed. It is more problematic to decide whether we can assume q_j are known. Strictly we do not know q_j. We do know how many cases there would have been in a random sample because of the way the data was collected, however, this random sample is not independent of the choice–based sample. For the CML estimation this will only cause a problem with the variance of the intercept parameters, but these will be more complex as (16) should also involve terms for $\mathrm{Cov}(\beta^*_{0j}, \log(t_j/T))$ and $\mathrm{Cov}(\beta^*_{0j}, \log(t_1/T))$. We could expand these terms using a Taylor series, but we would still need to know $\mathrm{Cov}(\beta^*_{0j}, t_j/T)$ and $\mathrm{Cov}(\beta^*_{0j}, t_1/T)$. We simply present here the results assuming (i) that q_j are known, and (ii) that q_j as estimated is from an independent sample. It is important to note that the results from (ii) are given simply to illustrate the technique, we do not really have an appropriate independent sample to estimate q_j.

With each approach we fitted the minimal model, that is the model to give just the intercept parameters, β_{0j}. We are effectively modelling a large contingency table which in parts is rather sparse. We cannot really assess the overall fit of a model, however we can assess the value of adding terms to the model using an F statistic,

$$F = \left[\frac{G^2(2|1)}{u}\right]\left[\frac{G^2}{I-s}\right]^{-1} \quad \sim F_{u,I-s}$$

where M_1 is a model with r terms, M_2 is the model with u further terms, $r+u=s$. G^2 is the deviance from fitting model M_2 and $I-s$ is the number of degrees of freedom for the deviance, $G^2(2|1)$ is the change in deviance from fitting M_2 rather than M_1 and u is the change in degrees of freedom. Roberts et al (1987) discuss the merits of this and other goodness of fit tests.

As the study design was set up to include six courts representative of different area effects we might expect there to be different court effects which should, if present, be allowed for in the model before we consider other explanatory variables. The court effects were found to be significant, as shown in Table 4. We then added OSCORE to the model, this has a highly significant effect. Finally employment status was added to the model, this also has a highly significant effect, that is, it appears that employment status has a significant influence on the choice of sentence, even after we allow for the court and OSCORE.

[Table 4 about here]

We are unable to make any claims about how well the model fits the data, but this criticism can also be made of the Crow and Simon (1987a, b) analysis. The only assessments of fit they made was to consider R^2 for each of their multiple linear regressions.

The parameter estimates from WESML estimation are given, with their standard errors, in Table 5. The results for CML estimation with standard errors from equation (17) are also given in Table 5. If we use the formulation of Hsieh et al (1985) we get slightly different results for the parameters β_{0j}. These are shown in Table 6. Although the parameter estimates appear different for CML in Tables 5 and 6, they will give the same values for P_{jz}. Table 6 shows how close the standard errors from the Taylor series expansion (15) are to those from the more complicated formulation of equation (13).

[Tables 5 and 6 about here]

Notice how close the standard errors given by GENSTAT are to those from any of the corrections, either with q_j known or unknown. From a practical point of view there is probably little value in making the corrections for this example. This is likely to be the situation in most examples with reasonably large data sets.

The easiest way to compare the results of the multinomial logit model with those of the Crow and Simon approach is to consider some graphs of the probability of selecting particular sentences. All the information necessary to construct such graphs is given in Tables 3 and 5. For illustration we show in Figure 1 the graph for the probability of fine in court 1, whilst Figure 2 shows the probability of probation in court 2. It is clear from the figures that although the linear probability model may give, on average, a reasonable approximation to the probabilities it cannot allow for a probability peak in the middle of the OSCORE range. This is particularly a problem with probation and CSO, which tend to be used most in the middle of the OSCORE range. For the sentence discharge, which tends to be given only for very low OSCORE, the linear probability model gives probability predictions very close to those of the logit model in all courts.

[Figures 1 and 2 about here]

As an illustration of the sentencing probabilities within a court Figure 3 shows the sentence probabilities for the unemployed in court 1, whilst Figure 4 shows the pattern for the employed in court 1. Broadly similar pictures are found in the other courts. As we might expect there is a tendency for courts 2, 4, and 6 to show higher probabilities for awarding custodial sentences, SI and ICS, than courts 1,3, and 5. There are also differences between the courts with different area employment levels, for example courts 3 and 4 have a higher probability of giving a fine to an employed offender than courts 1 and 2, for all values of OSCORE.

[Figures 3 and 4 about here]

We agree with the conclusions of Crow and Simon that employment status affects the choice of sentence. In any particular court, for offenders with the same OSCORE, an unemployed offender is less likely to be fined than an employed offender; is more likely to be discharged; is more likely to receive probation; is more likely to have a CSO; is less likely to get suspended imprisonment; is more likely to get ICS. In addition we note that particularly for CSO and ICS, the probability of receiving such a sentence gets reasonably large (say greater than 0.2) at lower OSCORE for the unemployed. For example, in court 1 for unemployed the probability of ICS exceeds 0.2 at OSCORE 17, whilst for the employed this does not occur until OSCORE 19.

Crow and Simon (1987b) draw conclusions about the possible consequences of the rising unemployment rate on the number of offenders given custodial sentences. This is easy for a linear probability model because of the simple interpretation of the coefficients in such models. Crow and Simon (1987b) provide evidence that an increase of 1% in the unemployment rate in the general population might be followed by a 3% rise in unemployment among the offending population. If the number of men sentenced in a particular court in a year is m and supposing that the employment coefficient for the ICS regression in the court is c, then for a 1% rise in the unemployment rate there will be $3cm/100$ more men receiving a custodial sentence, assuming all other variables are the same and the model still holds.

The interpretation of the logit model is not so simple. Suppose we consider a particular court, say 1, and OSCORE value, say s, then the expected number of offenders given ICS is

$$\left[\frac{\exp(\beta_{06})\exp(\beta_{16}s)}{1+\sum_{j=2}^{6}\exp(\beta_{0j})\exp(\beta_{1j}s)}\right]m_u + \left[\frac{\exp(\beta_{06})\exp(\beta_{16}s)\exp(\beta_{26})}{1+\sum_{j=2}^{6}\exp(\beta_{0j})\exp(\beta_{1j}s)\exp(\beta_{2j})}\right]m_e$$

where m_u is the number of unemployed offenders in court 1 with OSCORE s, and m_e is the number of employed offenders. Suppose m_c more lose their job, so the number employed is reduced to $(m_e - m_c)$, the change in expected number of offenders given ICS is

$$\left[\left(\frac{\exp(\beta_{06})\exp(\beta_{16}s)}{1+\sum_{j=2}^{6}\exp(\beta_{0j})\exp(\beta_{1j}s)}\right) - \left(\frac{\exp(\beta_{06})\exp(\beta_{16}s)\exp(\beta_{26})}{1+\sum_{j=2}^{6}\exp(\beta_{0j})\exp(\beta_{1j}s)\exp(\beta_{2j})}\right)\right]m_c$$

A more common interpretation of the parameters is to consider the odds ratio of choice j for individuals with different sets, $\mathbf{z}^*$ and $\mathbf{z}$ of risk variables, that is

$$\frac{P_{j\mathbf{z}^*}}{P_{j\mathbf{z}}} = \exp\left(\boldsymbol{\beta}_j'(\mathbf{z}^* - \mathbf{z})\right)$$

So $\exp(\beta_{ij})$ is the fraction the risk of j is increased (or decreased) for every unit change in z_i.

Also of interest might be the odds of sentence j compared to say sentence 1 (fines).

$$\frac{P_{jz}}{P_{1z}} = \exp(\boldsymbol{\beta}_j'\mathbf{z})$$

or the odds of sentence j relative to sentence k which is $\exp\left[(\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{z}\right]$. So, for example, if all other variables are the same, the odds ratio of ICS compared to fine is multiplied by $\exp(\beta_{26}) = 0.181$ for a change from unemployed to employed, or by $\exp(-\beta_{26}) = 5.507$ for a change from employed to unemployed. It is difficult to provide confidence intervals for P_{jz}, or for the relative risk. However, since $\boldsymbol{\beta}$ is asymptotically normally distributed, so $\exp(\beta_{ij})$ will be asymptotically lognormally distributed, thus we can calculate confidence limits for $\exp(\beta_{ij})$.

Alternatively we could consider the log odds, in which case

$$\log\left(\frac{P_{jz}}{P_{kz}}\right) = \sum_{i=1}^{H}(\beta_{ij} - \beta_{ik})z_i$$

where H is the number of explanatory variables. If all other variables are the same then the effect on the log odds of ICS relative to fine of a change from employed to unemployed is $-\beta_{26} = 1.70$. The standard errors given in Table 5 can be used to calculate a 95% confidence interval for this effect, which we find does not contain 1, so conclude that being unemployed significantly increases the log odds of ICS relative to fine. The effect on the log odds of sentence j relative to sentence k of a change in employment will be $(\beta_{2j} - \beta_{2k})$, the variance of this includes the term $\mathrm{Cov}(\beta_{2j}, \beta_{2k})$. For CML estimation this comes directly from GENSTAT whilst for WESML it will come from the adjusted covariance matrix.

6. Discussion

The multinomial logit model seems structurally more appropriate than the model of Crow and Simon (1987a, b) however, we cannot test whether the logit model actually fits the data any better. The conclusions about the effect of employment status are similar from both models; that is, in any court for offenders with the same OSCORE, fine and SI are more likely for an employed offender than an unemployed offender, whilst the other sentences are more likely for an unemployed offender. This raises a number of moral and political issues. However, from a practical point of view there is no point in imposing a large fine on someone who has little chance of being able to pay the fine, they will simply default and have to be dealt with again by the court.

It is important, if we have choice–based data, to make adjustments for this in the analysis. It is preferable that assumptions about the unsampled cases, such as those made by the 'doubling up' method, are avoided. No such assumptions are made by CML and WESML estimation and the sample size is preserved. The example considered in this paper perhaps raises doubts about the practical benefits of adjusting the covariance matrix, however it is hard to say under what circumstances the adjustment is necessary.

A further point of practical interest is whether we should use CML or WESML. Hsieh et al (1985) perform Monte Carlo trials to determine some properties of the CML and WESML estimators, but find little to suggest one method is preferable. If the model to be estimated is the multinomial logit then CML seems much easier to perform than WESML as the estimation can be carried out with any package capable of performing multiple logistic regression and the adjustments for choice–based sampling can be done with a calculator. To perform WESML requires a package allowing the use of a weight vector, and adjustment of the covariance matrix requires a computer program. We recommend using CML for the multinomial logit model.

Schmidt and White (1984) argue that the ordered logit model is appropriate for sentencing data. In such models it is assumed that there is an underlying, unobservable, latent variable and that what we observe is an ordered categorical variable which corresponds to a partitioning of the latent variable. If such a latent variable were to be interpreted as a measure of the offenders 'culpability' then the model might be considered appropriate for sentencing data. However, since the sentencing variable is incomplete, in the sense that some disposals have been excluded, and because there would be some dispute about the correct ordering of some of the disposals, we decided such a model would be inappropriate.

The econometric package LIMDEP has the capability to fit multinomial logistic models. There is an option available that supposedly makes the corrections for choice based sampling using WESML estimation and assuming q_j known. Unfortunately the standard errors produced by the option seem to be incorrect.

In most survey work a design–based approach, rather than a model–based approach, is usual. However, in this area most development has been in the medical or econometric fields in which stochastic modelling is second nature, so a model–based approach has become the convention. Recent work on a design–based approach is discussed by Scott and Wild (1989) for the binary outcome situation.

In epidemiological studies the outcome variable is frequently binary but the sample design may be complex. Breslow and Cain (1988) discuss the use of logistic regression for two-stage samples of case–control data and Breslow and Zhao (1988) consider logistic regression for stratified case–control data. The papers provide the necessary adjustments for the parameter estimates and their covariance matrix to make allowance for the choice–based sampling for a binary outcome variable. Both of these sampling structures are likely to be of use in choice surveys in which the choice set is more than two. Extensions of these papers to a polytomous outcome variable would be useful.

Acknowledgements

We are grateful to NACRO for allowing us access to the data. The views expressed in this paper are not necessarily those of NACRO. Thanks also to John Hinde and Geoff Watterson for helpful discussions.

References

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical Modelling in GLIM*. Oxford:Oxford University Press.

Amemiya, T. (1985) *Advanced Econometrics*. Oxford: Backwell.

Breslow, N.E. and Cain, K.C. (1988) Logistic regression for two–stage case–control data. *Biometrika*, **75**, 11–20.

Breslow, N.E. and Day, N.E. (1980) *Statistical Methods in Cancer Research, Volume 1 – The Analysis of Case–control Studies*. Lyon: International Agency for Research on Cancer.

Breslow, N.E. and Zhao, L.P. (1988) Logistic regression for stratified case–control studies. *Biometrics*, 44, 891–899.

Crow, I. and Simon, F. (1987a) *Unemployment and Magistrates' Courts*. London: NACRO.

Crow, I. and Simon, F. (1987b) *Unemployment and Magistrates' Courts Appendices*. London: NACRO.

Hsieh, D.A., Manski, C.F. and McFadden D. (1985) Estimation of response probabilities from augmented reterospective observations. *Journal of the American Statistical Associaition*, **80**, 651–662.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T.C. (1985) *The Theory and Practice of Econometrics*, 2nd edn. New York: J. Wiley and Sons.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Manski, C.F. and Lerman, S.R. (1977) The estimation of choice probabilities from choice based samples. *Econometrica*, 45, 1977–1988.

Manski, C.F. and McFadden, D. (1981) Alternative estimators and sample designs for discrete choice analysis. In *Structural Analysis of Discrete Data with Econometric Applications*, (eds. C. Manski and D. McFadden). Cambridge, Massachusetts: MIT Press.

Roberts, G, Rao, J.N.K. and Kumar, S. (1987) Logistic regression analysis of sample survey data. *Biometrika*, **74**, 1–12.

Schmidt, P. and Witte, A. (1984) *An Economic Analysis of Crime and Justice: Theory, Methods and Applications*. Orlando, Florida: Academic Press.

Scott, A.J. and Wild, C.J. (1989) Selection based on the response variable in logistic regression. In *Analysis of Complex Surveys*, (eds. C.J. Skinner, D. Holt and T.M.F. Smith). New York: J. Wiley and Sons.

Table 1: The research design, showing the characteristics and sample period for each court.

Use of custodial sentence	Area unemployment Low	High	Low to High
Below average	Court 1 (June 83 to May 85)	Court 3 (July 83 to Dec 84)	Court 5 (Jan 84 to May 84)
Above average	Court 2 (July 83 to Dec 84)	Court 4 (Jan 84 to June 84)	Court 6 (Jan 84 to Sept 84)

Table 2: The seven variables that are included in OSCORE. The table shows the score for each variable category. An individual's OSCORE is computed by adding up his seven variable scores.

Variable	Category	Score
Current offence	burglary	4
	theft, fraud, forgery, deception	2
	damage	1
Number of charges	1	1
	2 or 3	2
	4 or more	3
Value of property involved	less then £20	1
	£20 – £99	2
	£100 or more, or motor vehicles	3
Number of previous convictions	none	0
	1 or 2	1
	3 or 4	2
	5 or more	3
Time since last sentence or release from custody	less than 1 year	1
	no previous conviction, more than 1 year	0
Similarity of offence type	if any previous offence was of similar type	1
	no previous offence was similar	0
Most severe previous sentence or disposal (lowest in list)	None	0
	discharge, compensation only	1
	fine, probation, supervision, care order	2
	attendence centre, community service order	3
	any custodial sentence	4

Table 3: Regression coefficients from the series of multiple linear regressions performed for each court seperately as described in section 2, with the fines cases 'doubled up'.

Sentence type	Coefficient for	Court 1	Court 2	Court 3	Court 4	Court 5	Court 6
Fine	constant	0.9798	0.6930	1.0646	1.2000	0.8364	1.0173
	OSCORE	-0.0487	-0.0417	-0.0444	-0.0574	-0.0383	-0.0501
	employment	0.1628	0.3273	0.0786	0.0912	0.1176	0.2253
Discharge	constant	0.2988	0.3106	0.2986	0.2403	0.6065	0.2573
	OSCORE	-0.0168	-0.0170	-0.0141	-0.0118	-0.0334	-0.0169
	employment	-0.0729	-0.0915	-0.0479	-0.0792	-0.0591	-0.0529
Probation	constant	0.0333	0.1806	-0.0676	0.0135	0.0386	0.0669
	OSCORE	0.0121	-0.0013	0.0158	0.0026	0.0068	-0.0001
	employment	-0.0889	-0.0675	-0.0221	-0.0213	-0.0855	-0.0558
Community service order	constant	-0.1217	0.0489	-0.0981	-0.0723	-0.1388	0.0130
	OSCORE	0.0232	0.0015	0.0186	0.0131	0.0191	0.0187
	employment	-0.0288	-0.0160	-0.0305	-0.0262	0.0102	-0.1142
Suspended imprisonment	constant	-0.1082	-0.0640	-0.0629	-0.1305	-0.1845	-0.1922
	OSCORE	0.0176	0.0214	0.0069	0.0196	0.0244	0.0198
	employment	0.0329	-0.0422	0.0219	0.0487	0.0374	0.1103
Immediate custody	constant	-0.0820	-0.1692	-0.1335	-0.2508	-0.1582	-0.1619
	OSCORE	0.0126	0.0371	0.0173	0.0339	0.0214	0.0287
	employment	-0.0005	-0.1101	-0.0001	-0.0130	-0.0205	-0.1127

Table 4: Assessing the model fit. The figures in brackets are for the WESML approach, the unbracketed figures are for the CML approach

Model	Deviance G^2	Change in deviance, $G^2(2/1)$	Degrees of freedom for G^2, $I-s$	Degrees of freedom for $G^2(2/1)$, u	F
Minimal	1735.7 (1454)	–	1015	–	–
+ court effect	1543 (1303)	193 (151)	990	25	4.95 (4.59)
+ OSCORE effect	767 (632.7)	776 (670)	985	5	199 (208)
+ employment effect	696 (568)	71 (64)	980	5	20.0 (22.1)

Table 5: Parameter estimates and adjusted standard errors for both WESML and CML estimation. Note for both WESML and CML the estimates shown are after adjustments have been made. For CML with q_j known, s.e.$(\hat{\beta})$ are the same as those given by GENSTAT.

Parameter	WESML				CML		
	Estimate $\hat{\beta}$	GENSTAT s.e.$(\hat{\beta})$	Known q_j s.e.$(\hat{\beta})$*	Unknown q_j s.e.$(\hat{\beta})$	Estimate $\hat{\beta}$	GENSTAT s.e.$(\hat{\beta})$	Unknown q_j s.e.$(\hat{\beta})$**
Discharge β_{02}	-0.598	0.351	0.343	0.307	-0.574	0.328	0.337
OSCORE β_{12}	-0.092	0.027	0.026	0.025	-0.094	0.026	0.026
Employment β_{22}	-0.867	0.208	0.204	0.191	-0.882	0.192	0.192
Court 2 β_{32}	0.192	0.362	0.344	0.330	0.163	0.333	0.333
Court 3 β_{42}	0.167	0.305	0.297	0.284	0.183	0.281	0.281
Court 4 β_{52}	-0.203	0.346	0.335	0.317	-0.195	0.317	0.317
Court 5 β_{62}	0.994	0.304	0.295	0.289	1.001	0.286	0.286
Court 6 β_{72}	-0.668	0.402	0.379	0.359	-0.699	0.363	0.363
Probation β_{02}	-3.829	0.488	0.449	0.406	-3.835	0.441	0.449
OSCORE β_{12}	0.262	0.035	0.032	0.031	0.261	0.032	0.032
Employment β_{22}	-1.265	0.274	0.252	0.245	-1.241	0.246	0.246
Court 2 β_{32}	0.777	0.345	0.324	0.327	0.797	0.327	0.327
Court 3 β_{42}	-0.776	0.341	0.320	0.316	-0.802	0.317	0.317
Court 4 β_{52}	-1.615	0.443	0.402	0.403	-1.614	0.401	0.401
Court 5 β_{62}	-0.446	0.348	0.321	0.326	-0.459	0.328	0.328
Court 6 β_{72}	-1.066	0.404	0.367	0.374	-1.019	0.370	0.370
CSO β_{02}	-6.526	0.624	0.567	0.546	-6.500	0.560	0.566
OSCORE β_{12}	0.455	0.043	0.039	0.039	0.452	0.039	0.039
Employment β_{22}	-1.091	0.277	0.255	0.256	-1.104	0.254	0.254
Court 2 β_{32}	0.274	0.437	0.390	0.412	0.399	0.404	0.404
Court 3 β_{42}	-0.796	0.374	0.344	0.343	-0.853	0.349	0.349
Court 4 β_{52}	-0.867	0.402	0.367	0.367	-0.865	0.372	0.372
Court 5 β_{62}	-0.620	0.388	0.321	0.326	-0.651	0.363	0.363
Court 6 β_{72}	0.458	0.335	0.317	0.326	0.531	0.322	0.322

Table continued

Parameter	WESML Estimate $\hat{\beta}$	GENSTAT s.e.$(\hat{\beta})$	Known q_j s.e.$(\hat{\beta})$*	Unknown q_j s.e.$(\hat{\beta})$	CML Estimate $\hat{\beta}$	GENSTAT s.e.$(\hat{\beta})$	Unknown q_j s.e.$(\hat{\beta})$**
SI β_{02}	-8.699	0.776	0.688	0.768	-8.658	0.687	0.693
OSCORE β_{12}	0.579	0.052	0.046	0.050	0.575	0.046	0.046
Employment β_{22}	-0.238	0.269	0.246	0.262	-0.233	0.248	0.248
Court 2 β_{32}	1.292	0.395	0.364	0.403	1.390	0.374	0.374
Court 3 β_{42}	-2.325	0.621	0.523	0.519	2.402	0.550	0.550
Court 4 β_{52}	-0.105	0.394	0.360	0.380	-0.099	0.367	0.367
Court 5 β_{62}	-0.146	0.395	0.355	0.365	-0.180	0.370	0.370
Court 6 β_{72}	-0.346	0.411	0.370	0.389	-0.264	0.382	0.382
ICS β_{02}	-12.863	1.010	0.858	0.858	-12.774	0.886	0.890
OSCORE β_{12}	0.832	0.063	0.054	0.055	0.825	0.055	0.055
Employment β_{22}	-1.706	0.357	0.310	0.330	-1.700	0.321	0.321
Court 2 β_{32}	2.685	0.497	0.429	0.472	2.758	0.456	0.456
Court 3 β_{42}	-0.967	0.547	0.462	0.470	-1.052	0.492	0.492
Court 4 β_{52}	0.664	0.482	0.419	0.449	0.668	0.440	0.440
Court 5 β_{62}	-0.021	0.501	0.425	0.461	-0.065	0.455	0.455
Court 6 β_{72}	1.042	0.469	0.407	0.447	1.139	0.431	0.431

* Uses equation (6)

** Uses equation (17)

Figure 1: Probability of Fine in court one

Curves are probabilities from multinomial logit model.
Straight lines are probabilities from linear probability model.

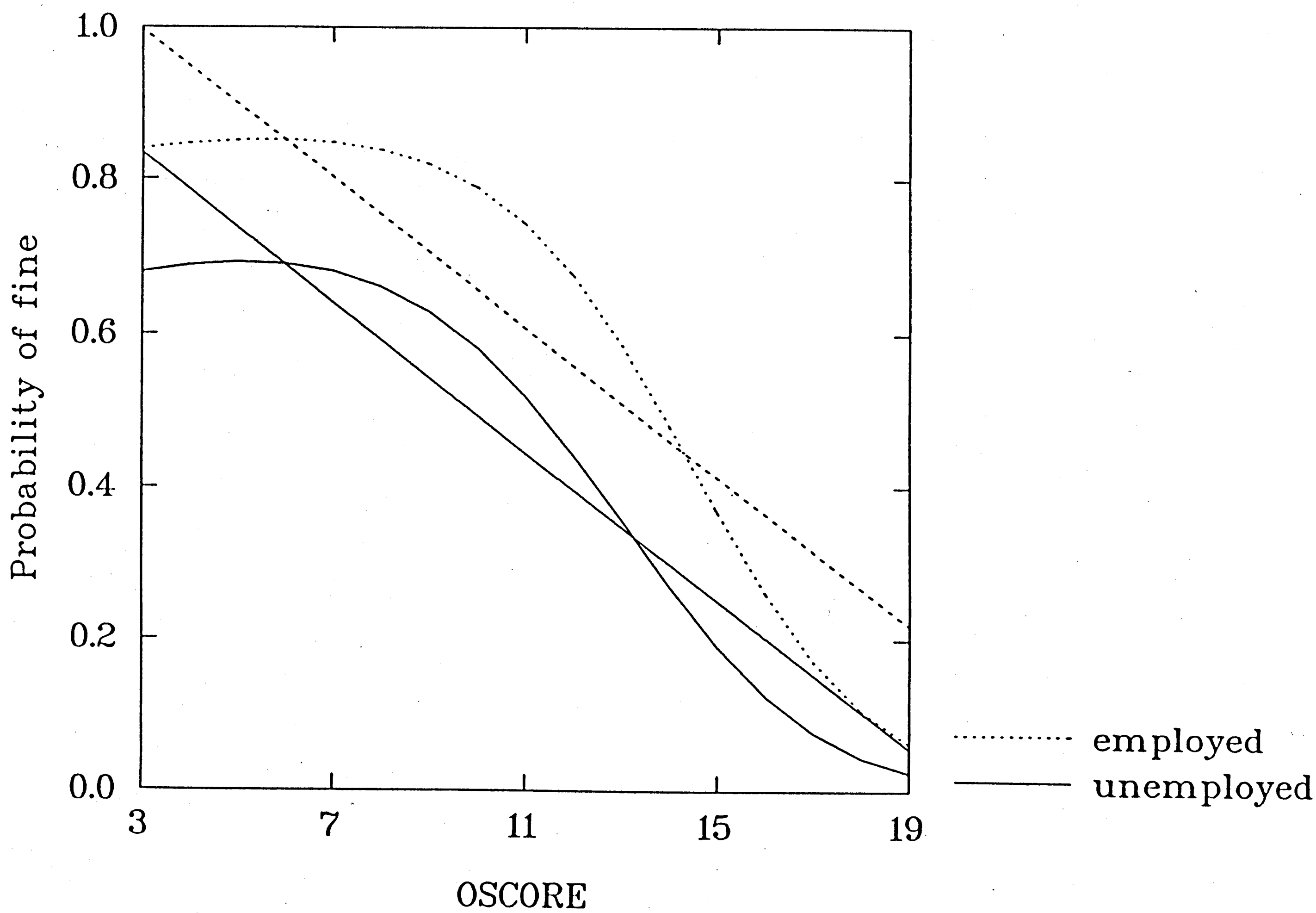


Table 6: The adjusted β_{0j} parameters and their standard errors for CML estimation and the population proportions q_j are unknown. Using the approach of Hsieh et al (1985).

Parameter	GENSTAT values		Adjusted	Adjusted s.e.	
	β^*_{0j}	s.e.(β^*_{0j})	parameter β_{0j}	Eqn (13)	Eqn (15)
β_{01}	0	0	0.3788	0.0214	0.0214
β_{02}	0.119	0.328	-0.195	0.3332	0.3343
β_{03}	-3.142	0.441	-3.456	0.4459	0.4477
β_{04}	-5.807	0.560	-6.121	0.5629	0.5646
β_{05}	-7.965	0.687	-8.279	0.6896	0.6916
β_{06}	-12.081	0.886	-12.395	0.8875	0.8891

Figure 2: Probability of Probation in court two

Curves are probabilities from multinomial logit model.
Straight lines are probabilities from linear probability model.

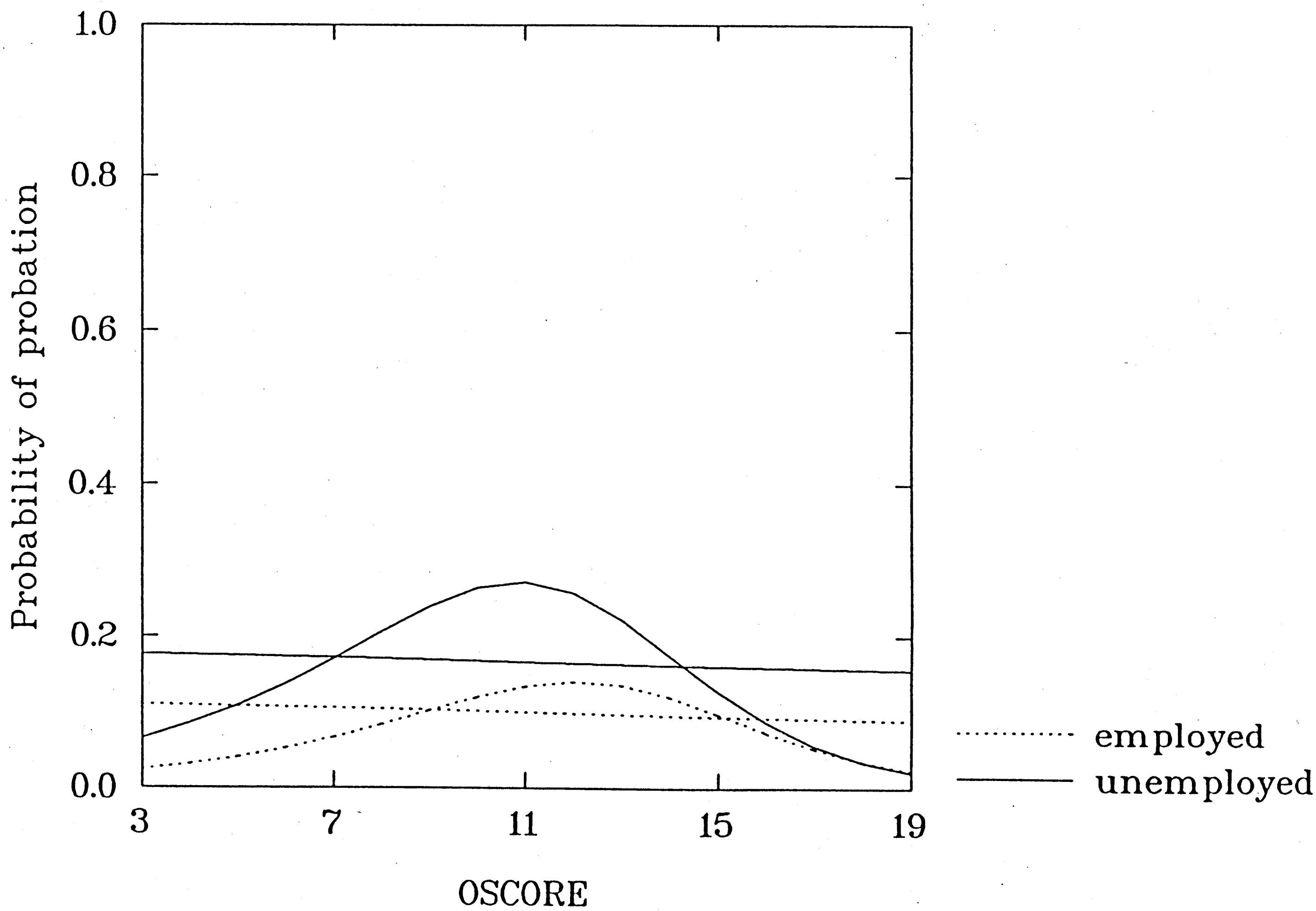


Figure 3: Probabilities for unemployed in court one

Probability predictions are from the multinomial logit model.

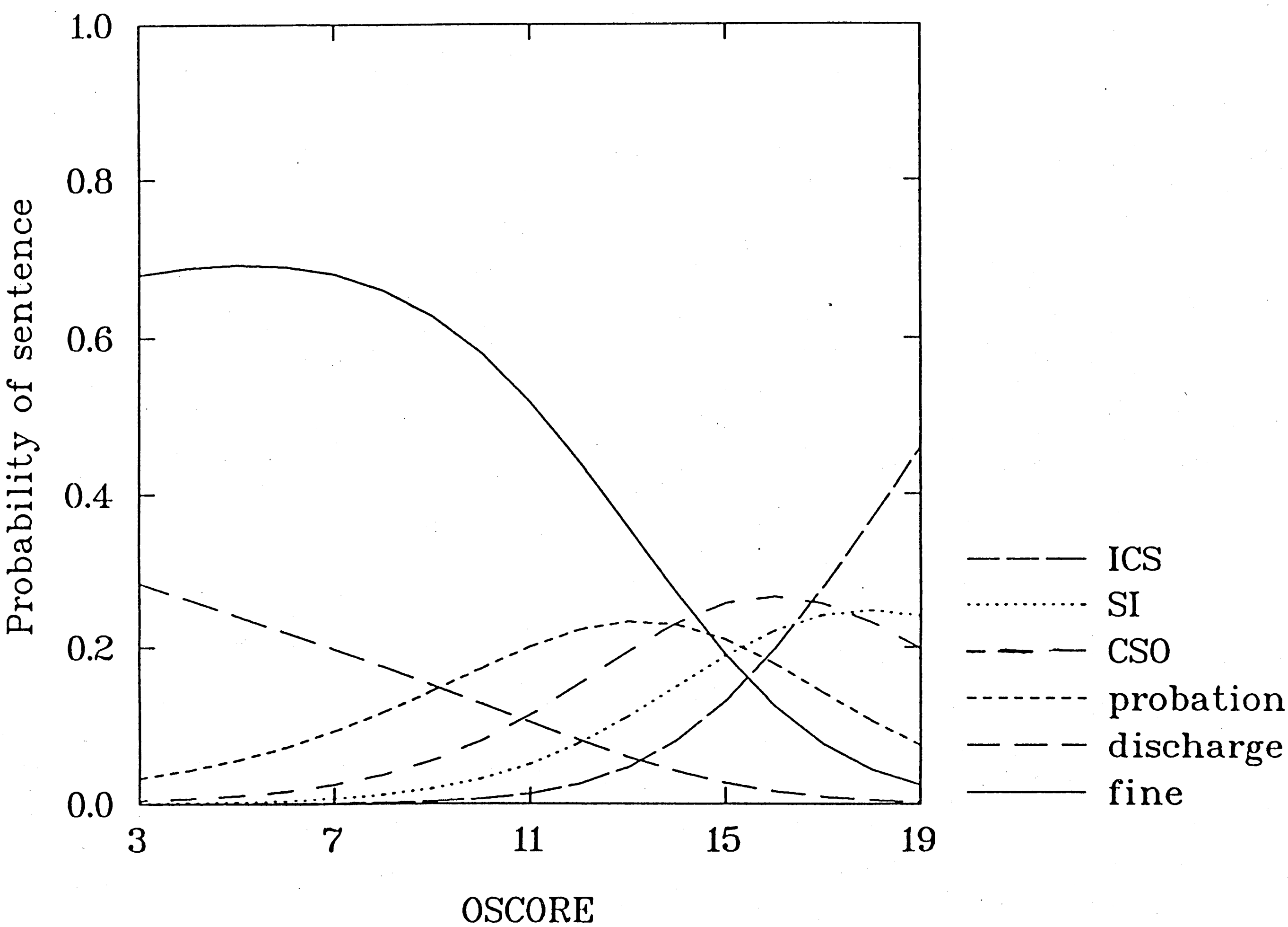
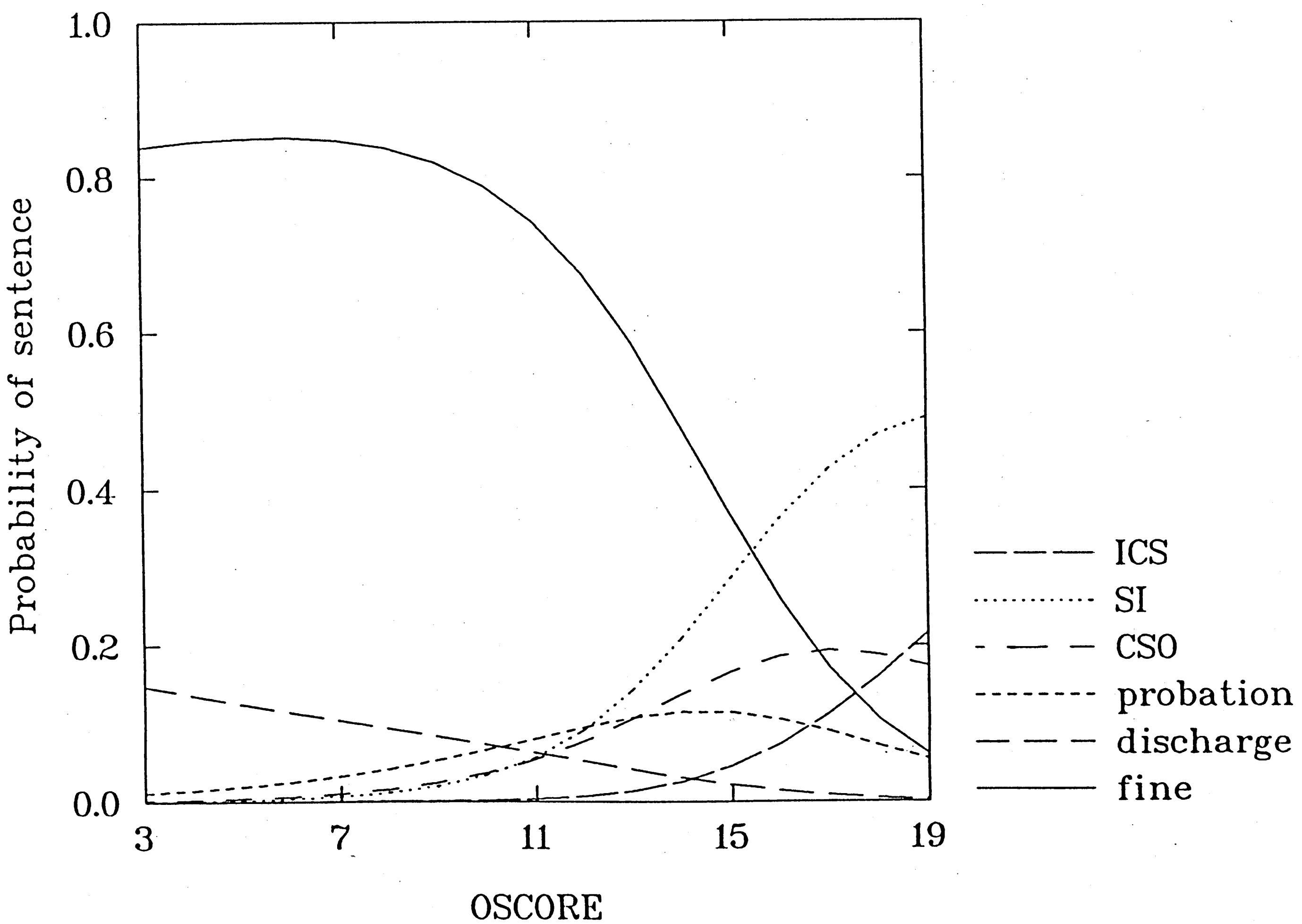


Figure 4: Probabilities for employed in court one

Probability predictions are from the multinomial logit model.



MONASH UNIVERSITY

DEPARTMENT OF ECONOMETRICS

WORKING PAPERS

1989

7/89 Russel J. Cooper, Dilip B. Madan and Keith R. McLaren, "A 'Gormanesque' Approach to the Solution of Intertemporal Consumption Models".

8/89 Keith R. McLaren, "A Variant on the Arguments for the Invariance of Estimators in a Singular System of Equations".

9/89 T.R.L. Fry, "Univariate and Multivariate Burr Distributions: A Survey".

10/89 Merran Evans, "Robustness and Size of Tests of Autocorrelation and Heteroscedasticity to Non-normality".

11/89 R.D. Snyder. "A Computerized System for Forecasting Spare Parts Sales: A Case Study".

12/89 Maxwell L. King and Ping X. Wu. "Small-Disturbance Asymptotics and the Durbin-Watson and Related Tests in the Dynamic Regression Model".

13/89 W.T.M. Dunsmuir and R.D. Snyder. "ABC Analysis in Inventory Control - The Issue of Stability".

1990

1/90 P. Burridge. "The Functional Central Limit Theorem: An Introductory Exposition with Application to Testing for Unit Roots in Economic Time Series".

2/90 Maxwell L. King and Ping X. Wu. "Locally Optimal One-Sided Tests for Multiparameter Hypotheses."

3/90 Grant H. Hillier. "On Multiple Diagnostic Procedures for the Linear Model."

4/90 Jean-Marie Dufour and Maxwell L. King. "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors."

5/90 Keith R. McLaren. "A Reappraisal of the Neoclassical Approach to Modelling Business Investment."

6/90 Francis Vella. "Non-Wage Benefits in a Simultaneous Model of Wages and Hours: Labor Supply Functions of Young Females."

7/90 Francis Vella. "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors."