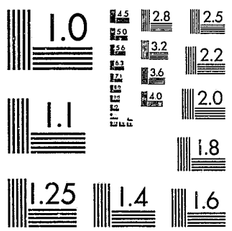

REPORT TO THE DEPARTMENT OF CORRECTION ON CONTRACT 33-045-275-12: DETERMINANTS OF CRIMINAL RECIDIVISM

Submitted By

Ann D. Witte and Peter Schmidt

September 15, 1976

TABLE OF CONTENTS

INTRODUCTION

This report describes the results of research on the determinants of criminal recidivism, which was performed under contract to the North Carolina Department of Correction. The report is composed of four chapters.

The first chapter describes an analysis of the timing of return to criminal activity, after release from prison. The second chapter analyzes the seriousness of post-release criminal activity, as measured by the length of recidivist prison sentences. The third chapter presents our analysis of various types of criminal activity; specifically, misdemeanors and felonies, and personal, property, and other crimes. Finally, chapter four contains our conclusions.

CHAPTER 1

ANALYSIS OF TIMING OF RETURN TO CRIMINAL ACTIVITY

1. INTRODUCTION

This report presents the results of an analysis of the timing of return to criminal activity after release from imprisonment. Section two of the report contains a discussion of the data set on which the analysis is based. Section three discusses the statistical methods used. Section four contains the actual analysis of the timing of the first conviction after release. Based on this analysis, we give illustrative probabilities of conviction after various lengths of time for selected types of individuals in section five. Section six contains an analysis of the length of time until return to a North Carolina prison, and section seven gives illustrative probabilities of return to prison in North Carolina after various lengths of time for selected types of individuals. Finally, section eight summarizes our findings.

2. THE NATURE OF DATA

The data to be used was collected originally in order to evaluate the North Carolina Prisoner Work Release program. The data set consists of information on a systematic random sample of 641 men who were inmates in prison units in the South Piedmont area of North Carolina in 1969 or 1971, and is unusually complete and extensive.

Due to the original purpose for which it was collected, this sample is not representative of all men in prison in North Carolina. (As noted below, it excludes men convicted of sex crimes or serious drug offenses, since they would not be eligible for work release.) However, it is fairly representative of men in medium and minimum custody prison facilities in North Carolina. These men account for approximately 66% of the North Carolina prison population in 1974.

The sample of inmates was chosen as follows. First, the population of prison inmates in the South Piedmont in 1969 and 1971 was divided into three groups: (1) a group which participated in work release in one of these years; (2) a group which was in prison in one of these years but never participated in the work release program, and (3) a group which was in prison in one of these years but participated in work release at some other time. Group 3 was dropped from consideration and Group 1 and Group 2 were considered for sampling. Before sampling the following adjustments were made. First, members of both groups who could not be followed up were eliminated. This group consisted of men who had not been released by June, 1973, men who died in prison and men who were listed as being on escape. Next all members of the non-work release group, Group 2, who were in prison for crimes which prevented their placement on work release in the 1969-1971 period were eliminated. This group consisted of all men in prison for sex crimes (D.O.C. crime codes 700, 701, 710, 711, 712) and serious drug offenses (D.O.C. crime code 804, 805, 806 resulting in a sentence of 4 years or more). In addition, all members of the non-work release population who were convicted of the public drunk offense (D.O.C. crime code 857) were eliminated as they were not in prison long inough to be processed for work release. From these adjusted groups, a random sample of 297 work releasees from Group 1 and 344 non-work releasees from Group 2 was drawn. The size of the two samples was set so that an estimate of a proportion (rate of recidivism) would not differ from the population proportion by more than 5 percent, 95 percent of the time.

Interviewing of sample members took place between the beginning of July 1973 and the end of June 1974. The project was able to locate and interview 71 percent of the total sample. A total of 453 former prison

inmates was interviewed. Of the 188 men the project did not interview, 19 percent were located but were impossible to catch up with, 7 percent were located but refused to be interviewed, and 60 percent were never located. The project was able to obtain partial information on 14 percent of the men who were not interviewed. Twenty three percent of the people whom the project was unable to locate were also wanted by some law enforcement agency. The men in the sample were followed up for an average of 37 months. The followup period ranged from 3 to 71 months.

A wide range of information on post-release activities was collected on the men that the project was able to interview. This information included, among other things, information on criminal activities, work activities, family situation, and results of psychometric attitude tests. For a copy of the questionnaire used, see Appendix B of A.D. Witte, Work Release in North Carolina: An Evaluation of Its Effects After Release from Incarceration (Raleigh: North Carolina Department of Correction, 1975).

Great pains were taken to ensure that complete post-release criminal activity records were obtained on both interviewed and non-interviewed members of the sample. Each member of the sample who was located was asked the date, location, and disposition of all arrests since his release from the sentence he was serving when chosen for the project (sample sentence). Also, in a much later portion of the questionnaire, he was asked to indicate all areas in which he had lived since his release from the sample sentence. The criminal record in all areas where a man claimed to have lived or to have been arrested were searched to determine the validity and completeness of the criminal histories elicited from each man. Unsurprisingly, many of the men in the sample claimed substantially fewer contacts with the criminal justice authorities than they actually had.

If the project could not locate a member of the sample after extensive search, an FBI check was run to determine if the man had any criminal record which had been reported to the FBI since his release from the sample sentence. If the FBI check indicated a record, the reporting agency (police department, court, etc.) was written requesting a complete criminal record and a current address.

Information obtained other than criminal history was checked only when the information appeared to be inconsistent. For instance, if a man claimed a substantial increase or decrease in salary over previous jobs, his employer was contacted and asked for verification of the man's wage claim.

If one compares the subjects the project staff was able to interview with those it was unable to interview, one finds significant differences. This comparison was made on those variables supplied by the Department of Correction and on wage and previous conviction information obtained by the project staff from the subjects' records in the Department of Correction. The subjects who the staff were unable to interview had significantly[1] shorter sentences for the incarceration during which they were sampled (28 months vs. 40 months), were more likely to be misdemeanants, had fewer rule violations, and were more likely to have been unsupervised when released (not on parole or conditional release) than were the subjects the staff were able to interview. The difference between those the staff was unable to interview and those it was able to interview in age at release (33 vs. 31), tested educational level (6.2 vs. 5.5) and convictions before the incarceration during which they were sampled (4 vs. 3) approached

[1]The term statistically significant indicates that a finding of statistical significance would have resulted if a two-tailed test at the .05 level ($\alpha = .05$) or a one-tail test at .025 level were used.

statistical significance.[2] There was no significant difference between the uninterviewed and interviewed subjects in IQ, grade claimed, marital status, race, type of crime (crimes against property, persons, etc.) for which they were incarcerated when sampled, or wage and time sentenced before their incarceration for the offense during which they were sampled. Overall, the information obtained during the interview tends to underrepresent the old, habitual, misdemeanant offender. Part of the reason for the underrepresentation of this group may be their death. The project staff obtained death records whenever possible but it is quite likely that a number of such records were not found.

3. THE STATISTICAL METHODOLOGY USED

The nature of the independent variables analyzed in this report, length of time until first conviction (LTFCV) or first prison conviction (LTFPCV), calls for some care in the statistical analysis. In the first place, the variables are non-negative, by definition. In the second place, most individuals will have values of either variable which are small or moderate, as evidenced by a median value of 8 months for LTFCV. However, there is a reasonable probability of rather large values, as some men will not be convicted for some time, or may never be convicted again in their lifetime. In statistical terminology, the distribution is "positively skewed."

These facts essentially rule out use of the normal distribution and techniques based on it (e.g. multiple regression), since a normal distribution can be negative, and is not skewed. However, a lognormal distribution does meet the two conditions mentioned, and is a reasonable candidate for fitting to LTFCV and LTFPCV.

[2]The term approached statistical significance indicates that a finding of statistical significance would have resulted if a two-tailed test at the .10 level ($\alpha = .10$) or a one-tailed test at the .05 level were used.

A further difficulty is that we can not observe values of LTFCV and LTFPCV in excess of the length of the followup period. (This varied from individual to individual, averaging 37 months.) In statistical terminology, the distribution has been "truncated" at the end of the followup period. This dictates use of a truncated lognormal distribution. The truncated lognormal distribution has recently been used in the social sciences in similar contexts--e.g., to predict duration of welfare dependency.[3] A mathematical description of the truncated lognormal technique is given in Appendix A.

The chief advantage of using a technique which takes the truncation into account is that it makes it possible to use all observations, even those on men not convicted during the followup period. Using alternative techniques (such as ignoring those not convicted, or setting LTFCV or LTFPCV equal to the length of the followup period for those not convicted) would tend to give results which are overly pessimistic--that is, biased.

The shape of the lognormal distribution largely determines the estimates of the probability of return shortly after release, say for the first four months, and also for the time period beyond the longest followup, 71 months. The estimates presented below are most reliable for the period 4 to 37 months (length of the average followup period). The estimates for the period 37 to 71 months after release can be considered fairly reliable since the activities of a number of individuals were observed in this period. Since estimates of activity prior to 4 months after release and after 71

---

[3] See T. Amemiya and M. Boskin, "Regression Analysis when the Dependent Variable is Truncated Lognormal, with an Application to the Duration of Welfare Dependency," International Economic Review, 15 (June, 1974), pp. 485-496.

months after release are largely dictated by the distribution assumed, estimates for these periods should be treated with considerable caution.

4. ANALYSIS OF THE LENGTH OF TIME UNTIL FIRST CONVICTION

The dependent variable used in the analysis of this section is LTFCV, defined as the length of time (in months) from release until the first arrest which ultimately results in a conviction. To this variable we fit a (truncated) lognormal distribution.

The mean of the distribution of the dependent variable is a linear function of the explanatory variables. The explanatory variables which were found to be significantly related to LTFCV are as follows.

A constant term, denoted CNST.

The number of rule violations during the prison sentence preceding release, denoted RULE.

The number of convictions prior to the one leading to the sample sentence, denoted CONVBS.

Age (in months) at release, denoted AAR.

A dummy variable equal to one if the individual's record indicates a serious problem with alcohol, or use of hard drugs, and equal to zero otherwise; denoted ALKY.

A dummy variable equal to one for whites, and equal to zero for non-whites; denoted RACE.

A dummy variable equal to one if the release from the sample sentence was supervised, and equal to zero if it was not; denoted SUPER.

The results for the specification including these explanatory variables are given in Table 1. For each variable, the first number given is its coefficient, which represents the partial effect of the variable on the

expected value (mean) of LTFCV. For example, the coefficient of RULE is -.7956, which indicates that, holding the other explanatory variables constant, an additional rule violation reduces an individual's mean value of LTFCV by .7956 months. Similarly, the coefficient of ALKY is -19.349, which indicates that an alcoholic will have a mean value of LTFCV which is 19.349 months lower than the mean value for a non-alcoholic of otherwise identical characteristics.

The second number given is the coefficient multiplied by $e^{-1/2\ \sigma^2}$. The point is that the median of a lognormal distribution equals the mean times $e^{-1/2\ \sigma^2}$. Therefore these "corrected" coefficients give the partial effect of each variable on the median of LTFCV. For a skewed distribution, the median is probably a better measure than the mean of what one intuitively thinks of as the "average" value of the distribution. If one wanted to make a statement like, "An alcoholic will, on the average, be convicted ____ months sooner than a non-alcoholic of otherwise similar characteristics," the number in the blank should probably be 7.2335 (the partial effect of ALKY on the median of LTFCV) rather than 19.349 (the partial effect of ALKY on the mean of LTFCV).

The final number given is the "t ratio" for each variable. These "t ratios" are distributed as N(0,1) in large samples, under the null hypothesis that the coefficient is zero. The significance of a variable is therefore determined by comparing the "t ratio" to the critical point of the N(0,1) distribution; a value in excess of 1.645 in absolute value indicates significance at the .05 level (for a two-tailed test).

The variables listed above are all statistically significant at the .05 level. They indicate that an individual will tend to have a longer time until conviction, the fewer rule violations he had, the fewer previous convictions he has had, the older he is, if he is not an alcoholic or a user of hard drugs, if he is non-white, and if he was supervised when released from prison. Conversely, an individual is likely to be reconvicted sooner the more rule violations he had, the more previous convictions he had, the younger he is, if he is an alcoholic or a user of hard drugs, if he is white, and if he was not supervised when released from prison.

The results of Table 1 constitute the basic results of our analysis of LTFCV. In the process of arriving at this final specification, a number of other explanatory variables were tried, and were dropped due to having coefficients which were not statistically significant. A list of these other variables follows, along with the "t ratios" of their estimated coefficients.

Number of years of schooling completed; "t ratio" = 0.68.

A dummy variable equal to one if the sample conviction was for a felony, and equal to zero if for a misdemeanor; "t ratio" = 1.36.

A dummy variable equal to one if the individual was married at the time of his release from imprisonment, and equal to zero if he was not; "t ratio" = 0.60.

A dummy variable equal to one if the individual participated in the work release program, and equal to zero if he did not; "t ratio" = -0.53.

The squared value of AAR; "t ratio" = 0.004.

A dummy variable equal to one if the sample conviction was for a crime against property, and equal to zero otherwise; "t ratio" = -0.35.

A dummy variable equal to one if the sample conviction was for a crime against a person, and equal to zero otherwise; "t ratio" = 0.98.

We also tried creating one dummy variable (ALKY1) which equalled one for individuals with a serious drinking problem, and which equalled zero otherwise; and a second dummy variable (ALKY2) which equalled one for individuals with a history of hard drug use, and equalled zero otherwise. The difference between the coefficients of ALKY1 and ALKY2 was not statistically significant; "t ratio" = 1.13.

Finally, because of the highly significant effect of RACE, we tried to find interactions between RACE and the other explanatory variables in the final specification (RULE, CONVBS, AAR, ALKY, and SUPER) by estimating separate equations for the whites and non-whites in the sample. None of the differences in the coefficients for whites and non-whites were statistically significant at the .05 level. The "t ratios" for the differences (white minus non-white) in the coefficients of RULE, CONVBS, AAR, ALKY and SUPER were respectively 0.78, 0.59, -0.22, 1.30, and -1.17.

5. PROBABILITIES OF CONVICTION AFTER VARIOUS LENGTHS OF TIME

The results of the previous section indicate positive and significant effect on LTFCV of AAR and SUPER, and a negative and significant effect on LTFCV of RULE, CONVBS, ALKY and RACE. The type of individual who is most likely to be reconvicted soon after release is a young, white alcoholic (hard drug user) with a lot of rule violations and previous convictions, and who is unsupervised when released. Conversely, an old, black non-alcoholic and non-hard drug user with no rule violations or previous convictions, and who is supervised when released, is the type of individual who is least likely to be reconvicted soon after release.

To see more precisely what this means, we list in Table 2 the frequency of the lognormal distributions implied by our results, for the above two types of individuals, as well as for the "average" individual (i.e.,

characteristics equal to the mean values of the explanatory variables in our sample). These probabilities are based on the parameter estimates given in Table 1.

Case 1 evaluates the probability of return after numbers of months from 1 to 100 for the "average" individual. The "average" individual is defined as the individual who has characteristics (values of the explanatory variable) equal to the sample means; i.e. RULE = .672, CONVBS = 2.735, AAR = 380.107, ALKY = .481, RACE = .503, and SUPER = .316. This case corresponds to an "average" set of probabilities. The mean time until conviction for this group is 37.05 months. Due to the marked skewness of the distribution used, the median is a better general indication of the "average" time until return, and is 13.85 months. This means that one would expect approximately half of the individuals with "average" characteristics to be reconvicted within 14 months. Indeed more than 25 percent are convicted by the end of six months. Before looking at the probabilities reported for individual months, it is well to repeat the limitations discussed above. One should view with caution the probabilities predicted for the first four months and for the period beyond the longest followup 71 months, since these probabilities are largely dictated by the distribution used. One can have greatest confidence in the prediction for the middle portion of the remaining period, say 12 to 48 months. Looking at the individual probabilities reported in this table, we see that 72 percent return only after more than 6 months have passed; that is, 28 percent have returned within six months. Forty six percent return within one year; that is, 54 percent return only after more than 12 months. The probabilities are certainly not encouraging and if we look further we see that by 45 months, 80 percent have been reconvicted.

The two other cases presented in Table 2 represent individuals with the most favorable characteristics (case 2) and least favorable characteristics (case 3). The most favorable case, case 2, corresponds to a non-white with no rule violations and no previous convictions, who is 50 years of age when released, who has no serious problem with alcohol or hard drug use, and who is supervised when released. Comparing the probabilities for an individual with these characteristics to the "average" individual reported as case 1, the median time until reconviction is approximately doubled (27.04 vs. 13.85 months). The proportion returning within six months drops from 28 percent to 14 percent and the proportion returning within 45 months drops from 80 to 64 percent.

The least favorable case, case 3, corresponds to a white with three rule violations and five previous convictions, who is 18 years of age when released, who has a serious alcohol problem or uses hard drugs, and who is unsupervised when released. If the two previous cases have not been encouraging, this case is downright discouraging. The median time until reconviction is only 1.98 months. Eighty eight percent of these individuals return within six months and 99 percent within forty five months.

6. ANALYSIS OF THE LENGTH OF TIME UNTIL FIRST PRISON CONVICTION

The dependent variable used in the analysis discussed in this section is the length of time (in months) from release until an arrest which ultimately results in a sentence in North Carolina of 15 days or more (LTFPCV). This dependent variable is more restrictive than the one discussed above because it deals only with convictions in North Carolina which result in a sentence of 15 days or more while the previous dependent variable

dealt with any conviction anywhere.[4] To this dependent variable, we fit a truncated lognormal distribution. The distribution of the dependent variable was assumed initially to be a linear function of the explanatory variables from the final specification of section 4. The only change made is that previous prison convictions (denoted CONPBS) was used instead of all previous convictions (CONVBS). That is, the mean of LTFPCV was assumed to be a linear function of a constant term (CNST); the number of rule violations during the prison sentence preceding release (RULE); the number of prison convictions prior to the one leading to the sample prison sentence (CONPBS); age (in months) at release (AAR); a dummy variable equal to one if the individual's record indicates a serious problem with alcohol or hard drug use, and zero otherwise (ALKY); a dummy variable equal to one for whites, and zero for non-whites (RACE); and a dummy variable equal to one if an individual was supervised when released and zero if he was unsupervised (SUPER). However, when this specification was estimated the coefficient of RULE was very insignificant ("t ratio" = -0.03). As a result we dropped RULE from our list of explanatory variables to get a "reduced" specification in which the mean of LTFPCV depends on CNST, CONPBS, AAR, ALKY, RACE and SUPER. The results for this specification are given in Table 3, which is of roughly the same form as Table 1.

Note that the coefficients reported here are much larger than the coefficients in Table 1, which is just a reflection of the fact that the length of time until the first prison conviction ought to be larger than

---

[4] In the sample of 641, 209 received a conviction which resulted in a sentence of 15 days or more. Of these 209 individuals, 22 received such sentences outside North Carolina and were not considered to be reconvicted according to the definition used in this section.

the length of time until the first conviction of any type. However, if we look at the coefficients multiplied by $e^{-1/2\ \sigma^2}$, which give the partial effects of the explanatory variables on the median (rather than the mean) of LTFPCV, the magnitudes are more reasonable. As argued previously in section 4, it is the median which corresponds more closely to the intuitive concept of the "average" value of the variable in question; as a result, the partial effects on the median of LTFPCV are more easily interpreted than the partial effects on the mean of LTFPCV.

We note that the signs of the coefficients of AAR, ALKY, RACE, and SUPER are the same whether the dependent variable is LTFPCV (Table 3) or LTFCV (Table 1). Also the sign of CONPBS in Table 3 is the same as the sign of CONVBS in Table 1. So, similarly to the conclusion of section 4, an individual is apt to return to prison in North Carolina more quickly the more previous prison convictions he has had, the younger he is at release, if he is an alcoholic or hard drug user, if he is white, and if his release was unsupervised. Conversely, he is less likely to return to prison in North Carolina the less previous prison convictions he has had, the older he is at release, if he is not an alcoholic or hard drug user, if he is black, and if his release was supervised.

As a final note, we can see that all variables which appear in Table 1 and in Table 3 have smaller "t ratios" in Table 3; they are statistically less significant. (CONPBS has a higher "t ratio" in Table 3 than CONVBS in Table 1, but it is a somewhat different variable.) In fact RACE and SUPER are not quite significant at the .05 level, though they are close to being so. They were left in to prevent the specification from being too small to be useful.

7. PROBABILITIES OF RETURN TO PRISON IN NORTH CAROLINA AFTER VARIOUS LENGTHS OF TIME

In analyzing the length of time until the first arrest which results in a conviction in North Carolina with a sentence of 15 days or more (LTFPCV), the results of the previous section indicate a clearly significant positive effect for AAR and negative effect for ALKY, and CONPBS. The results also indicate a marginally significant (significant at the 10 percent level with a one-tailed test) positive effect for SUPER and negative effect for RACE. This means that the type of individual most likely to return to the North Carolina prison system is a young, white, alcoholic or hard drug user with a large number of previous prison victions and who is unsupervised when released. The type of individual least likely to return is an old, non-white non-addict with no previous prison convictions and who is supervised when released.

To see more precisely what this means and to give an example of the type of output the Department of Correction can expect from the computer program supplied, we list in Table 4 the probabilities implied by our estimated lognormal distribution for the two above types of individuals and for the "average" individual (an individual with characteristics equal to the mean for the sample). The probabilities reported in this list are based on the parameter estimates given in Table 3.

Case 1 evaluates the probability of return after various lengths of time for the "average" individual. The "average" individual is defined as the individual who has characteristics (values of the explanatory variable) equal to the sample mean; i.e. CONPBS = 2.186, AAR = 380.107, ALKY = .481, RACE = .503, SUPER = .316. This case corresponds to an "average" set of probabilities. The mean time until return to the North Carolina prison system for this group is 540.79 months and the median is 103.70 months.

One would expect about half of the individuals with "average" characteristics to return to the North Carolina prison system within 8.6 years and approximately a quarter to return within 2.5 years. Looking at the probabilities for which we have reasonable confidence (probabilities for 4 to 71 months), we see that 12 percent of these "average individuals" return within one year and 32 percent within 45 months.

The two other cases presented in Table 4 represent individuals with the most favorable characteristics (case 2) and least favorable characteristics (case 3). The most favorable case corresponds to a non-white with no previous prison convictions, whose age is 50 at release, with no history of a serious alcohol problem or hard drug use, and who is supervised when released. Comparing the probabilities for individuals with these characteristics to the "average" individual reported as case 1, we see that the median time until return to a North Carolina prison is more than doubled (103.70 vs. 213.57 months). The proportion returning within a year drops from 12 to 6 percent and the proportion returning within 45 months drops from 32 to 20 percent.

The least favorable case, case 3, corresponds to a white with four previous prison convictions, who is 18 years of age when released, with a history of serious alcohol abuse or hard drug use, and who is unsupervised when released. The median time until return for individuals with these characteristics is only 84.18 months. Forty four percent of these individuals return within a year and 71 percent within 45 months.

8. SUMMARY

This report contains the results of our analysis of the determinants of the length of time until first conviction (LTFCV) and the length of time until receipt of a sentence of fifteen days or more in North Carolina

(LTFPCV). We investigated the determinants of these variables using the truncated lognormal distribution.

We found the major determinants of the timing of first conviction to be the number of rule violations during the prison sentence preceding release (sample sentence), the number of convictions prior to the sample sentence; age (in months) at release, a serious problem with alcohol or hard drug use, race, and whether or not an individual is supervised when released. A man, under this model, will be reconvicted less rapidly after release if he is non-white, has no problems with drugs or alcohol, is old, has no rule violations during his sample sentence or convictions prior to it, and is supervised when released.

The major determinants of the timing of return to the North Carolina prison system were found to be the number of prison convictions prior to the sample sentence, age (in months) at release, a serious problem with alcohol or hard drug use, race, and whether or not an individual is supervised when released. A man, under this model, will return to the North Carolina prison system less rapidly after release if he is non-white, has no problems with drugs or alcohol, is old, has no prison conviction prior to his sample sentence, and is supervised when released.

For both dependent variables (LTFCV and LTFPCV) we developed and presented the lognormal distribution implied by our estimates for three types of individuals. The "average" set of probabilities (all explanatory variables of their sample means) indicates that 46 percent of the men released from North Carolina medium and minimum custody prison units will be reconvicted within a year and that 12 percent of these men will return to the North Carolina prison system within a year. By the end of three years, approximately 38 percent of these men have returned to the North Carolina prison system.

Table 1

Results of Fitting Truncated Lognormal Distribution to LTFCV

Variable	Coefficient, representing $\frac{\partial E(LTFCV)}{\partial X}$	Coefficient$\cdot e^{-1/2\ \sigma^2}$, representing $\frac{\partial Median(LTFCV)}{\partial X}$	"t ratio"
CNST	32.061	11.984	4.73
RULE	-.7956	-.2974	-1.81
CONVBS	-.5062	-.1892	-4.69
AAR	.05489	.02052	3.36
ALKY	-19.349	-7.2325	-4.38
RACE	-13.851	-5.1774	-3.23
SUPER	7.3483	2.7468	1.66

Value of $\hat{\sigma}^2$ = 1.9681

Table 2

Predicted Probabilities P(LTFCV > N)

N	Case 1	Case 2	Case 3
1	.969	.991	.687
2	.916	.968	.497
3	.862	.941	.383
4	.812	.913	.308
5	.766	.886	.254
6	.724	.858	.215
7	.687	.832	.184
8	.652	.807	.160
9	.621	.784	.140
10	.592	.761	.124
11	.565	.739	.111
12	.541	.719	.099
14	.497	.681	.082
16	.459	.646	.068
18	.426	.614	.058
20	.397	.585	.050
22	.371	.558	.043
24	.348	.534	.038
26	.326	.511	.033
28	.308	.490	.029
30	.291	.471	.026
32	.275	.452	.024
34	.261	.435	.021
36	.248	.419	.019
38	.236	.404	.018
40	.225	.390	.016
45	.200	.358	.013
50	.180	.331	.011
55	.163	.306	.009
60	.148	.285	.008
65	.135	.266	.006
70	.124	.249	.006
75	.114	.234	.005
80	.106	.220	.004
85	.098	.207	.004
90	.091	.196	.003
95	.085	.185	.003
100	.079	.176	.003
Mean	37.047	72.343	5.293
Median	13.848	27.042	1.979
Mode	1.935	3.778	.276

Table 3

Results of Fitting Truncated Lognormal Distribution to LTFPCV

Variable	Coefficient, representing $\frac{\partial E(LTFPCV)}{\partial X}$	Coefficient$\cdot e^{-1/2\ \sigma^2}$, representing $\frac{\partial Median(LTFPCV)}{\partial X}$	"t ratio"
CNST	394.44	75.633	1.94
CONPBS	-9.2656	-1.7767	-2.96
AAR	.8471	.1624	2.27
ALKY	-337.23	-64.663	-2.12
RACE	-118.93	-22.804	-1.56
SUPER	211.08	40.475	1.52

Value of $\hat{\sigma}^2$ = 3.3031

Table 4

Predicted Probabilities P(LTFPCV > N)

N	Case 1	Case 2	Case 3
1	.995	.998	.937
2	.985	.995	.875
3	.974	.991	.823
4	.963	.986	.779
5	.952	.981	.740
6	.942	.975	.707
7	.931	.970	.677
8	.921	.965	.650
9	.911	.959	.626
10	.901	.954	.604
11	.891	.949	.584
12	.882	.943	.565
14	.865	.933	.531
16	.848	.923	.502
18	.832	.913	.476
20	.817	.904	.453
22	.803	.894	.432
24	.790	.885	.414
26	.777	.877	.397
28	.764	.868	.381
30	.753	.860	.367
32	.741	.852	.353
34	.730	.844	.341
36	.720	.836	.329
38	.710	.829	.319
40	.700	.822	.309
45	.677	.804	.286
50	.656	.788	.267
55	.636	.772	.250
60	.618	.756	.235
65	.601	.744	.222
70	.586	.730	.210
75	.571	.718	.199
80	.557	.705	.189
85	.544	.694	.180
90	.531	.683	.172
95	.519	.672	.165
100	.508	.662	.158
Mean	540.791	1113.778	84.185
Median	103.697	213.568	16.143
Mode	3.813	7.853	.594

APPENDIX A

Consider a set of independently distributed variables Y_t^*, t=1,2,...,T. The variable Y_t^* has a lognormal distribution with mean $X_t\beta$ and variance $\eta^2(X_t\beta)^2$, where X_t is a row vector of explanatory variables, and β is a vector of parameters to be estimated. It follows that $\log Y_t^*$ is normally distributed with mean $\log (X_t\beta) - 1/2\ \sigma^2$ and variance σ^2, where $\sigma^2 \equiv \log(1+\eta^2)$.

We further suppose that we observe not Y_t^* but Y_t, defined by

$$(1) \qquad Y_t = \begin{cases} Y_t^* \text{ if } Y_t^* \leq \alpha_t \\ \alpha_t \text{ if } Y_t^* > \alpha_t \end{cases},$$

where α_t is known. (In the present context it is the length of the followup period for the t^{th} individual.) We can then write the likelihood function of the sample as

$$(2) \qquad L = \prod_{t\varepsilon\theta_1} f(\log Y_t) \prod_{t\varepsilon\theta_2} P(\log Y_t^* > \log \alpha_t)$$

where

$$\theta_1 = \{t | Y_t^* \leq \alpha_t\}$$

$$\theta_2 = \{t | Y_t^* > \alpha_t\}$$

$$f(\log Y_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\log Y_t - \log X_t\beta + 1/2\ \sigma^2)^2\right]$$

$$P(\log Y_t^* > \log \alpha_t) = F\left[-\log \alpha_t + \log X_t\beta - 1/2\ \sigma^2)/\sigma^2\right],$$

F being the cumulative distribution function of the standard normal distribution. This likelihood function, or its logarithm, can be maximized with respect to β and σ^2, by computer using a numerical maximization routine.

Asymptotically valid tests of hypotheses can be constructed using the likelihood ratio principle. Alternatively, Amemiya and Boskin (1974) give an expression for the information matrix. The "asymptotic t ratio," formed by dividing an estimate by the square root of the appropriate diagonal element of the information matrix, converges in distribution to N(0,1) under the null hypothesis that the parameter is zero.

APPENDIX B

DESCRIPTION OF COMPUTER PROGRAM FOR PREDICTION

OF LENGTH OF TIME UNTIL FIRST PRISON CONVICTION

1. INTRODUCTION

The purpose of the program is to generate, for individuals with certain specified characteristics, a predicted frequency of the length of time until first prison conviction, denoted LTFPCV. This variable is more precisely defined in Section 2 below. The specified characteristics consist of values for five variables: CONPBS, AAR, ALKY, RACE, and SUPER. These variables are also defined in Section 2 below, and are chosen because they have been found, in the statistical analysis performed under this contract, to be significantly related to LTFPCV.

The frequencies are generated from a lognormal distribution, whose mean is a weighted sum of the five explanatory variables. The weights were estimated in the statistical analysis already referred to, by fitting a lognormal distribution to information on a sample of 582 former inmates. For more detail on the procedure, rationale for use of the lognormal distribution, etc., see Section 3 of Chapter 1.

2. DESCRIPTION OF VARIABLES

The variable whose frequency is being predicted, LTFPCV, is defined as the length of time after release until the first arrest which leads to a conviction and a sentence in North Carolina of 15 days or more. LTFPCV is measured in months.

CONPBS is the number of previous convictions resulting in a jail or prison sentence. It does not include the prison conviction just prior to the period of the follow-up. (For example, it equals zero, not one, for someone whose prison conviction prior to the follow-up was his first.) This definition corresponds to the current Department of Correction's definition for "prior conviction;" i.e., a conviction resulting in a sentence of one day or more.

AAR is chronological age at the time of release from the imprisonment after which the inmate's activities were followed. It is also measured in months.

ALKY is a dummy variable which is essentially set equal to one for individuals with a serious drinking problem and/or a history of hard drug use, and zero for other individuals. This is equivalent to Department of Correction's (D.O.C.) code 4 or 5 under "drinking habits" or code 6, 7 or 8 under "drug type."

RACE is a dummy variable which is set equal to one for whites, and equal to zero for non-whites.

SUPER is a dummy variable which is equal to one if the release from the imprisonment before follow-up was supervised, and equal to zero if it was not supervised.

3. INPUTS INTO AND OUTPUTS FROM THE PROGRAM

The only input which must be provided to the program is the set of values of CONPBS, AAR, ALKY, RACE and SUPER for the individual (or group) for which the predicted frequency of LTFPCV is to be generated. The program attempts to read these variables off an input data card, with FORMAT 5F10.0 . This means that the value of CONPBS must be punched in columns 1-10; AAR in columns 11-20; ALKY in columns 21-30; RACE in columns 31-40; and SUPER in columns 41-50. As long as an explicit decimal point is punched (e.g., punch 1. or 1.0, not just 1), it does not matter where in each 10-column field a value is punched.

If the value of CONPBS, AAR, ALKY, RACE and/or SUPER is not available, or if it is desired not to take into account the effects of one or more of these variables, it is necessary to punch <u>any negative</u> value (e.g., -1.0) for

the variable in question. The program will then automatically assign to the individual the average value (over our 582 observations) of that variable. (The average values of CONPBS, AAR, ALKY, RACE and SUPER are 2.166, 380.1, 0.481, 0.503, and 0.316, respectively.)

The first output provided is the set of values read for CONPBS, AAR, ALKY, RACE, and SUPER. When an average value has been substituted for any variable, as described above, this is so indicated.

The second output provided is the mean, median and mode of the implied lognormal distribution. Because the lognormal distribution is highly skewed, the median value is probably the best measure of what one intuitively thinks of as the "average" value of LTFPCV for that individual. The mean will typically be much larger, and the mode much smaller.

The last output, and the one of major interest, is the frequency of LTFPCV. For all numbers of months between 1 and 100, it displays the probability (frequency) that LTFPCV will be at least as large as that number of months.

4. CAUTIONARY NOTES

There are a few words of caution worth stressing. These have to do not with the program itself, but with the interpretation of the results.

1. The sample used in the statistical analysis upon which these projections are based is described in some detail in the report of the statistical analysis. Briefly, it is essentially a random sample of inmates who had not been convicted of crimes that would have prevented their placement on the work release program and who served time in minimum and medium custody prison units in the South Piedmont administrative area. It therefore did not contain individuals convicted of sex offenses, (D.O.C. crime codes

700, 701, 710, 711, 712) serious drug offenses (D.O.C. crime codes 804, 805, 806 resulting in a sentence of 4 years or more), or the public drunk offense (D.O.C. crime code 857). The sample is not representative of inmates who never served time in medium and minimum custody prison units but are confined throughout their imprisonment to the institutions of the prison system, i.e., the specialized youthful offenders and maximum and close custody institutions. The sample also contained no women. As a result, the projections made here should not be applied to women, individuals who do not serve time in medium and minimum custody unit, or individuals who have been convicted of the above mentioned offenses.

2. Strictly speaking, the program gives the probability distribution of LTFPCV for an individual with certain specified characteristics. It can of course be used to give the expected frequency of LTFPCV for a group of individuals, by using the average values of CONPBS, AAR, ALKY, RACE and SUPER for that group. However, its use in this way involves an approximation. (For a group of individuals, the average value of P(LTFPCV > N) is only approximately the same as the value of P(LTFPCV > N) for the "average" member of the group.) This approximation should be fairly accurate for groups of individuals whose values of the explanatory variables are similar, but it may be poor for groups of individuals with very different values of the explanatory variables.

3. It should be remembered that in the sample used in the statistical analysis, the average follow-up time was 37 months, and the maximum was 71 months. Although we give projections up to 100 months, those past 71 months should be interpreted with extreme care. It might well be possible to pick an alternative distribution (other than lognormal) that would fit the observations up to 71 months in a similar fashion, but would imply vastly different frequencies for points further in the future. Similarly, the

density (frequency) of a lognormal distribution has its mode at some point greater than zero. Our projections therefore must exhibit this fact, and this also could just be an artifact due to the particular distribution used. The projections for the first few months should therefore also be interpreted with extreme care. The frequencies from, say, 4 to 71 months should be fairly trustworthy; we can have the most confidence in those toward the middle of this range, and less confidence in those toward either end.

CHAPTER 2

ANALYSIS OF SERIOUSNESS OF CRIMINAL ACTIVITY

1. INTRODUCTION

This report presents the results of an analysis of the seriousness of criminal activity after release from imprisonment. Our measure of seriousness of criminal activity is the total length of all prison sentences received by an individual during a specific time period following his release from prison.

Section two of the report gives a brief discussion of the data set on which the analysis is based. Section three discusses the statistical methodology employed. Section four contains the actual analysis of the seriousness of recidivist criminal activity. Based on this analysis, section five contains predicted frequencies of various levels of seriousness for selected types of individuals. Section six concludes.

2. THE DATA

The data is described in some detail in our report, "Analysis of Timing of Return to Criminal Activity," previously submitted. We will therefore give here only a brief overview of a few relevant aspects of the data set.

The data set consists of information on a random sample of 641 men who were inmates in prison units in the South Piedmont area of North Carolina in 1969 or 1971. As a result it is not representative of all men in prison in North Carolina; however, it is fairly representative of all men in medium and minimum custody prison facilities in North Carolina.

Since the original purpose of the data set was to evaluate the North Carolina Prisoner Work Release program, it does not include individuals who were imprisoned for offenses which would prevent their placement on the work release program. This group consists of men in prison for sex

crimes (D.O.C. crime codes 700, 701, 710, 711, 712) or serious drug offenses (D.O.C. crime codes 804, 805, 806 resulting in a sentence of 4 years or more). In addition it does not include men convicted of the public drunk offense (D.O.C. crime code 857) since they were not in prison long enough to be processed for work release. The group also contains no women since there were no women in South Piedmont area prisons in 1969 and 1971.

3. THE STATISTICAL METHODOLOGY

The variable used to measure the seriousness of recidivist criminal activity is the total length (in months) of all N.C. prison sentences received by an individual during his followup period, and is denoted TTSENT. The followup period during which such sentences are compiled is the period from the individual's release from prison until the end of the period in which the data was gathered. The length of the followup period varied from 3 to 71 months, with an average of 37 months.

Approximately 70% of the individuals in the sample received no prison sentence during their followup period, and therefore had a value of TTSENT of zero. The remaining individuals had values of TTSENT ranging from 1 to 480 months, with a mean value of 45 months.

From a statistical point of view, the unusual features of TTSENT are that it is non-negative, and that zero is a common value. A suitable technique for the analysis of such a variable is the Tobit technique.[1] This technique has been previously applied to such variables as the amount that

---

[1] See J. Tobin, "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26 (1958), pp. 24-36; and T. Amemiya, "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica, 41 (1973), pp. 997-1025.

an individual spends on a car during a year, which is another example of a variable which is non-negative, and often zero. The Tobit technique essentially treats the positive observations in the same way as a least squares regression does, but it also attempts to account for the sizable number of observations which equal zero, which a least squares technique cannot do. Its chief advantage is that it enables us to use both the observations for which TTSENT is zero and those for which it is positive. A mathematical description of the technique is given in Appendix A.

4. RESULTS OF THE ANALYSIS

The dependent variable used in the analysis is TTSENT, as defined in section 3. The explanatory variables which were found to be significantly related to TTSENT are as follows:

A constant term, denoted CNST.

A dummy variable equal to one for whites, and equal to zero for non-whites; denoted RACE.

The number of prison convictions prior to the one leading to the sample sentence; denoted CONPBS.

A dummy variable equal to one if the individual participated in the work release program, and equal to zero if he did not; denoted WR.

A dummy variable equal to one if the individual's record indicates a serious problem with alcohol, or use of hard drugs, and equal to zero otherwise; denoted ALKY.

A dummy variable equal to one if the individual was married at the time of release from the sample sentence, and equal to zero if he was not; denoted MS.

Age (in months) at the time of the individual's first arrest; denoted AFA.

Age (in months) at release from the sample imprisonment; denoted AAR.

The results for the specification including these explanatory variables are given in Table 1. For each variable, the first number given is its coefficient, which represents the partial effect of the variable on the expected value (mean) of TTSENT. For example, the coefficient of RACE is approximately 19.3, which indicates that a white (RACE = 1) will have a mean value of TTSENT that is 19.3 months higher than a non-white (RACE = 0) of otherwise identical characteristics. The second number given is the "t ratio" for the variable. These "t ratios" are distributed as N(0,1) in large samples, under the null hypothesis that the coefficient is zero. The significance of a variable is therefore determined by comparing the "t ratio" to a critical point of the N(0,1) distribution; a value in excess of 1.96 in absolute value indicates significance at the .05 level (for a two-tailed test).

The variables listed above, except CNST, are all statistically significant at the .05 level. They indicate that an individual will tend to have a larger value of TTSENT (more serious recidivist criminal activity) the more previous prison convictions he has had, the younger he was at his first arrest, the younger he is at release, if he is white, if he was not on the work release program, if he is an alcoholic or drug addict, and if he is not married. Conversely, an individual will tend to have a smaller value of TTSENT the fewer previous prison convictions he has, the older he was at his first arrest, the older he is at release, if he is non-white, on the work release program, married, and not an alcoholic or drug addict.

The results of Table 1 constitute the basic results of our analysis of TTSENT. In the process of arriving at this final specification, a number of other explanatory variables were tried, and were dropped due to having coefficients which were statistically insignificant. A list of these other variables follows, along with the "t ratios" of their estimated coefficients.

Number of years of schooling completed; "t ratio" = 1.01.

The number of rule violations during the sample imprisonment; "t ratio" = 0.87.

A dummy variable equal to one if the sample conviction was for a felony, and equal to zero if for a misdemeanor; "t ratio" = 1.36.

A dummy variable equal to one if the sample conviction was for a crime against property, and equal to zero otherwise; "t ratio" = 1.49.

A dummy variable equal to one if the sample conviction was for a crime against a person, and equal to zero otherwise; "t ratio" = 1.36.

Age at release, squared; "t ratio" = 0.04.

We also included the length of the followup period (LENG) and its squared value (LENGSQ) to investigate the effect of length of followup on TTSENT. The "t ratio" of LENG was 0.11, and that of LENGSQ was 0.04. Dropping LENGSQ increased the "t ratio" of LENG only to 0.74, which is still far from being significant.

The fact that the length of the followup period had no discernible effect on TTSENT is somewhat surprising. Since TTSENT is the total time sentenced during the followup period we would have expected it to be larger for individuals who were followed up for longer periods. However, as just noted, the relationship we actually found was not statistically significant. This is probably because most individuals who return to criminal activity do so quite quickly, as noted in the previous report.

5. PREDICTED FREQUENCIES OF TOTAL TIME SENTENCED

The results of the previous section indicate positive and significant effects on TTSENT of RACE, CONPBS, and ALKY, and negative and significant effects on TTSENT of WR, MS, AFA, and AAR. The type of individual most likely to have a large value of TTSENT is a white, unmarried alcoholic (or hard drug user), who was first arrested at an early age, was young at the time of release, was not on the work release program, and has had many previous prison convictions. Conversely, a black, married non-addict, who was relatively old when first arrested and when released, has not had many previous prison convictions, and was on the work release program, is least likely to have a large value of TTSENT.

To see more precisely what this means, we have listed in Table 2 the frequencies of TTSENT implied by our results, for the above two types of individuals, as well as for the "average" individual (i.e., characteristics equal to the mean values of the explanatory variables in our sample). The entries given are the probability that TTSENT is greater than N months, for certain values of N between 0 and 360. Note that P(TTSENT > 0) is the probability of at least one prison sentence and corresponds to a typical definition of recidivism. It might also be worth pointing out that, as indicated earlier, our results do not indicate that these probabilities are very sensitive to the length of the followup period. Since the average length of the followup period was approximately 3 years (it was actually 37 months), we can roughly interpret these as projected frequencies of TTSENT for various types of individuals, during a followup period of approximately 3 years. However, as previously noted, the exact length of time used is not apt to be very important, since most individuals who return to criminal activity do so quite quickly. The rapidity of return is illustrated

by the fact that approximately 70 percent of those who were reconvicted during the followup period were reconvicted within one year of release, and 90% within two years of release.

Case 1 evaluates these probabilities for an "average" individual. This (hypothetical) "average" individual is defined as the individual whose values of the explanatory variables equal the sample means; i.e., RACE = 0.503, CONPBS = 2.19, WR = 0.462, ALKY = 0.481, MS = 0.311, AFA = 280.3, AAR = 380.1. This case corresponds to an "average" frequency for TTSENT. According to Table 2, such an individual has a probability of some sentence (TTSENT > 0) - that is, a probability of recidivism - of 0.243. (This corresponds reasonably closely to the actual proportion of individuals in our sample with TTSENT > 0, 0.296.) The probability of TTSENT being greater than 24 months is 0.169; for N = 48 months it is 0.112; for N = 96 months it is 0.041, and for N > 240 months it is very close to zero.

Case 2 corresponds to an individual with very favorable characteristics; i.e., an old (AAR = 600) non-white who is married, is not an alcoholic or user of hard drugs, was fairly old when first arrested (AFA = 480), was on work release, and did not have any previous prison convictions. Such an individual's probability of some time sentenced is 0.012, which is strikingly less than the value of 0.243 for an "average" individual. The other probabilities are also much lower than for the "average" individual; P(TTSENT > N) is very close to zero for N > 96 months.

Case 3 corresponds to an individual with very unfavorable characteristics; i.e., a young (AAR = 216) white who is not married, is an alcoholic or drug addict, was very young when first arrested (AFA = 156), was not on work release, and had four previous convictions. For such an individual there is a probability of 0.681 of some time sentenced, a probability of 0.481 that this amount will be greater than 48 months, and a probability of 0.204 that it will be greater than 120 months.

Of course, cases 2 and 3 are fairly extreme, and therefore so are the associated probabilities. Nevertheless they do illustrate the strong influence of the individual's characteristics on the likely seriousness of his recidivist criminal activity.

6. CONCLUSIONS

This report contains the results of our analysis of the determinants of the seriousness of recidivist criminal activity. Our measure of seriousness of criminal activity is the total time sentenced during the followup period (TTSENT). The analysis was carried out using the Tobit technique.

The variables which were found to exert significant influence on TTSENT were the number of previous prison convictions, race, age at first arrest, age at release, marital status, whether or not the individual is an alcoholic or user of hard drugs, and whether or not the individual participated in the work release program. Our results indicate that the seriousness of criminal activity is highest for a young, white, unmarried alcoholic or drug addict, who was first arrested at an early age, was not on the work release program, was young at time of his release from imprisonment, and has had many previous prison sentences.

We also presented the frequencies of TTSENT implied by our results, for various types of individuals, for a followup period of approximately three years. The "average" frequency indicates a probability of recidivism during this period of 24%; 17% of such individuals will have received sentences of more than two years, and 9% will have received sentences of more than five years. For individuals with more favorable (or less favorable) characteristics, these probabilities are correspondingly lower (or higher).

TABLE 1

Results of Tobit Analysis of TTSENT

VARIABLE	COEFFICIENT	"t ratio"
CNST	27.6339	1.46
RACE	19.2723	1.83
CONPBS	2.5317	2.50
WR	-21.1050	-2.04
ALKY	47.9032	4.39
MS	-36.4357	-3.18
AFA	-.1748	-2.27
AAR	-.1583	-2.64

Value of $\hat{\sigma}^2$ = 8561.95

TABLE 2

Predicted Frequency, P(TTSENT > N), For Various N

N	CASE 1	CASE 2	CASE 3
0	.243	.012	.681
12	.204	.009	.633
24	.169	.006	.583
36	.139	.004	.532
48	.112	.003	.481
60	.089	.002	.429
72	.070	.001	.379
84	.054	.001	.331
96	.041	.000	.285
108	.031	.000	.243
120	.023	.000	.204
180	.004	.000	.070
240	.000	.000	.017
360	.000	.000	.000

APPENDIX A

Consider a set of independent variables Y_t^*, $t = 1,2,\ldots,T$. The variable Y_t^* is assumed to be normally distributed with mean $X_t\beta$ and variance σ^2, where X_t is a row vector of explanatory variables, and β is a vector of parameters to be estimated.

We observe not Y_t^*, but Y_t, defined by

$$(1) \qquad Y_t = \begin{cases} Y_t^* \text{ if } Y_t^* > 0 \\ 0 \text{ if } Y_t^* \leq 0 \end{cases} .$$

Letting

$$\theta_1 = \{t | Y_t^* > 0\}$$

$$\theta_2 = \{t | Y_t^* \leq 0\} ,$$

the likelihood function of the sample can be written as

$$(2) \qquad L = \prod_{t\epsilon\theta_1} f(Y_t) \prod_{t\epsilon\theta_2} P(Y_t^* \leq 0) ,$$

where

$$f(Y_t) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-1}{2\sigma^2}(Y_t - X_t\beta)^2\right]$$

$$P(Y_t^* \leq 0) = F(-X_t\beta/\sigma)$$

and where F is the cumulative distribution function of the standard normal distribution.

This likelihood function (or its logarithm) can be maximized with respect to β and σ^2, using a computer, with a numerical maximization routine. This gives us the maximum likelihood estimates. Asymptotically valid tests can be constructed using the likelihood ratio principle. Alternately, we can

form the information matrix. The "t ratio", formed by dividing a parameter estimate by the square root of the appropriate diagonal element of the information matrix, converges in distribution to N(0,1) under the null hypothesis that the parameter is zero.

APPENDIX B

DESCRIPTION OF COMPUTER PROGRAM FOR PREDICTION OF FREQUENCY OF TIME SENTENCED AFTER RELEASE

1. INTRODUCTION

The purpose of the program is to generate, for individuals with certain specified characteristics, a predicted frequency of the total time sentenced during a specified period after release from prison, denoted TTSENT. This variable is defined more precisely in Section 2 below. The specified characteristics consist of values for seven variables: RACE, CONPBS, WR, ALKY, MS, AFA and AAR. These variables are also defined in Section 2, and are chosen because they have been found, in the statistical analysis performed under this contract, to be significantly related to TTSENT.

The frequencies are generated from a normal distribution, truncated from below at zero, whose mean is a weighted sum of the seven explanatory variable. The weights, and the variance of the distribution, were estimated in the statistical analysis already referred to; for details see Sections 3 and 4 of Chapter 2.

2. DESCRIPTION OF THE VARIABLES

The variable whose frequency is predicted is TTSENT, defined as the total length (in months) of all prison sentences received by an individual, during a period of approximately three years, following his release from prison.

RACE is a dummy variable which is set equal to one for whites, and equal to zero for non-whites.

CONPBS is the number of previous convictions resulting in a jail or prison sentence. It does not include the prison sentence just prior to the period of the followup. This definition corresponds to the current Department of Correction's definition for "prior convictions."

WR is a dummy variable which equals one if the individual participated in the North Carolina prisoner work release program, and equals zero if he did not.

ALKY is a dummy variable which is set equal to one for individuals with a serious drinking problem and/or a history of hard drug use (which is equivalent to the Department of Correction's code 4 or 5 under "drinking habits", or code 6, 7 or 8 under "drug type"), and equal to zero for other individuals.

MS is a dummy variable which equals one if the individual was married at the time of release, and equals zero if he was not.

AFA is the age (in months) at the time of the individual's first arrest.

AAR is the individual's age (in months) at release from the sample imprisonment.

3. INPUTS INTO AND OUTPUTS FROM THE PROGRAM

The only input which must be provided to the program is the set of values of RACE, CONPBS, WR, ALKY, MS, AFA and AAR for the individual (or group) for which the predicted frequency of TTSENT is to be generated. The program reads these variables off an input data card, with FORMAT 7F10.0. This means that the value of RACE must be punched in columns 1-10; CONPBS in columns 11-20; WR in columns 21-30; ALKY in columns 31-40; MS in columns 41-50; AFA in columns 51-60; and AAR in columns 61-70. As long as an explicit decimal point is punched (e.g., punch 1. or 1.0, not just 1), it does not matter where in each 10-column field a value is punched.

If the value of RACE, CONPBS, WR, ALKY, MS, AFA and/or AAR is not available, or if it is desired not to take into account the effects of one or more of these variables, it is necessary to punch <u>any negative</u> value (e.g., -1.0) for the variable in question. The program will then automatically assign to the individual the average value (over our 582 observations) of

that variable. (The average values of RACE, CONPBS, WR, ALKY, MS, AFA and AAR are respectively 0.503, 2.19, 0.462, 0.478, 0.311, 280.3 and 380.1.)

The first output provided is the set of values read for RACE, CONPBS, WR, ALKY, MS, AFA and AAR. When an average value has been substituted for any variable, as just described, this is so indicated.

The second output, and the one of major interest, is the predicted frequency of TTSENT. For numbers of months between 0 and 480, it displays the probability that TTSENT will exceed that number. Note in particular that the first entry therefore represents P(TTSENT > 0), which is the probability of some prison sentence, and is thus the probability of recidivism during the three years following release.

4. CAUTIONARY NOTES

There are a few words of caution worth mentioning. These have to do not with the program itself, but with the interpretation of the results.

1. The sample used in the statistical analysis upon which these projections are based is described in some detail in a previous report. Briefly, it is a random sample of inmates who had not been convicted of crimes that would have prevented their placement on the work release program, and who served time in minimum and medium custody prison units in the South Piedmont administrative area. It therefore did not contain individuals convicted of sex offenses (D.O.C. crime codes 700, 701, 710, 711, 712), serious drug offenses (crime codes 804, 805 or 806 resulting in a sentence of 4 years or more), or the public drunk offense (crime code 857). The sample is not representative of inmates who never served time in minimum and medium custody prison units but are confined throughout their imprisonment to the institutions of the prison system, i.e., the specialized

youthful offenders and maximum and close custody institutions. The sample also contained no women. As a result, the projections made here should not be applied to women, individuals who do not serve time in minimum and medium custody units, or individuals who have been convicted of the above mentioned offenses.

2. Strictly speaking, the program gives the frequency of TTSENT for an individual with specified characteristics. It can be used to give the expected frequency of TTSENT for a group of individuals by using the average values of RACE, CONPBS, WR, ALKY, MS, AFA and AAR for the group. However, its use in this way involves an approximation. This approximation should be fairly accurate for groups of individuals whose values of the explanatory variables are similar, but it may be poor for groups of individuals with dissimilar values of the explanatory variables.

3. In our sample, values of TTSENT in excess of 100 months were fairly rare; only 4.8% of our sample had values of TTSENT that large. Although we provide projections for values of TTSENT as large as 480 months, those for values over 100 months should be interpreted with care.

CHAPTER 3

ANALYSIS OF TYPE OF CRIMINAL ACTIVITY

1. INTRODUCTION

This report presents an analysis of the type of criminal activity after release from imprisonment. The analysis is in two parts, corresponding to different ways of characterizing the type of criminal activity. The first part analyzes the factors which influence whether the most serious conviction during a specified period after release is for a misdemeanor or for a felony. (The most serious conviction is defined as the one resulting in the longest prison sentence.) The second part analyzes the factors which influence whether the most serious conviction is for a crime against property, for a crime against a person, or for some other crime. The persons and property categories are as defined in Appendix G of Work Release in North Carolina: An Evaluation of Its Effects After Release from Incarceration except that robbery (Crime Codes 200, 201, 203, 204) has been moved from the persons to the property category. The other category used in this report includes all crimes falling in the following categories defined in Appendix G: crimes involving drugs and alcohol, crimes against the family, vehicular offences, and other crime types.

Section two of the report gives a brief description of the data set on which the analysis is based. Section three discusses the statistical methodology used. Section four presents the results of our analysis of the misdemeanor/felony breakdown of type of crime; based on this analysis, section five contains predicted probabilities of these two types of crimes for certain types of individuals. Section six presents the results of our analysis of the persons/property/other breakdown of type of crime; based on this analysis, section seven contains predicted probabilities of these three types of crimes for certain types of individuals. Section eight summarizes our findings.

2. THE DATA

The data is described in some detail in our report, "Analysis of Timing of Return to Criminal Activity," previously submitted. We will therefore give here only a brief overview of a few relevant aspects of the data set.

The data set consists of information on a random sample of 641 men who were inmates in prison units in the South Piedmont area of North Carolina in 1969 or 1971. As a result it is not representative of all men in prison in North Carolina; however, it is fairly representative of all men in medium and minimum custody prison facilities in North Carolina.

Since the original purpose of the data set was to evaluate the North Carolina Prisoner Work Release program, it does not include individuals who were imprisoned for offenses which would prevent their placement on the work release program. This group consists of men in prison for sex crimes (D.O.C. crime codes 700, 701, 710, 711, 712) or serious drug offenses (D.O.C. crime codes 804, 805, 806 resulting in a sentence of 4 years or more). In addition it does not include men convicted of the public drunk offense (D.O.C. crime code 857) since they were not in prison long enough to be processed for work release. The group also contains no women since there were no women in South Piedmont area prisons in 1969 and 1971.

3. THE STATISTICAL METHODOLOGY

Type of crime is a qualitative variable rather than a quantitative one; that is, there is no natural numerical value attached to it. The most widely used model for analysis of qualitative variables is the logit model, which has been used elsewhere to analyze such qualitative variables as type of occupation that a person is employed in, type of housing that a person chooses, survival or death of a bacteria following administration of an antibiotic, etc.

Consider the case in which we categorize crimes as misdemeanors or felonies. In this case there are three possibilities, for each individual: (1) no conviction resulting in imprisonment, during the follow-up period; (2) most serious conviction during the follow-up period is for a misdemeanor; (3) most serious conviction during the follow-up period is for a felony. For the i^{th} individual, we can represent the probabilities of these three possibilities as P_{1i}, P_{2i}, P_{3i}, respectively. Then the appropriate logit model specifies that

$$\ln (P_{2i}/P_{1i}) = \sum_{j=1}^{K} \beta_{2j} X_{ij}$$

$$\ln (P_{3i}/P_{1i}) = \sum_{j=1}^{K} \beta_{3j} X_{ij} \quad .$$

There are K explanatory variables, and X_{ij} represents the value of the j^{th} explanatory variable for the i^{th} individual. The β's are coefficients to be estimated.

The dependent variables are the logarithms of the ratios of probabilities. Note that, for example, $\ln (P_2/P_1)$ equals infinity when $P_1 = 0$, and equals minus infinity when $P_2 = 0$. More generally, $\ln (P_2/P_1)$ increases whenever P_2 increases relative to P_1; that is, when P_2 rises or P_1 falls. Therefore, a positive value for one of the β_{2j} above indicates that an increase in the corresponding explanatory variable will cause an increase in P_2 relative to P_1. Conversely, a negative value for one of the β_{2j} indicates that an increase in that explanatory variable causes a decrease in P_2 relative to P_1 (or, equivalently, an increase in P_1 relative to P_2).

Note that from the two equations above, it is possible to derive an equation explaining the relative odds of the second and third possibilities:

$$\ln (P_{3i}/P_{2i}) = \ln (P_{3i}/P_{1i}) - \ln (P_{2i}/P_{1i})$$

$$= \sum_{j=1}^{K} (\beta_{3j} - \beta_{2j}) X_{ij} \quad .$$

Similar comments apply to the case in which we catagorize crimes as crimes against persons, crimes against property, and other crimes. Then there are four possibilities for each individual: (1) no conviction resulting in imprisonment during the follow-up period; (2) most serious conviction during the follow-up period is for a crime against property; (3) most serious conviction during the follow-up period is for a crime against a person; (4) most serious conviction during the follow-up period is for any other crime. The basic model then specifies equations, like those above, explaining $\ln (P_2/P_1)$, $\ln (P_3/P_1)$, and $\ln (P_4/P_1)$. Also, from these one can derive equations for other comparisons; for example, $\ln (P_4/P_2) = \ln (P_4/P_1) - \ln (P_2/P_1)$.

The logit model is estimated by the method of maximum likelihood. A technical summary of the model and its estimation is given in Appendix A.

4. ANALYSIS OF MISDEMEANORS AND FELONIES

The variable to be analyzed in this section is the nature of the most serious crime during the follow-up period. The most serious crime is defined as the one resulting in the longest prison sentence. As mentioned in the previous section, there are three possibilities:

(1) NONE, indicating that the individual did not commit any crime during the follow-up period for which he received a North Carolina prison sentence;

(2) MISD, indicating that the most serious crime was a misdemeanor which resulted in a North Carolina prison sentence;

(3) FLNY, indicating that the most serious crime was a felony which resulted in a North Carolina prison sentence.

Eight explanatory variables were found to have a significant influence on the nature of the most serious crime. They are the following:

A constant term, denoted CNST.

A dummy variable equal to one for whites, and equal to zero for non-whites; denoted RACE.

A dummy variable equal to one if the sample sentence was for a felony, and equal to zero if it was for a misdemeanor; denoted MF.

A dummy variable equal to one if the individual's record indicates a serious problem with alcohol, or use of hard drugs; denoted ALKY.

Age (in months) at the time of the individual's first arrest; denoted AFA.

A dummy variable equal to one if the release from the sample sentence was supervised, and equal to zero if it was not; denoted SUPER.

Age (in months) at release from the sample imprisonment; denoted AAR.

The number of prison convictions prior to the one leading to the sample sentence; denoted CONPBS.

The results for the logit model with these explanatory variables are given in Table 1. The first column lists the above explanatory variables. The next three columns contain these variables' coefficients and "t ratios;" the "t ratios" are listed in parentheses under the coefficients, and are asymptotically distributed as N(0,1) under the null hypothesis that the coefficient equals zero. Therefore a "t ratio" in excess of 1.96 in absolute value indicates statistical significance at the .05 level, for the usual two-tailed test.

The coefficients and "t ratios" in the column labeled MISD/NONE are for the equation whose dependent variable is the logarithm of the ratio of the probability of MISD to the probability of NONE. So, for example, the positive coefficient of RACE in this equation indicates that to be white (RACE = 1) rather than non-white (RACE = 0) increases the probability that the most serious conviction is for a misdemeanor (MISD) relative to the probability of no prison conviction (NONE). Similarly, the heading FLNY/NONE indicates that the dependent variable is the logarithm of the ratio of the probability of FLNY to the probability of NONE, while the heading FLNY/MISD indicates that the dependent variable is the logarithm of the ratio of the probability of FLNY to the probability of MISD.

If we look in more detail at the effects of RACE, we see that to be white (RACE = 1) increases the probability of MISD relative to both NONE and FLNY, while its effect on the probabilities of FLNY and NONE relative to each other is too small to have any confidence about. This means that to be white increases the probability of conviction for a misdemeanor, and causes roughly equal (proportional) decreases in the probabilities of conviction for a felony and of no conviction.

The effects of MF are very straightforward. If the sample conviction was for a felony (MF = 1) rather than a misdemeanor (MF = 0), this increases the probability of FLNY relative to either MISD or NONE, but decreases the probability of MISD relative to NONE. All this really says is that individuals whose sample conviction was for a felony are more likely to commit a felony, and less likely to commit a misdemeanor, than are individuals whose sample conviction was for a misdemeanor.

Looking at the effects of ALKY, we find that to be an alcoholic or user of hard drugs (ALKY = 1) increases the probability of both FLNY and

MISD relative to NONE, and decreases the probability of FLNY relative to MISD. This means that to be an alcoholic or user of hard drugs increases the probability of both misdemeanors and felonies, but it increases the probability of a misdemeanor more than it increases the probability of a felony. This is quite reasonable considering that alcoholics predominate in our sample and that most alcohol related crimes are misdemeanors.

A larger value of AFA decreases the probability of FLNY relative to both NONE and MISD; its effect on the relative probabilities of MISD and NONE is too small to be sure of. This means that individuals who are older at the time of their first arrest are less likely to commit felonies, but this variable does not have a significant impact on the commission of misdemeanors.

If the release from the sample imprisonment was supervised (SUPER = 1), this decreases the probability of both FLNY and MISD relative to NONE, but has only a small effect on the relative probabilities of FLNY and MISD. So supervision at release decreases the probability of both misdemeanors and felonies, and we do not have strong evidence that it decreases the probability of either one more than the other.

The effects of AAR are similar to those of SUPER, in that a larger value of AAR decreases the probability of both FLNY and MISD relative to NONE, but has only a very small effect on the relative probabilities of FLNY and MISD. This means that to be older at time of release decreases the probabilities of both felonies and misdemeanors, and we do not have evidence to suggest that it decreases the probability of either one more than the other.

The effects of CONPBS are just the opposite of SUPER and AAR. A larger value of CONPBS increases the probability of both FLNY and MISD relative to NONE, but has only a very small effect on their probability

relative to each other. Therefore a larger number of previous prison convictions increases the probabilities of both felonies and misdemeanors, but does not increase the probability of either one more than the other.

In arriving at the results just reported, a number of other explanatory variables were tried, but were dropped due to having coefficients which were not statistically significant ("t ratio" of 1.96 or more) in any of the above three equations. A list of these variables follows, along with the largest (in absolute value) of their "t ratios" in the three equations.

A dummy variable equal to one if the individual participated in the work release program, and equal to zero if he did not, denoted WR; "t ratio" = -1.14.

A dummy variable equal to one if the individual was married at the time of his release from the sample sentence, and equal to zero if he was not, denoted MS; "t ratio" = -1.27.

Number of years of schooling completed, denoted SG; "t ratio" = 0.74.

The number of rule violations during the sample imprisonment, denoted RULE; "t ratio" = -0.75.

A dummy variable equal to one if the sample sentence was for a crime against a person, and equal to zero if it was not, denoted PERSON; "t ratio" = -1.37.

A dummy variable equal to one if the sample sentence was for a crime against property, and equal to zero if it was not, denoted PROPTY; "t ratio" = 1.91.

The length (in months) of the follow-up period, denoted LENG; "t ratio" = 1.07.

The fact that the length of the follow-up period had no significant effect is somewhat surprising. We would have expected that the probability of FLNY and/or MISD would increase with increasing length of the follow-up

period, since the probability of NONE ought to decrease with increasing length of the follow-up period. However, as just noted, the relationship we actually found was far from being statistically significant. This is analogous to the result in our (previously reported) analysis of total time sentenced, where LENG was also insignificant. As was the case there, this is presumably because most individuals who return to crime do so quite quickly. Since the average length of our follow-up period was approximately 37 months, we can interpret our results as those that hold for a follow-up period of approximately that length of time. However, the insignificance of LENG indicates that the actual length of the follow-up period is not apt to be very important, within reasonable limits, say one to five years.

Finally, it is interesting to note that many of the same variables that were found to be significant determinants of the nature of the most serious crime were also found, in our previous analysis, to be significant determinants of total time sentenced. Specifically, RACE, CONPBS, ALKY, AFA, and AAR were significant in the determination of both time sentenced and nature of the most serious crime. This is, of course, not surprising since both variables in some sense measure the seriousness of recidivist criminal activity.

5. PREDICTED PROBABILITIES FOR MISDEMEANORS AND FELONIES

In Table 2 we report the probabilities of NONE, MISD, and FLNY implied by our model for three hypothetical types of individuals. For each of these three cases are listed P(NONE), P(MISD), and P(FLNY), which are respectively the probability of no prison conviction during the follow-up period, the probability that the most serious conviction during the follow-up period is for a misdemeanor, and the probability that the most serious

conviction during the follow-up period is for a felony. As noted in the last section, these should be interpreted for a follow-up period of approximately three years, though the exact length of the follow-up period is not apt to be very important as long as it is at least one year, and preferably two years.

Case 1 in Table 2 corresponds to the hypothetical "average individual," defined as the individual whose values for all explanatory variables equal the sample means: RACE = 0.503, MF = 0.307, ALKY = 0.478, AFA = 280.3, SUPER = 0.316, AAR = 380.1, CONPBS = 2.186. We see that P(NONE) = 0.769, P(MISD) = 0.183, and P(FLNY) = 0.048.

Case 2 corresponds to an individual with very favorable characteristics; i.e., an old (AAR = 600) non-white who is not an alcoholic or user of hard drugs, who has no previous prison convictions (prior to the sample sentence), whose sample sentence was for a misdemeanor, whose release was supervised, and who was quite old when first arrested (AFA = 480). Such an individual has a very high probability of no conviction during the follow-up period -- P(NONE) = 0.972 -- and very low probabilities of conviction for a misdemeanor or especially for a felony -- P(MISD) = 0.028, P(FLNY) = 0.001

Case 3 corresponds to an individual with very unfavorable characteristics; i.e., a young (AAR = 216) white who is an alcoholic or user of hard drugs, who has four previous prison convictions, whose sample sentence was for a felony, whose release was not supervised, and who was very young when first arrested (AFA = 156). Such a person has a much lower than average probability of no prison conviction during the follow-up period -- P(NONE) = 0.306. He also has a higher than average probability of conviction for a misdemeanor -- P(MISD) = 0.275 -- and a much higher than average probability of conviction for a felony -- P(FLNY) = 0.419.

Cases 2 and 3 are fairly extreme, and therefore so are the associated probabilities. However, they do illustrate the strong influence of individual characteristics on the probabilities of various types of criminal convictions.

6. ANALYSIS OF PERSONAL, PROPERTY, AND OTHER CRIMES

The variable to be analyzed in this section is the type of crime resulting in the longest prison sentence received during the follow-up period; that is, the type of the most serious crime. As mentioned in section 3, there are four possibilities considered:

(1) NONE, indicating that the individual did not commit any crime during the follow-up period for which he received a North Carolina prison sentence;

(2) PROP, indicating that the most serious crime which resulted in a North Carolina prison sentence was a crime against property, as defined in section 1;

(3) PERS, indicating that the most serious crime which resulted in a North Carolina prison sentence was a crime against a person, as defined in section 1;

(4) OTHER, indicating that the most serious crime which resulted in a North Carolina prison sentence was for a crime other than a crime against property or a crime against a person.

Eight explanatory variables were found to have a significant influence on the type of the most serious crime. They are CNST, RACE, AAR, ALKY, CONPBS, SUPER, PERSON, and PROPTY, all of these being as defined previously in section 4. The results for the logit model with these explanatory variables are given in Table 3. The format of Table 3 is essentially identical to that of Table 1 (as explained in section 4) except for the differences due to the different dependent and explanatory variables. So,

for example, the results in the column labeled PROP/NONE are the coefficients and "t ratios" for the equation whose dependent variable is the logarithm of the ratio of the probability of PROP to the probability of NONE.

Consider first the effects of RACE. To be white (RACE = 1) rather than non-white (RACE = 0) increases the probability of OTHER relative to NONE, PROP and PERS, and has very small effects on all other relative probabilities. So the only noticable effect of RACE is that whites are more likely than non-whites to commit OTHER crimes -- that is, crimes which are neither against persons nor property.

A larger value of AAR decreases the probability of all three types of crimes (PERS, PROP, OTHER) relative to none. It also decreases the probability of PROP relative to both PERS and OTHER. What this means is that increased age decreases the probability of all three types of crime, but it causes the largest decrease in the probability of property crimes.

Looking at ALKY, we find that to be an alcoholic or user of hard drugs (ALKY = 1) increases the probability of all three types of crimes (PERS, PROP and OTHER) relative to NONE. However, it decreases the probability of PROP relative to PERS and OTHER. So while alcoholics or users of hard drugs are more likely to commit all three types of crimes, the increase in the probability of property crimes is less than the increase in the probability of crimes against persons and other crimes. The last effect is expected since crimes directly involving drugs or alcohol are included in the other category.

A higher value of CONPBS also increases the probability of all three types of crime relative to NONE. There is weak evidence that it increases the probability of PROP relative to PERS and OTHER, which means that a larger number of previous convictions causes a larger increase in the probability of property crimes than in the probability of crimes against a person or other crimes.

If the release from prison was supervised (SUPER = 1), this decreases the probability of all three types of crime relative to NONE. The largest decrease is in the probability of crimes against a person, as indicated by the fact that it decreases the probability of PERS relative to both PROP and OTHER.

If the sample conviction was for a crime against a person (PERSON = 1), this increases the probability of PERS relative to all other possibilities (PROP, OTHER, NONE). This is statistical confirmation of the obvious guess that an individual who has previously committed a crime against a person is more likely to commit such a crime in the future than is a person who previously committed some other type of crime. It is also true that if the sample conviction was for a crime against a person, this decreases the probability of PROP relative to all other possibilities. So an individual who has previously committed a crime against a person is less likely to commit a property crime than is a person who previously committed a property or other crime.

Interestingly, the main effect of the sample conviction being for a property crime (PROPTY = 1) is that it increases the probability of PERS relative to all other possibilities (PROP, OTHER, NONE), and increases the probability of PROP relative to OTHER and NONE. So for the sample conviction to be for a property crime increases the probability of both property crimes and crimes against persons, but it causes a bigger increase in the probability of crimes against persons than in the probability of property crimes. However, the evidence on this last point is not very strong, in the sense that several of the coefficients are significant only at fairly low confidence levels.

In arriving at the results just reported, a number of other explanatory variables were tried, but were dropped due to having coefficients which were

statistically insignificant ("t ratio" = 1.96 or less) in all of the above six equations. A list of these variables follows, along with the largest (in absolute value) of their "t ratios" in the six equations: MF, 1.58; AFA, 1.83; LENG, 1.14; RULE, 1.39; MS, -1.86; SG, 1.63; WR, -1.05. All of these variables are as previously defined in section 4.

Again, it is interesting to note that many of the same variables that were found to be significant here were also found to be significant in the analysis of section 4 and in the previously reported analysis of time sentenced. Specifically, RACE, CONPBS, ALKY, and AAR were found to have a significant influence in all three analyses; in addition, SUPER was found to have a significant influence in both analyses reported here, though not in the analysis of time sentenced.

7. PREDICTED PROBABILITIES FOR PERSONAL, PROPERTY AND OTHER CRIMES

In Table 4 we report the probabilities of NONE, PROP, PERS and OTHER implied by our model for three hypothetical types of individuals. The format of Table 4 is similar to that of Table 2 except that here we report P(NONE), P(PROP), P(PERS), and P(OTHER) rather than P(NONE), P(MISD), and P(FLNY).

Case 1 in Table 4 corresponds to the hypothetical "average individual," defined as the individual whose values for all of the explanatory variables equal the sample means: RACE = 0.503, AAR = 380.1, ALKY = 0.478, CONPBS = 2.186, SUPER = 0.316, PERSON = 0.171, PROPTY = 0.421. We see that P(NONE) = 0.790, P(PROP) = 0.098, P(PERS) = 0.031, and P(OTHER) = 0.081.

Case 2 corresponds to an individual with very favorable characteristics; i.e., an old (AAR = 600) non-white who is not an alcoholic or user of hard drugs, who has no previous prison convictions (prior to the sample sentence), whose release was supervised, and whose sample crime was neither a property

crime nor a crime against a person. Such an individual has a much higher than average probability of no prison conviction -- P(NONE) = 0.961. He correspondingly has much lower than average probabilities of property, personal and other crimes -- P(PROP) = 0.019, P(PERS) = 0.006, P(OTHER) = 0.015.

Finally, case 3 corresponds to an individual with very unfavorable characteristics; i.e., a young (AAR = 216) white who is an alcoholic or user of hard drugs, who has four previous prison convictions, whose release was not supervised, and whose sample conviction was for a crime against a person. Such an individual has a much lower than average probability of no prison conviction -- P(NONE) = 0.271 -- and an about average probability of conviction for a property crime -- P(PROP) = 0.087 -- but a much higher than average probability of conviction for a personal or other crime -- P(PERS) = 0.337, P(OTHER) = 0.304.

8. CONCLUSIONS

This report contains two similar but separate analyses. One is of the nature of the most serious conviction during the follow-up period, where the three possibilities are no prison conviction, a misdemeanor, and a felony. The other analysis is again of the type of crime for the most serious conviction during the follow-up period, where the four possibilities are no prison conviction, a crime against property, a crime against a person, and any other crime (against neither a person nor property).

Five variables were found to have significant effects in both analyses: race, age at release, number of previous prison convictions, whether the individual is an alcoholic or user of hard drugs, and whether the release from the sample conviction was supervised. Two additional variables had significant effects in the first analysis but not in the second: age at

first arrest, and nature of the sample crime (misdemeanor or felony). Conversely, the type of the sample crime (personal, property, or other) was found to have significant effects in the second analysis but not in the first.

No brief summary of the effects of the above variables can be entirely accurate. However, it is roughly true that we find that the type of individual most likely to engage in criminal activity is a young white alcoholic or user of hard drugs, with many previous prison convictions, and whose release from the sample sentence was unsupervised. The sense in which this generalization is roughly true is spelled out in some detail in the discussions of section 4 and 6 above.

We also presented the predicted probabilities of various types of convictions (for misdemeanors or felonies, and for personal, property and other crimes) implied by our results, for various types of individuals, for a follow-up period of approximately three years. Such probabilities can of course be generated for other types of individuals as well.

TABLE 1

Results of Logit Analysis of the Nature of the Most Serious Crime

VARIABLE	MISD/NONE	FLNY/NONE	FLNY/MISD
CNST	-.65401 (-0.73)	1.0266 (1.14)	1.6806 (1.74)
RACE	.61605 (2.66)	.10316 (0.30)	-.51289 (-1.36)
MF	-.94355 (-2.96)	.62333 (1.64)	1.5669 (3.44)
ALKY	1.2965 (5.41)	.60331 (1.74)	-.69318 (-1.79)
AFA	-.000728 (-0.45)	-.009319 (-2.32)	-.008592 (-2.08)
SUPER	-.57979 (-2.02)	-1.0108 (-2.40)	-.43100 (-0.90)
AAR	-.003302 (-2.42)	-.004141 (-1.71)	-.000839 (-0.33)
CONPBS	.10159 (3.16)	.07707 (1.44)	-.02452 (-0.48)

TABLE 2

Predicted Probabilities for Misdemeanors and Felonies

	CASE 1	CASE 2	CASE 3
P(NONE)	.769	.972	.306
P(MISD)	.183	.028	.275
P(FLNY)	.048	.001	.419

TABLE 3

Results of Logit Analysis of the Type of the Most Serious Crime

VARIABLE	OTHER/NONE	PROP/NONE	PERS/NONE	PROP/OTHER	PERS/OTHER	PERS/PROP
CNST	-2.8517 (-4.62)	.42505 (0.69)	-2.8217 (-2.98)	3.2768 (5.47)	-.030122 (-0.04)	-3.2468 (-3.05)
RACE	1.3423 (3.90)	.013101 (0.05)	.057987 (0.13)	-1.3293 (-3.28)	-1.2844 (2.45)	.04488 (0.09)
AAR	-.002183 (-1.73)	-.007253 (-4.36)	-.003882 (-1.95)	-.005070 (-2.62)	-.001699 (-0.76)	.003371 (1.37)
ALKY	1.6772 (4.87)	.58382 (2.09)	1.3266 (2.97)	-1.0933 (-2.65)	-.35051 (-0.65)	.74281 (1.49)
CONPBS	.10824 (3.44)	.14104 (4.01)	.10534 (2.41)	.032813 (1.13)	-.002900 (-0.07)	-.035708 (-0.88)
SUPER	-.48690 (-1.49)	-.90245 (-2.74)	-1.9964 (-3.14)	-.41555 (-0.95)	-1.5095 (-2.18)	-1.0939 (-1.57)
PERSON	-.01440 (-0.03)	-1.1587 (-1.80)	2.0718 (3.52)	-1.1443 (-1.55)	2.0862 (3.09)	3.2305 (3.87)
PROPTY	-.36812 (-1.10)	.31057 (1.04)	1.0725 (1.90)	.67870 (1.67)	1.4406 (2.32)	.76194 (1.25)

TABLE 4

Predicted Probabilities for Various Types of Crimes

	CASE 1	CASE 2	CASE 3
P(NONE)	.790	.961	.271
P(PROP)	.098	.019	.087
P(PERS)	.031	.006	.337
P(OTHER)	.081	.015	.304

APPENDIX A

Let there be N possible responses, with probabilities $P_1, P_2, \ldots, P_N$. Then the multiple logit model can be written as:

$$(1) \qquad \log_e \left[\frac{P_{jt}}{P_{1t}}\right] = X_t \beta_j , \quad j = 2,3, \ldots,N; \quad t = 1,2,3, \ldots,T;$$

where t is the observation index, T the number of observations, X_t the t'th observation on a $1 \times K$ vector of explanatory variables, and β_j a $K \times 1$ vector of (unknown) parameters.

The N-1 equations in (1), plus the requirement that the probabilities for every t sum to one, determine the probabilities uniquely. Explicitly, the solution is:

$$P_{1t} = \frac{1}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}}$$

$$(2) \qquad P_{it} = \frac{e^{X_t \beta_i}}{1 + \sum_{j=2}^{N} e^{X_t \beta_j}} \qquad i = 2, \ldots,N .$$

This model can be estimated by maximum likelihood by observing that the likelihood function is

$$(3) \qquad L = \prod_{t \varepsilon \theta_1} P_{t1} \prod_{t \varepsilon \theta_2} P_{t2} \cdots \prod_{t \varepsilon \theta} P_{tN}$$

where

$$(4) \qquad \theta_j = \{t \mid j^{th} \text{ response is observed}\} .$$

Hence:

$$L = \prod_{t\varepsilon\theta_1} \frac{1}{1 + \sum_{j=2}^{N} e^{X_t\beta_j}} \prod_{i=2}^{N} \prod_{t\varepsilon\theta_i} \frac{e^{X_t\beta_i}}{1 + \sum_{j=2}^{N} e^{X_t\beta_j}}$$

(5)

$$= \prod_{t=1}^{T} \frac{1}{1 + \sum_{j=2}^{N} e^{X_t\beta_j}} \prod_{i=2}^{N} \prod_{t\varepsilon\theta_1} e^{X_t\beta_i} \quad .$$

The maximization of this, or its logarithm, can be done using a non-linear maximization program.

To get asymptotic variances for the estimates of $\beta_2,\beta_3,\ldots,\beta_N$, it is necessary to form the information matrix. This turns out to be of the form

(6)
$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_{22} & \mathcal{J}_{23} & \cdots & \mathcal{J}_{2N} \\ \mathcal{J}_{32} & \mathcal{J}_{33} & \cdots & \mathcal{J}_{3N} \\ \vdots & \vdots & & \vdots \\ \mathcal{J}_{N2} & \mathcal{J}_{N3} & & \mathcal{J}_{NN} \end{bmatrix}$$

where

$$\mathcal{J}_{rr} = \sum_{t=1}^{T} P_{rt}(1 - P_{rt})\, X_t'X_t \qquad r = 2,\ldots,N$$

(7)

$$\mathcal{J}_{rs} = -\sum_{t=1}^{T} P_{rt}P_{st}X_t'X_t \qquad r,s = 2,\ldots,N \quad r \neq s \quad .$$

The inverse of $\mathcal{J}$ is then the asymptotic covariance matrix of $\hat{\beta} = (\hat{\beta}_2',\hat{\beta}_3',\ldots,\hat{\beta}_N')'$.

The point of obtaining the asymptotic variances of the elements of $\hat{\beta}$ is to allow testing of hypotheses concerning β. Naturally these tests are only valid asymptotically.

APPENDIX B

DESCRIPTION OF COMPUTER PROGRAM FOR PREDICTION
OF THE NATURE OF THE MOST SERIOUS CONVICTION

1. INTRODUCTION

This description is actually a description of two distinct programs. The programs are very similar, so much of the description need only be made once.

The purpose of the first program is to generate, for individuals with certain specified characteristics, probabilities of the three following possibilities:

(1) the individual receives no prison conviction during the follow-up period, denoted NONE;

(2) the most serious conviction (i.e., longest prison sentence) is for a misdemeanor, denoted MISD;

(3) the most serious conviction is for a felony, denoted FLNY.

The specified characteristics consist of values for seven variables: RACE, MF, ALKY, AFA, SUPER, AAR, and CONPBS. These variables are defined below in Section 2, and were chosen because they have been found, in our statistical analysis, to have a significant effect on the nature of the most serious conviction, as described by the three possibilities listed above.

The purpose of the second program is to generate similar probabilities for the following four possibilities:

(1) NONE, as defined above;

(2) the most serious conviction is for a crime against property, denoted PROP;

(3) the most serious conviction is for a crime against a person, denoted PERS;

(4) the most serious conviction is for a crime which is neither a crime against a person nor a crime against property, denoted OTHER.

In this case the characteristics of the individual which need to be specified are: RACE, AAR, ALKY, CONPBS, SUPER, PERSON, and PROPTY. These variables are defined in Section 2 below and were chosen because they have been found, in our statistical analysis, to have a significant effect on the type of the most serious conviction, as described by the four possibilities listed below.

In both cases, the probabilities are generated from the logistic distribution implied by the logit model. The parameters of the two logit models were estimated in the statistical analyses already referred to; for details see Chapter 3.

2. DESCRIPTION OF THE VARIABLES

The three possibilities (NONE, MISD, FLNY) whose probabilities are generated by the first program, and the four possibilities (NONE, PROP, PERS, OTHER) whose probabilities are generated by the second program have been defined in Section 1.

RACE is a dummy variable which is equal to one for whites, and equal to zero for non-whites.

AAR is age (in months) at the time of release from the sample sentence.

ALKY is a dummy variable which is set equal to one for individuals with a serious drinking problem and/or a history of hard drug use (which is equivalent to the Department of Correction's code 4 or 5 under "drinking habits," or code 6, 7, or 8 under "drug type"), and equal to zero otherwise.

CONPBS is the number of previous convictions resulting in a jail or prison sentence. It does not include the sample prison sentence (the one just prior to the follow-up period). This corresponds to the Department of Correction's definition for "prior convictions."

SUPER is a dummy variable which is equal to one if the release from the sample prison sentence was supervised, and equal to zero if it was not.

AFA is the age (in months) at the time of the individual's first arrest.

MF is a dummy variable equal to one if the sample sentence was for a felony, and equal to zero if it was for a misdemeanor.

PROPTY is a dummy variable equal to one if the sample sentence was for a crime against property, and zero otherwise.

PERSON is a dummy variable equal to one if the sample sentence was for a crime against a person, and zero otherwise.

3. INPUTS INTO AND OUTPUTS FROM THE PROGRAM

For the first program, the only input necessary is the set of values of RACE, MF, ALKY, AFA, SUPER, AAR, and CONPBS for the individual for whom the probabilities of NONE, MISD, and FLNY are to be calculated. The program reads these variables off an input data card with FORMAT 7F10.0. This means that the value of RACE must be punched in columns 1-10; MF in columns 11-20; ALKY in columns 21-30; AFA in columns 31-40; SUPER in columns 41-50; AAR in columns 51-60, and CONPBS in columns 61-70. As long as an explicit decimal point is punched (e.g., punch 1.0 or 1., not just 1), it does not matter where in each 10-column field a value is punched.

The first output provided is the set of values read for RACE, MF, ALKY, AFA, SUPER, AAR, and CONPBS. The second output, and the one of major interest, is the list of the probabilities of NONE, MISD, and FLNY.

The input into the second program is the set of values of RACE, AAR, ALKY, CONPBS, SUPER, PERSON, and PROPTY for the individual for whom the probabilities of NONE, PROP, PERS and OTHER are to be calculated. The program reads these variables off an input data card with FORMAT 7F10.0, as above.

The first output provided is the set of values read for RACE, AAR, ALKY, CONPBS, SUPER, PERSON, and PROPTY. The second output, and the one of major interest, is the list of the probabilities of NONE, PROP, PERS, and OTHER.

One feature that is common to both programs is that if the value of one or more of the explanatory variables is not available, or if it is desired not to take into account the effects of one or more variables, it is possible to punch any negative value for the variable in question. The program will then automatically assign to the individual the average value (over our 582 observations) of that variable.

4. CAUTIONARY NOTES

There are a few words of caution worth stressing. These have to do not with the program itself, but with the interpretation of the results.

1. The sample used in the statistical analysis upon which these projections are based is described in some detail in the report of the statistical analysis. Briefly, it is essentially a random sample of inmates who had not been convicted of crimes that would have prevented their placement on the work release program and who served time in minimum and medium custody prison units in the South Piedmont administrative area. It therefore did not contain individuals convicted of sex offenses (D.O.C. crime codes 700, 701, 710, 711, 712), serious drug offenses (D.O.C. crime codes 804, 805, 806 resulting in a sentence of 4 years or more), or the public drunk offense (D.O.C. crime code 857). The sample is not representative of inmates who never served time in medium and minimum custody prison units but are confined throughout their imprisonment to the institutions of the prison system, i.e., the specialized youthful offenders and maximum and

close custody institutions. The sample also contained no women. As a result, the projections made here should not be applied to women, individuals who do not serve time in medium and minimum custody unit, or individuals who have been convicted of the above mentioned offenses.

2. Strictly speaking, each program gives probabilities for an individual with specified characteristics. Either program can be used to generate probabilities for a group of individuals by using the average values of the explanatory variables for that group. However, use of the programs in this way involves an approximation. This approximation may be fairly accurate for groups of individuals whose values of the explanatory variables are similar, but it may be poor for groups of individuals with dissimilar values of the explanatory variables.

3. The length of the follow-up period for which the predicted probabilities are relevant is not specified explicitly, and the evidence in the statistical analysis suggests that it is not terribly important, within reasonable bounds. The average length of the follow-up period in our sample was approximately three years, so the predicted probabilities can be thought of as being relevant for follow-up periods of approximately that length. The predicted probabilities are probably not of much use for follow-up periods shorter than, say, one year.

CHAPTER 4

CONCLUSIONS

This report has described the results of four separate analyses of recidivist criminal activity. The first analysis, reported in Chapter 1, was of the timing of return to criminal activity. Two dependent variables were considered in this analysis: the length of time after release from prison until the first conviction, and the length of time after release from prison until the first conviction resulting in a North Carolina prison sentence; denoted LTFCV and LTFPCV, respectively. The second analysis, reported in Chapter 2, was of the seriousness of recidivist criminal activity. The dependent variable used was the total length of all North Carolina prison sentences received by an individual during the follow-up period after release; denoted TTSENT. The third and fourth analyses, reported in Chapter 4, were of the type of recidivist criminal activity. In the third analysis the three possibilities analyzed were that the individual received no North Carolina prison sentence during the follow-up period; that the most serious conviction resulting in a North Carolina prison sentence was for a misdemeanor; and that the most serious conviction resulting in a North Carolina prison sentence was for a felony. These possibilities were denoted NONE, MISD, and FLNY, respectively. In the fourth analysis the four possibilities analyzed were that the individual received no North Carolina prison sentence during the follow-up period; that the most serious conviction resulting in a North Carolina prison sentence was for a crime against property; that the most serious conviction resulting in a North Carolina prison sentence was for a crime against a person; and that the most serious crime resulting in a North Carolina prison sentence was for a crime other than a crime against property or a crime against a person. These four possibilities were denoted NONE, PERS, PROP, and OTHER, respectively.

There were four explanatory variables that were found to be significant determinants of the dependent variables in each of the four analyses. These four variables were clearly revealed to be important determinants of recidivist criminal activity.

The first such variable was ALKY, which had a value of one if the individual's record indicated a serious problem with alcohol, or use of hard drugs, and which had a value of zero otherwise. To be an alcoholic or user of hard drugs was found to decrease LTFCV and LTFPCV and to increase TTSENT. Interestingly, while being an alcoholic or user of hard drugs increased the probability of both MISD and FLNY (relative to NONE), it increased the probability of MISD more than that of FLNY. Similarly, while it also increased the probability of PROP, PERS and OTHER (relative to NONE), it caused smaller increases in the probability of PROP than in the probability of PERS and OTHER.

The second explanatory variable that was found to be important was RACE, which had a value of one for whites and zero for non-whites. To be white was found to decrease LTFCV and LTFPCV and to increase TTSENT. It also increased the probability of MISD (relative to NONE or FLNY), and increased the probability of OTHER (relative to NONE, PERS, or PROP).

The third important explanatory variable was CONPBS, defined as the number of previous prison convictions. An individual with more previous prison convictions will tend to have a smaller value of LTFCV and of LTFPCV, and a larger value of TTSENT. The number of previous prison convictions does not seem to have much effect on the type of recidivist activity. To have a larger value of CONPBS increases the probability of both MISD and FLNY (relative to NONE), but by roughly equal amounts; similarly, it increases the probabilities of PERS, PROP and OTHER (relative to NONE), but, again, by roughly equal amounts.

The fourth important explanatory variable was AAR, the individual's age at release. An individual who is older at release will tend to have a larger value of LTFCV and LTFPCV and a smaller value of TTSENT. To be older causes roughly equal decreases in the probabilities of MISD and FLNY (relative to NONE). It also decreases the probabilities of PROP, PERS, and OTHER (relative to NONE); the largest decrease is in the probability of PROP.

One other variable was found to be significant in three of the four analyses performed. This was SUPER, which took on a value of one if the release from imprisonment was supervised, and zero if it was not. SUPER did not have a significant effect on TTSENT. However, an individual whose release was supervised will tend to have a larger value of LTFCV and LTFPCV. He will also have a smaller probability of both MISD and FLNY (relative to NONE), and smaller probabilities of PROP, PERS and OTHER relative to NONE. Supervision appears to decrease the probability of PERS more than that of PROP or OTHER.

Based on these results, it would be roughly accurate to say that the type of releasee about whom we should be least optimistic is a young, white alcoholic or hard-drug user, whose release was unsupervised, and who had a large number of previous convictions. Conversely, the type of releasee about whom we should be most optimistic is an old black who is not an alcoholic or user of hard drugs, whose release was supervised, and who had no previous convictions (particularly prison convictions).

In these conclusions, we have tried to emphasize the effects of those variables which appeared to have the strongest influence on the timing, seriousness and type of recidivist criminal activity. A number of other variables were found to have significant effects in some analyses, but not

in others. These effects have already been described in the individual chapters of this report, and will not be repeated here.

END