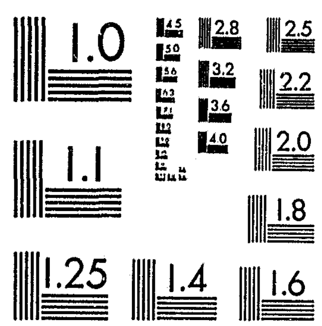

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

9/29/83

89617

Executive Summary

BANK ROBBERS - SURVEY OF BANKS*

William G. Saylor and Michael Janus

The increasing frequency of occurrence of armed bank robbery in the Washington, D.C. area and throughout the nation has caused grave concern among the nation's bankers and law enforcement personnel. As a result of this concern, all major financial institutions in the Washington, D.C. metropolitan area were surveyed in 1978 regarding various aspects of their experiences with the robbery situation. A great majority (246) of banks, savings and loans, and credit unions surveyed, responded. Initial figures indicate that bank robbery was a relatively common occurrence amongst financial institutions in the Washington area in the ten year period from 1968 to 1977. Over 62 percent of the banks responding were robbed at least once. Among the 246 financial institutions responding, there were 507 robberies in the ten year period. This averages to over two per institution.

In an attempt to ascertain the characteristics of banking institutions which contribute to their being robbed, three separate methodological approaches were utilized. The first, an ordinary least squares regression was used to analyze the rate of robberies per year. The second two were logit regression models which were used to analyze the probability of not being robbed versus the probability of being robbed one or more times. In all three models it generally appears that size and ease of accessibility are the two major factors associated with being robbed in the Washington, D.C. area. More specifically, we found that after controlling for the amount of time a facility had been open savings and loans had the highest probability of being robbed, followed by banks and then credit unions. We also found that facilities with more entrances (whether they were banks, savings and loans or credit unions) were more likely to be robbed, that facilities which had only direct entrances were more frequently robbed than those with other entrance arrangements, that facilities with only corridor

entrances were less likely to be robbed and that the robbery rate for facilities with both direct and corridor entrances mediated the robbery rates for direct only and corridor only facilities. Additionally, facilities with larger numbers of teller stations also tend to be associated with higher robbery rates. To reiterate, these findings seem to suggest that larger, more easily accessed and in general more convenient facilities have higher probabilities of being robbed. As far as security devices are concerned facilities with Security guards tend to be less frequently robbed while the findings relating to the number of visible surveillance cameras is confounded or inconclusive.

Since the extent of security measures was somewhat confounded with the physical and financial size of the institution, the present study could not contribute much useful knowledge to the efficacy of security devices. Future investigations should attempt to better distinguish these issues by obtaining realistic measures of institution size in terms of financial backing, cash and customer flow in order to better differentiate the importance of these factors. In addition, better characterization of the institution's geographic location, by such indicators as the local socio-economic conditions (possibly available from census information) would appear to be salient additions to the robbery equation.

*The data analyzed in this report was collected for the Federal Prison System's Office of Research by Peter L. Nacci in collaboration with Special Agent Clyde Whitson of the Federal Bureau of Investigation with the cooperation of the District of Columbia Bankers Association.

BANK ROBBERS - SURVEY OF BANKS

William G. Saylor and Michael Janus

Introduction

This study examines the characteristics of banks in the Washington, D.C. area which contribute to their being robbed. The study reflects a cooperative effort between the Federal Prison System and the Federal Bureau of Investigation which began in 1976.

The effort was a reaction to the alarming increase in the rate of bank robberies both in Washington and in the rest of the country at that time. The facets of bank robberies which will be examined are the characteristics of the banks which were robbed. This examination may contribute useful knowledge to the nature of the bank robbed or the robbery situation which may, in turn, contribute to the advancement of physical deterrence to reduce the number of bank robberies.

The analysis is based on responses to a survey distributed to all financial institutions (Banks, Savings and Loans, and Credit Unions) in the Washington, D.C. area in 1978. The great majority of institutions responded (N=246). Those not responding generally did not for reasons of inapplicability to the robbery situation (e. g. a private credit union which handles no cash.)

The substantive areas probed by the survey included:

- Architectural characteristics of the bank, such as how many customer entrances it has, or whether or not these entrances are on the ground floor.

- Accessibility to transportation such as parking availability, nearness to major traffic arteries.

- Security features, such as the presence of guards, bullet proof glass, and surveillance cameras.

- Personnel characteristics such as the proportion of male to female tellers.

- Robbery characteristics such as the number of robberies per year for the ten year period from 1968 to 1977.

All of the above areas have some intuitive relationship to bank robbery. FBI Special Agent Clyde Whitson noted in a "Washington Post Magazine" article in 1976 that the attitude of the "friendly banker" and the structurally more inviting aspects of banking may be contributors to the increased frequency of bank robberies. The FBI has noted that getaway cars are used in about 80 percent of all robberies lending credence to the idea that rapid getaway routes may be an important consideration contributing to the robbery situation. There is some indication that female tellers are more likely to be the target of predominantly male robbers. As far as bank security is concerned, banks have been under heavy pressure from police agenices to increase their protecion, while banking groups such as the District of Columbia Bankers Association have been faulting the Criminal Justice System for the higher incidents of robbery.

One resolution to the problem will be the determination of physical characteristics which provide the type of security that police organizations recommend at a minimal cost, while providing the kind of environment which banks find appealing to draw in business. What needs to be ferreted out therefore are the kinds of unobtrusive physical devices which do the best job of securing the facility. By focusing on the implementation and maintenance of these devices institutions will spend less money on mechanisms which are not as adept at protecting particular facilities given, for example, the type of facility, location and other characteristics. This would also allow for a concentration of efforts on the development of security devices which best meet the environmentally aesthetic and ease or accessibility characteristics desired by the banking organizations.

What we hope to determine from the present analysis is an indication of the physical characteristics which existed at the time this questionnaire was completed that have been either more or less frequently associated with banks which have been robbed. In this sample of banking facilities we expect to find (hypothesize) that more convenient facilities (larger facilities with large numbers of teller stations, walk up and drive in stations, longer hours, ample parking and quick access to highways, and a friendly less secure looking environment) will be more susceptible to robberies.

Some Univariate Descriptive Results

Some interesting descriptive characteristics excerpted from responses to the survey are described below:

- Among the 246 financial institutions surveyed, there were 507 robberies over the ten year period covered by the questionnaire. This averages to over two per institution.
- 62% of all banks surveyed were robbed at least once. 12% of the banks indicated that they had been robbed more than four times.
- Only 12% of the banks indicated that they used guards (either armed uniformed, unarmed uniformed or plain clothes).
- 73% of the banks utilized some form of visible surveillance cameras.
- 44% had bullet resistent barriers which separated tellers from customers.
- For the FY 1976 83% of all responding banks reported either a majority of female tellers or all female tellers.

Multivariate Modelling Strategy

The analysis which follows consists of one ordinary least squares regression and two logistic (logit) regression models. The design of the survey questionnaire precluded (without discarding a large amount of data) the parsimony achieved by one inclusive model. Each of the three models presented operationalizes the measure of bank robbery incidents in a different manner.

For the first two models we discuss, we have limited the scope of the characteristics to those of a physical nature since the questionnaire asked about (and these analyses model) bank characteristics which existed during the Fiscal Year 1976 while the robbery information was collected for a ten year period ranging

from 1968 to 1977. In order to avoid wasting all of the data relating to the number of robberies prior to Fiscal Year 1976 we opted to look at the relationship or association between banking characteristics and the number of robberies per year in the first model, and the occurrence of no robbery, as opposed to any number of robberies, in the second model we present. While it does not make any logical sense to talk about banking characteristics present in Fiscal Year 1976 causing robberies which occurred prior to that year these models do allow us to look at any association which might exist between banking characteristics and the occurrence of robberies, if we can assume that the characteristics of banks which we choose to look at have remained the same over the ten year period. For this reason, we chose to look only at the physical characteristics since it is more likely that these would have remained constant over the ten year period, whereas many of the other characteristics are much less stable and are thus less likely to have been the same (in previous years) as they were reported to be in Fiscal 1976 (e.g., ratio of male to female tellers). How realistic this assumption is (about the stability of the physical characteristics over a ten year period) can probably be best assessed by members of the banking community.

The variables were selected for their intuitive (theoretical) appeal, though it was additionally necessary to weight this selection criteria by the content quality of the items (i.e., the degree of ambiguity in the item, the amount of variance in the item, and the amount of missing data). Table 1 shows the variables used in the ordinary least squares regression analysis along with their distributional characteristics.

The Ordinary Least Squares Regression Model

While the logit models (described later) will address what contributes most to being robbed at least once as opposed to not being robbed at all, they will not tell us anything about what contributes most to being robbed multiple times.

To assess this we created a rate variable, the ratio of the number of times robbed to the number of years (out of the ten observed) the facility had been open. Since ratio dependent variables frequently relate to independent variables in a nonlinear (in the variables) manner we used the maximum likelihood method suggested by Box and Cox (1964) for specifying the correct functional form for which the parameters should be estimated. This nonlinearity stems from the positive skewness of ratio variables, the transformation reduces the skewness creating a more symmetric distribution. The independent variables (except for the number of cameras and the number of teller stations) were all categorical variables and were coded in an appropriate manner (see Table 1). The number of cameras and the number of teller stations were transformed by the natural log to alleviate the positive skewness in these measures. Table 2 shows the results of the estimation procedure. The equation fits the data reasonably well. The F ratio (the ratio of the explained variance to unexplained variance) indicates that the explained variance is eleven times larger than the unexplained variance. This ratio tells us that we can reject the null hypothesis that all the coefficients of the equation are not statistically different from zero, or in other words, the hypothesis that the independent variables make no statistically significant contribution to explaining the variance of the dependent variable. The critical F value at $p = .001$ for 4 and 213 degrees of freedom is 3.27. Thus, we are 99% certain that we are correct in rejecting the null hypothesis and hence that one or more of the independent variable coefficients is different from zero. The equation explains 29% of the variance in the dependent variable. From the T ratios (B/SE) in Column 2 of Table 2 we see that only three of the dummy variables - direct entrance, bank type facility, savings & loan type facility - and the interval variable indicating the number of teller stations are statistically (at $p < .05$) different from zero. Looking at Column 1 at the bottom of Table 2 we see that

the largest unstandardized affect was due to savings and loan facilities. The second largest affect was due to bank facilities. Since these are dummy variables, and the reference variable (vector) is the credit union category of the institution type (the parent variable), this indicates that the rate of robberies for banks and savings & loans is significantly greater than it is for credit unions. Taking the difference between the unstandardized coefficients for banks and savings & loans (.792 - .736 = .0561) gives the difference in mean robbery rates between these two types of institutions. Taking the ratio of this difference between the coefficients to the square root of the product of the mean square error of the regression equation and the sum of the inverse of the number of cases in each dummy variable category ($.0561 / \sqrt{.8092 * (1/105 + 1/68)} = .401$) yields a measure of the statistical significance of the difference between these two coefficients. The ratio is less than two and we therefore cannot reject the hypothesis (at $p = .05$) that there is no difference in the robbery rate for these two types of institutions. The facility types are followed in importance by facilities with direct entrances. Looking at the standardized affects in Column 3 we see that while the type of facility is still the most important factor the number of teller stations had a slightly greater impact on the robbery rate than did direct entrances. Since the number of teller stations is also significantly associated with robbery rates, this indicates that the greater the number of teller stations the greater the robbery rate. This seems to suggest that it is the size of the facility which makes a financial institution more susceptible to robbery, not the bank-savings and loan distinction. Thus, after we control for the size of the facility (as indicated by the number of teller stations) there is no statistical difference in the robbery rate for banks and savings and loans. However, controlling for the type of facility still leaves the size of the facility making a statistically significant impact on robbery rates. This appears to fit with the hypothesis that more convenient facilities, be they bank or savings & loan, will be more susceptible to robbery. It is also plausible that bank robbers associate larger banks (in as far as the number of teller stations is a proxy for a banking institution's physical size) with larger deposits, or with greater anonymity.

The Logistic (Logit) Regression Model

Alternatively, we could choose to look at how these characteristics tend to be associated with being robbed any number of times (regardless of the exact number of robberies) as opposed to not being robbed during the ten year period observed. The dependent variable is a simple dichotomy indicating whether or not the facility has been robbed. Thus the dependent variable takes only one of two values, one if the facility has been robbed one or more times, and zero if the facility has never been robbed during the observed ten year period. We propose this second model primarily as a comparative link between the first ordinary least squares model and the third model we present later (a logit analysis of whether facilities were robbed in 1975 or 1976). While the variation in the number of robberies over the ten year period is sufficient for analysis by ordinary least squares regression there is not a sufficient amount of variation in the number of robberies in the third model, where we limit the measure of robberies to those occurring in 1975 and 1976, to allow for analysis by ordinary least squares. Using logistic regression to re-estimate Model 1 simply as the occurrence or non-occurrence of a robbery over the ten year period allows for a rough (non-statistical) comparison of the occurrence or non-occurrence of robberies over the ten year period to robbery-nonrobbery over the two year period (each being controlled for time). The logit regression models will allow us to assess whether the factors which distinguish simply between facilities that have been robbed and those that have not been robbed are the same as the factors which are associated with multiple robberies.

The analytic procedure used is one developed for analysis of dichotomous

dependent variables such as whether a facility was robbed or not robbed. Before we discuss the findings of this model we will briefly describe the nature of the analytic technique. Logit analysis was developed to overcome the problems (violation of the normal distribution of errors assumption, i.e., the normal distribution of observed Y values about the value of Y predicted for each value of X) involved in using ordinary least squares regression analysis on dependent variables which can take only one of two values. This is accomplished by creating conditional odds ratios out of the dependent variable. The dependent variable then becomes the natural logarithm of the conditional odds of falling in one of the categories rather than the other. Odds are computed simply as the ratio of the number of cases in one of the categories to the number of cases not in that category. What makes the odds "conditional" is that the ratio of the number of observations in one of the categories of the dependent variable to the number of observations not in that category is computed for each distinct covariate pattern (i.e., for each group of one or more observations, where the groups are defined by the criterion that each observation in a group has an identical profile - exactly the same combination of values on all the independent variables.) Thus the conditional odds tell us the proportion of times we observe the occurrence of one category relative to the other category within each possible combination of independent variable values.

For example if we had 100 banks and 75 were robbed while 25 were not, the marginal odds (that is, the odds of being robbed or the robbery risk we would assume each bank had if we didn't know anything else about what determines whether or not a bank will be robbed) would be 75:25 = 3:1 and the odds ratio is therefore equal to 3, and so we would guess that three banks would be robbed to every one bank that is not robbed. However, if we can find some bank characteristics which are more or less frequently associated with banks that are robbed and if, for each distinct covariate pattern, we compute a separate odds ratio of the facilities which were robbed to those that were not robbed then, these odds ratios (the conditional odds) will differ for each group of observations (each covariate pattern) thereby creating an interval level dependent variable which is unbounded on either side. In other words, the scale created can theoretically take on any real numerical value from minus infinity to plus infinity. If the bank characteristic was say, for example, the bank's location and we recorded two possible values, urban and non-urban, sampling (for simplicity) an equal number of facilities within each characteristic (50 urban and 50 non-urban) and we observed say 45 robberies among the urban banks and 30 among the non-urban banks, the conditional odds would be 45:5 = 9:1 = 9 and 30:20 = 3:2 = 1.5 for urban and non-urban banks respectively. Thus, in this hypothetical example we would expect 9 urban banks to be robbed to every one which is not robbed and on the average 1 1/2 non-urban banks to be robbed for every one non-urban bank which is not robbed. If we did not know the location of the banks our best guess as to whether or not a particular bank would be robbed would be the marginal odds ratio, that is, we would predict for any bank that three banks would be robbed out of every four. This would obviously result in some bad predictions since it would under-predict robberies in urban areas while over-predicting those in non-urban areas. By knowing the bank's location, on the other hand, we would do a much better job of predicting whether each bank had or hadn't been robbed, since the odds of being robbed for urban banks would be 6 times greater than the odds for non-urban banks.

Given a set of plausible characteristics the stepwise logistic regression program will choose the explanatory variables one by one based on their ability to choose the appropriate category of the dependent outcome variable. Just as in a stepwise ordinary least squares regression, if one or more of the explanatory variables explain the same part of the dependent variable only one will be chosen. Those that discriminate best are chosen first. After each independent

variable is chosen all of the independent variables are re-evaluated by this discriminating criterion. If the discriminating ability of a particular independent variable is the same as one that is already chosen then that variable will not be selected since at some earlier point in the selection procedure one or more of the variables already selected explained all that the variable being evaluated could and more.

As in Model 1, we constrain our attention to only those physical characteristics which might be most likely to remain constant over the ten year period. Recall that while the questionnaire asked for the number of robberies over a ten year period some of the banks (about 30%) were not open the full ten years. Since we are dealing with a dichotomous dependent variable in logistic regression models we cannot control for the differing amounts of time facilities having been open by creating a rate (number of robberies per year). It was therefore necessary to include the number of years the bank was open out of the ten year period in order to control for the amount of time the bank was at risk. Stated differently, we are not interested in knowing how well we can explain the occurrence of a robbed facility by the number of years it has been open. Naturally the longer an institution is open the more likely it is to have been robbed. What we are interested in determining are the characteristics of banks which tend to be associated with being robbed or not robbed, aside from the amount of time the faciltity has been open. This was done by entering the variable which indicates the number of years each facility was open into the equation before any other variables. The model was specified in this manner since every variable is entered only if it can explain some aspect of the dependent variable which has not yet been explained by a variable which entered prior to it. Each coefficient represents the contribution of that variable to the explanation of the dependent variable after the shared contribution (the explained portion of the dependent variable which it shares or has in common with the other independent variables already selected) has been removed. In other words, all of the coefficients are partial coefficients, just as in typical ordinary least squares linear regression. In this way we let the number of years the bank is open explain as much as it can about whether a facility is robbed or not robbed, and hence, time at risk is held constant for all the other explanatory variables.

Since we are more interested in the impact of the characteristics of the institutions on robberies rather than the type of institution (Bank, Savings & Loan, or Credit Union) we also entered the institution type simultaneously in the first variable selection step along with the number of years open.

Table 3 displays the variables used in this model and their distributional characteristics. The bottom of Table 4a shows the stepwise results of the logistic regression. This Table shows two measures of the fit of the equation to the data at each step (after the independent variable making the greatest contribution to the explanation of the dependent variable has been included). The two measures of fit are the improvement chi square and the goodness of fit likelihood ratio chi square. The improvement chi square is computed as the difference between the goodness of fit likelihood ratio chi square at each step of the solution.

The improvement chi square tests the hypotheses that the term (variable(s)) entered at that step significantly improves the prediction of the correct category of the dependent variable. Stated differently, we are considering whether the addition of the variable yields a significant reduction in the chi square goodness of fit. The larger the improvement in prediction the larger the drop in the chi square goodness of fit and therefore the larger the chi square improvement. We interpret the chi square improvement probability (the P-value to the immediate right of the chi square improvement value at the bottom of Table 4a) in the usual significance test manner, that is, the smaller the value the better.

For this model we see that all the variables selected are significant at

the 5% level, we are 95% certain that these variables improve our prediction of the dependent variable in the population. Since the items in the stepwise procedure are selected one by one on the basis of whether they, independently, can significantly improve prediction at that selection step, as determined by the value of the improvement chi square and its accompanying P-value, and since the items are selected in order of the magnitude of their predictive improvement (larger improvements preceding smaller ones) the items not selected are rejected by the test of the hypothesis that they significantly improve the prediction of the category of the dependent variable with the degree of confidence (at the probability level) indicated by the chi square improvement P-value.

The chi square goodness of fit tests the hypothesis that the model at that step fits the data adequately. A small value of the chi square goodness of fit probability (the P-value to the right of the chi square goodness of fit value at the bottom of Table 4a) indicates that the model does not fit the data. Conversely, a large P-value indicates that the model is consistent with the data to within sampling fluctuations. The chi square goodness of fit statistic can be interpreted as the amount of variation in the log of the conditional odds of falling in the robbed rather than the not robbed category which is unexplained by the model. The chi square goodness of fit P-value can be interpreted as the probability that the differences obtained between the observed frequencies and the fitted frequencies (those specified by the model) could have arisen by chance, given that the model is correct.

We can see from the bottom of Table 4a that this model represents a very good fit to the data since we are likely to see differences (between the observed and fitted values) of this magnitude about 90% of the time. It is not possible to compute a coefficient of determination (R^2) for these types of models since that statistic is only computable for models which are linear in their parameters. A logistic regression is non-linear in the parameters. Dumouchel

(1976) has, however, come up with a statistic which is completely analogous to the multiple R^2 in ordinary least squares regression. This analog "R^2" indicates the percentage reduction in the magnitude of the likelihood value due to the specification of the model. From this analog "R^2" we can see that the constant, the number of years open out of the 10 observed, and the type of facility explain 25% of the prediction error under the null hypothesis (that all the independent variable coefficients equal zero). Since we are interested in the amount of variance we can explain over and above the 25% due to these factors we wish to statistically control for them by computing partial analog "R^2" values (partialing out this portion of the dependent variables explained variation). These partial "R^2s" are displayed to the right of the chi square measures at the bottom of Table 4a. These values indicate the percent of predictive error explained when we have statistically controlled for the terms entered at step 0. We see that the type of entrance, number of cameras and the interaction term between the type of entrance and the type of facility explain 3%, 2.5% and 3.5% of the prediction error, respectively. Cumulatively this amounts 3%, 5.5% and 9%. Again, the total "R^2" indicates that overall we can explain about 32% of the prediction error while the partial "R^2" indicates that after we have statistically controlled (held constant) the amount of time at risk and the type of facility the additional terms allow us to explain 9% more prediction error over and above that explained by the terms which we are statistically holding constant.

From the top of Table 4a (Column 2) we can see that all of the independent variable main effects, the number of years open, the type of institution, number of visible surveillance cameras and the type of entrance are statistically different from zero ($p < .05$ for a two tail test) the latter 3 variables contributing significantly to the explanation of which facilities are robbed after controlling for the amount of time the facilities have been open. With the exception of the

number of cameras the direction and magnitudes of the coefficients fit with our theoretical proposition that the more convenient and friendly looking facilities would be robbed more frequently. Savings & loans are in fact about 2 3/4 times more likely to be robbed than non-savings & loan institutions (Column 3 of Table 4a). Converting the raw metric coefficients (Column 1 of Table 4a) to probabilities, (Column 4 of Table 4a) we see that savings & loans are 25% more likely to be robbed than non-savings & loans. Banks in this sample have a 3% probability of being robbed while credit unions are 28% less likely to be robbed than noncredit unions. Looking at the type of entrance we see that again the notion of convenience seems supported since facilities which are accessible only through a corridor are 16% less likely to be robbed than are facilities which are accessible directly from the streets or from the streets and from a corridor, even after controlling for the amount of time the facility has been open and the type of facility (bank, savings & loan, or credit union). Table 4a also indicates that the coefficient for the number of visible surveillance cameras is the only item which is contrary to what we expected. While one might expect the number of cameras to be negatively related to the odds of being robbed this is not the case. A greater number of cameras tend to be associated with greater odds of being robbed. Column 3 of Table 4a indicates that with the addition of each camera the odds of being robbed increases nearly one and a half times. Probabilistically speaking (Column 4, Table 4a) with the addition of each camera the probability of being robbed increases by about 10%. It seems that the number of cameras, (as with the number of teller stations in Model 1) is most likely a proxy for some other physical, financial or historical characteristic(s) (e.g. the location of the banking facility, the number of previous robberies at that banking facility or of banking facilities in the same locality, the physical or financial size of the facility) which tends to be associated with facilities which have been robbed.

In Table 4a we see that some of the interaction effects between the type of institution and the type of facility are statistically significant. This indicates that the magnitudes of the coefficients for savings & loans and for credit unions (i.e., the effect being a credit union or savings & loan has on the odds that the facility will be robbed) differ depending on their type of entrance. Writing separate regression equations for each type of facility (Table 4b) we see that for direct entrance savings & loans the odds of being robbed are nearly two and one half times greater than the odds for banks with a direct entrance. In probabilities we see that the probability of being robbed for savings & loans and credit unions with direct entrances is about 20% while the probability for banks with direct entrances is 10% (we don't have much confidence in these coefficients from the separate facility type equations since they are based on the direct entrance main effects coefficient and the main effect was not very stable). The odds of being robbed are about equally likely for banks with either direct or direct and corridor entrances while the odds of being robbed are considerably less for banks with corridor entrances only. For savings & loans and credit unions with direct and corridor entrances the odds of being robbed are much smaller than the odds for these same type facilities with direct entrances (the probabilities of being robbed being about 30% smaller). The odds of being robbed for savings & loans and credit unions with only corridor entrances mediate the odds for this type of institution when they have either direct only or both corridor and direct entrances.

Although wasteful of data a more correct specification of the model would be the assessment of banking characteristics which predict only those robberies occurring during the Fiscal Year 1976 (July 1, 1975 through June 30, 1976). Recall that the questionnaire design, while asking for the number of robberies which occurred each year for ten years, asked for the characteristics of the banking facilities for only the Fiscal Year 1976. We could not precisely isolate

only those robberies occurring during Fiscal Year 1976, since the robberies were recorded in yearly aggregates, therefore we looked at those robberies which occurred during 1975 and 1976. We suspect that any differences in the characteristics of banking facilities between January 1, 1975 to June 30, 1975 and July 1, 1976 to December 31, 1976 and the characteristics which existed during Fiscal Year 1976 obtained by the survey would be minimal. The dependent variable in this third model is once again a dichotomy with zero indicating no robbery and one indicating those facilities which were robbed one or more times during 1975 or 1976.

Theoretically we would expect all the variables in Table 5 to have some impact on (or association with) whether or not facilities are robbed. It might initially seem redundant that we have included the number of years open out of the ten observed in addition to the proportion of 1975 and 1976 open for business. Our reasoning is that while the latter will allow us to control for the amount of time at risk to robbery the former will allow us to assess the impact of differences in architectural design and other systematic changes that might characterize buildings constructed at different periods in time. Unless there are redundancies among the independent variables, problems stemming from unrepresentativeness of banking facilities due to sample bias, or measurement errors in the variables due to the questionnaire design, we should observe all of the variables from Table 5 selected in the stepwise logistic regression. Table 6 displays the items selected out of the 20 described in Table 5 and indicates that this is not the case. The coefficients of the variables which were not selected were not statistically different from zero. The reasons for the item rejections might be: that a sample of only Washington Metropolitan Area banking facilities differ in some respect from banking facilities in other areas (and the population of banking facilities in general); that the measurement of the important issues and characteristics was too coarse or that measurement error was injected into our empirical measures in some other manner; or that the characteristics not selected in fact have no relationship with whether or not a facility is robbed.

Looking at the bottom of Table 6 we see the fit of the model. The chi square goodness of fit P-value of almost .17 indicates a mediocre fit for this number of observations. The total R^2 (R^2_3) indicates that we have explained about 13% of the predictive error. Controlling on the amount of time open for business during 1975 and 1976, knowing the type of facility, the number of customer entrances and the type of entrance we explain about 11% of the predictive error. Notice that while the coefficient for the variable indicating proportion of 1975 and 1976 open for business (time at risk) appears to be very large, it is very close to the constant in absolute magnitude. This is because the number of facilities open during the 1975 and 1976 period was relatively small and the item therefore did not contain much variation. Constraining this item in the solution from the initial step results in a close relationship between the constant term and the coefficient for the time open variable. Also note that the logit scale score (computed as the sum of the products of the raw metric coefficients and their associated variables) for all but the facilities which opened during 1975 or 1976 receive the difference between the constant and the time at risk coefficient (-12.858 + 10.674 = -2.184). This is due to the fact that the time at risk variable is an indicator of the proportion of time open and therefore the value is equal to one for all these observations. Those facilities which opened during that two year period received a score equal to something less than one and therefore the contribution of these two terms to the logit scores of these observations is smaller. This point appears even more lucid if we look at the transformation of the coefficients to probabilities. In the case of the multiplicative model the constant is near zero which (after taking the product of the constant and the time at risk coefficient) brings the very large term for time at risk down to a fractional value.

Substantively we find that this model, in terms of the items which had a statistically significant impact on the number of robberies in 1975 and 1976 is very similar to the preceding model in which we were predicitng robberies over the ten year period. Savings & loans are almost twice more likely to be robbed, banks about a third more likely. Probabilistically speaking this translates to savings and loans being 14% more likely, banks 6% more likely and credit unions 20% less likely of being robbed. Direct only entrances have greater odds of being robbed than direct and corridor entrances and direct and corridor entrances in turn have greater odds than corridor only entrances (direct only entrances multiplying the odds of being robbed about one and eight tenths times more than non-direct only facilities). Intuitively and logically it makes a great deal of sense that the odds of being robbed for facilities with both direct and corridor entrances mediates the odds of being robbed for facilities with direct only or corridor only type entrance. We also note from Table 6 that with each additional entrance the odds of being robbed increased by about a factor of two.

Our hypotheses dictate that all of the independent variables in Table 5 should be included in the model. Since a misspecification of the model (due to not including in the equation all relevant explanatory variables - items which in reality impact on the occurrence of robbery) could result in biased and inconsistent estimates of the coefficients in the model, we re-estimate the model simultaneously including all the theoretically important explanatory variables.[1] Although we included the type of facility in the model displayed in Table 6 we exclude this item from the simultaneous estimation of the model since we are more interested in the impact of the characteristics of banking

[1] Both bias and inconsistency in the estimated coefficients are undesirable properties. A biased estimate will not, on the average, lead to the correct value of the coefficient; an inconsistent estimate does not, on the average, tend more toward the true value of the coefficient as the sample size increases.

facilities on the occurrence of robbery than we are the impact of the type of facility. We see from Table 7 that excluding the type of facility results in statistically significant coefficients for two additional explanatory variables (number of teller stations and number of types of guards). All of the contribution of these two variables to the prediction of robbery in the model displayed in Table 6 was completely absorbed (accounted for) by the explanatory variable indicating the type of facility. This makes sense since certain types of bank facilities are more and others less likely to have guards or large numbers of teller stations.

The coefficients of the items selected in the stepwise estimation remain essentially the same, relative to one another, in the simultaneous solution. We also see that each additional teller station results in an expected increase of 3% in the probability of being robbed and each additional type of guard results in a 22% decrease in the probability of being robbed. The R^2 at the bottom of Table 7 indicates we have explained about 15% of the predictive error with this model. Since we have included more variables in this equation we would expect the R^2 to be larger than that found for the estimation of the equation displayed in Table 6. While the chi square goodness of fit for the model displayed in Table 7 is smaller than the chi square goodness of fit for the model displayed in Table 6 the larger number of variables requries an equally large number of degrees of freedom resulting in a poorer fit, a chi square goodness of fit P-value of only .09.

Table 8 displays the results obtained by re-estimating the model while including only those explanatory variables found statistically significant in Table 7. This was done to increase the sample size by allowing the inclusion of observations which had missing values on one or more of the items which did not attain statistical significance in the simultaneous estimation of the model (due to the listwise deletion of observations with missing values). These excluded cases could be different from those with complete information and the results displayed in Table 8 allow us to assess this possibility to some extent.

It seems from Table 8 that the observations which were excluded from the simultaneous estimation displayed in Table 7 did not have any noticeable impact on the relative values of the coefficients. Each additional customer entrance multiplies the odds of being robbed about twice. The direct entrance multiplies the odds of being robbed about two and one half times. The direct and corridor entrance again has the second highest impact of the type of entrances, followed by the corridor only entrance which again has the smallest odds of being robbed. Each additional teller station increases the probability of being robbed by about 2% while each additional guard type results in a 21% decrease in the probability of being robbed. These results are all almost identical to those of Table 7. The bottom of Table 8 displays the fit of the model and the increments to R^2 made by each of the explanatory variables. The type of entrance, number of customer entrances, number of teller stations and number of guard types each controlled for those variables which preceded its entry into the equation and all of the variables having been controlled for the amount of time open explain about 9%, 1.8%, .9% and 1.1% of the prediction error respectively. This results in a total R^2 of approximately 15% and an R^2 controlling on the amount of time at risk of about 13%. The P-value associated with the chi square goodness of fit for the final step, equal to .34, indicates a reasonably good fit for this number of observations.

Summary and Conclusions

The increasing frequency of occurrence of armed bank robbery in the Washington, D.C. area and throughout the nation has caused grave concern among the nation's bankers and law enforcement personnel. As a result of this concern, all major financial institutions in the Washington, D.C. metropolitan area were surveyed in 1978 regarding various aspects of their experiences with the robbery situation. A great majority (246) of banks, savings and loans, and credit unions surveyed, responded. Initial figures indicate that bank robbery was a relatively common occurrence amongst financial institutions in the Washington area in the ten year period from 1968 to 1977. Over 62 percent of the banks responding were robbed at least once. Among the 246 financial institutions responding, there were 507 robberies in the ten year period. This averages to over two per institution.

In an attempt to ascertain the characteristics of banking institutions which contribute to their being robbed, three separate methodological approaches were utilized. The first, an ordinary least squares regression was used to analyze the rate of robberies per year. The second two were logit regression models which were used to analyze the probability of not being robbed versus the probability of being robbed one or more times. In all three models it generally appears that size and ease of accessibility are the two major factors associated with being robbed in the Washington, D.C. area. More specifically, we found that after controlling for the amount of time a facility had been open savings and loans had the highest probability of being robbed, followed by banks and then credit unions. We also found that facilities with more entrances (whether they were banks, savings and loans or credit unions) were more likely to be robbed, that facilities which had only direct entrances were more frequently robbed than those with other entrance arrangements, that facilities with only corridor entrances were less likely to be robbed and that the robbery rate for facilities with both direct and corridor entrances mediated the robbery rates for direct only and corridor only facilities. Additionally, facilities with larger numbers of teller stations also tend to be associated with higher robbery rates. To reiterate, these findings seem to suggest that larger, more easily accessed and in general more convenient facilities have higher probabilities of being robbed. As far as security devices are concerned facilities with Security guards tend to be less frequently robbed while the findings relating to the number of visible surveillance cameras is confounded or inconclusive.

Since the extent of security measures was somewhat confounded with the physical and financial size of the institution, the present study could not

contribute much useful knowledge to the efficacy of security devices. Future investigations should attempt to better distinguish these issues by obtaining realistic measures of institution size in terms of financial backing, cash and customer flow in order to better differentiate the importance of these factors. In addition, better characterization of the institution's geographic location, by such indicators as the local socio-economic conditions (possibly available from census information) would appear to be salient additions to the robbery equation.

References

~~Box, G.E.P. and D.R. Cox "An Analysis of Transformations,"~~ Journal of the Royal Statistical Society, Series B, Vol. 26, 1964, pp.211-243.

Dixon, W.J. and M.B. Brown (Eds.) (1979) Biomedical Computer Programs P-Series 1979. University of California Press.

DuMouchel, William H. (1976) "On the Analogy Between Linear and Log-Linear Regression." Technical Report No. 67, University of Michigan, Department of Statistics.

Huang, Chung-Liang and Leonard C. Moon (1978) "A Box-Cox Procedure Estimation Routine User Guide." Department of Agricultural Economics, Georgia Experiment Stations

TABLE 1

CODING SCHEME FOR CATEGORICAL VARIABLES IN MODEL 1

Variable	Code	Category Label	Frequency
Parking	-2	Poor	54
	-1	Fair	79
	1	Good	74
	2	Excellent	38
Bullet Resistent Barrier Separating Tellers from Customers	0.	No	138
	1.	Yes	109
Possible to Jump Over Teller Counter	0.	No	98
	1.	Yes	238
Direct Entrance to Street (Only)	0.	No	134
	1.	Yes	111
Bank	0.	No	142
	1.	Yes	105
Savings & Loan	0.	No	179
	1.	Yes	68
Residential/Commerical Area	0.	No	151
	1.	Yes	95

TABLE 2

RESULTS OF MODEL 1

Variable	Means	Standard Deviations
Robbery Rate	.405	.296
Parking Availability	.002	1.689
Bullet Resistent Barrier	.534	.500
Jump Counter	.583	.494
Direct Entrance	.480	.501
Bank	.453	.499
Savings & Loan	.287	.453
Residential/Commercial Area	.391	.490
Ln (No. Cameras)	.201	1.169
Ln (No. Teller Stations)	1.516	.866

Variable	B	B/S.E.	β
Parking Availability	- .017	- .449	-.027
Bullet Resistent Barrier	- .115	- .769	-.054
Jump Counter	.149	1.016	.069
Direct Entrance	.377	2.650*	.177
Bank	.736	3.116*	.344
Savings & Loan	.792	3.580*	.336
Residential/Commercial Area	.017	.132	.008
Ln (No. Cameras)	.005	.075	.006
Ln (No. Teller Stations)	.240	2.431*	.195
Intercept	-2.796	-14.195*	

* Significant at or below the 5% level for a one tail test.

$F = 11.0872$ $p < .001$ $F_{critical,\ 9,213} = 3.27$ $p = .001$

Standard Error of Estimate = .899

R^2 adjusted for Degrees of Freedom = .29

$\lambda = -.5$ (λ is the maximum likelihood estimate of the Y (robbery rate) transformation parameter (i.e., $Y^\lambda = \beta X + \varepsilon$). The power transformation parameter of -.5 corresponds to an inverse square root transformation.)

TABLE 3

ALL VARIABLES CONSIDERED FOR INCLUSION IN MODEL 2

Dependent Variable	Frequency		Total N	Marginal Proportion Robbed
One or More Robberies during Ten Year Period	No	96	243	.605
	Yes	147		

Interval Independent Variables	Min.	Max.	Mean	Standard Deviation
Years Open Out of Ten Observed	1.000	10.000	8.480	2.769
No. of Visible Cameras	0.000	9.000	1.683	1.458

Categorical Independent Variables		Frequency
Bullet Resistent Barrier	No	134
	Yes	109
Type of Entrance	Direct Only	111
	Corridor Only	84
	Direct & Corridor	48
Type of Facility	Bank	103
	Savings & Loan	68
	Credit Union	72

TABLE 4a

RESULTS OF MODEL 2

	B	B/S.E.	e^B (Multiplicative Model)	P (p=.5) (Probability Model)
Constant	-3.356	-4.698*	.35	
Years Open	.419	5.582**	1.52	.10
Type Institution				
Bank	.140	.408	1.15	.03
Savings & Loan	1.001	2.431**	2.72	.25
Credit Union	-1.141	-2.623**	.319	-.28
V23 Cameras	.39	2.485*	1.48	.10
Type Entrance				
Direct Only	.396	1.228	1.49	.10
Direct & Corridor	.252	.628	1.29	.06
Corridor Only	- .648	-1.689**	.523	-.16
Interaction Terms for Type Ent. X Type Inst.				
Direct X Bank	.010	.025	1.01	.00
Direct & Corr. X Bank	.218	.456	1.24	.05
Corridor X Bank	- .228	- .485	.796	-.06
Direct X S & L	.480	1.041	1.62	.12
Direct & Corr. X S & L	-1.622	-2.862*	.20	-.40
Corridor X S & L	1.142	1.740	3.133	.28
Direct X C.U.	- .49	- .966	.613	-.12
Direct & Corr. X C.U.	1.404	2.08 *	4.071	.35
Corridor X C.U.	- .914	-1.799	.401	-.23

* $p < .05$ Two Tailed Test ** $p < .05$ One Tailed Test

Step	Variable Selected	df	Improvement	P-Value	df	Goodness of Fit	P-Value	R^2 Value
Marginal	Constant (Null HYP. $B_1=\ldots=B_5=0$)				113	326.293	.000	
0	Constant, Years Out of 10 observed, type of Facility	3	215.511	.000	110	110.782	.461	"R^2_0"=.254
1	Type of Entrance	2	8.956	.011	108	101.825	.649	"R^2_1"=.276
								"$R^2_{Y1.0}$"=.030
2	No. of Cameras	1	6.847	.009	107	94.978	.791	"R^2_2"=.294
								"$R^2_{Y2.0}$"=.055
3	Type of Entrance X Type of Facility (Interaction Term)	4	9.577	.048	103	85.401	.896	"R^2_3"=.321
								"$R^2_{Y3.0}$"=.029

TABLE 4b

REGRESSION EQUATIONS FOR MODEL 2 WRITTEN SEPARATELY FOR EACH TYPE OF FACILITY

	B Bank	B S & L	B C.U.	Multiplicative Model e^B Bank	e^B S & L	e^B C.U.	Probability Model P (For P = .5) Bank	S & L	C.U.
Constant	-3.216	-2.355	-4.497	.040	.095	.011			
Years Open	.419	.419	.419	1.52	1.52	1.52	.10	.10	.10
Type Institution									
Bank									
Savings & Loan									
Credit Union									
V23 Cameras	.39	.39	.39	1.48	1.48	1.48	.10	.10	.10
Type Entrance									
Direct Only	.406	.876	.886	1.501	2.401	2.425	.10	.22	.22
Direct & Corr.	.470	-1.37	-1.152	.60	.254	.316	.12	-.34	-.29
Corr. Only	- .876	.494	.266	.416	1.639	1.305	-.22	.12	.0[illegible]

TABLE 5

ALL VARIABLES CONSIDERED FOR INCLUSION IN MODEL 3

Dependent Variable	Frequency		Total N	Marginal Proportion Robbed
One or More Robberies during 1975 or 1976	No	149	215	.307
	Yes	66		

Interval Independent Variables	Min.	Max.	Mean	Standard Deviation
Years Open Out of Ten Observed	1.000	10.500*	8.479	2.758
Average Number of Robberies Prior to 1975	0.000	1.640	.172	.261
Number of Customer Entrances	1.000	3.000	1.395	.586
Number of Teller Stations Inside Main Bldg.	0.000	20.000	5.712	3.594
Number of Visible Cameras	0.000	9.000	1.767	1.447
Number of Types of Guards**	0.000	3.000	.163	.450
Portion of 1975 & 1976 Open for Business	.443	1.000	.975	.094

Categorical Independent Variables		Frequency
Type of Facility	Bank	98
	Savings & Loan	62
	Credit Union	55
Location of Facility	Residential/Commercial	87
	Commercial	128
Parking Availability	Poor	48
	Fair	67
	Good	65
	Excellent	35
Physical Structure	Single	64
	Part of Larger Complex	151
Teller Area on Ground Floor	No	41
	Yes	174
Alley W/In 300 Ft of Bldg	No	92
	Yes	123
Walk Up Windows	No	159
	Yes	56
Drive In Teller Stations	No	178
	Yes	37
Type of Entrance	Direct Only	105
	Corridor Only	67
	Direct & Corridor	43
Bullet Proof Barrier	No	114
	Yes	101
Possible to Get Over Top of Teller Counter	No	92
	Yes	123
Ratio of Male to Female Tellers	Mostly or All Female	178
	Otherwise	37

* Facilities open more than ten years were assigned a value of 10.5.

** None or one or more of the following: plain clothes, unarmed uniformed, armed uniformed.

TABLE 6

RESULTS OF MODEL 3: STEPWISE SELECTION FROM ALL VARIABLES DISPLAY IN TABLE 5

Variable	B	B/SE	e^B (Multiplicative Model)	P (p=.3) (Probability Model)
Type of Facility				
Bank	.291		1.338	.061
Credit Union	- .957	-2.154*	.384	- .201
Savings & Loan	.666	2.217*	1.946	.140
Number of Customer Entrances	.776	2.202*	2.173	.163
Type of Entrance				
Direct Entrance(s) Only	.586	2.052*	1.797	.123
Corridor Entrance(s) Only	- .582		.559	- .122
Direct & Corridor Entrances	- .004	- .011	.996	- .001
Proportion of 1975 & 1976 Open For Business	10.674	1.954*	43217.47	2.241
Constant	-12.858	-2.345*	.261E-5	-2.700

* Significant at or below the 5% level for a one tailed test

Step	Variable Selected	df	Improvement	P-Value	df	Goodness of Fit	P-Value	R^2 Value
Marginal	Constant (Null HYP. $B_1=\ldots=B_{20}=0$)				208	265.199	.003	
0	Constant Proportion Open 1975 & 1976	1	8.11	.004	207	257.089	.010	"R^2_0"=.022
1	Type of Facility	2	27.141	.000	205	229.948	.112	"R^2_1"=.100
								"$R^2_{Y1.0}$"=.08[illegible]
2	No. of Customer Entrances	1	3.694	.055	204	226.254	.136	"R^2_2"=.111
								"$R^2_{Y2.0}$"=.091
3	Type of Entrance	2	4.956	.084	202	221.298	.167	"R^2_3"=.126
								"$R^2_{Y3.0}$"=.106

TABLE 7

MODEL 3 SIMULTANEOUS INCLUSION OF ALL VARIABLES

Variable	B	B/SE	e^B (Multiplicative Model)	P (P=.3) (Probability Model)
No. of Years Open During 10 Observed	-.063	-.757	1.065	.004
Average No. of Robberies in 8 Years Prior to 1975	-.551	-.835	1.735	-.116
Facility Location	-.023	-.116	1.023	-.005
Parking Convenience				
Excellent Parking	-.346	-.938	.707	-.073
Good Parking	.342	1.203	1.408	.072
Fair Parking	-.104	-.366	.901	-.022
Poor Parking	.108		1.114	.023
Physical Structure	-.254	-1.131	.776	-.053
Location of Teller Station	.425	1.090	1.530	.089
Alley W/In 300 Ft	-.105	-.574	.900	-.022
No. Customer Entrances	.701	1.742*	2.02	.147
Walk Up Windows	-.156	-.704	.855	-.033
No. Teller Stations	.131	1.961*	1.140	.027
Drive Up Windows	.059	.212	1.061	.012
Type of Entrance				
Direct Only	.934	3.040*	2.545	.196
Direct & Corridor	-.150	-.410	.861	-.031
Corridor Only	-.784		.457	-.165
Bullet Proof Barrier	-.248	-1.130	.780	-.052
Possible to Jump Teller Counter	.198	.914	1.219	.042
No. Cameras	-.069	-.367	.933	.014
No. of Types of Guards	-1.068	-1.722*	.344	-.224
Ratio of Female to Male Tellers	.001	.003	1.00	.000
Proportion of 1975 & 1976 Open	12.857	1.994*	383463.61	2.700
Constant	-15.230	-2.437*	.24E-6	-3.198

* Significant at or below the 5% level for a one tailed test

Step	Variable Selected	df	χ^2 Improvement	P-Value	df	χ^2 Goodness of Fit	P-Value	R^2 Value
Marginal	Constant (Null HYP. $B_1=\ldots=B_{21}=0$)				208	265.199	.003	
1	All 19 Variables Included	21	52.292	.000	187	212.907	.094	"R^2_1"=.152

TABLE 8

INCLUSION OF ONLY SIGNIFICANT VARIABLES FROM TABLE 7

Variable	B	B/SE	e^B (Multiplicative Model)	P (p=.3) (Probability Model)
No. Customer Entrances	.687	2.066*	1.988	.144
No. Teller Stations	.105	2.178*	1.111	.022
Type of Entrance				
Direct Only	.917	3.539*	2.502	.193
Direct & Corridor	.174	.055	1.190	.036
Corridor Only	-1.091	-3.532*	.336	-.229
No. of Guards	-1.014	-1.834*	.363	-.213
Portion of 1975 & 1976 Open	10.989	2.005*	59219.135	2.308
Constant	-13.476	-2.448*	.140E-5	

* Significant at or below the 5% level for a one tailed test

Step	Variable Selected	df	Improvement	P-Value	df	Goodness of Fit	P-Value	R^2 Value
Marginal	Constant (Null HYP. $B_1=\ldots=B_7=0$)				152	285.782	.000	
0	Constant Proportion Open 1975 & 1976	1	87.279	.000	151	198.503	.006	"R^2_0"=.024
1	Type of Entrance	2	32.682	.000	149	165.823	.164	"R^2_1"=.112
								"$R^2_{Y1.0}$"=.09
2	No. of Customer Entrances	1	5.963	.015	148	159.859	.239	"R^2_2"=.129
								"$R^2_{Y2.0}$"=.108
3	No. Teller Stations	1	3.277	.070	147	156.582	.279	"R^2_3"=.138
								"$R^2_{Y3.0}$"=.117
4	No. of Guards	1	4.191	.041	146	152.391	.342	"R^2_4"=.149
								"$R^2_{Y4.0}$"=.129

END