ADJUSTING UCR DATA FOR RESEARCH:

FINAL REPORT FOR NIJ GRANT 85-IJ-CX-0029

WILLIAM M. HOLMES

ROBERT GRANFIELD

WILLIAM BOWERS

GLENN PIERCE

Center for Applied Social Research

Northeastern University

Boston, Massachusetts 02115

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1

PROBLEMS IN UCR DATA

INTRODUCTION

For a half century, the FBI's Uniform Crime Reports (UCR) have served as the principal basis for the study of crime in America. UCR data has been used in literally hundreds of published studies of crime, in the work of the National Commission on the Causes and Prevention of Violence in America (Mulvihill et al., 1969), and in the movement toward the econometric modeling of crime rates (e.g., Ehrlich, 1971). Moreover, in recent years with the publication of state-level crime estimates, the UCR has become a critical resource for the evaluation of innovative criminal justice policies and programs in various states. Indeed, these data have been the chief basis for policy-relevant research on such controversial issues as the impact on crime rates of capital punishment (Bowers, 1984) and gun control legislation (Wright et al., 1983).

OVERVIEW OF PROBLEMS

With the application of increasingly sophisticated statistical techniques to UCR data, there has come growing recognition that measurement problems pose a threat to the validity and reliability of research findings. For example, investigators have demonstrated a biasing effect of unreliability in the measurement of criminal homicide on analytic results (Bowers and Pierce, 1975; Klien et al., 1978; Bowers, 1984: Chapter 9). More generally, the National Academy of Sciences Task Force on Deterrence and Incapacitation identified measurement problems as a major methodological difficulty with existing deterrence studies (Blumstein et al., 1978). Others have echoed and elaborated upon this judgment (Brier and Feinberg, 1978), and still others have explicitly incorporated assumptions about measurement error into their analyses (e.g., Hellman and Fox, 1984; Parker, 1985).

In research on the impact of criminal justice policy and programs in specific states and localities, investigators have begun to work with agency specific, temporarily disaggregated UCR data, conducting interrupted time-series analyses with ARIMA (auto regressive integrated moving average) modeling techniques (Deutsch, 1978; McCleary, 1980). Here, too, questions have been raised about the appropriateness or adequacy of these models in view of UCR measurement problems (Barnett, 1984). Indeed, divergent results where attempts have been made to estimate the impact of gun control legislation may be due in part to such problems. Thus, parameters for the "noise" factor in ARIMA models or gun crime vary by jurisdiction and by type of offense (Bowers et al., 1984; Loftin and McDowell, 1982; Margarita, 1984). In some cases, different researchers using ostensibly the same data, though obtained from different sources, have produced different noise models and estimated effects (cf. Bowers et al., 1984 and Loftin and McDowell, 1982).

Moreover, in the course of the research, one of these investigations (Bowers et al., 1984) found five reporting agencies with anomalous patterns of gun related crime; where data were missing or unreliable. Thus, in the estimation of intervention effects for jurisdictions in Massachusetts outside of Boston, Springfield had to be deleted from the analysis because of the erratic reporting pattern. In identifying control jurisdictions for comparison with Boston, a search of cities of 250,000 or more inhabitants revealed that New York was missing data on gun homicides (from the Supplementary Homicide Reports, which provide data on the use of guns in homicide); Detriot reported gun robberies as roughly 95 percent of all armed robberies, far exceeding the proportion for other jurisdictions; and Buffalo and Rochester reported patterns of gun assault so disparate from those of other agencies they appeared unreliable. That these anomalies were found in data aggregated annually for a restricted set of offenses in a limited set of agencies clearly point to the need for a systematic analysis of measurement problems in the UCR data.

OBJECTIVES

The study focusses on three particular problems that affect analysis of UCR data: definitional uncertainty, missing data, and classification variability. Each of these problems affect the accuracy of statistics on the level of crime.

Definitional uncertainty means that users of UCR data cannot clearly or easily tell what some of the UCR data codes mean. Definitional uncertainty occurs because of the difficulty in finding definitions in the documentation for some variables, the ambiguity of some definitions, and changes in definitions.

Data missing from the UCR Return A means that a researcher cannot construct a complete record of all UCR offenses in a jurisdiction for a given month. Missing data occurs because some jurisdictions do not report every single month or reports may not be complete.

Classification variability implies that persons filling out Return A may not be consistent in how crime reports are categorized as UCR offenses. Classification variability occurs because information on some offenses may be sufficiently ambiguous to allow classification in more than one category, despite the hierarchy rule and other criteria in the FBI UCR Coding Handbook (FBI, 1980). It also occurs because of changes in the definitions of the measures and because specific events in a local jurisdiction may influence how reports are categorized, whether intentionally or unintentionally.

These problems directly lead to the objectives of this study: (1) greater clarification of the definitions of the UCR measures (2) identification of the likelihood and nature of missing information and classification variability in UCR data, (3) description of the agencies, time periods, and types of offenses for which such problems exist, (4) evaluation of the extent to which these data problems may bias the results of various kinds of statistical analyses, and (5) recommendations for the most efficient and effective methods for dealing with such threats to the validity and reliability of research findings.

DEFINITIONAL UNCERTAINTY

Over time the FBI has sought to improve the definitions of its indicators and the quality of the UCR data (Lejuns, 1957). Its published national and state level annual crime estimates are adjusted for the biasing effects of undercoverage of reporting agencies and for agencies periodically joining or dropping out of the system (UCR, 1980). Presently, the FBI and the Bureau of Justice Statistics (BJS) are engaged in a review and redesign project which is attempting to identify the ways in which definitional and classification variability of crime enter the reports which are ultimately forwarded from local agencies to the national UCR program (Poggio, 1984). Hopefully, this redesign effort will eventually improve the accuracy and reliability of reported UCR data. In the meantime, clarifying the definitions of UCR measures will allow more careful, and more correct, interpretations of the data that are available.

MISSING DATA

A major problem in the analysis of monthly UCR data is that there is a fair amount of data missing. This is due to the fact that not all reporting agencies consistently submit their reports. Even when reports are submitted, they may be incomplete, laking information on some offenses.

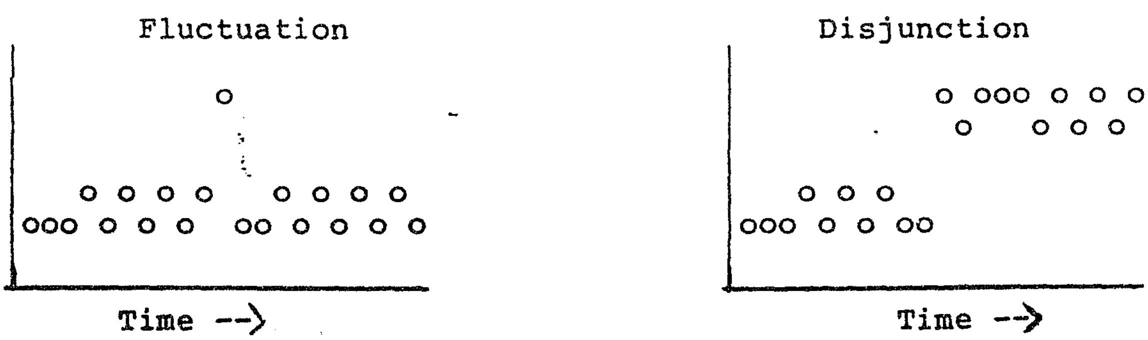
Missing data is also due to the addition of agencies in recent years that were not contributing to the system in earlier years. There simply are fewer agencies in the earlier years. Problems in filling out the forms and transferring the information to data tapes also result in missing data.

CLASSIFICATION VARIABILITY

Related to the problems of missing data, figures reported by an agency may fluctuate wildly from month-to-month or have significant disjunctions in trends. This complicates the process of estimating credible statistical models as these fluctuations and disjunctions can undermine the reliability and the validity of parameter estimates. Illustrations of such a fluctuation and a disjunction are given in Figure 1-1.

FIGURE 1-1

FLUCTUATION AND DISJUNCTION



Such fluctuations and disjunctions can be caused by changes in procedures for the classifying or recording of offenses, as well as changes in the "true" crime rate. For smaller reporting jurisdictions fluctuation may also result from their sensitivity to shocks from the environment.

One pattern may especially indicate shifts in classification procedures. Trends of crime subtypes that are negatively correlated can be produced by classification changes in which an increase in one subtype is associated with a decrease in another. An illustration of such divergence or convergence is given in Figure 1-2.

FIGURE 1-2

NEGATIVELY CORRELATED TRENDS

```
|
|                     xxx
|        xx          x   x       o
|    xxxx  xxxxxxxx        xxxxxo
|    oooo  oooooooo        ooooox
|        oo          o   o       x
|                     ooo
|
|________________________________
         Time  --->
```

Studies of individual agencies have found evidence that such variability in crime may sometimes be a result of manipulation of the data. Systematic changes in classifications of reports from rape to assault or from aggravated assault to simple assault are examples of such manipulation (Ferracuti et al., 1962; Center and Smith, 1973; Chilton, 1979).

APPROPRIATENESS OF ANALYTICAL PROCEDURES

Procedures are needed to detect when these problems undermine the analysis and to compensate for the problems when they occur. Failure to address these problems lead to meaningless conflicting and misleading results. However, a framework is needed to coordinate strategies for dealing with these multiple problems.

A framework for examining these problems is presented in Table 1. It shows analytical issues associated with a given combination of source and type of error. This framework implies that the appropriate technique for detecting a problem depends on the source of error contributing to the problem and on the nature of the errors effect.

TABLE 1
ANALYTICAL ISSUES BY SOURCE OF ERROR
AND ITS EFFECTS

EFFECT OF ERROR	SOURCE OF ERROR: Classification Variability	SOURCE OF ERROR: Missing Data	SOURCE OF ERROR: Confounding Trends
<u>Unreliability</u>			
Internal Inconsistency	Inequality of Measures	Correlated Missing Data Across Measures	Correlated Error Terms
Instability	Inequality of Parameters	Correlated Missing Data Across Time	Hetergenity of Error Terms
<u>Invalidity</u>			
External Inconsistency	Criterion Validity	Correlated Missing Data	Replication Across Groups
Misspecification	Grouping Validity	Correlated Missing Data Across Errors	Replication Across Models

The table indicates that change in the source or effect of error alters the analytical issues. Solving only one source of error does not necessarily improve the validity of one's analysis. For example, removing confounding trends without adjusting for classification variability may merely shift the analytical problem from difficulty in replication to that of replicable models with invalid (biased) parameters. Confirmatory factor analysis alone or ARIMA time-series analysis alone cannot address the problems simultaneously. Each gives only a partial solution to the problem and the resulting estimates still retain more unreliability and invalidity (bias) than is necessary. This is especially true when making comparisons across groups, time points, or alternative models. To deal with the interdependence of these issues, a generalized error modeling procedure is necessary that represents variation in the source and effects of these errors.

PROPOSED STRATEGY

OVERVIEW

The research we propose here is designed to address these issues through the development of a measurement modeling methodology for merged cross-section, time-series multinomial distributions and through the application of this methodology to data sampled from the National Time-series Community level Crime Database (NTCCD) containing the monthly reports of some twenty-seven forms of crime filed by 10,000 and 15,000 police agencies over the period 1967 through 1982 -- a data resource of roughly 67 million elements.

The NTCCD Database

The FBI has made data on crime available to researchers, at first as published national and state level annual crime estimates adjusted for undercoverage and reporting discontinuity, and since the mid 1960's, through computer tapes of the agency specific monthly crime reports (used to prepare the FBI's annual UCR publication). Since the FBI's original purpose was to facilitate the publication of the annual UCR report on Crime in America, not to provide computerized data for research and evaluation, the formating of the tapes has made the data relatively difficult to work with.

To make these data more accessible, the Center for Applied Social Research (CASR) at Northeastern University undertook the development of a National Time-series Community level Crime Database (NTCCD), in conjunction with an NIJ funded evaluation of the Massachusetts Bartley-Fox gun law. CASR obtained the agency specific crime report tapes for the years 1967 through 1980 and merged the data to form a cross-section, time-series multinomial data file (ICPSR, 1984). This NTCCD database is now being updated through 1982 with support from the Bureau of Justice Statistics (BJS).

This massive database now contains sixteen years of monthly ~~time-series data on~~ 27 crimes for all of the more than 15,000 police agencies that made such reports to the UCR during the period. The NTCCD is readily available to researchers, administrators, and policymakers through the Inter-University Consortium for Political and Social Research (ICPSR, 1984). It represents an unrivaled resource for the study of local, state, and national criminal justice initiatives. And what is more, it provides us with a powerful tool for detecting the presence of likely measurement error in these data, for estimating the extent of such error, and for evaluating approaches for ameliorating the effects of such error.

The Measurement Modeling Strategy

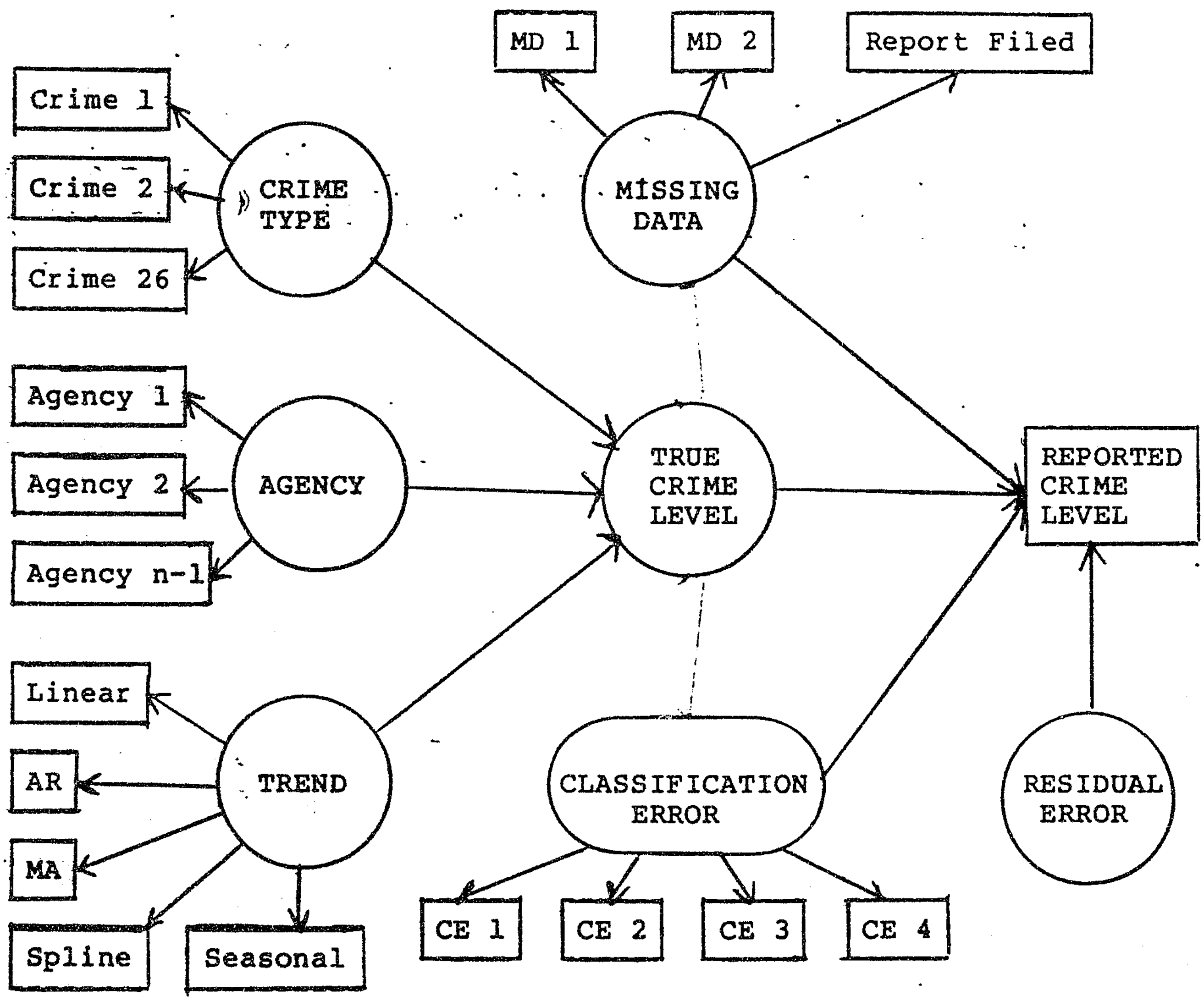
The error modeling capacity of Confirmatory Factory Analysis (LISREL) is precisely such a procedure when it is combined with time-series analysis. LISREL measurement models can be generated for each of the sources of error in reported crime listed in Table 1 as well as for other sources of variation in crime level. Including sources in Table 1, five variables seem most important to model: crime type, reporting agency, trend factors, classification errors, and missing data. Once the measurement models are estimated, they can then be used to identify outliers from the expected values and to adjust the suspect data for suspected sources of error. The identified suspect data points can also be targeted for further investigation to verify factors affecting reported values for a specific offense, agency, and time period.

Joreskog has presented LISREL models appropriate for longitudinal data in which the errors have serial correlation and for multi-group, multiple indicator data (Joreskog and Sorbom, 1979). The extension of these models to time-series data in which the errors are affected by autoregressive and moving average processes provides a basis for simultaneous modeling of error components of merged cross-sectional, time-series, multinomial distributions.

Figure 1 represents the preliminary measurement models and their relationship to reported crime levels.

FIGURE 1-3

STRUCTURAL MODEL OF REPORTED CRIME LEVEL



Indicators of Agency and Crime Type are dummy variables for data from a given agency and type of crime. Indicators for

Missing Data indicate crime-agency-month for which information is absent and whether a Return A report was filed by that agency for that month. Indicators for Classification Effor reflect shifts in the recording of a crime from one category to another via changing ratios between such categories. Indicators AR and MA for the Trend variable are measures of autoregressive and moving average processes, respectively. The spline indicator in the Trend variable is to measure disjunctions in trends. Linear and seasonal indicators are functions computed from the monthly sequence of the data. The Reported Crime Level is the actual UCR values. Residual Error is error variance unaccounted by the specified influences. A more detailed discussion of the indicators used to measure the unobserved variables is presented below in the section "Data and Measurement."

Unobserved variables in the model are circled. Observed or computed indicators are enclosed in boxes. The observed indicators of variables may themselves be correlated. It is traditional to assume this possibility, rather than drawing in many curved double-headed arrows between the indicators. Such intercorrelations between classification error and trends would seem especially likely. These intercorrelations are additional reasons why simultaneous controls for error sources are needed.

If monthly data is used in estimating the above model, then the unit of analysis is the level of a type of crime in a given agency for a given month, the crime-agency-month. If there were no error, these values would be the true counts for each crime type for a given agency in a given month. The True Crime Level plus the error is the reported level of each type of monthly crime for an agency. The difference between the reported crime level and the true level is the miscount.

The most important aspect of organizing the data along these three dimensions of agency, crime type, and time is that it allows analysis of error factors that could not previously be estimated. While Box and Tiao have estimated time-series multinomial distributions and merged cross-sectional time-series (Box and Tiao, 1979), to date the merged cross-sectional, time-series, multinomial distribution has been intractable. This stems from the complexity of issues in specifying a reasonable representation of the measurement models for errors in the variables, as well as models of the interrelations of the variables themselves. ARIMA modeling has been able to deal with noise models for the errors in times series. Confirmatory factor analysis has been able to address multiple indicator models of errors such as classification variability and missing data. By merging these two procedures it is possible to deal with both problems simultaneously. Organizing the data along these three dimensions allows, for the first time, simultaneous use of these techniques. Because of the complexity of the model and the interrelatedness of the issues (see Table 1), a multi-stage process is needed to lay the foundation for reliable estimation. The sequence needed is described in the section on data analysis.

The model in Figure 1 is a preliminary model of these sources of errors. It is presented as a way of organizing the discussion of factors relevant to the study and as a tool for integrating the univariate, bivariate, and multivariate analysis. Diagnostic evidence of its goodness-of-fit of the preliminary model with the data would be used for revising the model, criteria for which are reported below.

PROJECT DESIGN

The project utilizes UCR data from 1965 to 1983 in a merged cross-sectional time-series analyses. One hundred ninety-eight jurisdictions have been selected for analysis. With monthly data for nineteen years this results in 45,144 units of analysis, of which 45,060 have sufficiently complete returns for use. This is a figure large enough to satisfy the requirements of confirmatory factor maximum likelihood analysis. The sample includes fifty-two cities with populations over 250,000 and one hundred and forty-six cities with populations between 100,000 and 250,000. These 198 cities represent more than 40 percent of all reported offenses in 1980 and more than 50 percent of all offenses in cities.

Multiple indicator confirmatory factor analytic models will be utilized to estimate source, strength, and nature of biases resulting from systematic missing data and discontinuities in trends and coverage. The contribution of possible changes in classification and reporting to these patterns will also be examined. Sensitivity analysis will compare model parameters by city size, region, and consistency of reporting to the UCR system to verify reliability and stability of the results.

Data

The study uses UCR Return A data as the basis for the analysis. The NTCCD database provides this information. Data are organized in a monthly time series for each agency, which means that the units of analysis are agency-months.

The five unobserved variables in Figure 1 are measured using multiple indicators -- missing data, crime type, agency, trend, classification error. True crime level is estimated as an unobserved variable intervening between reported crime level and the other variables (Heise, 1972). The measurement of these variables is discussed in subsequent chapters dealing with those variables.

Plan for Analysis

Data analysis will proceed in three stages. The first stage will construct work files for alternative subsets of reporting agencies, according to criteria discussed below. The second stage will perform a set of bivariate analyses in which measurement models of each construct and its indicators will be estimated. The bivariate analyses will not control for correlation of the indicators with other constructs or other indicators; that will reserved for the multivariate analysis. The third stage will involve multivariate estimation of the model, revision of the model, and investigation of the effects of alternative strategies to ameliorate the error upon the model estimates.

File Construction

Work files are created for the Return A data containing the indicators specified. Contents of these files are compared with known frequency distributions of the data to verify valid construction. Work files are constructed that vary in their composition of consistently reporting agencies, size of reporting agency, and region of the country to allow subsample comparisons of estimates as part of subsequent sensitivity analysis described in the bivariate and multivariate sections of the plan for analysis.

Univariate frequency distributions of the indicators are produced as part of the file construction to verify accurate construction and check distributional assumptions of the indicators. All indicators to be used in the LISREL model are checked for normality of distribution or, in the case of dichotomous indicators, adequacy of their split for use as dummy variables.

Bivariate Analysis

The bivariate associational analysis will examine the effects of missing and suspect data upon the estimation of bivariate relationships. Residual analysis and missing data analysis are a central component of the bivariate analysis. Sensitivity analysis of bivariate relationships will also be conducted. Factor analytic measurement models of the influence of each source of error are estimated separately. This will specify major issues in the modeling of these errors. Information from the bivariate analysis is used in designing the multivariate analysis.

Development of measurement models for the respective sources of error separately will simplify and allow more orderly development of the multivariate models. While some indicators will remain in the measurement model at this stage that subsequently fall out of the multivariate model, this process provides an initial screening of indicators so that fewer need to be considered at the multivariate stage. It also identifies some of the stronger measurement issues so that the multivariate stage can focus on them. It will make the multivariate analysis more manageable.

The fact that different ARIMA models of trends have been found for different jurisdictions and offenses, raises the possibility of an interactive effect for the trend factor with agency and/or offense type. This situation would require adding multiplicative terms to the equations representing the model. The number of possible multiplicative terms is quite large, since it increases geometrically with the number of agencies, offense types, and trend functions in the study. Thus, the bivariate analysis also includes a screening procedure to detect significant associations of trend interactions. All trend interactions not significant at the bivariate level are excluded from the multivariate model to make the estimation more managable. In addition, trend interactions having significant bivariate association with crime level are checked to see if this association remains after controlling for the additive effects of agency or crime and trend components in the interactive terms. This further reduces the number of interactive terms in the model.

If the number of interactive terms at that point still exceeds the capacity of the LISREL program, two strategies are adopted to further reduce the number of indicators in the model. First, the remaining interactive trend terms are reduced to their principal components (via a principal components factor analysis) and the components used instead of the separate indicators. Secondly, insertion of the interactive terms will begin with the most significant terms and sequentially add less significant terms until the new terms added do not produce significant effects or the capacity of the program is exhausted. Significant interactive terms that might remain out of the model as a result of this process are the least significant interactive effects. Thus, their exclusion, if necessary, has minimal effects on the model compared with excluding other indicators. Since the number of indicators that can be included is quite large, it is not anticipated that many, if any, of the significant interactive terms will remain out of the model. In any case, the resulting model will still represent the best possible estimates of the error factors in UCR crime data given existing technology.

The estimation of spline functions to capture disjunctions in the trends is sequential process in which the slopes and interrupts of adjacent segments are compared. Extreme difference in either imply a possible disjunction has occurred. Tests of significance screen out smaller disjunctions. Spline functions are estimated for the larger disjunctions. They are then screened according to the strategy discussed for the trend interactions.

Tests of ill-conditioning of the measurement models utilize measures of multicollinearity (Rockwell, 1975), dependence and independence (Holmes, 1982, 1983). They provide evidence of the reliability of the covariance matrices from which the parameters are estimated.

Multivariate Analysis

The multivariate modeling of errors utilizes confirmatory factor analysis to estimate measurement models that control for the errors specified in the bivariate analysis and for the systematic missing data. Measures having significant coefficients in the measurement models have the best prospect for being used to adjust estimates in removing bias resulting from the errors and the missing data. Comparisons of measurement models for various subgroups of agencies (sensitivity analysis) also identify effects of dropping cases from the sample.

The multivariate modeling is a sequential process in which a separate measurement model for each unobserved variable is estimated. Measurement models are refined according to the significance of their coefficients, the magnitude of standardized residuals in the corresponding covariance matrix, and the size of derivatives and Lagrangian multipliers of each model. Chi-square "badness-of-fit" statistics (Joreskog and Sorbom, 1979) are also used in refining the model. After each separate measurement model is refined, they are merged in a single model of the effects of these variables on crime level.

Developing this type of a multivariate model allows for detecting suspect values in the data, adjusting for them, or flagging cases having such values. Flagging such cases would allow targeting them for closer inspection, dropping the cases, or issuing cautionary remarks about unusual variations in the data. Detection of suspect values occurs by comparing the values estimated by the model with the observed values and by identifying statistical outliers in the data. When the standardized errors (the standardized difference of estimated and observed values) are large, those cases would be suspect as outliers. Once the suspect cases are identified, one can either adjust the values using parameters of the model or chose to drop the cases. Comparisons between these two strategies of adjustment or dropping cases are made to determine the efficiency and effectiveness of these alternatives for large versus small outliers, as well as types of crime and reporting agencies.

Estimation of the multivariate model simultaneously controls for crime type, agency, trends, classification variability, and missing data. To the extent that other sources of error in UCR data are correlated with these variables, those other factors are also controlled. Those factors independent of variables in the model are remain as unexplained residual or error. While it may be possible to reduce the residual error by adding still other variables and indicators to the model, such is left for a future study. To allow an orderly and reliable development and testing of the model, it is important not to indicators; that was reserved for the multivariate analysis. The third stage involved multivariate estimation of the model, revision of the model, and investigation of the effects of alternative strategies to ameliorate the error upon the model estimates.

OVERVIEW OF REPORT

The results of this work include: (1) evaluation of the effects of such problems on the time-series analysis of UCR data and the advantage of alternative procedures for ameliorating these effects; (2) description and specification of the methodology developed for assessing measurement problems in the UCR data; and (3) the preparation of a Suspect Data Key for all agencies with population coverage of 100,000 or more in the NTCCD database that will flag suspect data points, indicate the source(s) of likely error, and provide adjusted replacement values.

Chapter 2 discusses UCR definitions that contibute to these issues. This includes problems of documentation, ambiguity of definitions, and changes in definitions. Advantages of these definitions will also be discussed.

Chapter 3 discusses missing data in the UCR file. It locates areas of missing data and strategies for dealing with it. Tradeoffs between dropping cases, estimating missing values, and adjusting parameters are elaborated.

Chapter 4 examines classification variability. It considers issues of misclassification, fluctuation in classification, and changes resulting from variable definitions.

A particular focus of this chapter is on outliers in the data and procedures for dealing with them, especially tradeoffs between dropping extreme cases and adjusting results by modeling the fluctuations.

Chapter 5 discusses analytical procedures for analysis of UCR data. It identifies problems and strengths of each procedure. It also proposes a new procedure, Structural Equation Time Series Analysis (SETSA), and demonstrates its use.

UCR DEFINITIONS

Although there have been ambiguities and changes in UCR definitions, a careful search of available documentation has clarified their meaning and the points at which major changes in definitions have occurred. Additional work is needed to further clarify remaining points of uncertainty.

A simple and direct action can be take to make the meaning of "adjusted return" more clear and more useful for analysis. When the computer tape for a state is revised, only those localities having changed information should be labeled "adjusted." It should not require a major change in the computer program to leave unchanged the Card Type variable when the information for that agency has not been changed. At present it is impossible to tell how many agencies have their reports "adjusted," except that it can be no more than 30 percent of all agencies (the approximate proportion of all reports that are classified as "adjusted" in this sample).

A second action that would clarify the meaning of UCR data would be to examine the "estimated" data values for bias in the estimation. The clustering of missing reports in adjacent months combined with seasonal fluctuations may introduce bias into the estimation procedure. Documentation of the extent and severity of this bias would help establish how much these estimated values can be trusted.

MISSING DATA

An important finding of this study is that so few reports are missing from the larger agencies. Many of the larger agencies consistently file UCR reports. When a researcher is faced with missing UCR data, it is important to consider its likely impact. The missing data may not be relevant to the particular analytical issue at hand. Even when the missing data is relevant, this study has found it is often feasible to track down the missing reports and obtain "adjusted" values. This would be the preferable strategy when working with data from larger agencies. Estimating or dropping missing cases are likely to be less reliable and less valid strategies than adjustment, since more complex assumptions must be made about potential sources of bias resulting from these procedures; but the decrease in reliability and validity resulting from using estimated values or dropping cases has not been demonstrated to be severe, particularly for these larger agencies. Even in the absence of adjusted or estimated data, the amount of reports missing from larger agencies is relatively small and may not be a major source of bias in one's analysis.

CLASSIFICATION VARIABILITY

The findings show that UCR reports can be highly variable from month to month, but most of the time they are not. Indeed, instability in the reported values was more a function of the agency size. The smaller of these large agencies had more outliers than the larger ones. The anchoring effects of having many months with zero offenses reported for a crime type was more of a problem for statistical estimation than that of a few large increases. This does not reduce the substantive importance of crime increases. It does suggest that interpretation of such increases needs to be in light of the baseline and trend values from which the deviations occur.

The findings also show that the choice of criteria for outliers has little effect on the findings. Although the amount of variability is effected by the criteria, excluding the more variable cases by either criteria did not greatly alter the findings. Consequently, strategies of dropping cases are unlikely to resolve problems of classification variability.

MODEL REVISION

The findings suggest further revisions of the crime level model and underscore a strength of the SETSA procedure. The correlations between observed indicators of different latent variables was a major source of the "badness-of-fit" of the model. This, combined with the extensive negative intercorrelations of dummy variables implies that additional latent variables need to be added to the structural model and that the measurement models need revisions. Specifically, indicators of agency, missing data, and trend variables should be thought of as antecedent to the latent variables. This results in reversing the direction of the arrows between the indicators and the construct.

Such reversal has significant substantive and statistical implications. This change implies that these variables are ex-post facto constructs imputed to the indicators. They are not underlying causes of the indicators. These latent variables, however, do serve heuristic utility. Statistically, the change means that there needs to be greater use of block variable estimation of latent variables that are the consequence of the indicators, rather than estimating them with a confirmatory factor measurement model that assumes the latent variables cause the observed indicators.

APPLICATIONS OF SETSA

How can true change in crime levels resulting from a legal intervention be separated from change associated with classification variability, trends, or missing data? This depends on whether or not analysis of UCR data uses a comparison or control group design. Estimation of true changes is much more reliable when comparison or control groups are used than when using a single time-series of data.

When comparison or control groups are used, change associated with the error factors estimated in the control group can be subtracted from change observed in the jurisdictions of the "treatment" group. Remaining change can be imputed to be true changes, depending on how "good" a control group is used.

When control groups are absent, more restrictive assumptions must be made about the nature of fluctuations associated with the error factors. For those crime types thought to be affected by an intervention, adjustment for all classification error change would remove true change as well. The researcher is faced with not adjusting for any classification error, removing only error associated with crime types thought to be unaffected by the intervention, or assuming that change of a given duration or magnitude will be regarded as true change. For example, the researcher may assume that change lasting more than six months represents true change and that change lasting less than six months is "error." Error associated with missing data can still be removed for this type of data.

These limitations when a single time-series is used also apply to traditional ARIMA models of impacts. The difference is that traditional ARIMA with a single series always assumes that there is no missing data and that classification error is not present or that it is random. The proposed procedure allows removing classification error in some circumstances and removing missing data error in all circumstances. Consequently, the advantages of using comparison or control groups with this method are so great that it would be much the preferred strategy when applying this method to the study of criminal justice interventions.

The SETSA procedure has utility for predicting crime levels. When well-developed constructs are used, the predicted values should prove more reliable and valid. This gain in reliability and validity, however, depends on how good is the model with which one begins. Bad theories will produce badly fitting models with this procedure. A strength of SETSA is the ease with which poorly fitting models are rejected. This allows the researcher to more quickly reject implausible models and should lead to better conceptualization and predictive equations.

REFERENCES

Blumstein, A., J. Cohen and D. Nagin, eds.
1978 Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates.

Bowers, William J., Glenn L. Pierce and John F. McDevitt.
1984 Legal Homicide: Death as Punishment in America, 1864 - 1982, Northeastern Univeristy Press, Boston, Massachusetts.

Bowers, William J., Glenn L. Pierce, and Raymond Hyatt.
1984 "Assessing Local Policies to Deter Crime," Final Report for National Institute of Justice Grant 82-IJ-CX-0008. Boston: Center for Applied Social Research, Northeastern University.

1985 "The Illusion of Deterrence in Isaac Ehrlich's Research on Capital Punishment", Yale Law Journal, December - January.

Brier, Stephen S. and Stephen E. Feinberg.
1978 "Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis?", Univeristy of Minnesota,

Chilton, Roland
1979 "Politics and the Production of Crime Statistics," Paper presented at the annual meeting of the Society for the Study of Social Problems, Boston MA.

Chilton, Roland and Adele Spielberger
1972 "Increases in Crime: The Utility of Alternative Measures," Journal of Criminal Law, Criminology and Police Science, 63: 68-74.

Center, L. and Thomas Smith
1973 "Crime Statistics--Can They Be Trusted?" 11 American Criminal Law Review, 1045.

Deutsch, S.J. and LuAnn Sims.
1978 Identification Algorithm for Dynamic Intervention Modeling with Application to Gun Control, 29(78), Georgia Institute of Technology, School of Industrial and Systems Engineering: Atlanta.

Ehrlich, I.
1972 "Deterrent Effect of Criminal Law Enforcement", Journal of Legal Studies, 1:259.

Federal Bureau of Investigation
1980 Uniform Crime Reporting Handbook. Washington, D.C.: U.S. Department of Justice.

Ferracuti, F., Rosita Hernandez and Marvin Wolfgang
1962 "A Study of Police Errors in Crime Classification," Journal of Criminal Law, Criminology, and Police Science, 53:113-119

Heise, David
1972 "Employing Nominal Variables, Induced Variables, and Block Variables in Path Analysis," Sociological Methods and Research 1,2:147-173.

Hellman, Daryl A. and James Alan Fox.
1984 Urban Crime Control and Property Values: Estimating Systematic Interactions Washington, D.C.: National Institute of Justice, April.

Holmes, William
1982 "Ill-Condition Correlation Analyzer," Behavior Research Methods and Instrumentation 14, 2: 77.

1983 "Detection of Ill-Conditioning in Small Samples." Presented at the annual meetings of the American Sociological Association, Detroit, Michigan.

1985 "Identification of Missing Data Bias in the Uniform Crime Reporting System Using Confirmatory Factor Analysis." Proceedings of the American Society of Criminology, San Diego, California.

1986 "LISREL Program Summary: LISREL Version VI." Submitted for publication to Scientific Software, Inc., Chicago, Illinois.

forthcoming "Structural Equation Time Series Analysis and the Forcasting of Crime. To be presented at the annual meetings of the International Society of Forcasters, Boston, Massachusetts, 1987.

Inter-University Consortium for Political and Social Research
1984 Uniform Crime Reports: National Time Series Community-Level Database, 1967-1980, (Baird, James, William J. Bowers, and Glenn L. Pierce.

Joreskog, Karl G. and Dag Sorbom
1979 Advances in Factor Analysis and Structural Equation Models. Cambridge, Massachusetts: Abt Books.

Kituse, John and Aaron Cicourel
1963 "A Note on the Uses of Official Statistics," Social Problems 11(Fall): 131-139.

Klien, L.R., B.E. Forst and V. Filator.
1978 "The Deterrent Effect of Capital Punishment: An Assesment of the Estimates", in Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates, A. Blumstein, J. Cohen, and D. Nagin, eds., Washington, D.C.:

National Academy of Sciences.

Loftin, Colin and David McDowall
1982 "The Deterrent Effects of the Florida Felony Firearm Law," unpublished.

Magidson, Jay
1977 "Toward a Causal Model Approach for Adjusting for Preexisting Differences in the Non-equivalent Control Group Situation: A General Alternative to ANOVA", Evaluation Quarterly, 1(3):399-420.

Malinvaud, P.
1970 Statistical Methods of Econometrics. The Hague, Netherlands: Mouton.

Margarita, Mona
1984 The 1980 New York Gun Law: An Evaluation of Its Implementation and Impact. Final Report of National Institute grant 80-IJ-CX-0100. Washington, D.C.: Police Foundation.

McCleary, R. and R.A. Hay, Jr. with E.E. Meidinger and D. McDowall.
1980 "Applied Time Series Analysis for the Social Sciences", Beverly Hills, California: Sage.

Mulvihill, D and M.M. Tumin.
1969 Crimes of Violence, Volume II, National Commission on the Causes and Prevention of Violence.

Parker, Robert
1985 Forthcoming, "A Demonstration of Aggregation, Ratio, Variable and Homicide Measurement Error Effects in Modeling" in Criminal Journal of Quantitative Criminology.

Pierce, Glenn and James A.Fox
1984 "UCR Data as a Resource for Research: Problems and Opportunities." Grant funded by the Bureau of Justice Statistics.

Poggio, Eugene
1984 "The Accuracy of Uniform Crime Report Data: An Analysis of Adult Data Presented at American society of Criminology, Cincinnatti.

Rockwell, Richard C.
1975 "Assessment of Multicollinearity: The Haitovsky Test of the Determinant," *Sociological Methods and Research 3,3: 308-320.

Schneider, Vicki and Wiersema, Brian.
1985 "Estimating Crime: The Impact of Imputation in the Uniform Crime Reports." Proceedings of the American

Society of Criminology, San Diego, California.

Seidman, David and Michael Couzins
1974 "Getting the Crime Rate Down: Political Pressure and Crime Reporting," Law and Society (Spring): 457-489.

Uniform Crime Reporting (See Federal Bureau of Investigation.)

Wright, James D. and Rossi, Peter H.
1981 "Weapons and Violent Crime: Executive Summary." Amherst, Massachusetts: Social and Demographic Research Institute.

Wright, James, Peter H. Rossi, and Kathleen Daly
1983 Under the Gun, Aldine Publishing.

Zimring, Franklin
1972 "The Medium as the Message: Firearm Caliber as a Determinant of Death from Assault." Journal of Legal Studies. 1:94-124.

APPENDIX A

PRODUCTS

PUBLICATIONS

Holmes, William M. 1985. Identification of Missing Data Bias in the Uniform Crime Reporting System Using Confirmatory Factor Analysis. Proceedings of the American Society of Criminology, San Diego, California

1986. LISREL Program Summary: LISREL Version VI. Submitted for publication to Scientific Software, Inc., Chicago, Illinois.

forthcoming. Structural Equation Time Series Analysis and the Forcasting of Crime. To be presented at the annual meetings of the International Society of Forcasters, Boston, Massachusetts, 1987.

TABLE A-1

LIST OF MOST EXTREME SUSPECT AGENCY MONTHS

	ORICODE	YEAR	MONTH	GRAND TOTAL	TOTAL CHANGE	BURG	BURG CHANGE
1	AZ00713	65	1.00	82	.06	26	.05
2	AZ00729	65	1.00	159	.02	27	.06
3	AZ01003	77	8.00	2908	.89	906	13.00
4	CA00103	65	1.00	526	.16	133	.04
5	CA00106	72	6.00	0	.00	0	.33
6	CA00106	72	6.00	572	573.00	183	2.00
7	CA00109	65	1.00	1609	1.01	398	16.00
8	CA00109	65	1.00	1512	.93	402	.06
9	CA00704	65	1.00	207	.05	42	.04
10	CA00704	79	10.00	0	.00	0	.16
11	CA00704	79	10.00	731	732.00	192	4.00
12	CA00704	79	10.00	-9	-0.01	-9	.
13	CA00704	80	10.00	762	-95.37	214	.
14	CA01925	77	8.00	0	.00	0	.50
15	CA01925	77	8.00	1358	1359.00	412	8.00
16	CA01953	65	1.00	643	.02	150	.02
17	CA01953	76	8.00	927	1.01	236	.04
18	CA03001	72	5.00	1079	.88	334	19.00
19	CA03001	73	6.00	1174	.83	382	.09
20	CA03001	73	6.00	1198	1.02	408	11.00
21	CA03610	65	1.00	571	.19	173	.09
22	CA03801	66	2.00	4564	1.03	1069	.06
23	CA03801	66	2.00	4054	.88	868	11.00
24	CA03905	65	1.00	515	.09	109	.05
25	CA04316	65	1.00	155	.04	34	.05
26	CA05604	76	8.00	-9	.00	-9	.
27	CA05604	77	8.00	782	-97.87	270	.
28	CO03004	70	1.00	0	.00	0	.12
29	CO03004	70	1.00	416	417.00	83	3.00
30	CT00015	65	9.00	320	.08	141	.07
31	CT00064	81	8.00	2374	.97	573	.07
32	CT00093	65	9.00	250	.12	63	.07
33	FL00603	65	9.00	586	.10	167	.04
34	FL00603	79	7.00	1341	.89	334	.09
35	FL00603	79	7.00	1684	1.25	495	11.00
36	FL00603	79	7.00	1546	.95	604	.07
37	FL01307	74	4.00	-9	-0.01	-9	.
38	FL01307	75	4.00	681	-85.25	227	.
39	FL02902	72	2.00	1703	1.01	540	11.00
40	FL04804	65	9.00	333	.10	81	.08
41	HI00200	65	9.00	1214	.28	423	.02
42	IA05701	65	9.00	158	.03	34	.04
43	IA07703	79	6.00	1341	.98	218	.09
44	IL08402	73	3.00	0	.00	0	.20
45	IL08402	73	3.00	744	745.00	152	2.00
46	IL08402	75	4.00	-9	.00	-9	.
47	IL08402	76	4.00	717	-89.75	224	.

48	ILCPD00	65	9.00	11905	13.18	2628	30.33
49	IN00201	65	9.00	464	.01	78	.01
50	KS08703	65	9.00	829	.34	187	.09
51	KS08703	73	3.00	1546	.92	455	.08
52	KS08901	75	4.00	1028	1.17	261	14.00
53	KY03402	79	6.00	983	1.04	258	12.00
54	KY05602	76	5.00	1835	1.08	627	.07
55	KY05602	82	9.00	1922	.96	575	.09
56	MA00308	65	9.00	153	.04	71	.03
57	MA01460	65	9.00	324	.06	102	.05
58	MA01460	81	8.00	-9	.00	-9	.
59	MA01460	82	8.00	863	-108.00	301	.
60	MI25398	65	9.00	721	.09	160	.13
61	MI33519	83	9.00	906	1.06	277	16.00
62	MI41436	75	4.00	1195	1.06	340	11.00
63	MI50765	67	9.00	0	.00	0	.06
64	MI50765	67	10.00	233	234.00	65	1.00
65	MI50765	67	10.00	0	.00	0	1.00
66	MI50765	67	10.00	129	130.00	41	2.00
67	MI50765	67	10.00	-9	-0.06	-9	.
68	MI50765	69	11.00	145	-18.25	63	.
69	MI50765	69	11.00	-9	-0.05	-9	.
70	MI50765	69	11.00	136	-17.12	37	.
71	MI50765	69	11.00	-9	-0.05	-9	.
72	MI50765	69	11.00	151	-19.00	46	.
73	MI50765	69	11.00	-9	-0.05	-9	.
74	MI50765	69	11.00	151	-19.00	37	.
75	MI50765	69	11.00	-9	-0.05	-9	.
76	MI50765	69	11.00	136	-17.12	21	.
77	MI50765	69	11.00	-9	-0.05	-9	.
78	MI50765	69	11.00	202	-25.37	38	.
79	MI50765	69	11.00	-9	-0.03	-9	.
80	MI50765	69	11.00	180	-22.62	34	.
81	MI50765	69	11.00	-9	-0.04	-9	.
82	MI50765	69	11.00	206	-25.87	42	.
83	MI50765	69	12.00	-9	-0.03	-9	.
84	MI50765	69	12.00	202	-25.37	29	.
85	MI50765	69	12.00	-9	-0.03	-9	.
86	MI50765	69	12.00	184	-23.12	39	.
87	MI50765	69	12.00	-9	-0.04	-9	.
88	MI50765	69	12.00	162	-20.37	56	.
89	MI50765	69	12.00	-9	-0.04	-9	.
90	MI50765	69	12.00	143	-18.00	31	.
91	MI50765	80	7.00	0	.00	0	1.00
92	MI50765	80	7.00	1366	1367.00	177	2.00
93	MI50765	82	8.00	0	.00	0	.33
94	MI50765	82	8.00	2408	2409.00	426	14.00
95	MI50765	83	8.00	470	.98	82	12.00
96	MI50765	83	8.00	524	1.11	67	.08
97	MI50806	82	8.00	0	.00	0	.33
98	MI50806	82	8.00	5179	5180.00	676	10.00
99	MI73717	74	3.00	0	.00	0	.08
100	MI73717	74	3.00	1177	1178.00	257	2.00
101	MI81218	80	7.00	0	.00	0	.50

102	MI81218	80	7.00	1310	1311.00	217	4.00
103	MI81218	82	8.00	0	.00	0	.16
104	MI81218	82	8.00	1771	1772.00	281	8.00
105	MI81218	82	8.00	0	.00	0	.50
106	MI81218	83	8.00	756	757.00	156	3.00
107	MI81218	83	8.00	0	.00	0	.50
108	MI81218	83	8.00	764	765.00	153	2.00
109	MI81218	83	8.00	0	.00	0	.50
110	MI81218	83	8.00	900	901.00	202	4.00
111	MI82343	79	6.00	3	.01	0	1.00
112	MI82343	79	6.00	1443	1444.00	143	5.00
113	MI82343	80	7.00	0	.00	0	.50
114	MI82343	80	7.00	1512	1513.00	122	2.00
115	MI82343	82	8.00	0	.00	0	1.00
116	MI82343	82	8.00	1830	1831.00	274	4.00
117	MI82343	83	8.00	0	.00	0	1.00
118	MI82343	83	9.00	628	629.00	74	2.00
119	MI82349	65	9.00	4957	7.15	1358	49.00
120	MI82538	65	9.00	169	.01	40	.00
121	MI82538	81	7.00	0	.00	0	1.00
122	MI82538	81	7.00	2644	2645.00	565	6.00
123	MI82538	82	8.00	0	.00	0	.16
124	MI82538	82	8.00	2484	2485.00	519	8.00
125	MI82538	82	8.00	0	.00	0	.12
126	MI82538	82	8.00	3017	3018.00	794	9.00
127	MN02711	71	1.00	1974	.99	699	.04
128	MN02711	71	1.00	2411	1.22	813	25.00
129	MN02711	80	6.00	0	.00	0	.06
130	MN02711	80	6.00	7572	7573.00	2148	72.00
131	MN06209	73	2.00	0	.00	0	.20
132	MN06209	73	2.00	19425	19426.00	7329	93.00
133	MN06209	74	2.00	0	.00	0	.01
134	MN06209	74	3.00	22508	22509.00	7502	92.00
135	MN06209	75	3.00	0	.00	0	.01
136	MN06209	75	4.00	24433	24434.00	7666	93.00
137	MN06209	76	4.00	0	.00	0	.01
138	MN06209	76	4.00	26340	26341.00	8479	97.00
139	MN06209	77	4.00	0	.00	0	.01
140	MN06209	77	5.00	23305	23306.00	7608	119.00
141	MN06209	78	5.00	0	.00	0	.00
142	MN06209	78	6.00	22257	22258.00	7145	140.00
143	MN06209	79	6.00	0	.00	0	.00
144	MN06209	79	6.00	23172	23173.00	7135	126.00
145	MN06209	80	6.00	0	.00	0	.00
146	MN06209	80	7.00	24526	24527.00	7297	141.00
147	MN06209	81	7.00	0	.00	0	.00
148	MN06209	81	7.00	25215	25216.00	7964	198.00
149	MN06209	82	8.00	0	.00	0	.00
150	MN06209	82	8.00	26725	26726.00	8341	217.00
151	MN06209	83	8.00	0	.00	0	.00
152	MN06209	83	9.00	24292	24293.00	8005	221.00
153	MN06906	65	9.00	180	.00	37	.00
154	MS02501	65	9.00	258	.05	59	.06
155	NC06001	70	12.00	1487	1.01	417	.09

156	NC06001	71	1.00	1182	.78	347	.06
157	NJ01225	65	9.00	92	.09	36	.14
158	NJNPD00	65	9.00	2051	.89	628	.06
159	NJNPD00	65	9.00	2471	1.20	830	18.00
160	NM00101	80	7.00	538	.20	94	.08
161	NV01601	65	9.00	296	.08	74	.16
162	NY01401	81	7.00	-9	.00	-9	.
163	NY01401	82	8.00	2161	-270.25	691	.
164	NY01451	65	9.00	81	.03	27	.12
165	NY01451	81	7.00	-9	-0.03	-9	.
166	NY01451	82	8.00	168	-21.12	22	.
167	NY01455	77	5.00	-9	-0.02	-9	.
168	NY01455	78	5.00	266	-33.37	88	.
169	NY01455	81	7.00	-9	-0.02	-9	.
170	NY01455	82	8.00	247	-31.00	70	.
171	NY03102	65	9.00	141	.00	33	.00
172	NY03102	81	7.00	-9	-0.01	-9	.
173	NY03102	82	8.00	361	-45.25	99	.
174	OH01842	65	1.00	64	.08	22	.25
175	OH04807	82	12.00	-9	.00	-9	.
176	OH04807	83	1.00	3803	-475.50	792	.
177	OH05009	65	1.00	271	.09	59	.04
178	OH05702	72	10.00	-9	.00	-9	.
179	OH05702	73	1.00	0	-0.12	0	.
180	OH05702	73	12.00	20112	20113.00	6006	119.00
181	OH05702	74	1.00	1727	.08	636	.08
182	OH05702	80	12.00	2613	1.02	719	.09
183	OHCOP00	65	1.00	1210	.27	368	.08
184	OK07205	65	1.00	900	.29	234	.09
185	OR02002	65	1.00	244	.10	50	.09
186	PA00614	65	1.00	110	.02	32	.03
187	PAPEP00	65	1.00	3680	8.60	1065	24.00
188	RI00409	65	1.00	635	.22	194	.04
189	TN03301	65	1.00	342	.13	148	.09
190	TNMPD00	83	1.00	0	.00	0	.02
191	TNMPD00	83	12.00	56026	56027.00	18224	764.00
192	TX05711	65	1.00	180	.00	84	.00
193	TX05715	66	12.00	1	.00	0	.25
194	TX05715	67	1.00	259	130.00	44	1.00
195	TX10115	65	1.00	124	.04	17	.10
196	TX22012	67	1.00	1808	1.07	409	11.00
197	TX22012	70	12.00	2393	1.02	633	12.00
198	TX22012	72	12.00	1859	.93	642	14.00
199	TX22101	65	1.00	244	.06	67	.03
200	TX22701	82	7.00	0	.00	0	.03
201	TX22701	82	12.00	16960	16961.00	4607	98.00
202	TX24001	65	1.00	134	.05	52	.16
203	TX24001	66	1.00	0	.00	0	1.00
204	TX24001	66	2.00	124	125.00	44	1.00
205	TX24001	66	3.00	0	.00	0	1.00
206	TX24001	66	4.00	126	127.00	41	1.00
207	TX24001	66	8.00	0	.00	0	1.00
208	TX24001	66	9.00	140	141.00	45	1.00
209	TXHPD00	80	11.00	-9	.00	-9	.

210	TXHPD00	81	1.00	12827	-1603.50	4482	.
211	TXHPD00	81	10.00	-9	.00	-9	.
212	TXHPD00	82	1.00	14010	-1751.37	4807	.
213	VA00701	65	1.00	365	.23	64	.09
214	VA12800	74	10.00	884	1.11	144	.09
215	WA03204	65	1.00	271	.16	51	.05
216	WA03204	77	10.00	1067	1.04	228	14.00
217	WA03204	82	8.00	1317	1.05	273	11.00
218	WI01301	65	1.00	242	.05	46	.05

TOTAL is total UCR Part 1 Offenses. BURG is total burglary. Change measures are the ratio of the current year divided by the preceeding year, excluding January, 1965. Periods indicate missing or undefined values.

(Deutsch, 1978; McCleary, 1980) But here too questions have been raised about the appropriateness or adequacy of these models in view of UCR measurement problems (Barnett, 1984). Indeed, divergent results where attempts have been made to estimate the impact of gun control legislation may be due in part to such problems. Thus, parameters for the "noise" factor in ARIMA models or gun crime vary by jurisdiction and by type of offense (Bowers et al., 1984b; Loftin and McDowell, 1982; Margarita, 1984). And in some cases, different researchers using ostensibly the same data, though obtained from different sources, have produced different noise models and estimated effects (cf. Bowers et al., 1984b and Loftin and McDowell, 1982).

A framework for examining these analytical problems was presented in table 1-1. It showed that the analytical issues associated with a given combination of source and type of error. This framework implies that the appropriate technique for detecting a problem depends on the nature of the errors and their effect.

Table 1-1 indicates that change in the source or effect of error can alter the analytical issues. Solving only one source of error does not necessarily improve the validity of one's analysis. For example, removing confounding trends without adjusting for classification variability may merely shift the analytical problem from difficulty in replication to that of replicable models with invalid (biased) parameters. Confirmatory factor analysis alone or ARIMA time-series analysis alone cannot address the problems simultaneously. Each gives only a partial solution to the problem and the resulting estimates still retain more unreliability and invalidity (bias) than is necessary. This is especially true when making comparisons across groups, time points, or alternative models. To deal with the interdependence of these issues, a generalized error modeling procedure is necessary that represents variation in the source and effects of these errors.

ANALYTICAL PROCEDURES

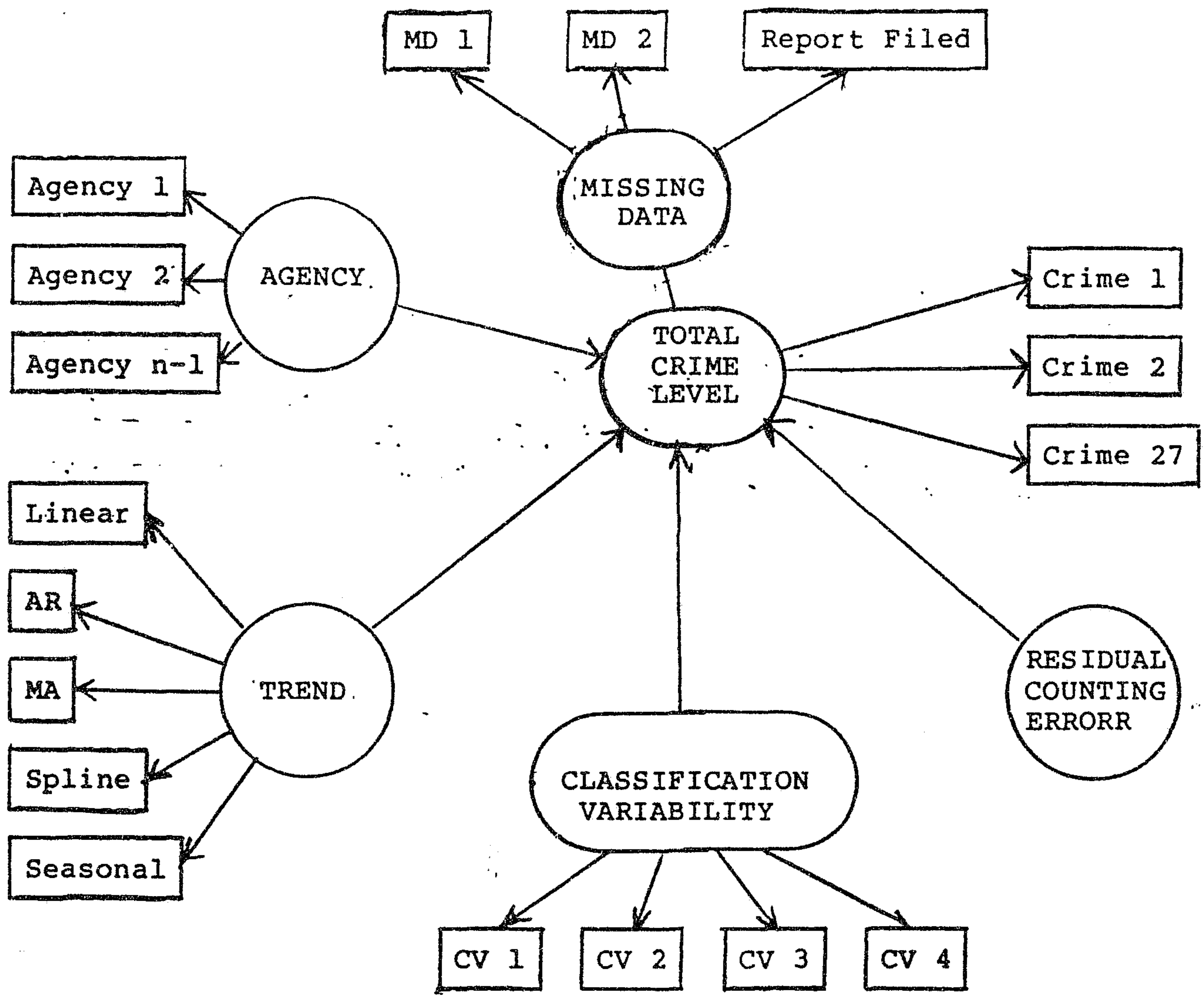
Joreskog has presented LISREL models appropriate for longitudinal data in which the errors have serial correlation and for multi-group, multiple indicator data (Joreskog and Sorbom, 1979). The extension of these models to time-series data in which the errors in indicators of latent variables are affected by autoregressive and moving average processes provides a basis for simultaneous modeling of error components of merged cross-sectional, time-series, multinomial distributions. This procedure has been called Structural Equation Time Series Analysis, SETSA (Holmes, 1985).

The intent of SETSA is to achieve model estimates that approximate Three Stage Least Squares (3SLS) estimation of latent variables. Direct computation of such model parameters is not possible because the model is underidentified. Such estimates can be approximated by introducing _a priori_ assumptions and constraints on the model

Figure 1-1 presented the preliminary measurement models and their hypothsized structural relationship to reported crime levels. Figure 5-1 presents a revision of that model in which numbers of crimes, by type, are treated as components of the reported crime level. This change more accurately reflects literature that suggests choice of crime type is an endogenous variable in models of criminal behavior. The change also shifts the units of analysis from that of crime-agency-months (i.e., crimes) to that of agency-months (i.e., agencies). This corresponds more closely to the research goals of examining sources of variability in agency reports. It is also more easily estimated, since it eliminates the need for dummy variables to indicate crime type and attendent heterscedasticity of errors.

FIGURE 5-1

REVISION OF STRUCTURAL MODEL



The meaning of symbols in the figure follows conventions discussed in chapter 1. Indicators for the unobserved variables remain the same.

To evaluate the adequacy of this strategy a design is proposed that will utilize UCR data in a merged cross-sectional time-series analyses. One hundred fifty-five jurisdictions will be selected for analysis, using data from 1965 to 1983. With monthly data this results in 45,060 observation points, a figure large enough to satisfy the requirements of confirmatory factor maximum likelihood analysis, which will be used for the structural equation time-series analysis (SETSA).

SETSA models will be estimated using multiple indicator confirmatory factor analysis to produce three stage least squares estimates of an unobserved structural model. This procedure will estimate source, strength, and nature of biases resulting from systematic missing data and discontinuities in trends and coverage. The contribution of possible changes in classification and reporting to these patterns will also be examined. Sensitivity analysis will compare model parameters by city size, region, and consistency of reporting to the UCR system to verify reliability and stability of the results.

<u>Reporting Agency</u>: will be measured by a set of dummy variables indicating whether or not the data is from a given agency. With one hundred fifty-five agencies to be included in the study, one hundred fifty-four dichotomous indicators of agency will be used.

Because dummy variables are negatively correlated and because the strength of that negative correlation is an artifact of the number of cases in each category, reporting agency is not a latent variable whose measurement parameters are free for estimation. Attempting to estimate the parameters of such a latent variable results in inconsistent estimates (see below in the section presenting results of cross-sectional estimates). Indeed, in this case the indicators must be treated as separate, observed variables.

. __Crime__ __Trend__: will be a reflection of functions computed from the data. Linear and seasonal trends will be computed from the monthly sequence of the data. The linear trend will correspond to a monthly counter for the one hundred ninty-two months of data in the study. The seasonal trend will be measured by dummy variables for the months of the year. The autoregressive and moving average functions will be computed from results of a correlogram analysis of time-series trend residual after removing first differences, linear, and seasonal trends and variation associated with crime type, agency, and missing data from the trend following the procedures of Malinvaud (1970).

Structural Equation Time Series Analysis can be regarded as three stage least squares regression (3SLS) in which unobserved variables are substituted for observed variables. Such a model is always underidentified unless sufficient constraints or restrictive assumptions are made. The constraints that allow estimating a SETSA model use estimates of time series trends and measurement models to obtain initial estimates of residuals of predicted endogenous variables. These, in turn, are used to refine the original estimates of trends and measurement models. This iterative procedure is similar to that used in refining time series (ARIMA) models with observed indicators, except that the residuals analyzed are estimates of latent variable residuals.

TIME SERIES COMPONENT

The time series latent variable is estimated by specifying linear, polynomial, or seasonal trend factors and autoregressive or moving average trends in the error terms. Linear, polynomial, and seasonal trends can be estimated by computing appropriate time series or dummy variables and checking to see if they are significantly related to variables in the model. Autoregressive or moving average trends in the error terms can be diagnosed by means of correlogram analysis.

A major difference between SETSA and the LISREL 2SLS estimation of Joreskog and Sorbom is that it adds ARIMA parameters to the equations. This implies that time series data must be "de-trended" before estimating the structural equations. Removing these trends can significantly alter the parameter estimates in the model.

An example of the importance of "pre-whitening" (removing trends) is found in Table 5-1. This table compares the correlations of the total crime level with its components using raw scores, first differences of raw scores, and first differences with moderate outliers removed.

TABLE 5-1

CORRELATION STABILITY OF TOTAL CRIME WITH INDIVIDUAL CRIMES

CRIMES	RAW WITH OUTLIERS	RAW WITHOUT OUTLIERS	DIFFERENCED WITH OUTLIERS	DIFFERENCED WITHOUT OUTLIERS
Murder	.93	.94	.31	.23
Manslgtr	.36	.37	.14	.03
Total Rape	.93	.94	.51	.24
Forced Rape	.94	.94	.48	.22
Atmpt Rape	.82	.82	.33	.13
Total Rob	.95	.96	.61	.71
Gun Rob	.88	.88	.42	.48
Knife Rob	.74	.75	.34	.36
Oth Wepn Rob	.75	.76	.35	.41
Str. Arm Rob	.95	.96	.66	.56
Total Aslt	.96	.97	.59	.51
Gun Aslt	.93	.94	.62	.42
Knife Aslt	.95	.96	.58	.43
Other Aslt	.94	.95	.55	.41
Hands Aslt	.85	.88	.30	.21

Simp Aslt	.70	.69	.73	.37
Total Burg	.98	.98	.94	.82
Forced Burg	.97	.97	.88	.67
Nforced Brg	.85	.86	.81	.43
Atmpt Burg	.80	.81	.70	.35
Total Larceny	.98	.99	.97	.89
Motor Theft	.97	.98	.81	.69
Auto Theft	.96	.97	.76	.64
Truck Theft	.45	.54	.30	.16
Other V Theft	.54	.53	.46	.21
Larceny [$50	.41	.43	.25	.29

It is apparent that substantial differences in the correlations appear when comparing the raw crime data with the differenced crime data. In contrast, fewer differences occur among the differenced measures when moderate outliers are dropped from the data. This underscores the importance of some form of detrending as part of the statistical analysis and supports a strategy of developing a multiple indicator equivalent of Three Stage Least Squares estimation, SETSA.

Correlogram Analysis

A correlogram is a matrix of correlations of residuals lagged across time. It shows whether the residual at a given point is correlated with the resudual for a different point in time. To specify an autoregressive integrated moving average (ARIMA) model of the correlations of error terms, correlograms were constructed of the first differences of the crime measures (first differences result when the crime value in a given month has subtracted from it the value for that crime in the preceeding month). To examine the effect of outliers on estimation of the time series trends, correlograms of the first differences were constructed with and without the cases of moderate outliers. The results are presented in tables 5-2 and 5-3.

TABLE 5-2
CORRELOGRAM OF FIRST DIFFERENCES, INCLUDING OUTLIERS

VARIABLE	TIME LAG OF FIRST DIFFERENCE							
	1st	2nd	3rd	4th	5th	6th	11th	12th
Murder	-.45	-.04	.07	-.02	.00	-.02	-.03	-.02
Manslaughter	-.47	.07	.00	.00	-.01	.01	-.01	-.02
Total Rape	-.47	.06	-.02	.01	-.03	-.04	.00	-.06
Forcible Rape	-.24	-.04	-.02	.01	-.03	-.04	.00	-.04
Attempted Rape	-.28	-.04	-.01	.00	-.01	-.02	.01	-.03
Armed Robbery	-.24	-.04	-.00	-.06	-.00	-.15	-.02	-.01
Strong Armed Robbery	-.28	-.01	.02	-.05	-.01	-.08	-.01	-.02
Total Assault	-.15	.11	.06	-.17	-.14	-.31	.12	.06
Total Burglary	-.42	.01	.00	-.01	.00	-.06	-.01	-.21
Forcible Entry	-.40	-.01	.00	-.00	.00	-.06	-.01	-.14
Non-forcible Entry	-.37	.01	.00	-.01	-.02	-.05	.00	-.25
Attempted Entry	-.22	-.01	-.00	-.00	-.02	-.03	.02	-.14
Total Larceny	-.37	.03	-.00	-.01	-.01	-.10	-.00	-.17
Motor Vehicle Theft	-.34	.03	-.01	-.04	-.01	.00	-.02	-.02
Auto Theft	-.27	.03	-.01	-.04	-.01	-.13	-.03	-.01
Truck/Bus Theft	-.40	-.02	-.03	.00	-.01	.03	-.00	.01
Other Vehicle Theft	-.29	.05	.00	-.07	-.06	-.14	.06	.01
Grand Total Part I	-.35	.04	-.00	-.02	-.00	-.10	-.01	-.17
Larceny under $50	-.18	.08	-.01	-.01	-.02	.00	-.01	.03

TABLE 5-3
CORRELOGRAM OF FIRST DIFFERENCES, EXCLUDING OUTLIERS

VARIABLE	TIME LAG OF FIRST DIFFERENCE							
	1st	2nd	3rd	4th	5th	6th	11th	12th
Murder	-.44	-.06	.08	-.07	-.03	.01	-.03	-.01
Manslaughter	-.48	-.00	-.06	-.07	-.06	.00	-.07	-.06
Total Rape	-.45	.09	-.08	-.08	-.05	-.05	-.03	-.06
Forcible Rape	-.46	.07	-.07	-.07	-.05	-.05	-.03	-.07
Attempted Rape	-.47	.05	-.07	-.06	-.06	-.02	-.02	-.05
Armed Robbery	-.13	-.05	-.02	-.02	.01	-.16	-.01	-.05
Strong Armed Robbery	-.16	-.03	-.04	-.08	-.06	-.07	-.00	.01
Total Assault	-.07	.14	.01	-.16	-.19	-.32	.11	.14
Total Burglary	-.15	.02	-.03	-.06	-.01	-.12	-.02	-.05
Forcible Entry	-.16	-.03	-.03	-.05	-.00	-.09	-.05	-.04
Non-forcible Entry	-.10	.01	-.03	-.05	-.08	-.09	.01	-.01
Attempted Entry	-.00	-.01	-.04	-.04	-.01	-.02	.02	-.10
Total Larceny	-.08	.06	-.03	-.10	-.05	-.19	.03	.02
Motor Vehicle Theft	-.19	.09	-.05	-.09	-.07	.00	.01	-.00
Auto Theft	-.15	.03	-.05	-.06	-.06	-.17	.01	.00
Truck/Bus Theft	-.33	-.06	-.15	-.13	-.13	-.02	-.13	-.13
Other Vehicle Theft	-.16	.05	-.02	-.13	-.12	-.15	.07	.07
Grand Total Part I	-.11	.07	-.03	-.11	-.03	-.20	.01	-.02
Larceny under $50	-.02	.09	-.06	-.02	-.08	.00	-.01	.04

Tables 5-2 and 5-3 demonstrate two important points regarding the correlograms used to estimate ARIMA models of UCR data. First, the autocovariance parameters are not greatly affected by deletion of outliers in the data. The corresponding coefficients in each table are very similar. The exceptions are infrequent enough that either outliers have no significant effect on parameter estimation or the effects of outliers are highly specific for some crimes or jurisdictions and not others. Second, the pattern in the correlograms strongly suggests that a first order moving average process characterizes the trend "noise." Coefficients after the first lag drop off precipitously. There is the possibility of a six month or twelve month seasonal lag effect, but they appear to be relatively weak compared to the first order effect. Indeed, the bivariate autocorrelation coefficients for sixth month and twelve month lags generally become non-significant when the effect of the first order lag is controlled.

This result has important implications for previous studies that used ARIMA modeling. Some of those studies have found a twelve month lag effect in the ARIMA parameters. All of those studies were based on a relatively few agencies. Given the smaller power of those studies, the finding of the lag may be a "false positive," although the effect could be present for some types of crimes and not others. Much closer attention needs to be given to the specification of the noise models used in the

ARIMA based studies. The sensitivity of those findings to presence or absence of a twelve month lag component in the noise model especially needs examination.

CLASSIFICATION COMPONENT

A measurement model for the classification variability latent variable was estimated using four indicators:
One of these indicators, , had a non-significant parameter from the indicator to the classification variability latent variable (a non-significant lambda coefficient). Consequently, it was dropped from the measurement model. With only three indicators remaining, the measurement model is exactly identified and and had a 1.0 goodness-of-fit index. Future studies will need to specify more indicators of classification variability to evaluate the measurement properties of this variable.

MISSING DATA EFFECTS

A measurement model for the missing data latent variable was not estimated ecause it is composed primarily of dummy variables. The use of dummy variables to indicate the presence or absence of missing data poses an obstacle for estimation of a latent missing data variable. Dummies that are coded from categorical variables are negatively correlated, with the magnitude of the correlation depending on the proportion of cases in a given category. Such a pattern of correlations produces an inconsistent measurement model when the indicators are measures of the same latent variable. Consequently, dummy variables for missing data have to be treated as separate indicators.

The solution to this problem lies with block variable estimation (Heise, 1972). Using this procedure the collective effects of groups of indicators can be estimated, even though the measures are treated as separate indicators. Thus, the effects of latent variables measured by dummy variable coding of categorical variables can be estimated by pooling the effect of each indicator into a multiple-partial regression coefficient (a "sheaf coefficient").

AGENCY DIFFERENCES

Estimation of agency effects in the SETSA model is also confounded by the necessity of using dummy variables to stand for between agency differences. Similar to the problem with missing data dummy variables, those that stand for each agency are negatively correlated, which means that a measurement model for a latent agency effect cannot be estimated. The dummy variables for each agency have to be used separately to estimate agency effects.

The solution to this problem also lies with use of sheaf coefficients to estimate the joint effects of the separate indicators. However, because of the large number of agency dummy variables, a screening process is needed to limit the number of dummy variables. Dummy variables for agencies that do not have a significant bivariate relationship with the crime measures are not included in the model.

ERROR ESTIMATION

Errors in the model are detected by inspecting the iterations required for estimates to stablize and by examining measures of the goodness-of-fit of the model overall and for specific parameters within the model. In addition, differences between the observed correlations between the indicators and those estimated by assumptions of the model also provide evidence of possible errors in the model.

No structural model specified had parameter estimates converging within sixty (60) iterations of the procedure. Neither maximum liklihood no unweighted least squares estimation produced stable structural models. Consequently, no measures of goodness-of-fit nor modification indices were calculated. Apparently the structural relationships of these variables are sufficiently more complex than the model specified that this model cannot be satisfactorily estimated.

A contribution of the SETSA procedure in this instance is to document deficiencies of the hypothesized model, which leads to revision of the conceptualization of how these variables are related to each other and to their indicators. A strength of SETSA when the conceptualization of a model is in dispute is that it allows empirical confrontation of hypothetical models. Even when the fit is not good, it provides a means of examining the consequences of one's assumptions. In effect, SETSA is conservative in accepting a model as having a good fit with the data. This requires the user to give serious thought to the formulation and specification of the model to be estimated. It is not easy to achieve a good model by merely ransacking around the covariance matricies.

MODEL REVISION

Revisions of the model concentrate on those parameters that have the strongest evidence of error. Those parameters that are not significant are deleted. Those that have large modification indices are either fixed at zero or freed for estimation, depending on the prior status of the parameter. Parameters involving large residual correlations are also fixed or freed to improve the goodness-of-fit of the model.

Since stability of the estimates was not achieved after sixty (60) iterations, revisions concentrated on parameters having implausible or zero values. Indicators and variables whose estimated parameters were zero were dropped from the model. Even so, stable estimates of the structural relationships were not achieved. While the resulting models may have predictive ability with respect to crime levels, the theoretical interpretation of the coefficients remains problematic.

The findings suggest further revisions of the crime level model. The correlations between observed indicators of different latent variables was a major source of the "badness-of-fit" of the model. This, combined with the extensive negative correlations of dummy variables implies that additional latent variables need to be added to the structural model and that the measurement models need further revisions. Specifically, indicators of agency, missing data, and trend variables should be thought of as antecedent to the latent variables. This results in reversing the direction of the arrows between the indicators and the construct. Such reversal has significant substantive and statistical implications. This change implies that these variables are ex-post facto constructs imputed to the indicators. They are not underlying causes of the indicators. These constructs do, however, serve heuristic utility. Statistically, the change means that there needs to be greater use of block variable estimation of latent variables that are the consequences of the indicators.

CLEANING AND ADJUSTING UCR DATA

Cleaning UCR data includes checking for outliers in the data and for missing values. The most extreme outliers among the larger agencies are listed in the appendix. Less extreme outliers may be identified using the procedures outlined in this study. Missing values may be replaced either by obtaining the data from the agencies, resulting in "adjusted" data, or the missing values may be estimated using SETSA or some other multivariate technique.

The findings in this study demonstrate that using adjusted data is préferible to using estimated data. The amount of data missing from the reports in larger jurisdictions is relatively small. Getting the actual values from the agencies involved is likely to be less error prone than trying to achieve stable parameters for some of the possible estimation models.

IDENTIFYING TRUE CHANGE

How can true change in crime levels or in the ratio of crime levels be separated from change associated with classification variability, trends, or missing data? This depends on whether or not analysis of UCR data uses a comparison or control group design. Estimation of true changes is much more reliable when comparison or control groups are used than when using a single time-series of data.

When comparison or control groups are used, change associated with the error factors estimated i the control group can be subtracted from the jurisdictions in the "treatment" group. Remaining change can be imputed to be true changes, depending on how "good" a control group is used.

When control groups are absent, more restrictive assumptions must be made about the nature of fluctuations associated with the error factors. For those crime types thought to be affected by an intervention, adjustment for all classification error change would remove true change as well. The researcher is faced with not adjusting for any classification error, removing only error associated with crime types thought to be unaffected by the intervention, or assuming that change of a given duration or magnitude are regarded as true change. For example, the researcher may assume that change lasting more than six months represents true change. Error associated with missing data can still be removed for this type of data.

These limitations when a single time-series is used also apply to traditional ARIMA models of impacts. The difference is that traditional ARIMA with a single series always assumes no missing data or classification error is present or that it is random. The proposed procedure allows removing classification error in some circumstances and removing missing data error in all circumstances. Even so, the advantages of using comparison or control groups with this method are so great that it would be much the preferred strategy when applying this method to the study of criminal justice interventions.

APPLICATION TO INTERVENTION ANALYSIS

How can true change in crime levels resulting from a legal intervention be separated from change associated with classification variability, trends, or missing data? This depends on whether or not analysis of UCR data uses a comparison or control group design. Estimation of true changes is much more reliable when comparison or control groups are used than when using a single time-series of data.

When comparison or control groups are used, change associated with the error factors estimated in the control group can be subtracted from change observed in the jurisdictions of the "treatment" group. Remaining change can be imputed to be true changes, depending on how "good" a control group is used.

When control groups are absent, more restrictive assumptions must be made about the nature of fluctuations associated with the error factors. For those crime types thought to be affected by an intervention, adjustment for all classification error change would remove true change as well. The researcher is faced with not adjusting for any classification error, removing only error associated with crime types thought to be unaffected by the intervention, or assuming that change of a given duration or magnitude will be regarded as true change. For example, the researcher may assume that change lasting more than six months represents true change and that change lasting less than six months is "error." Error associated with missing data can still be removed for this type of data.

These limitations when a single time-series is used also apply to traditional ARIMA models of impacts. The difference is that traditional ARIMA with a single series always assumes that there is no missing data and that classification error is not present or that it is random. The proposed procedure allows removing classification error in some circumstances and removing missing data error in all circumstances. Consequently, the advantages of using comparison or control groups with this method are so great that it would be much the preferred strategy when applying this method to the study of criminal justice interventions.

CHAPTER 6

EXECUTIVE SUMMARY

OBJECTIVES

The study focusses on three particular problems that affect analysis of UCR data: definitional uncertainty, missing data, and classification variability. Each of these problems affect the accuracy of statistics on the level of crime.

Definitional uncertainty means that users of UCR data cannot clearly or easily tell what some of the UCR data codes mean. Definitional uncertainty occurs because of the difficulty in finding definitions in the documentation for some variables, the ambiguity of some definitions, and changes in definitions.

Data missing from the UCR Return A means that a researcher cannot construct a complete record of all UCR offenses in a jurisdiction for a given month. Missing data occurs because some jurisdictions do not report every single month or reports may not be complete.

Classification variability implies that persons filling out Return A may not be consistent in how crime reports are categorized as UCR offenses. Classification variability occurs because information on some offenses may be sufficiently ambiguous to allow classification in more than one category, despite the hierarchy rule and other criteria in the FBI UCR Coding Handbook (FBI, 1980). It also occurs because of changes in the definitions of the measures and because specific events in a local jurisdiction may influence how reports are categorized, whether intentionally or unintentionally.

These problems directly lead to the objectives of this study: (1) greater clarification of the definitions of the UCR measures (2) identification of the likelihood and nature of missing information and classification variability in UCR data, (3) description of the agencies, time periods, and types of offenses for which such problems exist, (4) evaluation of the extent to which these data problems may bias the results of various kinds of statistical analyses, and (5) recommendations for the most efficient and effective methods for dealing with such threats to the validity and reliability of research findings.

SUMMARY OF PROCEDURES

Data analysis proceeded in three stages. The first stage constructed work files for alternative subsets of reporting agencies, according to criteria discussed below. The second stage performed a set of bivariate analyses in which measurement models of each construct and its indicators were

estimated. The bivariate analyses did not control for correlation of the indicators with other constructs or other appropriate classification.

CONCEPTUALIZATION

Variability in classification implies that for any crime report there may be uncertainty regarding the appropriate category into which it is placed. This uncertainty may stem from incomplete information regarding the crime, ambiguities in the definitions of specific crime categories, pressure to under- or over-report certain crimes, and other sources.

For several years Uniform Crime Report statistics have been the object of persistent criticism from legal, political, and sociological scholars. many of these critics have challenged the aura of infallability that has surrounded these crime statistics. While UCR statistics have been purported to provide accurate indices of crime rates, many have pointed to confounding influences which serve to undermine the validity of officially reported crime rates. The criticisms relating to the assumed unreliability of UCR statistics may be collapsed into two separate categories. First, definitional problems are said to influence reporting of crimes (see Kituse and Cicoural, 1964). For instance, it is commonly the case that violence stemming from domestic disputes are not reported by police officers as assualtive crimes. Subsequently, crime rates regarding assualt may be therefore be subject to report bias and not accurate reflections of the frequency of its occurance. A second problem associated with these statistics relates to the issue of classification variability. While this report cannot address unreliability involving the former, the issue of classification variability is central to our objectives. Indeed, it is critical to assess the potential impact which artifically induced measurement characteristics have upon the rates of reported criminal offenses.

Classification variability may be operationalized as the degree of fluctuation within crime statistics which are artificial by-products of recording or accounted, and/or reporting practices by law enforcement officers. Misclassification of criminal offenses may occur as a result of variability in either recording of crimes or in the actual reporting decisions made by police officers. While the reasons for recording and reporting variability may differ it has been generally assumed that such practices restrict the reliability and subsequent utility of UCR statistics (Center and Smith, 1973).

Recording or accounting variability refers to changes in the official classification of crimes. Slight changes in the way crimes are accounted for such as the implementation of new bookkeeping practices or through technological innovations in crime recording may produce "paper" fluctuations in crime rates. Whenever a UCR reports a significant increase or decrease in crime trends one must question whether such changes are not simply the result of alterations in the manner in which crimes are officially recorded. For instance, Chilton and Spielburger (1972) in their analysis of alledged crime rate increases in six metropolitan areas, found that changes in rates reflected changes in the operation of police departments combined with changes in record keeping procedures. Similar artificially induced fluctuations in crime rates occured in New York City in 1966. New methods of recording and reporting crimes created the illusion of fluctuation (Dusheck, 1966). In a related analysis of shilfing crime trends in Baltimore, Twigg (1972) found that the implementation of a modern computerized system of crime reporting created the impression of rising crime rates. These fluctuations were seen as the direct result of new departmental procudures.

Police administrative procedures have been reported elsewhere as creating an image of increasing crime rates. In Seiman and Couzens (1974) remarkable study of crime reporting it was found that changes in larceny statistics in Chicago from about 10,000 to about 30,000 were associated with a change in bookkeeping procedures instituted by Orlando Wilson as part of his organizational reform program. As these authors point out, administrative changes in departmental guidelines and record keeping are common and may have a substantial impact on the image of crime conveyed in official statistics. Such studies subsequently challenge the assumed face-validity of crime statistics as being able to accurately portraying rates of criminal offenses.

Misclassification of crimes may also occur as a result of uncertainty and ambiguity which surrounds the accurate identification and classification of an offense. That police officers possess a great deal of discretion bears witness to the possibility that reporting practices can be arbitrary, variable, and even manipulated. Indeed, such cases of classification variability are often a result of both unintentional and intentional misclassification. The presence of such confounding influences could have inflationary or deflationary effects on crime statistics thereby restricting their validity.

Perhaps the most innocently motivated form of misclassification occurs as result of errors in translating facts of a crime into an official report. In their study of classification biasing of crime statitics, Ferracuti and his associates (19) found that a considerable number of reporting errors occured for certain crimes. Robbery was correctly identified by only 59of the subject while manslaughter was correctly identified by 87of the subjects. This study concludes that while the amount of actual errors may be unknown, it is probable that errors made in the identification and classification of crimes constitute a potential source of misinformation in crime statistics.

While misclassification often results from perceptual bias, if also frequently occurs in intentional ways. It has been demonstrated, for instance, that political and organization pressures can result in either intentional upgrading or downgrading of crimes. In most instances of intentional misclassification police hav deliberately downgraded criminal offenses. Downgrading occurs when a police officer intentionally reduces an index crime to a lesser index or non- index crime. In such cases criminal statistics fail to reflect actual crime trends and the meaning of the crime statistics is uncertain.

The result of this uncertainty is that errors will occur in the classification. Reports may be put in the wrong category. Such error in classification may be simple random error or it may be a systematic consequence of influence variables. For large data sets, a small amount of random error in classification is likely to average out and not distort the findings. Systematic error may not.

The use of a measurement modeling strategy is intended to document sources of variability and document their impact. If their impact is small, they may be ignored. Otherwise, action will have to be taken to minimize the effects of the sources of variability.

MEASUREMENT OF VARIABILITY

The measurement of variability is influenced by pressures on police departments to alter reported crime statistics. In the words of one critique:

> Sometimes the pressure is to show crime is being reduced. Sometimes the pressure is to increase the number of crimes. These pressures impinge upon the data gathering system, the police departments, and in some cases affect the statistics, entirely apart from the effects of the number of crimes which are actually committed. Consequently, those indicators almost invariably used for these purposes are highly misleading for what they are said to measure (Seidman and Couzens,1974:484).

These pressures produce variability in the reported crimes. Certain patterns of variability may be more indicative of these pressures. Consequently, measures of these patterns will be used to identiy months in which counting variability is more likely.

Classification variability is measured by a set of indicators suggested by information from the FBI UCR Coding Handbook (1983). In addition, statistical measures of outliers in the data are also used to identify extreme variability in classification.

For certain classes of crimes, classification variability is more likely. For example, an assault might be classified as either simple assault or as aggravated assault. Sudden changes in the ratio of simple assault to aggravated assault may indicate a change in classifying the cases, in as much as the driving process that produces assaults is likely to be similar for both types. Consequently, changes in the ratios of such crimes may indicate classification changes, rather than true changes. Any sudden shift in the level of a crime reported from one month to the next may also indicate that variability in the classification has occurred (see, for example, Diagrams 1-1 and 1-2 in Chapter 1).

While each such ratio change is a less than perfect indicator of classification error, a set of such indicators can provide evidence that such changes might be occurring. Crime types whose change ratios will be computed are: simple and aggravated assault, felony and non-felony robbery, burglary and larceny-theft, rape and aggravated assault, as well as homicide and manslaughter.

MISCLASSIFICATION

When a crime report is classified in a category contrary to FBI guidelines, the report is <u>misclassified</u>. The misclassification may be unintentional or intentional. In the former case the error can introduce either random or systematic variance into the measurement of variation in crime. In the latter case it is usually systematic error that is introduced. Unintentional misclassification introduces systematic error when the failure to code crime in one category (say, burglary) occurs with a tendency for the report to be erroneously classified in a particular other category (say, attempted entry).

Several instances of reporting variabilty have been identified that may be examples of misclassification. Chilton (1979) reports several cases in which decreases in crime were the result of deliberate misclassification. The author cites San Fransico for illustrative purposes. Early in 1970, Mayor Alioto appointed a new police chief and held him personally responsible for reducing crime. When UCR statistics for the first quarter were released San Fransico appeared to have experienced a 7% reduction in crime. Chilton concludes however, that political pressures lead to manipulation in reporting that gave the impression of crime reductions. Indeed, police officers themselves confirm the manipulability of these statistics (Chilton, 1979).

In their article on misclassification, Seidman and Couzens (1974) explored administrative efforts to downgrade larceny, robbery, and burglary. Evidence of such downgrading was found in several cases. Egregious misvaluation of property was used to downgrade serious larcenies to the lesser charge of larceny under 50 dollars. Police- determined values, often roughly $49 dollars, were found to be seriously discrepent with insurance claims. From this the authors draw the inference that the

Washington District police were deliberately misvaluing property in order to produce an artificial decline in the total number of index crimes. Motivation for such practices were found to be related to organizational policies. In addition to Washington, these authors found five other separate instances in major metropolitian areas where larceny showed similar patterns.

FLUCTUATIONS IN CLASSIFICATION

Variation in classification may be small or gradual. When there is a sudden change followed by a return to the previous level, a <u>fluctuation</u> can be said to have occurred. Fluctuations are the result of misclassification, short term crime events, changes in the definitions of crimes, or various combinations of these factors. Whatever the cause, the essence of fluctuation is a significant short term change in a crime level. Determining what constitutes a significant change and what is a short term change is difficult, since there are many different opinions of what is significant and what is the short term.

The major issue for identifying fluctuations is the uncertainty of criteria for differentiating small incremental variations from large fluctuations. Change is a continuum. Any specific cutting point for calling something a large change can be debated. Consequently, two alternative criteria are here used: one that is fairly restrictive in calling something a fluctuation, a criterion for extreme changes, and another one which includes more moderate changes. In a situation where the criteria are debatable, using several criteria allows one to examine the consequences for one's results of using the different criteria. If the results are robust in the face of different criteria used, then confidence can be had in the conclusions. If the results disagree, then greater attention will be needed to the justification and reasonableness of the criteria.

The criteria apply to two types of month-to-month change. Extreme change was defined as a ten-fold increase or decrease (a change ratio greater than 10.0 or less than 0.10). Moderate change was defined as a twenty percent increase or decrease (a cange ratio greater than 1.20 or less than 0.80). When the effect of criteria for extreme or moderate changes is being examined later in this study, these are the criteria being used.

DISJUNCTIONS IN CLASSIFICATIONS

Whatever the criteria chosen for defining the "short term," some changes will persist beyond that time. These changes are disjunctions. Fluctuations differ from disjunctions in that two significant changes are required for a fluctuation to occur, the original deviation and the return to a baseline, whereas disjunctions only require one change with no return to the original baseline (however the baseline is defined).

Criteria for identifying significant deviations from the baseline are also ambiguous. However, a strategy similar to that for fluctuations can be followed. If fluctuation in one month is not followed by a reverse fluctuation within three months, then a disjunction can be said to have occured. Such disjuctions were studied using both criteria for fluctuations to see how stable the results are for disjunctions.

OUTLIERS

As noted above, outliers may be identified using traditional z-score criteria, by inspection of the moments of the distributions, using Fisher's g1 (skewness) and g2 (kurtosis) statistics, and by applying substantive criteria to define some changes in values as "extreme" or "moderate."

Outliers can affect the "moments" of a distribution. The second, third, and fourth moment of a distribution are the basis of measures of variance, skewness, and kurtosis of distributions. If outliers significantly alter the distrbution of a variable, it must affect the moments of that distribution and these statistics derived therefrom. Tables 4-1 and 4-2 examine the effect outliers on the distributions by comparing statistics using these moments with and without the outliers included. If there are no differences in the variances, skewness, and kurtosis when outliers are dropped, the outliers cannot be said to have significantly altered the distribution.

Table 4-1 examines the effect on means and variances when outlier is defined as an extreme change ratio (a ten-fold increase or decrease).

TABLE 4-1
EFFECTS OF OUTLIERS ON MEANS AND VARIANCES

CRIME	WITH OUTLIERS MEAN	VARIANCE	WITHOUT OUTLIERS MEAN	VARIANCE	F-RATIO
Murder	4.1	136.4	4.0	130.1	1.04
Manslaughter	0.9	5.7	0.9	5.5	1.03
Total Rapes	11.7	776.7	11.5	734.2	1.05
Forced Rape	8.8	478.6	8.7	451.3	1.06
Attmpt Rape	2.8	46.0	2.8	44.1	1.04
Total Robbery	120.5	246,101.4	117.1	228795.7	1.07
Gun Robbery	58.1	61,739.5	56.3	56874.1	1.08
Knife Robbery	10.2	6,103.2	9.9	5672.8	1.07
Other Weapon Robbery	7.0	3,158.4	6.8	2966.9	1.06
Strong Arm Robbery	44.6	25,964.5	43.6	24477.8	1.06
Total Assaults	68.5	46080.9	67.4	45090.9	1.02
Gun Assault	23.3	3790.2	23.0	3621.8	1.04
Knife Assault	26.6	7356.3	26.1	7043.0	1.04
Other Assault	27.7	7949.7	25.3	7696.2	1.03
Hand/Feet Assault	12.5	1157.0	12.3	1115.7	1.03
Simple Assault	130.9	98885.0	129.4	95186.6	1.03
Total Burglary	499.2	1357806.4	491.1	1268773.6	1.07
Forcible entry	381.6	807912.7	375.4	752922.6	1.07
No Force	81.2	36068.4	80.2	34155.4	1.05
Attempted Entry	33.8	20149.3	33.1	18860.2	1.06
Total Larceny	865.1	2409035.3	855.9	2284328.5	1.05
Total Motor Theft	220.1	390414.0	216.1	370780.4	1.05
Attempted M. Theft	199.7	356060.8	195.9	337451.7	1.05
Truck/Bus Theft	11.5	2269.9	11.4	2244.4	1.01
Other Veh. Theft	8.1	436.0	8.1	425.9	1.02
Grand Total Part I Larceny under $50	199.0	260912.7	197.4	252780.4	1.03

This table reveals that when outliers are measured using this definition, little effects of the outliers on the means and variances appear to be present. The means of the data with and without the outliers are very similar. The F-ratios comparing the two variances are also very small. Although not shown here, similar results were obtained when moderate outliers were also excluded. These findings imply that with large data sets presence of outliers does not significantly bias the means or standard deviations. For small sub-sets of UCR data, however, such bias might still occur.

Table 4-2 shows the effect of outliers on summary measures of the distribution using Fisher's g1 and g2 measures of skewness and kurtosis, respectively.

TABLE 4-2
SKEWNESS, KURTOSIS, AND OUTLIERS*

CRIME CLASSIFICATION	WITH OUTLIERS SKEW	WITH OUTLIERS KURT	WITHOUT EXTREME OUTLIERS SKEW	WITHOUT EXTREME OUTLIERS KURT	WITHOUT MODERATE OUTLIERS SKEW	WITHOUT MODERATE OUTLIERS KURT
Murder	7.7	78.1	7.2	65.8	7.1	61.6
Manslaughter	5.8	52.1	6.6	60.5	6.5	58.0
Total Rapes	7.2	73.0	6.6	53.4	6.3	49.4
Number Forcible Rapes	7.4	78.2	6.7	56.5	6.5	52.6
Number Attempted Rapes	6.8	60.8	6.3	49.2	6.1	44.8
Total Robbery	11.4	165.7	10.1	117.9	9.7	107.7
Gun Robbery	13.3	230.2	11.4	166.6	11.0	152.6
Knife Robbery	18.0	346.1	15.0	239.9	14.3	216.5
Other Weapon Robbery	17.2	322.0	14.4	222.4	13.7	201.5
Strong Arm Robbery	10.6	142.9	9.2	104.0	8.9	96.2
Total Assaults	9.4	109.2	8.1	78.6	7.6	70.0
Gun Assaults	8.0	83.5	7.3	64.9	7.0	60.0
Knife Assaults	9.8	116.8	8.4	84.2	8.1	77.8
Other Weapon Assaults	11.3	153.2	9.5	107.4	9.1	97.6
Hands and Feet Assault	8.8	102.0	8.2	83.0	7.9	76.3
Simple Assault	8.0	83.5	7.3	64.2	7.2	61.6
Total Burglary	9.3	106.6	8.4	83.1	8.0	75.9
Forcible Entry	9.3	109.6	8.5	86.5	8.2	79.3
Non-forced Entry	8.2	91.7	7.7	75.8	7.4	69.2
Attempted Entry	18.6	396.8	15.7	278.0	15.0	252.2
Total Larceny	7.6	77.0	7.0	62.4	6.8	58.4
Motor Vehicle Theft	8.6	92.4	7.7	71.5	7.5	66.5
Auto Theft	8.9	99.2	7.9	75.0	7.6	69.6
Truck and Bus Theft	17.4	590.0	23.5	1291.5	11.3	174.3
Other Vehicle Theft	7.7	89.5	7.5	77.8	7.3	71.8
Grand Total Part I	8.7	95.5	7.8	73.8	7.6	68.4
Larceny Under $50	7.1	71.4	7.7	63.3	6.5	54.5

*Extreme outliers are those in which the ratio change from month-to-month is outside a 0.1 - 10.0 range. Moderate outliers are those in which the ratio change from month-to-month is outside a 0.8 - 1.2 range.

This table shows that the UCR raw data are significantly skewed and leptokurtic. Such a distribution is primarily a function of having zero as a lower boundry of the distribution. The principle role of zeros in making the distribution skewed and leptokurtic is revealed by examining what happens when the outliers are dropped from the sample. Using two different definitions of outliers, dropping them does not eliminate extreme skewness and leptokurtosis, even though it is reduced.

The non-normality of the distributions is exaserbated by the presence of outliers. Dropping the outliers does make the distributions more normal. However, the improvement in the normality of the distributions is not enough to make them close to normality. The problem of outliers with large samples cannot be solved solely by dropping cases. There may be exceptions with individual agencies in which there are few zero values and the outliers dominate the variance of crime for that agency. For most agencies this is not the case.

ADJUSTING CLASSIFICATION VARIABILITY

Statistically adjusting the data to remove classification variability is an alternative to dropping outliers. If one can accurately estimate variation associated with fluctuation, disjunction, or misclassification, then observed values and parameter estimates can be adjusted to remove those estimated effects. The problems in measuring classification variability mentioned above and the estimation difficulties mentioned in the discussion of adjusting for missing data (see chapter 3) indicate this is no easy task.

CONCLUSIONS

The findings show that UCR reports can be highly variable from month to month, but most of the time they are not. Indeed, instability in the reported values was more a function of the agency size. The smaller of these large agencies had more outliers than the larger ones. The anchoring effects of having many months with zero offenses reported for a crime type was more of a problem for statistical estimation than that of a few large increases. This does not reduce the substantive importance of crime increases. It does suggest that interpretation of such increases needs to be in light of the baseline and trend values from which the deviations occur.

The findings also show that the choice of criteria for outliers has little effect on the findings. Although the amount of variability is effected by the criteria, excluding the more variable cases by either criteria did not greatly alter the findings. Consequently, strategies of dropping cases are unlikely to resolve problems of classification variability.

CHAPTER 5

STRUCTURAL EQUATION TIME SERIES ANALYSIS

With the application of increasingly sophisticated statistical techniques to UCR data, there has come growing recognition that measurement problems pose a threat to the validity and reliability of research findings. For example, investigators have demonstrated a biasing effect of unreliability in the measurement of criminal homicide on analytic results (Bowers and Pierce, 1975; Klien et al., 1978; Holmes, 1983; Bowers et al., 1984a: Chapter 9). More generally, the National Academy of Sciences Task Force on Deterrence and Incapacitation identified measurement problems as a major methodological difficulty with existing deterrence studies (Blumstein et al., 1978). Others have echoed and elaborated upon this judgment (Brier and Feinberg, 1978), and still others have explicitly incorporated assumptions about measurement error into their analyses (e.g., Hellman and Fox, 1984; Parker, 1985).

In research on the impact of criminal justice policy and programs in specific states and localities, investigators have begun to work with agency specific, temporarily disaggregated UCR data, conducting interrupted time-series analyses with ARIMA (auto regressive integrated moving average) modeling techniques report (missing-available). A report may be filed having partial information (incomplete). The data missing from a report may be subsequently available, resulting in an adjustment for the report in which the data belongs (adjusted). The data may be unavailable, but it can be estimated from other data (estimated).

Each type of missing data may have a different effect on the crime counts. Hidden crime creates a greater problem for some crimes than others. Structural barriers may result in missing reports due to "underreporting," the failure of people to file reports of crimes they experience. This type of missing data is thought to be more concentrated in crimes of rape, burglary, simple assault, or petty larceny. The underreporting of murder, manslaughter, robbery, major larceny, or aggravated assault is believed to be substantially less. Analysis of data involving these latter crimes may not be greatly distorted by data missing from underreporting. With the former crimes, distortion of change measures will be small unless there is evidence of significant changes in the reporting practices for these crimes during the period of one's study.

Unavailable data creates problems in some jurisdictions, but not in others. Missing, but available, and adjusted data are more of a problem in recent reports that have not yet had time enough for the missing data to become available or to be adjusted than is true for earlier reports. Data missing from incomplete reports may be confined to specific types of crimes or periods of reporting and may be estimable. Before giving up on UCR data, a researcher must first decide whether the missing data creates problems for the particular analytical issue to be dealt with and also decide that none of these strategies will resolve the problem.

SOURCES OF MISSING DATA

Four sources of missing data are particularly important: structural, organizational, technological, and political. Structural sources result in missing data because there may be barriers at the societal level to recognizing or reporting the crime. Organizational sources of missing data are concentrated in procedures or personnel within criminal justice agencies that result in lost, incomplete, delayed, or absence of UCR Return A reports. Technological sources of missing data are lie mainly with the procedures for data storage and retrieval in an agency. Politics may also result in absence of UCR data. In some jurisdictions local political pressure may temporarly disrupt the reporting of crime data that is politically unpopular.

Organizational reasons of missing UCR reports often are associated with changes in personnel or agency procedures. For example, the person responsible for sending in the Return A report may quit and it may take a period of time before the job is filled, during which no one may take responsiblity for sending in the reports. When inspecting some cases in which missing reports clustered together in a series of months, some agencies involved were called and asked about the clustering. The most common reason given was that the difficulty of filling personnel positions meant there were periods when filing the UCR reports had lower priority because of a shortage of personnel.

The technology used in producing these reports is quite varied. In some cities it is highly automated. In others a great deal of manual tabulation is involved. In either case, a computer may bread down or file cabinets in a basement may be flooded.

MEASUREMENT OF MISSING DATA

Indicators of missing data will include a set of dichotomous variables denoting presence or absence of crime count data of a specific crime for a given agency month. A different indicator will identify whether a UCR report was filed for that month. Counts of the number of crimes an agency is missing data in a month and of the number of months an agency is missing data will provide additional indicators of missing data. Information for these indicators will come from the Card Type variables and from the missing value codes for each crime.

DESCRIPTION OF MISSING DATA

The vast majority of the American population (94.2% in 1980) is covered by UCR reports. However, there is considerable variation between states in the extent of coverage. In 1980 six states had all of their population covered by twelve monthly reports (see table 3-1, adapted from Schneider and Wiersema, 1985). Arizona, Delaware, District of Columbia, Maine, North Dakota, and Hawaii had 100 percent coverage of their population in 1980.

TABLE 3-1

POPULATION COVERAGE BY UCR REPORTS, 1980*

STATE	POP WITH COMPLETE COVERAGE	POP WITH INCOMPLETE COVERAGE	AGN WITH COMPLETE COVERAGE	AGN WITH INCOMPLETE COVERAGE	PCT POP COMPLETE COVERAGE	PCT AGEN COMPLETE COVERAGE
AL	3,732,184	129,282	309	51	96.6	85.8
AZ	2,715,357	0	112	0	100.0	100.0
AR	2,148,649	135,388	193	25	94.1	88.5
CA	23,491,824	40,856	593	9	99.8	98.5
CO	2,211,091	667,316	140	107	76.8	56.7
CT	2,607,372	487,852	95	9	84.2	91.3
DE	594,779	0	55	1	100.0	98.2
DC	635,233	0	3	0	100.0	100.0
FL	9,561,730	5,382	692	9	99.9	98.7
GA	4,623,950	776,901	305	220	85.6	58.1
ID	938,649	4,980	103	3	99.5	97.2
IL	10,651,869	703,193	515	155	93.8	76.9
IN	4,392,192	1,068,911	263	76	80.4	77.6
IA	2,574,896	332,908	205	23	88.6	89.9
KS	2,307,976	46,807	224	10	98.0	95.7
KY	3,641,128	351	467	4	99.9	99.2
LA	3,194,180	1,005,362	107	70	76.1	60.5
ME	1,123,670	0	130	1	100.0	99.2
MD	4,189,945	2,266	166	5	99.9	97.1
MA	5,028,259	700,029	292	127	87.8	69.7
MI	9,033,372	194,756	615	60	97.9	91.1
MN	4,043,733	17,502	277	5	99.6	98.2
MS	1,482,955	1,028,536	107	92	59.0	53.8
MO	4,094,973	806,315	204	85	83.6	70.6
MT	679,572	102,020	75	24	87.0	75.8
NB	1,465,032	98,889	165	44	93.7	78.9
NV	799,090	1,222	33	2	99.8	94.3
NH	877,493	41,621	116	8	95.5	93.5
NJ	7,335,695	6,469	512	89	99.9	85.2
NM	1,126,965	168,509	86	29	87.0	74.8
NY	17,393,554	1,131,136	748	119	99.4	86.3
NC	5,730,015	113,650	413	254	98.1	61.9
ND	652,473	0	78	2	100.0	97.5
OH	9,594,384	1,172,424	404	104	89.1	82.3
OK	2,848,687	152,565	255	48	94.9	84.2

TABLE 3-1 (CONTINUED)

STATE	POP WITH COMPLETE COVERAGE	POP WITH INCOMPLETE COVERAGE	AGN WITH COMPLETE COVERAGE	AGN WITH INCOMPLETE COVERAGE	PCT POP COMPLETE COVERAGE	PCT AGEN COMPLETE COVERAGE
OR	2,363,804	246,673	212	50	90.6	80.9
PA	11,619,778	204,442	1,130	79	98.3	93.5
RI	936,723	9,112	42	10	99.0	80.8
SC	3,043,170	21,387	190	74	99.3	72.0
SD	508,299	163,182	60	34	75.7	63.8
TN	4,284,409	261,181	251	44	94.2	84.8
TX	12,551,930	1,617,899	730	17	88.6	97.7
UT	1,433,377	25,352	115	4	98.3	96.6
VT	44,975	286,450	6	32	13.6	15.8
VA	5,313,482	9,930	385	11	99.8	97.2
WA	3,972,915	140,416	161	32	96.6	83.4
WV	1,885,881	44,906	333	34	97.7	90.7
WI	4,677,422	3,495	268	3	99.9	98.9
WY	454,561	14,393	66	3	96.9	95.7
AK	397,637	2,505	25	4	99.4	86.2
HI	964,680	0	8	0	100.0	100.0
US	211,975,933	13,176,721	13,076	2,477	94.2	84.1

*Agencies and associated populations that were "covered by" others are not included. Agencies under 100,000 population are included.

Two states had much less of their populations covered by UCR reports. Vermont had reports covering only 13.6 percent of its population and 15.8 percent of its law enforcement agencies; Mississippi, 59 percent of its population and 53.8 percent of its agencies. Other states with significant gaps in their coverage include colorado (76.8% of population covered) and Louisiana (76.1%). Generally, most of the population is covered (94.2%). The fact that a smaller percentage of agencies file complete reports than population is covered implies data is more likely to be missing from smaller agencies. Legal impact studies for smaller agencies are more subject to problems of missing data.

Much less data is missing for agencies covering populations greater than 100,000. Table 3-2 is a summary of types of UCR reports for each of the different crime categories for the agencies in our sample. A very small percentage of the reports were missing data among these agencies for this period.

Two significant facts stand out in table 3-2. First, the overwhelming majority of reports for these agencies were complete reports. This is because of the significant number of normal 44,868) or adjusted (12,986) reports. Few reports for these agencies were estimated (3) or missing (46 or 115, depending on the type of crime). Second, few differences in completeness occur for the types of crime. There is only a slight tendency for larceny crimes to be the crimes excluded from incomplete reports.

TABLE 3-2
REPORT TYPES BY CRIME

CRIME	REPORT MISSING	REPORT ESTIMATED	REPORT NORMAL	MONTHS COMPLETE	MONTHS INCOMPLETE
Total	115	3	31,956	44,868	192
Murder	46	3	31,956	44,856	204
Manslaughter	46	3	31,956	44,856	204
Total Rape	46	3	31,956	44,856	204
Forced Rape	46	3	31,956	44,856	204
Attempted Rape	46	3	31,956	44,856	204
Total Robbery	46	3	31,956	44,856	204
Gun Robbery	46	3	31,956	44,856	204
Knife Robbery	46	3	31,956	44,856	204
Other Weapon Robbery	46	3	31,956	44,856	204
Strong Armed Robbery	46	3	31,956	44,856	204
Total Assault	46	3	31,956	44,856	204
Gun Assault	46	3	31,956	44,856	204
Knife Assault	46	3	31,956	44,856	204
Other Weapon Assault	46	3	31,956	44,856	204
Hands Assault	46	3	31,956	44,856	204
Simple Assault	46	3	31,956	44,856	204
Total Burglary	46	3	31,956	44,856	204
Forced Entry	46	3	31,956	44,856	204
Non-Forced Entry	46	3	31,956	44,856	204
Attempted Entry	46	3	31,956	44,856	204
Total Larceny	46	3	31,956	44,856	204
Motor Vehicle Larceny	115	3	31,956	44,868	192
Other Theft	115	3	31,956	44,868	192
Truck/Bus Theft	115	3	31,956	44,868	192
Other Vehicle Theft	115	3	31,956	44,868	192
Larceny Under $50	115	3	31,956	44,868	192

Among the 45,060 agency months of reports, three were recorded as containing "estimated" data. Agencies with over 50,000 population are not supposed to have any estimated values. This apparent anomaly occurred because one agency with missing reports had a population under 50,000 in 1965 and over 100,000 in 1980.

EFFECTS OF MISSING DATA

Missing data may function as "white noise" or it may systematically distort information in the UCR system. If extreme values are missing from a set of reports, possibly due to work overload in the extreme situation, this could distort one's findings. Personal discussions with personnel who file such reports suggest that this is possible. With a larger number of agencies studied, however, the effects of the missing data on analysis using these agencies is likely to be small. When one is examining a small number of agencies, however, this may be a problem.

Missing data is especially likely to complicate analysis of smaller agencies. Nearly two-thirds of the missing data (64.3%) was from agencies whose population base was under 100,000 in 1980. All of the estimated months of data (3) were also for these smaller agencies. The presence of missing or estimated data may undermine the reliability of studies using smaller agencies. Although, most of the smaller agencies did provide complete reports when they were submitted. A judicious selection of agencies may get around this problem.

DROPPING MISSING DATA

Because so few of the larger agencies had missing data, a strategy of dropping agencies missing data relevant to a given study is fesible. However, this could on rare occasions result in losing cases needed for examining legal impacts in a given jurisdiction. If done with caution, dropping cases of larger agencies appears to be a viable strategy.

Dropping cases from smaller agencies is another question. Given the higher incidence of missing data, dropping agencies may significantly distory one's findings. Further exploration of this with smaller agencies than those in this study is warranted.

ESTIMATING MISSING DATA

There are occasions when the agencies one wishes to study are agencies that are missing data. In those circumstances some of the missing data may be estimated. Indeed, the FBI will estimate values for an agency if there are fewer than three months of data missing in a year and the population served by the agency is under 50,000. If the number of missing data points is small enough, it may be possible to contact the agencies involved and obtain the missing data from their local records.

How credible are the estimation procedures and the results? The use of multiple criteria for evaluating the results of one's models for estimation discussed in chapter 1 will strengthen the credibility of adjustment procedures. If one does not obtain models using SETSA that meet the tests for plausible and consistent results, estimation cannot be recommended.

Table 3-2 shows that for larger agencies there is not a lot of missing UCR reports. Studies that use primarily larger agencies should find it feasible to track down the missing values from the state UCR offices or the local agencies themselves. Given the finding that these larger agencies are missing relatively few reports, an estimation procedure should be used with these agencies only after trying to obtain the actual missing values.

ADJUSTING MISSING DATA

"Adjusting" the data for missing values involves obtaining it from the jurisdiction involved. It is an alternative to estimating values or dropping missing cases. The validity and reliability of one's adjustments depends on the accuracy of the revised values from the local agencies. Almost thirty percent of the reports in this study were classified as "adjusted." However, Chapter 2 explains why the actual number of agency months having adjusted data is considerably smaller than the 12,986 found in this study. Discovery of the true amount of adjusted data awaits more detailed information regarding which agencies had data that were adjusted.

CONCLUSIONS

An important finding of this study is that so few reports are missing from the larger agencies. Many of the larger agencies consistently file UCR reports. When a researcher is faced with missing UCR data, it is important to consider its likely impact. The missing data may not be relevant to the particular analytical issue at hand. Even when the missing data is relevant, this study has found it is often feasible to track down the missing reports and obtain "adjusted" values. This would be the preferable strategy when working with data from larger agencies. Estimating or dropping missing cases are likely to be less reliable and less valid strategies than adjustment, since more complex assumptions must be made about potential sources of bias resulting from these procedures; but the decrease in reliability and validity resulting from using estimated values or dropping cases has not been demonstrated to be severe, particularly for these larger agencies. Even in the absence of adjusted or estimated data, the amount of reports missing from larger agencies is relatively small and may not be a major source of bias in one's analysis.

CHAPTER 4

CLASSIFICATION VARIABILITY

This chapter examines the classification of crime reports into UCR offence categories. Factors influencing the classification are examined: FBI guidelines, misclassification, fluctuation in classification, and chages in definitions of the classifications.

The notion of outliers in the data is introduced. Evidence of such outliers and a description of the consequences of such outliers is provided. Strategies for dealing with outliers are examined, especially tradeoff between dropping extreme cases and adjusting results by modeling the fluctuations.

VARIABILITY IN CLASSIFICATION

The classification of crime reports into UCR offence categories is sometimes an uncertain process. Reports of crimes committed are classified into one of the UCR crime categories. When multiple crimes are committed as part of the same act, the most severe offence is chosen as as the category into which the crime is classified. Guidelines are provided by the FBI for determining the appropriate category (FBI, 1980). Even so, there are always some reports for which it is uncertain as to the overload the estimation process. Other variables thought to be important sources of error not included in the current model can be added subsequently in a systematic fashion as our knowledge of criminal justice increases.

Instability and Small Agencies

For less populous jurisdictions, the rate of change in crime levels can be unstable when the base values are relatively small in a month. To reliably estimate adjustment parameters and the "true crime level" it may be necessary to aggregate groups of these small jurisdiction, either across time or according to some similarity criteria. Such aggregated adjustments are likely to have more error than disaggregated ones. Although, if the aggregated units are very similar, the increase in error may be small. In the absence of any knowledge on the likely instability, reasonable strategy its compensate for the instability cannot be proposed. We conclude the complexity of these issues would better be addressed in a separate study, which is why this study only uses cities over 100,000 population.

OVERVIEW OF REPORT

The results of this work include: (1) evaluation of the effects of such problems on the time-series analysis of UCR data and the advantage of alternative procedures for ameliorating these effects; (2) description and specification of the methodology developed for assessing measurement problems in the UCR data; and (3) the preparation of a Suspect Data Key for all agencies with population coverage of 100,000 or more in the NTCCD database that will flag suspect data points, indicate the source(s) of likely error, and provide adjusted replacement values.

Chapter 2 will address UCR definitions that contibute to these issues. This includes problems of documentation, ambiguity of definitions, and changes in definitions. Advantages of these definitions will also be discussed.

Chapter 3 will address missing data in the UCR file. It will locate areas of missing data and strategies for dealing with it. Tradeoffs between dropping cases, estimating missing values, and adjusting parameters are elaborated.

Chapter 4 will examine classification variability. It will consider issues of misclassification, fluctuation in classification, and changes resulting from variable definitions. A particular focus of this chapter will be on outliers in the data and procedures for dealing with them, especially tradeoffs between dropping extreme cases and adjusting results by modeling the fluctuations.

Chapter 5 discusses analytical procedures for analysis of UCR data. It identifies problems and strengths of each procedure. It also proposes a new procedure, Structural Equation Time Series Analysis (SETSA), and demonstrates its use.

The final chapter summarizes major findings of the study. It also presents recommendations and conclusions regarding procedures for handling identified problems in UCR data and for dissemination of the findings. Since the NTCCD database has been made available to the Inter-University Consortium for Political and Social Research by the Center for Applied Social Research, documentation materials are being revised by us to reflect relevant findings of this study. In addition, persons who obtain data directly from CASR will also be provided this information with the data.

CHAPTER 2

CLARIFICATION OF UCR DEFINITIONS

This chapter concerns three basic problems that create the need for greater clarification of UCR definitions: documentation, ambiguity, and changes in definitions. These problems lead to missing data and classification variability. The advantages and disadvantages of these definitions (and changes therein) are discussed. Implications of these definitions for research are also presented.

DOCUMENTATION

Uncertainty of meaning arises partly because documentation for variables in the UCR data set does not occur in a single location. Even when documentation occurs in a single location, the meaning of a variable may be unclear. To simplify documentation for and clarify the meaning of these variables the following summary is provided in Table 2-1.

TABLE 2-1

DEFINITIONS OF SELECTED UCR VARIABLES

TERM	VARIABLE NAME	DEFINITION
ORI Number	ORICODE	Alphabetic code assigned by FBI to identify 5 digit number which places cities in alphabetical order regardless of state.
Core City	CORECITY	"Y" if agency is core city of an SMSA.
Covered by	COVERED	If blank, city is not covered by county which has submitted returns. Otherwise the total for the city is included in that for the county.
Population 1-3	POP1-POP3	Population data is reported in three fields because some cities are located in as many as three different counties. The population in the county having the largest area will be given first.
Follow-up	FRETURN	Request was made for

TERM	VARIABLE NAME	DEFINITION
		follow-up on report(s) not submitted for previous month(s). Field contains "Y" if agency should be sent follow-up.
Month Included	MOINCL	Used to indicate that data missing from one month in a submitted return may be found in the return for another month. Code is the month in which the missing data have been included. If reports are complete or missing data are unavailable, field contains zeros.
Card Type 1	CARD1	Contains type of return info submitted by the agency for category "Number of Actual Offenses, Including Attempts." Possible codes: 1 Incomplete Return 2 Adjustment 3 Estimated Return 4 Not Available 5 Normal Return 6 Not Updated

		Ambiguities in the meaning of these codes are discussed in the next section.
Card Type 2	CARD2	Contains type of return info submitted for category "Total Offenses Cleared by Arrest or Exceptional Means." Codes same as CARD1.
Card Type 3	CARD3	Contains type of return info submitted for category "Number of Clearances Involving Only Persons Under 18." Codes same as CARD1.
Card 1 P/T	VAR1	Indicates whether return submitted by agency for category "Number of Actual Offenses, Including Attempts," contains either breakdowns of offenses or totals.
Card 2 P/T	VAR2	As noted for VAR1 for category "Total Offenses Cleared by Arrest or Exceptional Means."
Card 3 P/T	VAR3	As noted for VAR1 for category "Number of

Clearences Involving Persons Under 18."

AMBIGUITY

Card Type

Several ambiguities exist for the Card Type variable. The Card Type variable identifies the type of information in the Return A card. The codes for the CARD1 variable require clarification.

Code "1" (Incomplete Return) summarizes the completeness of the reporting for the calander year. If monthly returns are less than twelve (12) and more than two and no adjustment or estimation has occurred, the annual description is listed as incomplete.

Code "2" (Adjustment) identifies returns in which there has been a revision in the return for some agency in the same state as this agency. This ambiguity occurs because adjustments for one locality affect the adjustment status for all localities in the same state. Adjusted returns are made only for those states submitting magnetic tape returns. When a locality makes an adjustment of a previous return, the entire state, by default, is designated as adjusted because the tape for the whole state must be rewritten. This does not imply that the entire state has been adjusted, only that some editing in one or more localities in the state as taken place. Subsequently, the entire state is designated as adjusted. It is impossible to determine which locality has actually adjusted its return without contacting the agencies involved. Adjusted returns should, therefore, be considered comparable to "normal returns." Inspection of the data should reveal several adjustments, since the N of adjustments can potentially equal the N of localities within states submitting magnetic tapes during this time frame (24 states) times the number of years being studied.

Code "3" (Estimated) applies only to cities under 50,000 population. An estimated type of return is not valid for cities over 50,000. Returns are estimated if at least nine months of data are available for that agency in that year. There is some uncertainty over the estimation process, but Schneider and Wiersema (1985) have demonstrated that it is essentially a substitution of average monthly values based on available months of the year in question. For agencies in which missing reports are not concentrated in a group of months in which crime reports tend to be very high or very low, this is an adequate procedure. However, when missing reports correlated with seasonal fluctuations in crime, this estimation procedure will systematically under- or overestimate the reported crimes. Findings in Chapter 3 demonstrate that missing reports do tend to cluster together and tend to be concentrated in the summer and fall months. Further analysis is needed of the amount of error likely to be introduced in the estimates from from the current procedure when applied to the the non-random clustering of missing reports.

Code "4" (Not available) indicates agencies for which no information is available for that month in that year. Less than three months of reports were filed in that year or the population size of the city was greater than 50,000 (or both), so estimation was not done either.

Code "5" (Normal Return) are returns in which reports are available for all twelve months of that year. In addition, no adjustments must have been made to the data.

Code "6" (Not Updated) is a defunct machine code that was used in D.C. It is no longer valid. We have be unable to determine the date when it became defunct because this code was not found in the 1965 to 1984 period. Apparently, it became defunct prior to 1965.

Month Included

The month included variable is intended to identify where data missing in one month are located in another month's report, if the missing data is included in a subsequent month's report. The code is the month in which may be found the current month's report. When data are missing at the end of the year, it might be found in a month for the following year. If the month included is less than the current month, the data is in next year's reports. If the month included is more than the current month, it is in the current year's reports. If month included is zero, the missing data is not available in the UCR computer file. For the agencies in this study, all missing reports included in a subsequent month were included in December of the year in which the report was missing (43 reports). There were sixty-nine reports that were missing and not available, all from Michigan agencies that were not reporting agencies in 1965.

CHANGES IN DEFINITIONS

Three types of crimes have had significant changes in definition during the 1965 to 1983 period: robbery, larceny, and manslaughter. Changes in the definition of assault and auto theft have also occurred.

The definition of robbery changed in 1974. Prior to 1974 robbery had two categories: armed and strongarm. From 1974 to the present robbery has three subcategories of armed robbery: firearm, knife or other cutting weapon, and other dangerous weapon. For armed robbery post 1973 one muse add the three subcategories.

Larceny also underwent a change in definition in 1974. Prior to 1974 larcenies involving property values less than $50 counted as a distinct offense (simple larceny). This classification was discontinued from Return A beginning in 1974. No Return A data is available for simple larceny after 1973. The variable location in the data for these years is coded as "inappropriate."

Manslaughter remained consistent until 1978 when police departments were directed to exclude traffic deaths from the count of negligent manslaughter. Prior to 1978 some traffic fatalities were included in the definition of negligent manslaughter. Some local jurisdictions still treat some traffic fatalities as negligent manslaughter (especially if they involve drugs or alcohol), but they are told not to count those in their UCR Return A reports. Consequently, their local data may not agree with the UCR totals. Analysis that examines assault data pre- and post-1974 must reconcile the measure so it means the same thing before and after 1974.

Total assaults is another variable that changed its meaning in 1974. Prior to 1974 "total assaults" was actually the same as "aggravated assaults." Simple assaults were not included in the total until 1974.

Changes in the definition of auto theft have varied between jurisdictions. In some jurisdictions the crime "unauthorized use of motor vehicle" (not a Return A offense) has replaced motor vehicle theft when the car has been found within a short period of time (24 hours or so). When analyzing motor vehicle theft for a small number of agencies, some adjustment may be necessary if those jurisdictions have such an offense not included in the Return A reports.

FURTHER CLARIFICATION

Although there have been ambiguities and changes in UCR definitions, a careful search of available documentation has clarified their meaning and the points at which major changes in definitions have occurred. Additional work is needed to further clarify remaining points of uncertainty.

A simple and direct action can be take to make the meaning of "adjusted return" more clear and more useful for analysis. When the computer tape for a state is revised, only those localities having changed information should be labeled "adjusted." It should not require a major change in the computer program to leave unchanged the Card Type variable when the information for that agency has not been changed. At present it is impossible to tell how many agencies have their reports "adjusted," except that it can be no more than 30 percent of all agencies (the approximate proportion of all reports that are classified as "adjusted" in this sample).

A second action that would clarify the meaning of UCR data would be to examine the "estimated" data values for bias in the estimation. The clustering of missing reports in adjacent months combined with seasonal fluctuations may introduce bias into the estimation procedure. Documentation of the extent and severity of this bias would help establish how much these estimated values can be trusted.

CHAPTER 3

STRATEGIES FOR MISSING UCR DATA

This chapter examines the problem of data missing from the UCR information system. The nature and consequences of the missing data is explained. Influences that encourage missing data are described. Strategies for dealing with missing UCR data are examined. The tradeoffs between these strategies are also explored.

PROBLEMS OF MISSING DATA

The basic problems with missing UCR data stem from the fact that there different types of missing data, different sources, and different consequences. The types and sources may result in diverse consequences, so that it is sometimes difficult to tell whether missing data confounds one's analysis or not.

TYPES OF MISSING DATA

There are six situations in which data may be said to be missing. It might never have been reported to the police in the first place or never recognized as a crime if reported (hidden crime). It may be absent from any UCR report (missing-unavailable). It may be absent from the UCR report in which it is supposed to be located, but present in a subsequent